Mass Transportation Problems in Probability Theory

J.A. Cuesta-Albertos*, *Universidat de Cantabria*
C. Matrán**, *Universidad de Valladolid*
S.T. Rachev***, *University of California, Santa Barbara*
L. Rüschendorf****, *Universität Freiburg*

Abstract

This paper surveys some recent results on the mass transportation problem (MTP). We examine the dual and explicit solutions of MTP and its various versions. A variety of possible applications of MTP in the context of probability theory are listed: probability metrics, stochastic algorithms, limit theorems, distributions with given marginals, computer tomography and others.

Key words: mass transportation, mass transshipment, probability theory, stochastic algorithms, method of moments, computer tomography.

AMS 1991 Subject Classification: Primary 60K30, 90C08

1 Introduction

Extensive surveys ([28], [61], [70], [85]) and two monographs ([30], [69]) have appeared in the early nineties treating various mass transportation problems and listing additional references on them. More recently, the subject has developed intensively and interesting new applications have been found. The scope of this survey is to review the most recent contributions.

* Departamento de Matemáticas, Estadística y Computación. Universidad de Cantabria. Spain.
** Departamento de Estadística e Investigación Operativa. Universidad de Valladolid. Spain.
*** Department of Statistics and Applied Probability. University of California. Santa Barbara. USA.
**** Institut für Mathematische Stochastik. Universität Freiburg. Germany.

Running title: Mass Transportation

The first two authors were partially supported by the Spanish DGICYT under grants PB91-0306-C02-00, 01 and 02, while the third was supported by the Alexander von Humboldt Research Award for U.S. Scientists

The paper is divided in three parts. Section 2 is devoted to the Monge-Kantorovich problem; Section 3 deals with the Kantorovich-Rubinstein problem, and Section 4 considers transportation problems with additional or relaxed constraints.

2 The Monge-Kantorovich problem

The mass transportation problem was formulated by Monge in 1781 [60]. Monge was interested in the minimization of the cost of transportation of soil from one location to another, which is fixed previously. Starting with the assumption that soil consists of small grains, the problem is to give the final location of every grain in such a way that the costs of transportation are as low as possible. Note that in this formulation it is not possible to divide the grains, so that mass which share the same initial location must also share the same final location.

The so-called Kantorovich formulation, [45], of the problem is similar except for the fact that here one is allowed to "divide grains". In this way we do not need to answer the question of what the final location of the grains should be, but the following alternative one: Given the sets A and B, what is the portion of the mass initially located in A which should be transferred to B? We will denote this quantity by $P_{1,2}(B|A)$ and say that the family $\{P_{1,2}(\cdot|\cdot)\}$ is a transportation plan. It is reasonable to assume that the mass does not vary in the transportation process. So, without loss of generality, we can assume that the total mass is one and identify the functions giving the initial and final distributions of the mass as probability measures P_1 and P_2. It is evident that not every transportation plan is admissible. We define a probability measure P on the product space $U \times U$ by using P_1 as the first marginal and $\{P_{1,2}(\cdot|\cdot)\}$ as the transition probability. The transportation plan given by $P_{1,2}$ is admissible if and only if the second marginal distribution of P coincides with P_2. To optimize the cost of transportation we need to know the cost of transporting a unit from one location to another. We denote by $c(x, y)$ the cost of transportation of a unit of mass initially located at x to its final destination y. We shall assume that c is positive. (The motivation for this problem and some basic results are taken from [69]. An extensive review on the discrete Monge problem is given in [15]).

Finally we can state the Monge-Kantorovich transportation problem (MKP) in the following terms:

Let (U, d) be a metric space, P_1 and P_2 be two probability measures defined on its Borel σ-algebra and let $\mathcal{M}(P_1, P_2)$ be the set of all probability measures defined on the product σ-algebra with marginal distributions P_1 and P_2 respectively. The problem is to compute the functional

$$C(P_1, P_2) := \inf \left\{ \int_{U \times U} c(x, y) P(dx, dy) : P \in \mathcal{M}(P_1, P_2) \right\}, \tag{1}$$

where $c : U \times U \to \Re^+$ is measurable.

If $P \in \mathcal{M}(P_1, P_2)$ is a solution of (1), then we say that P determines an optimal transportation plan (OTP) with respect to c for P_1 and P_2 or, in other words we say that P determines an OTP(c) for P_1 and P_2. $C(\cdot, \cdot)$ is called the Kantorovich Functional. Evidently the OTP(c) would be the family of conditional probabilities determined by P.

Equation (1) can also be written in terms of random variables (r.v.s). With this terminology, the MKP is equivalent to the problem of finding a pair of r.v.s (X, Y), such that the distributions of X and Y are P_1 and P_2 respectively and verifying that

$$C(P_1, P_2) = E[c(X, Y)].$$

With a slight abuse of notation, if the pair (X, Y) satisfies this equality then we say that it is an OTP(c) for their marginal distributions.

The first problem is to check whether the infimum in the definition is attained, or equivalently, whether an OTP(c) for P_1 and P_2 exists. This can be resolved under rather general conditions. If P_1 and P_2 are tight probability measures, then every probability distribution on $U \times U$ with P_1 and P_2 as marginals is also tight. Then a standard argument permits us to conclude that if c is continuous or at least lower semicontinuous, an OTP(c) for P_1 and P_2 exists.

An interesting topological property of $C(P_1, P_2)$ arises from the fact that, under suitable conditions on the function c, $C(P_1, P_2)$ induces a distance equivalent to the convergence in distribution together with the convergence of integrals (see Theorem 2.1 below).

For instance, if $H : [0, \infty) \to [0, \infty)$ is a continuous, nondecreasing function such that $H(0) = 0$ and H satisfies Orlicz's condition

$$\sup_{t>0} \frac{H(2t)}{H(t)} < \infty,$$

and $c(x, y) = H[d(x, y)]$, then C induces a distance between P_1 and P_2 which is in fact a minimal distance in the terminology of [69].

An important particular case is when $H(t) = t^r, r > 0$, because then

$$d_r(P_1, P_2) := C^{r^*}(P_1, P_2), r^* = \min\{1, 1/r\} \tag{2}$$

is a metric on the space of probability measures with finite moment of r-th order. In this paper we use the notation $c_r(x, y) := d^r(x, y)$, and C_r for the associated Kantorovich functional.

When $c(x, y) = H[d(x, y)]$, with H as above, the following result relates the convergence to zero of $\{C(P_n, P)\}$ to weak and moment convergence.

Theorem 2.1 *Let $\{P_n, n \geq 0\}$ be probability measures on the Borel σ-algebra on U. Assume that*

$$\int_U c(x, a) P_n(dx) < \infty, n = 0, 1, \ldots \textit{ for some } a \in U.$$

Then $\{C(P_0, P_n)\}$ converges to zero if and only if $\{P_n\}$ converges weakly to P_0 and

$$\lim_n \int_U c(x,b)(P_n - P_0)(dx) = 0$$

for some (and therefore for any) $b \in U$.

This result was proved in [65], [67], and in the special cases $H(t) = t^r, r \geq 1$ in [12], [66], [95], $r \geq 1$, for a bounded metric d in [38] and for $r = 2, U = \Re$ with the usual distance in [59].

These metrics have been employed in several ways in the literature, and many applications appear in [69]. We later discuss some more recent applications. Among them the following ones:

- In [36], [44], [74] metrics $d_r, r > 0$, have been used to measure the order of convergence of a sequence of approximations to the solution of a stochastic differential equation.
- In [80], [81] the metric d_2 is employed to measure the stability of a stochastic program model with respect to the underlying distribution. This situation is relevant when the underlying distribution is not known exactly.

An interesting metric related to d_r is the total variation metric:

$$\sigma(P_1, P_2) := \sup_{A \in B(U)} |P_1(A) - P_2(A)|,$$

where $B(U)$ is the Borel σ-algebra in U. In fact σ can be viewed as the limiting case for d_r:

$$\sigma(P_1, P_2) = \lim_{r \to \infty} d_r(P_1, P_2).$$

It has a representation as the minimal metric given by

$$\sigma(P_1, P_2) = \inf_{X,Y} \mu[X \neq Y],$$

where X and Y are r.v.s with distributions P_1 and P_2 respectively. For properties and applications of σ see [69].

An important problem is to determine the conditions under which the solution for the MKP coincides with that one for the Monge problem with the same marginals. Or, equivalently, under what conditions can every OTP for P_1 and P_2 be written as $(X, T(X))$?¿ It is not difficult to find examples with degenerate P_1 in which such a function does not exist because this property is related to a kind of continuity of the distribution P_1. In fact it is known that both solutions coincide in the following cases:

(i) If P_1 and P_2 are defined on a bounded subset of the finite dimensional space $\Re^k$ and are absolutely continuous with respect to the Lebesgue measure and the cost functional is C_2 (proved in [90]).

(ii) For the d_2-metric the result in (i) was extended to the case where only P_1 has a Lebesgue density. It has also been extended to the case of a separable Hilbert space U but under additional stronger continuity assumptions (cf. [3], [20], [24]).

(iii) For a general U and c under the only assumption that the infimum in (1) is attained, the support of P_2 is finite and $P_1\{x : c(x,a) - c(x,b) = h\} = 0$ for every $a, b \in U$ and $h \in \Re$ (see [27]).

Note that the condition in (iii) above is not fulfilled if $U = \Re^k$ endowed with the Euclidean norm, P_1 is absolutely continuous with respect to the Lebesgue measure and $c(x,y) = \|x - y\|$. Under the assumption that P_1 is atomless, the support of P_2 is discrete, U is a separable Banach space, and $c(x,y) = \|x-y\|^r, r \geq 1$, it is shown in [3] that there exists an OTP for P_1 and P_2 of the form $(X, T(X))$.

2.1 Duality theorems, explicit results and uniqueness

As in linear programming in finite dimensional spaces an important tool for studying and solving the transportation problem (1) is the use of duality theorems. The history of these duality results for the transportation problem goes back to 1942, when Kantorovich [45] considered the special case where the cost $c(x,y) = d(x,y)$ is a distance in a compact metric space. There are a lot of contributions to and extensions of the duality theorem (cf. [30], [46], [69], [83], [53]). Recently a quite general duality theorem has been proved in [78]. Before stating this result, we recall that a probability space $(\Omega, \mathcal{A}, P)$ is called perfect if for any measurable function $f : \Omega \to \Re$, one can find a Borel set $B \subset f(\Omega)$ such that $P(f^{-1}(B)) = 1$. Perfectness is a very weak regularity condition on Ω, P.

Theorem 2.2 *(General Duality Theorem)*
Let $(\mathcal{U}_i, \mathcal{A}_i, P_i)$, $i = 1, 2$ be probability spaces such that P_1 is perfect and let $c : \mathcal{U}_1 \times \mathcal{U}_2 \to \Re$ be product measurable and upper majorized (i.e. $c(x,y) \leq f_1(x) + f_2(y)$ for some $f_i \in \mathcal{L}^1(P_0)$) then the following duality theorem holds:

$$C(P_1, P_2) = \sup \left\{ \int h_1 dP_1 + \int h_2 dP_2;\ h_i \in \mathcal{L}^1(P_i),\ h_1(x) + h_2(y) \leq c(x,y) \right\}.$$

Duality theorems are the basis of many of the properties of the Monge Kantorovich functional $C(P_1, P_2)$ and very often lead to explicit results for the transportation problems. They also lead to the construction of optimal plans (cf. the introduction of this section), and under some conditions solutions for the dual problem in Theorem 2.2 exist. In this framework, an optimal plan $\mu^* \in \mathcal{M}(P_1, P_2)$ is characterized by the existence of $h_i^* \in \mathcal{L}^1(P_i)$, $i = 1, 2$ with

$$h_1^*(x) + h_2^*(y) \leq c(x,y)$$

such that with respect to μ^*

$$c(x,y) = h_1^*(x) + h_2^*(y) \quad \text{a.s.} \tag{3}$$

For special cost functions, e.g. for $c = d$ or d^r, more specific versions of the duality theorem have been established in the literature, see [69]. Later in this section we derive explicit characterizations for an OTP pair (X, Y) with marginals P_1 and P_2 and prove uniqueness.

Both problems are solved in full generality only in the one-dimensional case where it is known that OTPs coincide with increasing arrangements. The first known results in this direction are in early papers of Dall'Aglio [28], [59]. The following result from [26] includes and completes known characterizations of OTP(c_2)s.

Proposition 2.3 *Let X_1 and X_2 be real, square integrable r.v.s defined on the probability space $(\Omega, \mathcal{A}, \mu)$ with d.f.s F_1 and F_2. Then:*

a) The following are equivalent:

- *a.i) (X_1, X_2) is an OTP(c_2).*
- *a.ii) $F_{(X_1,X_2)}(x, y) = \min\{F_1(x), F_2(y)\}, \ \forall x, y.$*
- *a.iii) There exists a r.v. Z uniformly distributed on $(0, 1)$, such that for some nondecreasing functions ϕ_1, ϕ_2:*

$$X_1 = \phi_1(Z) \quad \textit{and} \quad X_2 = \phi_2(Z), \qquad \textit{a.s. with respect to } \mu$$

- *a.iv) $\mu \otimes \mu\{(\omega, \omega') : (X_1(\omega) - X_1(\omega')) \times (X_2(\omega) - X_2(\omega')) \geq 0\} = 1.$*

b) The functions ϕ_i in a) are essentially unique, $\phi_i = F_i^{-1}$ a.s. with respect to Lebesgue measure.

c) Given $\alpha \in (0, 1)$ and $x \in \Re$, we define

$$\mathcal{F}(x, \alpha) := \mu(X_1 < x) + \alpha\mu(X_1 = x).$$

If Z is a r.v. uniformly distributed on $(0, 1)$, independent of X_1, then the pair $\big(X_1, F_2^{-1} \circ \mathcal{F}(X_1, Z)\big)$ is an OTP(c_2) for P_1 and P_2.

d) If P_1 is nonatomic and (X_1, Y_1) is an OTP(c_2) for P_1 and P_2 then

$$Y_1 = F_2^{-1} \circ F_1(X_1), \qquad \textit{a.s. with respect to } \mu$$

e) If $Y_1 = \phi_1(X_1)$ with ϕ_1 non-decreasing, then (X_1, Y_1) is an OTP(c_2).

If $c(x, y) = \phi(|x - y|)$, then a uniqueness result holds essentially only if ϕ is convex (see Remark 1 (3) in [26]). So in the one-dimensional case the OTP does not depend on the cost of transportation functions c (in the class of functions considered above). This is not the case if the dimension is greater than one. In the following example (see [21]) the OTP(c_r) depends on r. Consider the points in $\Re^2$:

$$m_0 = (0, 0), m_1 = (1, 0), m_2 = (-1/2, \sqrt{3}/2),$$

and the probabilities $P_i, i = 1, 2,$ which allocate probability $1/2$ to m_0 and $m_i, i = 1, 2,$ respectively.

It is easy to show that a probability measure P giving an OTP(c_2) for P_1, P_2 is

$$P\{(m_0, m_2)\} = P\{(m_1, m_0)\} = \frac{1}{2}.$$

However for $r < (\log 4)/(\log 3)$ an OTP(c_r) is given by the probability P^* defined by

$$P^*\{(m_0, m_0)\} = P^*\{(m_1, m_2)\} = \frac{1}{2}.$$

In general spaces most results providing explicit OTPs have been found in cases in which a solution for the Monge problem gives an OTP (see the preceding section); i.e. for a function $T : U \to U$ the pair $(X, T(X))$ is an OTP.

The problem has two aspects, the construction and the uniqueness of the solution. A complete characterization is known for a pair (X, Y) to be an OTP. The following theorem is proved in [87] (see also [51], [89]).

Theorem 2.4 *Assume that $(U, \|\cdot\|)$ is a separable Hilbert space and that P_1 and P_2 are two probability measures such that $\int \|x\|^2 P_i(dx) < \infty, i = 1, 2.$*

If X, Y are two r.v.s with distributions P_1 and P_2 respectively, then (X, Y) is an OTP(c_2) if and only if $Y \in \partial f(X)$ a.s. for some lower semicontinuous convex function f, where $\partial f(x)$ denotes the subgradient of f in x:

$$\partial f(x) = \{y : f(x') - f(x) \geq \; < y, x' - x >, \textit{ for all } x' \textit{ in the domain of } f\}.$$

An interesting consequence of Theorem 2.4 is that the optimality of a certain map depends only on the map and not on the distributions P_1 and P_2. More precisely,

Corollary 2.5 *Let $T : U \to U$ be a measurable map such that $(X, T(X))$ is an OTP(c_2). Let X^* be a r.v. whose support is contained in that of X. Then $(X^*, T(X^*))$ is also an OTP(c_2).*

It is known (see [79]) that a function T satisfies $T(x) \in \partial f(x)$ for some semicontinuous convex function if and only if T is cyclically monotone; i.e.

$$\sum_{0 \leq i \leq m-1} < x_{i+1} - x_i, Tx_i > \leq 0, \text{ for } x_0, \ldots, x_m = x_0 \in U. \tag{4}$$

As a consequence, functions giving OTP(c_2) coincide a.s. with cyclically monotone functions.

Theorem 2.4 is particularly useful since the subgradients of convex functions or equivalently cyclically monotone functions are well studied in convex analysis (cf. [79]). They are basic for the solution of convex optimization problems. This allows the construction of many examples of optimal transportation functions and optimal pairs (X, Y) of the transportation problem. Examples of optimal functions are positive semidefinite, symmetric linear functions, radial transformations, projections on convex sets etc. (cf. [26], [85], [87]). A continuously differentiable function ϕ is an optimal function with respect to c_2 if and only if

$$\frac{\partial \phi_j}{\partial x_i} = \frac{\partial \phi_i}{\partial x_j} \quad \text{for} \quad i \neq j \quad \text{and} \quad \phi \quad \text{is monotone,} \tag{5}$$

i.e. $\langle x - y, \phi(x) - \phi(y) \rangle \geq 0$ (cf. [85]).

The symmetry of the derivatives is a consequence of the Poincaré lemma. The necessity of the monotonicity property of optimal functions was first established in [20] (see also [24]).

Monotone functions (also called Zarantonello-monotone) enjoy several good analytic properties. They are continuous a.s. with respect to the Lebesgue measure (and, therefore, measurable), they are continuous on each point such that the image lies in the interior of the range, etc. (cf. [24], [92]).

An immediate application of Theorem 2.4 is to the case of Gaussian probability measures (see [38], [51], [62], [87]).

Proposition 2.6 *Let P_1 and P_2 be two n-dimensional, centered Gaussian probability measures with covariance matrices Σ_1 and Σ_2 respectively (regular or not), then*

$$C_2(P_1, P_2) = \ trace \ \left[\Sigma_1 + \Sigma_2 - 2 \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right]. \tag{6}$$

Moreover, if X is a r.v. with distribution P_1, Σ_1 is non-singular and

$$A = \left(\Sigma_1^{1/2} \right)^{-1} \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \left(\Sigma_1^{1/2} \right)^{-1},$$

then (X, AX) is an OTP(c_2) for P_1 and P_2.

Equality (6) has been extended to general separable Hilbert spaces in [24], [35]. The extension in [25] includes an expression for the operator which corresponds to A in Proposition 2.6. Moreover it is shown in [25], [35], [87] that (6) provides an universal lower bound for the cost of transportation between distributions P_1 and P_2 with covariances Σ_1, Σ_2.

Proposition 2.7 *Let P_1 and P_2 be two n-dimensional probability measures centered in mean and with covariance matrices Σ_1 and Σ_2 respectively. Then*

$$C_2(P_1, P_2) \geq \ trace \ \left[\Sigma_1 + \Sigma_2 - 2 \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right] \tag{7}$$

This proposition has been generalized to separable Hilbert spaces in [24]. Moreover, in that paper two families of lower bounds are provided for $C_2(P_1, P_2)$ which depend on the orthogonal basis under consideration. In this way, for each orthogonal basis on U two lower bounds are found. One of them depends on just the first two moments of the one-dimensional marginal distributions of P_1 and P_2. The other lower bound (which is more precise) is the sum of the C_2-costs of transportation between the one dimensional marginals of P_1 and P_2. The lower bound in (7) is the supremum of the first family. The second family of lower bounds can be improved if one considers the sum of costs of transportation between marginals with dimension greater than one.

In general, however, the problem of determining optimal c_2-couplings remains difficult and leads to the problem of finding the solution of a partial differential equation of Monge-Ampère type. For example it is not known in general how to find the optimal plan for a pair P_1 and $P_1 T_\alpha^{-1}$, where T_α is a rotation by the angle α in $\Re^2$ (or more generally an orthogonal transformation in $\Re^k$). It has been proved (see [92]) that if the support of the distribution of X contains an open set, then $(X, T_\alpha(X))$, $\alpha > 0$ is not an OTP(c_2).

Theorem 2.4 has been extended to general cost functions c in [84], [85], [86]. The role played by convex functions, their subgradients, and cyclically monotone functions in convex analysis is replaced by c-convex functions, c-subgradients and c-cyclically monotone functions. The last three notions are introduced in non-convex optimization theory, but unfortunately have not been well studied till present. Several explicit examples of optimal c-plans (functions) have been found. For $c(x,y) = \|x-y\|^r, r > 1$ pairs $(X, \phi(X))$ are OTPs where

$$\phi(x) = |h(x)|^{-\frac{r-2}{r-1}} h(x) + x, \tag{8}$$

with h any cyclically monotone function. In particular

$$\phi(x) = (x^T A^2 x)^{-\frac{r-2}{2(r-1)}} Ax + x, \tag{9}$$

with A a positive semidefinite, symmetric linear function and $\phi(x) = g(\|x\|)\frac{x}{\|x\|}$ a radial transformation are optimal. Also norms different from Euclidean norms are considered in these papers. A general construction for optimal c-plans is yet to be found.

Some approximating algorithms have been established in the case when P_2 has finite support:

1. In [5] an algorithm is provided to find the OTP(c_2) between two distributions P_1 and P_2 defined in $\Re^k$ if P_1 is continuous with bounded support and the support of P_2 is finite.

2. In [1], [2] an algorithm is given to construct OTP(c_r) approximately if the cardinality of the support of P_2 is finite. P_1 is required to satisfy

$$P_1\{x : \|x-a\|^r - \|x-b\|^r = h\} = 0, \forall a, b \in U \text{ and } h \in \Re. \tag{10}$$

   For the algorithm one has to compute an element $\alpha \in \Re^{n-1}$, where n is the cardinality of the support of P_2, and then solve a certain equation $h(t) = \alpha, t \in \Re^{n-1}$. The solution of this equation is not easy to find and a procedure to construct a sequence of approximate solutions is given when P_1 satisfies certain Lipschitz conditions. The range of applications of these algorithms has yet to be investigated.

The problem of uniqueness of OTPs has been solved in some particular cases. In [3], [24] the uniqueness of the OTP(c_2) is proved in Hilbert spaces, if one of the probabilities involved verifies certain continuity condition. In fact the technique employed in [24] can be used to extend this result to the case in which it is known that the solutions of the Monge and Monge-Kantorovich problem coincide and $c(x, y) = H[d(x, y)]$ with H strictly convex.

Furthermore, in [27] the uniqueness of the OTP(c) for general U and c was proved under the condition that the infimum of (1) is reached, the support of P_2 is finite and $P_1\{x : c(x, a) - c(x, b) = h\} = 0$ for every $a, b \in U$ and $h \in \Re$; but the technique employed is of a different nature. This result has been extended in [1] to the case where the support of P_2 is denumerable but under the assumption of a c_r-cost function. The uniqueness result implies in particular that OTPs are continuous with respect to weak convergence (see [24]):

Theorem 2.8 *Let $P, Q_n, n = 0, 1, \ldots$ be probability measures defined on $\Re^k$ such that $\int \|x\|^2 dQ_n < \infty, n = 0, 1, \ldots$ and $\int \|x\|^2 dP < \infty$. If P is absolutely continuous with respect to the Lebesgue measure and $(X, T_n(X)), n = 0, 1, \ldots$ are OTP(c_2) between P and $Q_n, n = 0, 1, \ldots$ respectively, and Q_n converges weakly to Q_0, then*

$$T_n(X) \to T_0(X), \text{ a.s.}$$

The proof is based on the analytic properties of monotone functions mentioned before. So, the algorithm in [1], [4] together with the above result provides an approximate solution for the OTP(c_2) between two given probability measures if one of them is continuous. An extension of Theorem 2.8 to more general cost functions would be of a considerable interest.

2.2 Convergence of empirical measures, Laws of large numbers

We now investigate the topological properties of the metrics under consideration to obtain some classical results in probability theory.

Minimal metrics (like $d_r(P, Q)$) have been applied to prove various versions of the Central Limit Theorem (see e.g. [20], [69], [71], [73], [91]). The proofs are based on regularity properties of the minimal metrics as well as on Theorem 2.1.

It is not surprising that these metrics allow us to obtain a simple proof of Mourier's Strong Law of Large Numbers (SLLN) in Banach spaces as well. Consider a sequence, $\{X_n\}$, of U-valued, independent, identically distributed random elements with distribution Q defined on the probability space $(\Omega, \mathcal{A}, \mu)$. Assume that U is a separable Banach space with norm $\| \cdot \|$ and consider the function c_1 as the cost of transportation.

Let for $\omega \in \Omega$ P_n^ω be the empirical probability measure allocating probability $1/n$ to each of the points $X_1(\omega), \ldots, X_n(\omega)$. If we assume that $E\|X_1\| < \infty$, then the SLLN for real random variables implies that

$$\int \|x\| P_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \to \int \|x\| Q(dx), \quad \text{a.s. with respect to } \mu$$

Moreover, Varadarajan's extension of the Glivenko-Cantelli theorem (which only requires the SLLN for real, bounded r.v.s for its proof) states that the sequence of probabilities $\{P_n^\omega\}$ converges weakly to Q a.s. with respect to μ. By Theorem 2.1 we conclude that

$$\lim_n C_1(P_n^\omega, Q) = 0, \quad \text{a.s. with respect to } \mu, \tag{11}$$

a version of Varadarajan's theorem (cf. [22]). Finally, if (U_n^ω, V_n^ω) is an OTP(c_1) between P_n^ω and Q this implies that

$$\|\frac{1}{n}\sum X_i(\omega) - EX_1\| = \|EU_n^\omega - EV_n^\omega\| \leq E\|U_n^\omega - V_n^\omega\| = C_1(P_n^\omega, Q) \tag{12}$$

and we have proved that the sequence $\{\frac{1}{n}\sum X_i(\omega)\}$ converges a.s. with respect to μ to EX_1 (cf. [22]). For far reaching extensions of this result we refer to [58].

Moreover, the same kind of reasoning leads to results on the a.s. stability of sums of r.v.s in Banach spaces. For instance consider the case of weighted sums. Assume that U is a separable Banach space and that $\{X_k\}$ is a sequence of independent, integrable U-valued random elements with finite expectations. We seek conditions implying

$$\sum_{k\geq 1} a_{n,k}(X_k - EX_k) \overset{a.s.}{\longrightarrow} 0. \tag{13}$$

Here $\{a_{n,k}\}$ is a Toeplitz sequence of real numbers, i.e.

$$\lim_n a_{n,k} = 0 \text{ for each } k \geq 1$$

and

$$\sum_{k\geq 1} |a_{n,k}| \leq C \text{ for each } n \geq 1.$$

Assume that $C = 1$, and reduce the problem to the case in which all weights are positive. Define $X_0 := 0$ and $a_{n,0} := 1 - \sum_{k\geq 1} a_{n,k}$. Consider the probability measures $Q_n = \sum_{k\geq 0} a_{n,k} P_{X_k}$ and $P_n(\omega, \cdot), n = 1, 2, \ldots$ allocating mass $a_{n,k}$ to the point $X_k(\omega), k = 0, \ldots$. Argue as in the SLLN to obtain a variety of known results on almost sure stability of weighted sums in Banach spaces. This derivation does not require geometric conditions on the space but reduces the problem to the corresponding result for real r.v.s.

In the same way one can also cover cases in which the weights are not constants but random, or in which they are given by linear operators. Also one can obtain results on further summation methods such as Cesàro, Abel, and others (see [22]). Central limit theorems for summability methods by means of ideal metrics have been given in [71]. A generalization to operator-stable summation schemes is outlined in [58].

2.3 Simultaneous representations. Skorohod-Lebesgue spaces.

Here we are interested in the extension of some properties of the quantile functions to abstract spaces. For a probability measure P on the real line with distribution function F, its quantile function is defined as

$$T_P(t) := \inf\{u : F(u) \geq t\},\ t \in (0,1).$$

Two key properties of the quantile function are that $T_P(X)$ has distribution P if X is uniformly distributed on $(0,1)$ and that the mappings T_P are the typical example of the Skorohod representation for weak convergence of probability mesures. This is described in the following proposition.

Proposition 2.9 *Let X be a r.v. uniformly distributed on $[0,1]$. If $\{P_n\}$ is a sequence of probability measures defined on $\Re$ which converges weakly to P, then $\{T_{P_n}(X)\}_n$ converges a.s. to $T_P(X)$.*

Another important property is that quantile functions provide a simultaneous representation for the the Kantorovich functional in the one-dimensional case: According to Proposition 2.3, if X is a fixed r.v. with uniform distribution on $(0,1)$ then for every probabilities P_1 and P_2

$$C(P_1,P_2) = E[c(T_{P_1}(X),T_{P_2}(X))]. \tag{14}$$

These properties have the following interesting application. Let $L_r(\mu)$ be the set of all real r.v.s X such that $\int |X|^r d\mu < \infty$ and let $H : L_r(\mu) \to \Re$ be a functional which depends only on the distribution of the r.v.s, i.e. if $X,Y \in L_r(\mu)$ satisfy that $P_X = P_Y$ then $H(X) = H(Y)$. With a slight abuse of notation, we can write $H(P_X)$ instead of $H(X)$. Assume that H is continuous with respect to the L_r-norm on $L_r(\mu)$. By employing the same notation as in preceeding section, we want to show that strong pointwise consistency of $H(P_n^\omega)$,

$$\lim_n H(P_n^\omega) = H(Q), \quad \text{a.s. with respect to } \mu. \tag{15}$$

Taking into account that for every fixed ω, $T_{P_n^\omega}$ is a r.v. defined on the interval $(0,1)$ whose distribution is P_n^ω and that by the Glivenko-Cantelli theorem, Proposition 2.9, and the Strong Law of Large Numbers respectively, we have that a.s. with respect to μ

$$T_{P_n^\omega}(t) \to T_Q(t)$$

for almost all t in $[0,1]$, and

$$\int |t|^r P_n^\omega(dt) \to \int |t|^r Q(dt).$$

Then applying Theorem 2.1 we get that

$$\lim_n \int |T_{P_n^\omega}(t) - T_Q(t)|^r dt = 0, \quad \text{a.s. with respect to } \mu.$$

This gives the consistency result (15) by the continuity of H with respect to the L_r-norm. Von Mises functionals related to L_r-norms, as in the case of the r-means, are a typical example of application of this method (see [18], [19]), which can be generalized to cover further functionals related to Orlicz spaces (see [52]).

The interesting point in this method is the fact that the r.v.s T_Q and $T_{P_n}, n \in \mathcal{N}$ are defined on the same probability space which allows to simplify the usual arguments for a result of this type.

Therefore it would be interesting to have a simultaneous representation result as in (14) in more general spaces in order to obtain strong consistency for this kind of functionals in these spaces. Regrettably, it is well known that this representation does not exist even in the two-dimensional case. In [26] it is shown that certain families of probability distributions (distributions with the same dependence structure) admit a simultaneous representation with respect to c_2-costs but this is not enough to apply the arguments given above.

To avoid this problem in [23] the so-called Skorohod-Lebesgue spaces were introduced. These spaces can be considered as a general (and minimal) framework to develop the previous scheme in abstract spaces.

The idea is the following. As stated in (14), in the one-dimensional case, the quantile function provides simultaneous representations for OTPs and, by Proposition 2.9 also gives a simultaneous Skorohod representation for weak convergence. In [13] it is shown that Proposition 2.9 can be generalized to separable Banach spaces. Then for a separable Banach space U there exists a U-valued r.v. X, such that for any probability measure Q on U, there exists a fixed function T_Q with $T_Q(X) \stackrel{d}{=} Q$ and such that the weak convergence of $\{P_n\}$ to P implies

$$T_{P_n}(X) \to T_P(X), \text{ a.s.}$$

These simultaneous Skorohod representations are not uniquely defined. But, once one of them is fixed, a distance between P_1 and P_2 can be defined by

$$SL_r(P_1, P_2) := E[\|T_{P_1}(X) - T_{P_2}(X)\|^r]^{r^*}, r^* = \min(1, 1/r),$$

for all probability measures P_1 and P_2 on U with $\int \|x\|^r dP_i < \infty$, $i = 1, 2$. If we denote by $M_r(U)$ the family of all probability measures on U with finite expectation of the r-th power, then $(M_r(U), SL_r)$ is a separable metric space which is called the Skorohod-Lebesgue space of order r. The d_r metric (which is the C_r-Kantorovich functional) is topologically equivalent to SL_r (see [23]) and with this construction the arguments which were sketched for the one-dimensional case can be carried out exactly in the same way to prove the strong a.s. continuity of functionals of the form $H(P_n^\omega)$. This scheme includes the results related to almost sure stability of sums of r.v.s mentioned in the preceding section. A new application still to be developed is the fact that Skorohod-Lebesgue spaces provide a common framework for the comparison of all d_r-distances.

While in the one-dimensional case the Skorohod representation problem and the OTP share the same solution namely the quantile function, for higher dimensions the solution is not be the same. This can be seen from the fact, that C_r-functionals do not admit simultaneous representations while Skorohod representations can be chosen simultaneously. In [93] the conditions under which OTPs can be used to obtain Skorohod representations are analyzed. In other words, if $(X, T_n(X))$ is an OTP between P and $Q_n, n = 0, 1, \dots$ and if $\{Q_n\}$ converges weakly to P, when does $\{T_n\}$ converge to the identity?. This problem is solved in the case of a separable Hilbert space U with the cost function c_2. If the space is finite dimensional, then $\{T_n\}$ converges almost surely to the identity without additional assumptions. However, if the dimension is infinite, then this convergence is only in probability and a counter-example for the a.s. convergence is readily constructed. Theorem 2.8 generalizes this result in the finite dimensional case. This result has been extended in [42].

2.4 Rate of convergence in the Central Limit Theorem

An interesting application of the minimal l_p-metrics, defined as solutions of the transportation problems with respect to $c(x, y) = d^p(x, y)$ (cf. the introduction to section 2), are to general forms of the Central Limit Theorem and to the rate of convergence related to it. Recently some versions of these results have been found for martingales with values in separable Banach spaces U (cf. [73]).

We explain the idea at first in the independent case. Let (X_i) be iid, U-valued random variables and define

$$Z_n = n^{-1/\alpha} \sum_{i=1}^{n} X_i \tag{16}$$

the normalized sequence, assuming that X_i is centered at zero. Let ϑ be a (symmetric) α-stable random variable with values in U, i.e.

$$n^{-1/\alpha} \sum_{i=1}^{n} \vartheta_i \stackrel{d}{=} \vartheta, \vartheta \stackrel{d}{=} -\vartheta,$$

where (ϑ_i) are iid copies of ϑ. We considere the convergence of Z_n to ϑ with respect to the Kantorovich metric (the minimal l_1-metric identical to d_1):

$$l_1(P_1, P_2) = \inf \left\{ \int \|x - y\| dP(x, y); P \in M(P_1, P_2) \right\}. \tag{17}$$

To formulate a rate of convergence theorem for Z_n, we introduce the following smoothed (of order r) version of l_1:

$$\overline{l_r}(P_1, P_2) := \sup_{h>0} h^{r-1} l_1(X + h\vartheta, Y + h\vartheta) \tag{18}$$

and, similarly, for the total variation metric σ, we define the smoothed metric:

$$\overline{\sigma_r}(P_1, P_2) = \sup_{h>0} h^r \sigma(X + h\vartheta, Y + h\vartheta) \tag{19}$$

where $\sigma(X, Y) := \sigma(P^X, P^Y)$; in (18) and (19) ϑ is assumed to be independent of X and Y.

For the rate of convergence result we need the finiteness of the following distances:

$$l_1 := l_1(X_1, \vartheta), \overline{l_r} := \overline{l_r}(X_1, \vartheta)$$

$\sigma := \sigma(X_1, \vartheta)$ and $\overline{\sigma_r} := \overline{\sigma_r}(X_1, \vartheta)$. We have the following theorem describing the estimates of right orderunder the above conditions.

Theorem 2.10 *(Stable Limit Theorem - Rate of Convergence)*
Let $1 \leq \alpha \leq 2$ and assume that

a) $E\|\vartheta\| < \infty$ and

b) $l_1 + \overline{l_r} + \overline{\sigma_1} + \overline{\sigma_r} < \infty$ for some $r > \alpha$, then

$$l_1(Z_n, \vartheta) \leq C\left(\overline{l_r} n^{1-r/\alpha} + \tau_r n^{-1/\alpha}\right) \tag{20}$$

where $\tau_r := \max(l_1, \overline{\sigma_1}, \overline{\sigma_r}^{\frac{1}{r-\alpha}})$.

The proof of Theorem 2.10 is based on a generalization of the Bergström convolution method. It uses essentially the ideality properties of the metrics $\overline{l_r}, \overline{\sigma_r}$, e.g. $\overline{l_r}(X + Z, Y + Z) \leq \overline{l_r}(X, Y)$ for Z independent of (X, Y) and $\overline{l_r}(\alpha X, \alpha Y) = \alpha^r \overline{l_r}(X, Y)$ for $\alpha > 0$; also by definition $l_r(X, Y) \geq h^{r-1} l_1(X + h\vartheta, Y + h\vartheta)$. Furthermore, basic ingredients of the proof are the following smoothing inequalities:

$$l_1(X, Y) \leq l_1(X + \varepsilon\vartheta, Y + \varepsilon\vartheta) + 2\varepsilon E\|\vartheta\| \tag{21}$$

and for X, Y, Z, W independent

$$l_1(X + Z, Y + Z) \leq l_1(Z, W)\sigma(X, Y) + l_1(X + W, Y + W). \tag{22}$$

These properties, together with $m = [\frac{n}{2}]$, yield the following decomposition

$$\begin{aligned} l_1(Z_n + \vartheta) &\leq l_1(Z_n + \varepsilon\vartheta, \vartheta_1 + \varepsilon\vartheta) + C \cdot \varepsilon \\ &\leq l_1\left(Z_n + \varepsilon\vartheta, \frac{\vartheta_1 + X_1 + \ldots + X_n}{n^{1/\alpha}} + \varepsilon\vartheta\right) \\ &\quad + \sum_{i=1}^{m} l_1\left(\frac{\vartheta_1 + \ldots + \vartheta_r + X_{j+1} + \ldots + X_n}{n^{1/\alpha}} + \varepsilon\vartheta,\right. \\ &\quad \left.\frac{\vartheta_1 + \ldots + \vartheta_{j+1} + \ldots + X_n}{n^{1/\alpha}} + \varepsilon\vartheta\right) \\ &\quad + l_1\left(\frac{\vartheta_1 + \ldots + \varepsilon_{m+1} + X_{m+2} + \ldots + X_n}{n^{1/\alpha}} + \varepsilon\vartheta, \vartheta_1 + \varepsilon\vartheta\right). \end{aligned}$$

The terms in this expression can be estimated by the metric properties above and by using induction on the number of terms (for details cf. [73]). The finiteness condition has been established for several examples including certain stochastic processes.

A similar approximation result has been given for martingales in [73] where the quantities in the bounds are replaced by distances involving the conditional distributions, as e.g.

$$\overline{\tau_r} = \sup_j E\overline{l_r}\left(P_{X_j|\mathcal{F}_{j-1}}, P_{\vartheta_j}\right), \tag{23}$$

where $(X_j, \mathcal{F}_j)$ is the martingale. For the proof, it is necessary to introduce $\mathcal{G}$-dependence metrics defined by

$$\mu(X, Y/\mathcal{G}) = \sup_{V\in\mathcal{G}} \mu(X+V, Y+V), \tag{24}$$

where μ is a metric and the supremum is over all $\mathcal{G}$ measurable r.v.s V and to study the smoothing versions and the regularity properties of these metrics. In the one dimensional case one obtains as a consequence for the first time a rate of convergence result for martingales with respect to the Prohorov distance.

2.5 Convergence of algorithms

The main approaches to the asymptotic analysis of algorithms in the literature deal with transformation methods (moment generating functions, Mellin transformations, etc.), the martingale method, the method of branching processes and, for a more restricted class of stochastic algorithms, the method based on stochastic approximations. The analysis of algorithms is an important application of stochastics in computer science which poses difficult questions and problems; it has also led to some new developments in stochastic theory (cf. [4] and the references therein).

Based on the properties of minimal metrics introduced at the beginning of this chapter, a promising new method for asymptotic analysis has recently been introduced. In [82] Rösler gave an asymptotic analysis of the quicksort algorithm based on the minimal l_p-metric. His proof has been generalized by Rachev and Rüschendorf [72] to a general "contraction method" with a wide range of possible applications. A series of examples and further developments of the method may be found in some recent work [16], [17].

The contraction method (in its basic form) uses the following sequence of steps:

1. Find the correct normalization of the algorithms. (Typically by studying the first moments or tails.)

2. Determine the recursion for the normalized algorithm.

3. Determine the limiting form of the normalized algorithms. The limiting equation is typically defined via a transformation T on the set of probability measures.

4. Choose an ideal metric μ such that T has good contraction properties with respect to μ. This ideal metric has to reflect the structure of the algorithm.

It also has to have good bounds in terms of interpretable other metrics, and must allow the estimation of bounds (in terms of moments usually). As a consequence one obtains the following result.

5. The conjectured limiting distribution is the unique fixed point of T. Finally one should ensure that the recursion is stable enough for the contraction in the limit to be made use of in order to establish contraction properties of the recursion itself for $n \to \infty$. This is technically the most involved step in the analysis.

6. Establish convergence of the algorithm to the fixed point.

Applications of this method to several sorting algorithms, to the communication resolution interval (CRI) algorithm, to generalized branching type algorithms, to bootstrap estimators, to iterated function systems and to learning algorithms have been given, as well as to others algorithms.

We explain the contraction method for the example of the quicksort algorithm (cf. also [82]). The defining recursion is given by

$$L_n \stackrel{d}{=} n - 1 + L_{I_n} + \overline{L}_{n-I_n} \tag{25}$$

where I_n is uniformly distributed on $\{1, \ldots, n\}$, L_n is the number of steps needed by the quicksort algorithm to sort n numbers and $\overline{L_n}$ is an independent copy of L_n. The randomness in this problem arises from the assumption that the order of the numbers is uniform on the set of all permutations. A number is picked up randomly, all other $n-1$ elements are compared with this number and are divided into two groups, the group of smaller elements and the group of larger elements.

It is easy to establish the asymptotics of the mean $l_n = EL_n$,

$$l_n = 2n \log n + n(2\gamma - 4) + 2 \log n + 2\gamma + 1 + o(1)$$

where γ is Euler's constant. Also it can be seen that $\mathrm{Var}(L_n) = cn + o(n)$. Define the normalization

$$Y_n = \frac{L_n - l_n}{n} \stackrel{d}{=} \frac{I_n - 1}{n} Y_{I_n - 1} + \frac{n - I_n}{n} \overline{Y}_{n-I_n} + c_n(I_n) \tag{26}$$

with $c_n(j) = \frac{n-1}{n} + \frac{1}{n}(l_{j-1} + l_{n-j} - l_n)$. Then taking $c(x) := 2x \log x + 2(1-x)\log(1-x) + 1$, the following inequlity holds

$$\sup_{x \in (0,1)} |c_n([nx]) - c(x)| \le \frac{4}{n} \log n + 0\left(\frac{1}{n}\right).$$

Since $\frac{I_n}{n} \to \tau$, a random variable uniformly distributed on $(0,1)$ one obtains the limiting fixed point equation:

$$Y \stackrel{d}{=} \tau Y + (1-\tau)\overline{Y} + c(\tau). \tag{27}$$

The right hand side of (27) defines the transformation T on the set of all distributions (with finite p-th moments and expectation zero). It is easy to establish that T is a contraction with respect to the minimal l_p-metric, with contraction factor smaller that 1.

One can readily prove that

$$l_p(Y_n, Y) \to 0 \tag{28}$$

where Y is the unique solution of (27) with finite p-th moment.

The fixed point equation is not so easy to analyze and an exact solution of it is still unknown. But it was recently found in [16] that an extremely good approximation to the distribution of Y can be found in the class of lognormal distributions. The following simulation from [16] shows that the smoothed empirical density of a simulation and the lognormal fit show hardly any difference, so that the lognormal can be used in practice.

Figure 1
Quicksort: Lognormal approximation and smoothed empirical density for $n = 5000$. The two curves show hardly any difference.

2.6 Numerical approximation of stochastic differential equations

In this section we present numerical solutions of a multi-dimensional stochastic differential equation (SDE) following the results in [35], [36] and applying them in econometric models for asset returns. The method consists in determining the drift and diffusion coefficients at grid points, and then to combine the time discretization of the SDE with the discretization of the stochastic input-in our case the Wiener process. We start with the description of the grid, which as we shall see will give us an "almost" optimal approximation of the SDE.

On the interval $[t_0, T]$ define an equidistant grid H with points $t_0 = \widehat{t}_0 < \widehat{t}_1 < \cdots < \widehat{t}_{\widehat{n}} = T$ with step size $\widehat{h}$. H will be the minimal set of time points at which values are available for the method, and $\widehat{h}$ will be the period between two neighbouring observations in the past which influence the present drift and diffusion coefficients at any time. For any $t \in [t_0, T]$ we define $i_H(t) := \max\{i : \widehat{t}_i \leq t\}$ as the number of time steps $\widehat{t}$ one can go back into the past from t. As we shall see this is a standard framework in the so called ARCH (GARCH) modeling of asset returns. We consider a SDE where the drift and diffusion depend on the present and the past states:

$$
\begin{aligned}
(E) \qquad x(t) - x_0 &= \int_{t_0}^{t} b(x,s)ds + \int_{t_0}^{t} \sigma(x,s)dw(s) \\
&= \int_{t_0}^{t} b(x,s)ds + \sum_{j=1}^{q} \int_{t_0}^{t} \sigma_j(x,s)dw_j(s), \\
& \qquad t \in [t_0, T], x_0 \in \mathbb{R}^d.
\end{aligned}
$$

Here, $w = (w_1, \ldots, w_q)^T$ is a q-dimensional standard Brownian motion, and we use the notations

$$
\begin{aligned}
b(x,s) &:= b^{i_H(s)}(x(s), x(s-\widehat{h}), x(s-2\widehat{h}), \ldots, x(s - i_H(s)\widehat{h})), \\
\sigma(x,s) &= (\sigma_1(x,s), \ldots, \sigma_q(x,s)) \\
&:= \sigma^{i_H(s)}(x(s), x(s-\widehat{h}), x(s-2\widehat{h}), \ldots, x(s - i_H(s)\widehat{h}))
\end{aligned}
$$

with $b^\nu \in C(\mathbb{R}^{\nu+1)d}; \mathbb{R}^d)$ and $\sigma^\nu \in C(\mathbb{R}^{\nu+1)d}; L(\mathbb{R}^q; \mathbb{R}^d)), \nu = 0, \ldots, i_H(T)$, where $\sigma_j^\nu \in C(\mathbb{R}^{\nu+1)d}; \mathbb{R}^d), j = 1, \ldots, q$, denote the columns of the matrix function $\sigma^\nu = (\sigma_1^\nu, \ldots, \sigma_q^\nu)$. As usual, we denote by C spaces of continuous functions, by L spaces of linear mappings, and by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^n (n \in \mathbb{N})$ and the corresponding induced norm on L.

For a random variable ζ on a probability space (Ω, A, P) with values in a separable metric space (X, d) with the Borel σ-algebra $B(X)$, the notation $D(\zeta)$ mean the distribution $P \circ \zeta^{-1}$ of ζ. $P(X)$ is the set of all Borel probability measures (probabilities) on X.

For $p \in [1, \infty)$ we define on the set

$$
M_p(X) := \left\{ \mu \in P(X) : \int_X d(x,\theta)^p d\mu(x) < \infty, \theta \in X \right\}
$$

a metric W_p by

$$W_p(\mu,\nu) := \inf\left[\int_{X\times X} d(x,y)^p d\eta(x,y)\right]^{1/p} \quad (\mu,\nu\in M_p(X))$$

where the infimum is taken over all measures $\eta\in P(X\times X)$ with marginal distributions μ and ν. $W_p = d_p$ is the L_p-Wasserstein metric or L^p-Kantorovich metric (see (2)). We shall state the convergence results for a sequence of approximations to the solution x of (E) in terms of W_p, which is the "ideal" metric for this type of approximation problems.

The approximate solution of x in (E) can be viewed as a framework for studying asset pricing models known in ecometrics literature as *Autoregressive Conditional Heteroscedasticity (ARCH) or Generalized ARCH (GARCH)* models. We give a brief description of these models:

Consider an equidistant grid $t_0 = \widehat{t}_0 < \widehat{t}_1 < \cdots < \widehat{t}_{\widehat{n}} = T$ with step size $\widehat{h}$ on the time interval $[t_0, T]$. A univariate ARCH model is defined as a discrete time stochastic process $(\varepsilon_{\widehat{t}_i})i = 0,\ldots,\widehat{n}$ of the form

$$\varepsilon_{\widehat{t}_{i+1}} = \widehat{\sigma}_{\widehat{t}_i}\delta_{\widehat{t}_i}$$

where $\widehat{\sigma}_{\widehat{t}_i}$ is a positive measurable function of the time points $\widehat{t}_0,\widehat{t}_1,\ldots,\widehat{t}_i$ and the $\delta_{\widehat{t}_i}$ are i.i.d. r.v.s with zero mean and variance one. In a *linear ARCH* (ψ) the variances $\sigma_{\widehat{t}_i}$ depend on the squares of the past ψ values of the process:

$$\widehat{\sigma}^2_{\widehat{t}_i} := \omega + \sum_{r=0}^{\psi-1}\alpha_r\varepsilon^2_{\widehat{t}_{i-r}}$$

whereas in the more general *linear GARCH* (ϕ,ψ) they may also depend on the ϕ recent variances:

$$\widehat{\sigma}^2_{\widehat{t}_i} := \omega + \sum_{r=0}^{\psi-1}\alpha_r\varepsilon^2_{\widehat{t}_{i-r}} + \sum_{r=1}^{\phi}\beta_r\widehat{\sigma}^2_{\widehat{t}_{i-r}}.$$

In these models it is assumed that $\omega > 0, \alpha_r \geq 0, \beta_r \geq 0$ for all r. One can embed these models into the constructed approximation for the SDE (E). (see [36]).

We need the following general assumptions concerning (E):

(A1) There exists a constant $M > 0$ such that
for all $j = 1,\ldots,q, \nu = 0,\ldots,i_H(T)$ and $x_0,\ldots,x_\nu\in \mathbb{R}^d$
$\|b^\nu(x_0,\ldots,x_\nu)\| \leq M(1+\max\|x_\rho\|)$ and
$\|\sigma_j^\nu(x_0,\ldots,x_\nu)\| \leq M.$

(A2) There exists a contant $L > 0$ such that
for all $j = 1,\ldots,q, \nu = 0,\ldots,i_H(T)$ and $x_0,\ldots,x_\nu, y_0,\ldots,y_\nu\in \mathbb{R}^d$
$\|b^\nu(x_0,\ldots,x_\nu) - b^\nu(y_0,\ldots,y_\nu)\| \leq L\max_{0\leq\rho\leq\nu}\|x_\rho - y_p\|,$ and
$\|\sigma_j^\nu(x_0,\ldots,x_\nu) - \sigma_j^\nu(y_0,\ldots,y_\nu)\| \leq L\max_{0\leq\rho\leq\nu}\|x_\rho - y_p\|.$

(A1) and (A2) assure the existence and uniqueness of the solution of (E). The boundedness of σ_j in (A1) seems to be essential for the proof of Theorem 2.11.

As mentioned above, the approximate solutions are based on a "double grid" – a *coarse* grid for the time discretization and a *fine* grid, for the chance discretizaiton which yields a lower convergence speed than the time discretization. In fact, we consider a grid class $G(m, \alpha, \beta)$. Here $m : (0, T - t_0] \to [1, \infty)$ is a monotone decreasing function, and $\alpha, \beta > 0$ are constants. Then each element G of $G(m, \alpha, \beta)$ consists of two kinds of grid points:

(i) the time discretization points $t_k, k = 0, \ldots, n$, with $t_0 < t_1 < \cdots < t_n = T$ and

(ii) the chance discretization points $u_i^k, i = 0, \ldots, m_k, k = 0, \ldots, n-1$ with $t_k = u_0^k < u_1^k < \cdots < u_{m_k}^k = t_{k+1}, k = 0, \ldots, n-1$.

Now G is required to satisfy the following assumptions:

(G1) $t_k - t_{k-1} = \frac{T-t_0}{n} =: h \leq 1$ for all $k = 1, \ldots, n$ and $\widehat{h}/h \in \mathbb{N}$,

(G2) $1 \leq m_k \leq m(h)^\alpha$ for all $k = 0, \ldots, n-1$,

(G3) $u_i^k - u_{i-1}^k = \frac{h}{m_k} \leq \beta \frac{h}{m(h)}$ for all $k = 0, \ldots, n-1, i = 1, \ldots, m_k$.

Here (G1) means that the coarse grid is equidistant with step size h and contains the master grid H. (G2) ad (G3) say that each interval of the coarse subgrid is subdivided in an equidistant way by the points u_i^k, both the number of the subdivisions and the step size of the full grid being bounded by functions of h.

For a grid G of $G(m, \alpha, \beta)$ we define

$$[t]_G := t_k \text{ and } i_G(t) := k, \quad \text{if } t \in [t_k, t_{k+1}), k = 0, \ldots, n-1, \quad \text{and}$$

$$[t]_G^* := u_i^k \quad \text{if } t \in [u_i^k, u_{i+1}^k), i = 0, \ldots m_k - 1, k = 0, \ldots, n-1.$$

We construct the approximate solution of (E) in three steps. The first step is a pure time discretization. (Here only the coarse subgrid is involved.)

(E1) $y^E(t) = x_0 + \int_{t_0}^t b(y^E, [s]_G) ds + \sum_{j=1}^q \int_{t_0}^t \sigma_y(y^E, [s]_G) dw_j(s), t \in [t_0, T]$.

In the second step, a continuous and piecewise linear interpolation of the trajectories in (E1) between the points of the whole fine grid yields the method (E2):

(E2) $\widehat{y}^E$ is continuous, and linear in the intervals $(u_{i-1}^k, u_i^k]$, $i = 1, \ldots, m_k$, $k = 0, \ldots, n-1$, with $\widehat{y}^E(u_i^k) = y^E(u_i^k), i = 0, \ldots, m_k, k = 0, \ldots, n-1$.

In the third step, the Wiener process increments over the fine grid are replaced by other i.i.d. r.v.s: Let $\mu \in P(\mathbb{R})$ be a measure with mean value 0 and variance 1, and let

$$\{\xi_{js}^k : j = 1, \ldots, q; s = 1, \ldots, m_k; k = 0, \ldots, n-1\}$$

be a family of i.i.d. r.v.s with distribution $D(\xi_{11}^0) = \mu$. Then we can define the following method (E3) yielding continuous trajectories which are linear between neighbouring grid points:

$$\text{(E3)} \quad \begin{aligned} z^E(u_0^0) &= x_0, \quad \text{and} \\ z^E(u_i^k) &= x_0 + \textstyle\sum_{r=0}^{k-1} hb(z^E, t_r) + h\frac{i}{m_k} b(z^E, t_k) \\ &\quad + \textstyle\sum_{j=1}^{q} \left[\sum_{r=0}^{k-1} \sqrt{\frac{h}{m_r}} \sigma_j(z^E, t_r) \sum_{s=1}^{m_r} \xi_{js}^r + \sqrt{\frac{h}{m_k}} \sigma_j(z^E, t_k) \sum_{s=1}^{i} \xi_{js}^k \right] \\ &\quad \text{for all } i = 1, \ldots, m_k, k = 0, \ldots, n-1. \end{aligned}$$

For this last step, the Wiener process w and the r.v.s ξ_{ji}^k will have to be defined anew on a *common* probability space.

According to the evolution of the method (E3) via (E1) and (E2), each step will be represented by one convergence theorem, yielding then immediately the main result given in terms of the W_p-metric.

Theorem 2.11 *Suppose (A1) and (A2) hold. Suppose also that* $p \in [1, \infty)$ *and* $\mu \in P(\mathbb{R})$ *has the properties:*

$$\int_{-\infty}^{\infty} x d\mu(x) = 0, \int_{-\infty}^{\infty} x^2 d\mu(x) = 1$$

and

$$\int_{-\infty}^{\infty} e^{tx} d\mu(x) < \infty \text{ for all } t \text{ with } |t| \leq \tau, \tau > 0.$$

Moreover, let $(w(t))_{t\in[t_0,T]}$ *be a* q*-dimensional standard Wiener process and* $\{\xi_{ji}^k : j = 1, \ldots, q; i = 1, \ldots, m_k; k = 0, \ldots, n-1\}$ *a set of i.i.d. r.v.s with distribution* $D(\xi_{11}^0) = \mu$.

Then for the solution x *of the SDE and its numerical analog (E3), we have the following rate-of-convergence result:*

$$W_p(D(x), D(z^E)) \leq C\{h^{1/2} + \frac{1 + \ln m(h)}{\sqrt{m}(h)}\};$$

where C *is an absolute constant.*

The bound in Theorem 2.11 gives convergence rates with respect to h for the method (E3) and for any grid sequence in $G(m, \alpha, \beta)$. These rates consist of two summands, one depending on h and the other depending on $m(h)$, representing the rates of time and chance discretization, respectively. Obviously, it is not desirable that one of both summands converges faster than the other for this would only increase the costs in relation to the effect. Namely, if the second summand converged faster than the first, this would mean that $m(h)$ increases too fast and consequently

– because of (G3) – to have too small step sizes of the whole *fine* grid, i.e. to have too many points u_i^k in relation to the t_k in each grid and, therefore, to use a random number generator too often. If the first summand converged faster than the second, then $m(h)$ would increase too slowly, i.e. the intervals $[t_k, t_{k+1}]$ would not have enough intermediate grid points u_k^k, so that the chance discretization would not keep up with the time discretization. Therefore, it is desirable to tune the rates of both summands, i.e. to equal the powers of h in both summands. This means to choose $m(h)$ to be increasing like $1/h$.

Theorem 2.12 *Under the assumptions in Theorem 2.11 and with* $\max\{\sup_{0<s\leq 1} sm(s), \sup_{0<s\leq 1} \frac{1}{sm(s)}\} \leq K$ *we have:*

$$W_p(D(x), D(z^E)) \leq C \cdot h^{1/2}(1 - \ln h).$$

This result is almost optimal; the right order-bound should be $h^{1/2}$, see the discussion in [36].

3 Mass-transshipment problems

3.1 Dual representation and topological properties

In this section we shall study the Kantorovich-Rubinstein mass-transshipment problem. This can be stated as follows: *Given a topological space U, a Radon measure ρ with total mass* 0 *on U, and a cost function c: $U \times U \to \overline{\Re}$, it is required to find the minimum*

$$\overset{\circ}{C}_c(\rho) := \min \int_{U\times U} c(x,y)dQ(x,y) \tag{29}$$

over the set $\mathcal{D}(\rho)$ of finite nonnegative Borel measures Q on the product $U \times U$, subject to the balancing condition $\pi_1 Q - \pi_2 Q = \rho$, i.e.

$$Q(B \times U) - Q(U \times B) = \rho(B) \text{ for all Borel sets } B \subset U.$$

Recall that a finite Borel measure ρ on S is called a Radon measure if it is inner regular, i.e. $\rho(B) = \sup P(C)$ where the supremum is taken over all compact sets $C \subset B$. For any probabilities P_1 and P_2 on U with $\rho := P_1 - P_2$

$$\overset{\circ}{C}_c(P_1, P_2) := \overset{\circ}{C}_c(\rho)$$

is called the Kantorovich-Rubinstein functional. A relation between the Kantorovich functional

$$C_c(P_1, P_2) = \min\left\{\int_{U\times U} c(x,y)dQ(x,y) : \pi_1 Q = P_1, \pi_2 Q = P_2\right\}, \tag{30}$$

and the Kantorovich-Rubinstein functional $\overset{\circ}{C}_c$ can be obtained in the following way. For a symmetric cost function $c(x,y) \geq 0$, define the reduced cost function

$$\widetilde{c}(x,y) = \inf\left\{\sum_{i=1}^{n-1} c(x_i, x_{i+1}); n \in \mathbb{N}, x_i \in U, x_1 = x, x_n = y\right\}; \tag{31}$$

$\widetilde{c}(x,y)$ is the minimal cost of a transshipment from x to y which is carried out in several steps. Obviously, $\widetilde{c}(x,y) \leq c(x,y)$ and $\widetilde{c}$ satisfies the triangle inequality: $\widetilde{c}(x,y) \leq \widetilde{c}(x,z) + \widetilde{c}(z,y)$ for all $x,y,z \in U$. Furthermore, $\widetilde{c}$ is a (semi-)metric and it is obviously the largest (semi-)metric dominated by c. By a slightly modified form of Theorem 2.2 for the case of a semi-metric cost function, $C_{\widetilde{c}}$ admits a dual representation in the form of the Kantorovich metric

$$\begin{aligned} C_{\widetilde{c}}(P_1,P_2) &= \sup\Big\{\int_U f d(P_1 - P_2) : f : U \to R \text{ bounded and continuous, and} \\ &\qquad f(x) - f(y) \leq \widetilde{c}(x,y), \forall x, y \in U\Big\}. \end{aligned}$$

Moreover, as we shall see later in this section $\overset{\circ}{C}_c$ has the same dual representation (under some regularity conditions on U and c) implying

$$\begin{aligned} \overset{\circ}{C}_c(P_1,P_2) &= \sup\left\{\int f d(P_1 - P_2) : f(x) - f(y) \leq c(x,y), \forall x, y\right\} \\ &= \sup\left\{\int f d(P_1 - P_2) : f(x) - f(y) \leq \widetilde{c}(x,y), \forall x, y\right\} \\ &= \sup\left\{\Big|\int f d(P_1 - P_2)\Big| : \Big|f(x) - f(y)\Big| \leq \widetilde{c}(x,y), \forall x, y\right\} \\ &= C_{\widetilde{c}}(P_1,P_2). \end{aligned} \tag{32}$$

This gives a natural explanation of the relevance of $\overset{\circ}{C}_c$ for transportation problems. A somewhat different interpretation of $\overset{\circ}{C}_c$ can be found in Kemperman (1983) (multistage shipping). In linear programming the discrete analog is known as the network flow problem. Kantorovich and Rubinstein (1957) studied a modification with exactly n-stages of transportation. In terms of r.v.s we may also give the following representation:

$$\begin{aligned} &\overset{\circ}{C}_{\widetilde{c}}(P_1,P_2) = C_c(P_1,P_2) \\ &= \inf\{E\widetilde{c}(X_1,X_2)\ \forall \text{pairs of r.v.s } (X_1,X_2) \text{ with marginals } P_1 = P_{X_1} \text{ and } P_2 = P_{X_2}\} \\ &= \inf\{E[c(X_1,X_2) + c(X_2,X_3) + \ldots + c(X_{n-1},X_n)] :\ \forall \text{ r.v.s } X_1,\ldots,X_n \text{ with} \\ &\qquad P_{X_1} = P_1, P_{X_n} = P_2 \text{ and } P_{X_i} \text{ arbitrary for } 2 \leq i \leq n-1\} \\ &= \inf\left\{\int c(x,y) Q(dx,dy); Q \in \mathcal{D}(P_1 - P_2)\right\}, \end{aligned} \tag{33}$$

where $Q = \sum_{i=1}^{n-1} P_{X_i, X_{i+1}}$. Obviously $\overset{\circ}{C}_c$ is a semi-metric on $\mathcal{P}(U)$ if c is symmetric. If $c(x,y) = d^p(x,y)$, $p > 1$, $U = \Re^k$, then $\widetilde{c}(x,y) = 0$ and, therefore, $\overset{\circ}{C}_c (P_1, P_2) = 0$. This indicates a striking difference between C_c and $\overset{\circ}{C}_c$. The duality theorem for the Kantorovich-Rubinstein problem

$$\overset{\circ}{C}_c (P_1, P_2) = \\ \sup\left\{ \int f d(P_1 - P_2); f : U \to \Re^1 f(x) - f(y) \le c(x,y), \forall x, y \in S \right\} \quad (34)$$

has been proved by Kantorovich in [45] for the case that S is compact and c is continuous. In [53] this result is proved for the case that U is homeomorphic to a Baire subset of a compact space, $c : U \times U \to (-\infty, \infty]$ is bounded from below, for all α the sets $\{(x,y) \in U \times U; c(x,y) \le \alpha\}$ are analytic, i.e. are the projection of a Borel set in $(U \times U) \times Y$ for a Polish space Y and $\overset{\circ}{C}_c = \lim_{N\to\infty} C^{\circ}_{c\wedge N}$.

Finally, a strengthened version of the duality theorem for symmetric, nonnegative cost functions $c(x,y)$ on a separable metric space (U,d) is given in [77] under the following boundedness and continuity conditions:

C.1 $c(x,y) = 0$ if $x = y$,

C.2 $c(x,y) \le \lambda(x) + \lambda(y), \forall x, y$, for some $\lambda : U \to \Re_+$ mapping bounded sets into bounded sets,

C.3 $\sup\{c(x,y); x, y \in B_\epsilon(a), d(x,y) \le \delta\} \to 0$ as $\delta \to 0$ for each $a \in U$, $B_\epsilon(a)$ the ϵ-ball with center a.

Defining

$$\|f\|_c := \sup\left\{ \frac{|f(x) - f(y)|}{c(x,y)}; x \ne y \right\}$$

for $f : U \to \Re$, the following strengthened representation holds:

$$\begin{aligned} \overset{\circ}{C}_c (P_1, P_2) &= \sup\left\{ \left| \int f d(P_1 - P_2) \right|; \|f\|_c \le 1 \right\} \\ &= \sup\left\{ \int f d(P_1 - P_2); f(x) - f(y) \le c(x,y), \forall x, y \right\}, \end{aligned} \quad (35)$$

assuming $\int |x| dP_i(x) < \infty, i = 1, 2$. While obviously in general

$$\overset{\circ}{C}_c (P_1, P_2) \le C_c(P_1, P_2),$$

it follows that for $c(x,y) = d(x,y)$, $C_d(P_1, P_2) = \overset{\circ}{C}_d (P_1, P_2)$.

The cost function

$$c_p(x,y) = d(x,y) \max\left[1, |y - a|^{p-1}\right], p \ge 1, x, y \in \Re,$$

satisfies C.1 - C.3. From the above dual representation for $\overset{\circ}{C}_c$ one obtains the explicit representation

$$\overset{\circ}{C}_{c_p}(P_1, P_2) = \int_{-\infty}^{\infty} \max\left(1, |x-a|^{p-1}\right) |F_1(x) - F_2(x)| dx, \tag{36}$$

where the F_i are the distribution functions (d.f.s) of P_i. Except for $p = 1$, an optimal measure Q^* satisfying

$$\int c_p(x, y) Q^*(dx, dy) = \overset{\circ}{C}_{c_p}(P_1, P_2)$$

is not known. For $p \geq 1$ $\overset{\circ}{C}_{c_p}$ is identical to the Fortet-Mourier metric (cf. [69])

$$FM_p(P_1, P_2) = \sup\left\{ \left| \int_U f d(P_1 - P_2) \right|; f \in C^p \right\}, \tag{37}$$

where

$$C^p = \left\{ g : U \to \Re^1; \sup_{r \geq 1} r^{1-p} \sup\left\{ \frac{|g(x) - g(y)|}{d(x, y)}, x \neq y, d(x, a) \leq r, d(y, a) \leq r \right\} \leq 1 \right\}.$$

Inequalities between the L_p-minimal metric $l_p = C_{c_p}^{1/p}$, the Fortet-Mourier metric $\overset{\circ}{C}_{c_p}$ and other metrics on $\mathcal{P}(U)$ are studied in [69]. In particular for any $P_0 \in \mathcal{P}(U), \mathcal{D}(P_0, l_p) := \{P \in \mathcal{P}(U); l_p(P, P_0) < \infty\}$ is l_p-complete and for $P_0 = \delta_a$ the following convergence criterion holds in $\mathcal{D}(\delta_a, l_p)$:

$$\begin{aligned} & l_p(P_n, P) \to 0 \Leftrightarrow \overset{\circ}{C}_{c_p}(P_n, P) \to 0 \\ & \Leftrightarrow P_n \to P \text{ (weakly) and } \int d^p(x, y)(P_n - P)(dx) \to 0. \end{aligned} \tag{38}$$

For $P_0 \neq \delta_a$, no corresponding characterization is known.

3.2 Explicit representation

In the one dimensional case the explicit representation of the Kantorovich-Rubinstein functional in (36) has been extended to the following general result in [75].

Theorem 3.1 *Assume that* $c(x, y) = |x - y| \xi(x, y), x, y \in \Re^1$ *and that for* $x < t < y$ $\xi(t, t) \leq \xi(x, y)$ *holds. Furthermore, assume that* $\xi(x, y)$ *is symmetric, continuous on the diagonal and* $t \to \xi(t, t)$ *is locally bounded, then under the conditions of the duality theorem*

$$\overset{\circ}{C}_c(P_1, P_2) = \int \xi(t, t) |F_1(t) - F_2(t)| dt \tag{39}$$

It is interesting to note that the solution (39) depends only on the behaviour of the cost function on the diagonal. There are indications that an exact optimal transshipment plan does not exist for these kind of problems (cf. [10]). In the multivariate case an analogous explicit result has been obtained in [54] for the case of differentiable cost functions.

Theorem 3.2 *Suppose that U is a domain in $\Re^n$ and $c : U \times U \to \Re^1$ a bounded function with analytic level sets $\{c \leq \alpha\}, c(x,x) = 0, \forall x \in U$, c being continuously differentiable in some open neighbourhood of the diagonal. If $\overset{\circ}{C}_c (P_1, P_2) > -\infty$ then*

$$\overset{\circ}{C}_c (P_1, P_2) = \int u_0 d(P_1 - P_2) \tag{40}$$

with $u_0(x) = \int_{\gamma(x_0,x)} \operatorname{grad}_\xi c(\xi,\eta)|_{\eta=\xi} d\xi$, where γ is a piecewise smooth curve from x_0 to x.

Again the optimal value depends only on the gradient $\operatorname{grad}_\xi c(\xi,\eta)$ of the cost function $c = c(\xi,\eta)$ at the diagonal $\eta = \xi$. The differentiability of c at the diagonal is crucial for the derivation of this result. It excludes the important case $c(x,y) = \|x - y\|$.

In [54] the following upper bound for the transportation cost has been found.

Theorem 3.3 *Let $c_p(x,y) = \|x-y\|_p = (\sum |x_i - y_i|^p)^{1/p}$, $x,y \in \Re^n$, then for probability measures P_1, P_2 with Lebesgue densities f, g,*

a)
$$\overset{\circ}{C}_{c_p} (P_1, P_2) \leq \int_{\Re^n} \|y\|_p |I_H(y)| dy \tag{41}$$

holds, with $h := f - g$ and $I_H(y) := \int_\circ^1 t^{-(n+1)} h(\frac{y}{t}) dt$;

b) If there exists a continuous function $g : \Re^n \to \Re^1$ almost everywhere differentiable and satisfying

$$\begin{aligned} \nabla g(y) &= \quad (\operatorname{sgn}\ (y_i I_H(y))) \quad a.s. && \text{for } p = 1 \\ \nabla g(y) &= \left(\operatorname{sgn}\ (y_i I_H(y)) \left(\tfrac{|y_i|}{\|y\|_q} \right)^{q/p} \right) && \text{for } p > 1 \end{aligned} \tag{42}$$

then in (41) the equality holds.

Condition (42) is fulfilled in dimension one. A simple sufficient condition in the case $p = 1$ for (42) is $I_H \geq 0$ a.s. which is a stochastic ordering condition. From the derivation it seems that the bound in (41) in the case $p = 2$ should be much sharper than the classical inequality

$$C^\circ_{c_2}(P_1, P) \leq 4 \int \|y\|^2 |f(y) - g(y)| dy \tag{43}$$

due to Zolotarev [95].

3.3 Applications of the Kantorovich-Rubinstein problem

From the series of applications of the Kantorovich-Rubinstein theorem we single out two recent types, one in the field of mathematical economics and the other in the fields of representations of metrics as minimal metrics.

3.3.1 Utility functions

Let $(S, \leq)$ be a topological space with closed preorder $\leq$, i.e. $\{(x,y); x \leq y\}$ is closed in $S \times S$. Define the strict order relation $x \prec y$ if $x \leq y$ and if not $y \leq x$ and call an isotone function $u : S \to \Re^1$ a utility function if $x \prec y$ implies $u(x) < u(y)$. A fundamental result in mathematical economics due to Debreu asserts the existence of a continuous utility function for a closed, total preorder on separable metrizable space. It is not difficult to show that the assumptions that S is metrizable and separable, and the preorder is closed cannot be abandoned. The following result proved in [53], [55] shows that the restricting assumption that the preorder is total can be omitted for locally compact spaces.

Theorem 3.4 *(Utility Representation)*
Let $(S, \leq)$ be a separable, metrizable locally compact space with a closed semiorder $\leq$, then S admits a continuous utility function.

For the proof, the duality theorem is used to establish the following extension theorem (cf. [53]):

Suppose that S is compact, $F \subset S$ is closed and c is lower semicontinuous, such that $v(x) - v(y) \leq c(x,y)$ for $x, y \in F$ and some $v \in C(F)$. Then, a continuous extension of v to S exists with $v(x) - v(y) \leq c(x,y)$ for $x, y \in S$ if $c'(x,y) := \min\{c(x,y), a_v(x) - b_v(y)\}$ is lower semicontinuous, on $S \times S$, where $a_v(x) := \inf\{v(z) + c(x,z), z \in F\}$, $b_v(x) := \sup\{v(z) - c(x,z); z \in F\}$.

The following parametrized version of this result has also been established in [53].

Theorem 3.5 *If S is metrizable, separable locally compact, Ω is a metrizable topological space and for $\omega \in \Omega$, $\leq_\omega$ is a preorder on S such that $\{(\omega, x, y); x \leq_\omega y\}$ is closed in $\Omega \times S \times S$, then there exists a continuous utility function $u : \Omega \times S \to [0,1]$.*

3.3.2 Minimal representation of metrics

An important property of a probability metric is the possibility of finding a minimal representation of it. Consider for example

$$\kappa(P_1, P_2) = \int_{\Re^1} |F_1(x) - F_2(x)| dx,$$

for the probability measures P_i on $\Re^1$, $i = 1, 2$.

Then κ has the representation

$$\kappa(P_1, P_2) = l_1(P_1, P_2) := \inf\{E|X - Y|, X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2\}$$

as a minimal l_1-metric.

This representation allows us to obtain rate of convergence results in limit theorems for the κ-metric based on the inherent regularity structure of minimal metrics. A further example is the representation of the Fortet-Mourier metric (cf. (37)).

The metric

$$\xi_n(P_1, P_2) := \int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} \frac{(x-t)^{n-1}}{(n-1)!}(P_1 - P_2)(dt) \right| dx$$

was introduced by Zolotarev [95]. It is an ideal metric of order n. For $n = 1$, $\xi_1 = l_1$ but for $n > 1$, ξ_n does not allow a representation as a minimal metric with respect to a Monge-Kantorovich transportation problem. A representation for ξ_n as a minimal metric with respect to a Kantorovich-Rubinstein type problem has, however, recently been found in [68].

For a signed Borel measure m on $\Re^k$ with $m(\Re^k) = 0$, with finite n-th moments and with $\int (x_1 \dots x_k)^j m(dx) = 0, j = 1, \dots, n$, define the signed measure m_n as

$$m_n\left(\prod_{j=1}^{n}(-\infty, x_j]\right) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} \prod_{i=1}^{k} \frac{(x_i - t_i)^n}{n!} m(dt_1, \dots, dt_k)$$

for $x_j \le 0$ and by the corresponding "survival function" for components $x_j \ge 0$.

$$m_n\left(\prod_{j\in J}(-\infty, x_j] \times \prod_{j \in J^c}[x_j, \infty)\right)$$

$$= \int_{(-\infty, x_J]} \int_{[x_{J^c}, \infty)} \prod_{j\in J} \frac{(x_j - t_j)^n}{n!} \prod_{j\in J^c} \frac{(t_j - x_j)^n}{n!} m(dt_1, \dots dt_k)$$

where $x_J = (x_j)_{j\in J} \le 0$ and $x_{J^c} = (x_j)_{j \in J^c} > 0$. Also denote $B_n(m) := \{b; b$ a non-negative Borel measure on $\Re^k \times \Re^k$ with $b(A \times \Re^k) - b(\Re^k \times A) = m_n(A)$ for all $A\}$. Then any $b \in B_n(m)$ has an absolutely continuous marginal difference measure $\triangle b = b(\cdot \times \Re^k) - b(\Re^k \times \cdot)$ with

$$\frac{\partial^{(n-1)k}}{\partial x_1^{(n-1)} \dots \partial x_k^{(n-1)}} p_{\triangle b}(x) = F_m,$$

where $p_{\triangle b}$ is the density of $\triangle b$, and F_m the distribution function of m.

One can now introduce a version of the Kantorovich-Rubinstein norm

$$\|m\|_n := \inf\left\{\int c db; b \in B_n(m)\right\}.$$

The following duality theorem holds (see [68]).

Theorem 3.6 *The norm* $\|m\|_n$ *is given by* $\|m\|_n = \sup\{|\int f dm|; f \in L_n\}$*, where* L_n *is the class of* n*-th integrals* $g_n(x) := \int_0^x \prod \frac{(x_j - t)^{n-1}}{(n-1)!} g(t) dt$ *of Lipschitz functions* g.

In the case $k = 1$ and for the cost function $c(x, y) = |x - y| \max(h(|x - a|), h(|y - a|))$, h an increasing function on $t \ge 0$, $h(t) > 0$, this leads to

$$\|m\|_n = \int_{\Re^1} \left| \int_{-\infty}^{x} \frac{(x-t)^n}{n!} dF_m(t) \right| h(|x - a|) dx.$$

In particular, we obtain a minimal representation for the ξ_n-metric; namely, if $c(x,y) = |x-y|, x, y \in \Re$,

$$\xi_n(P_1, P_2) = \|P_1 - P_2\|_n.$$

If $c(x,y) = |x-y|, x, y \in \Re^k (k \geq 1)$,

$$Z_{k,n}(X,Y) := \|P^X - P^Y\|_n$$

is an ideal metric of order $kn+1$. This dependence of the order of the ideality upon the dimensionality may be considered as a drawback of $Z_{k,n}$. In [40], [41] a different approach was proposed leading to the following dual representation for ideal metrics of order $r > 0$ independent of the dimensionality of $\Re^k$. For a given $\alpha \in \mathbb{N}$, and any $s \geq \alpha$, let M_s° be the set all signed Borel measures μ on $\Re^n$ such that

$$\int_{\Re^n} x_1^{\alpha_1} \cdots x_n^{\alpha_n} d\mu(x_1, \ldots, x_n) = 0 \text{ and } \int |x|^s d|\mu|(x) < \infty$$

for every multiple index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$ such that $\alpha_1 + \ldots + \alpha_n \leq k-1$.

Next, let Γ_μ be the set of signed Borel measures Ψ on $\Re^{2n}$, viewed as "transshipment plans", satisfying the balancing condition

$$\int_{\Re^n} f d\mu = \int_{\Re^{2n}} \triangle_h^k f(x) d\Psi(x,h),$$

where

$$\triangle_h^k f(x) = \sum_{i=0}^{k} (-1)^{k-i} \tbinom{k}{i} f(x+ih)$$

is the k-th difference of f with step h for $x, h \in \Re^n$. Define the following minimal functional on M_r°

$$\|\mu\|_{k,r} = \inf_{\psi \in \Gamma_\mu} \int \|h\|^r d|\Psi|(x,h).$$

To state the dual representation for $\|\mu\|_{k,r}$ let Λ_r^k be the set of all locally bounded functions f on $\Re^n$ such that for some $C \geq 0, |\triangle_h^k f(x)| \leq C\|h\|^r$ over all $x, h \in \Re^n$. Λ_r^k is endowed with the seminorm $\|f\|_{\Lambda_r^k} = \inf C$. In [40], [41] the following duality theorem is proved:

Theorem 3.7 *Let $n \in \mathbb{N}, k \in \mathbb{N}, s = n+k-1$ and $\alpha - 1 < r \leq k$. Then for every $\mu \in M_r^\circ$*

$$\|\mu\|_{k,r} = \sup \left\{ \int_{\Re^n} f d\mu : \|f\|_{\Lambda_r^k} \leq 1 \right\}.$$

Moreover, the above supremum is attained; there is an $f \in \Lambda_r^k$ with $\|f\|_{\Lambda_r^k} = 1$ such that $\|\mu\|_{k,r} = \int f d\mu$.

The minimal functional $\|\mu\|_{k,r}$ defines an ideal metric of order r regardless of the dimensionality on $\Re^n$. In fact, let $K_r(P,Q)$ be the analog :

$$K_r(P,Q) = \sup\left\{ \Big| \int_{\Re^n} f d(P-Q) \Big| : \|f\|_{\Lambda_r^k} \le 1 \right\}.$$

of the Kantorovich metric on $\Re^n$. From Theorem 3.7 one can easily check the following:

(i) K_r is an ideal metric of order r and, for $k-1 < r \le k$, $\zeta_r \le c_1 K_r \le c_2 \zeta_r$ for some positive constants c_1 and c_2;

(ii) if $P - Q \in M^\circ_{n+k-1}$, then $K_r(P,Q)$ admits the dual representation

$$K_r(P,Q) = \|P-Q\|_{k,r}$$

and moreover,

$$K_r(P,Q) \le A \int_{\Re^n} \|x\|^r d|P-Q|(x) < \infty.$$

3.3.3 Stability of stochastic programs

In this part we study the stability of the following stochastic optimization problem:

$$P(\mu): \qquad \min\left\{ \int_{\mathbb{R}^s} f(x,z)\mu(dz) : x \in C \right\}$$

where $f : \mathbb{R}^m \times \mathbb{R}^s \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ is a normal integrand (i.e. $f(\cdot, z)$ is lower semicontinuous, for all $z \in \mathbb{R}^s$, and f is Borel measurable), $f(\cdot, z)$ is continuous on $C, \forall z \in \mu, C \subset \mathbb{R}^m$ is nonempty, closed and μ is a Borel probability measure on $\mathbb{R}^s$. The optimal value of $P(\mu)$ is defined by

$$\varphi(\mu) := \inf \{ \int_{\mathbb{R}^s} f(x,z)\mu(dz) : x \in C\},$$

and the corresponding solution set is

$$\psi(\mu) := \operatorname{argmin} \{ \int_{\mathbb{R}^s} f(x,z)\mu(dz) : z \in C\},$$

see [80], [81] and the references therein. Typically the probability μ is incompletely determined.

We start with an example of an "ideal" metric[1] for studying quantitative stability of $P(\mu)$. Let (Z,d) be a separable metric space, $P(Z)$ the set of all Borel probability

[1]The results in this section represent the main part of the lecture "Quantitative Stability of Stochastic Programs via Probability Metrics", by S.T. Rachev and W. Römisch given at the 3rd Int. Conf. on "Approximation and Optimization" in the Caribbean, Puebla (Mexico), Oct. 8-13, 1995.

measures on Z and $\Theta \in Z$ a fixed element playing the role of an "origin". For any $h : Z \to \mathbb{R}$ and any $r > 0$, define the Lipschitz norm:

$$\text{Lip}_h(r) := \sup\left\{ \frac{|h(z) - h(\tilde{z})|}{d(z,\tilde{z})} : z \neq \tilde{z}, d(z,\Theta) \leq r, d(\tilde{z},\Theta) \leq r \right\}.$$

Given a nondecreasing function $H : \mathbb{R}_+ \to \mathbb{R}_+$ with $H(0) = 0$ define the semi-norm of h as

$$\|h\|_H := \sup\{ \text{ Lip }_h(r)(\max\{1, H(r)\})^{-1} : r > 0\}.$$

Now we are ready so define the Fortét-Mourier metric

$$FM_H(P,Q) := \sup\left\{ \left| \int_Z h(z)(P-Q)(dz) \right| : \|h\|_H \leq 1 \right\}$$

in

$$P_H(Z) := \left\{ P \in P(Z) : \int_Z c_H(z,\Theta)P(dz) < \infty \right\},$$

where $c_H(z,\Theta) = d(z,\Theta)\max\{1, H(d(z,\Theta))\}$. The Fortét-Mourier metric arises in a natural way from the Kantorovich-Rubinstein Mass-Transshipment problem:

$$FM_H(P,Q) = \inf\left\{ \int_{Z\times Z} c_H(z,\tilde{z})\eta(dz,d\tilde{z}) : \eta \in D(P,Q) \right\}$$

for any $P, Q \in P_H(Z)$, where $D(P,Q)$ denotes the set of all bounded Borel measures η on $Z \times Z$ satisfying the "balancing" constraint $\eta(\cdot \times Z) - \eta(Z \times \cdot) = (P - Q)(\cdot)$. If the function H satisfies the property

$$\Delta_H := \sup_{t\neq s} \frac{|\, t\max\{1, H(t)\} - s\max\{1, H(s)\}\,|}{|\, t - s \,|\max\{1, H(t), H(s)\}} < \infty,$$

then

$$FM_H(P_n, P) \to 0 \text{ for } P_n, P \in P_H(Z)$$

if and only if (P_n) converges weakly to P and $\int_Z c_H(z,\Theta)(P_n - P)(dz) \to 0$ (cf. Corollary 4.3.4 in [69]). For example, $\Delta_H < \infty$ is valid for $H(t) = t^a, a > 0$. On the real line $Z = \mathbb{R}, d(z,\tilde{z}) =|\, z - \tilde{z}\,|$, the Fortét-Mourier metric admits an explicit representaiton (Theorem 5.4.1 in [69], see also [70]).

$$FM_H(P,Q) = \int_{-\infty}^{\infty} H(|\, z - \Theta \,|) \,|\, F_P(z) - F_Q(z) \,|\, dz,$$

where F_P denotes the distribution function of P.

In the next theorem, we use the Fortét-Mourier metric to evaluate the stability of $P(\mu)$ with respect to perturbations of the original distribution P.

Theorem 3.8 *Let $H : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing function with $H(0) = 0, P \in P_H(\mathbb{R}^s)$ and $\psi(P)$ be nonempty and bounded. Assume that*

(i) *the function* $f(\cdot, \xi)$ *is convex for each* $\xi \in \mathbb{R}^s$, *and*

(ii) *there exists an open, bounded subset* V *of* $\mathbb{R}^m$ *and a constant* $L_0 > 0$ *such that* $\psi(P) \subset V$ *and*

$$|f(x,\xi) - f(x,\widetilde{\xi})| \leq L_0 \max\{1, H(\max\{\|\xi\|, \|\widetilde{\xi}\|\})\} \quad \|\xi - \widetilde{\xi}\|$$

whenever $x \in V$ *and* $\xi, \widetilde{\xi} \in \mathbb{R}^s$.

Then the solution set mapping ψ *from* $(P_H(\mathbb{R}^s), FM_H)$ *is upper semicontinuous at* P *and there exist constants* $L > 0$ *and* $\delta > 0$ *such that*

$$|\varphi(P) - \varphi(Q)| \leq FM_H(P,Q) \text{ whenever } Q \in P_H(\mathbb{R}^s), FM_H(P,Q) < \delta.$$

The stability results can be applied to the empirical analysis of $P(\mu)$. Consider for example the approximation of $P(\mu)$ by its sample version

$$P_n(\mu): \qquad \min\left\{F_n(x) = \frac{1}{n}\sum_{i=1}^n f(x,\xi_i) : x \in C\right\},$$

where (ξ_i) are i.i.d. copies of ξ. Let φ and φ_n denote the optimal values of $P(\mu)$ and $P_n(\mu)$, respectively, and let ψ and ψ_n denote the corresponding solution sets. Applying the rate-of-convergence results for empirical measures in terms of the Fortét-Mourier metric (see [69]), Theorem 3.8 provides bounds for the distance between φ and φ_n. The stability analysis can then be used to estimate the sensitivity of a portfolio of asset returns having minimal risk with preassigned mean returns.

4 Transportation problems with additional or relaxed constraints

Several modifications of the transportation problem have been studied allowing bounds for the admissible supply and demand distributions, or capacity constraints for the admissible transportation plans (cf. [75]). Let P_1, P_2 be probability measures on $\Re^1$ with distribution functions (d.f.s) F_1, F_2 and let $\mathcal{F}(F_1, F_2)$ denote the class of joint d.f.s F with marginals F_1, F_2. The classical Hoeffding-Fréchet characterization of $\mathcal{F}(F_1, F_2)$ states that a d.f. F is in $\mathcal{F}(F_1, F_2)$ if and only if

$$F_-(x,y) := (F_1(x) + F_2(y) - 1)_+ \leq F(x,y) \leq \min\{F_1(x), F_2(y)\} =: F_+(x,y). \quad (44)$$

If $c(x,y)$ satisfies the "Monge" conditions, i.e. c is right continuous and

$$c(x',y') - c(x,y') - c(x',y) + c(x,y) \leq 0 \quad \text{for} \quad x' \geq x, y' \geq y, \qquad (45)$$

then for all $F \in \mathcal{F}(F_1, F_2)$

$$\int c dF_+ \le \int c dF \le \int c dF_-. \tag{46}$$

An equivalent form in terms of the random variables X, Y with $F_X = F_1, F_Y = F_2$ is

$$Ec\left(F_1^{-1}(U), F_2^{-1}(U)\right) \le Ec(X,Y) \le Ec\left(F_1^{-1}(U), F_2^{-1}(1-U)\right) \tag{47}$$

where U is uniformly distributed on $(0,1)$ and $F_i^{-1}(u) = \inf\{y : F_1(y) \ge u\}$ is the generalized inverse of F_1 (the quantile function).

Consider for given d.f.s F_1, F_2 the set

$$\mathcal{H}(F_1, F_2) := \{F; F \text{ is a d.f. on } \Re^2 \text{ with marginals } \widetilde{F_1}, \widetilde{F_2}, \text{ where } \widetilde{F_1} \le F_1, \widetilde{F_2} \ge F_2\} \tag{48}$$

with bounds on the marginal d.f.s.

We study the transportation problem:

$$\text{minimize} \int c(x,y) dF(x,y), \text{ subject to } F \in \mathcal{H}(F_1, F_2), \tag{49}$$

or, equivalently,

$$\text{minimize } Ec(X,Y), \text{ subject to } F_X \le F_1, F_Y \ge F_2. \tag{50}$$

Theorem 4.1 *(cf. [75])*
Suppose that $c(x,y)$ is symmetric, c satisfies the Monge condition (45) and $c(x,x) = 0$ for all x and define $H^*(x,y) := \min\{F_1(x), \max\{F_1(y), F_2(y)\}\}$, then $H^* \in \mathcal{H}(F_1, F_2)$ and H^* solves the relaxed transportation problem (49). Furthermore,

$$\int c dH^* = \int_0^1 c\left(F_1^{-1}(u), \min\left(F_1^{-1}(u), F_2^{-1}(u)\right)\right) du.$$

We remark that Theorem 4.1 suggests a greedy algorithm for the solution of the corresponding discrete transportation problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \\ \text{subject to:} & x_{ij} \ge 0 \\ & \sum_{s=1}^j \sum_{r=1}^n x_{rs} \ge \sum_{s=1}^j b_s =: G_j \\ & \sum_{r=1}^j \sum_{s=1}^n x_{rs} \le \sum_{r=1}^i a_r =: F_i \end{array} \tag{51}$$

where the sum of the "demands" $\sum_{s=1}^n b_s$ equals the sum of the "supplies" $\sum_{r=1}^n a_r$. We assume that the (c_{ij}) are symmetric, $c_{ii} = 0$ and c satisfies the discrete Monge condition

$$c_{i,j} + c_{i+1,j+1} - c_{i,j+1} - c_{i+1,j} \le 0. \tag{52}$$

The restrictions describe production and consumption processes based on priorities with capacities $s_1, \ldots, s_n$ such that what remains in stage i of the production (or consumption) process can be transferred to some of the subsequent stages $i+1,\dots,n$.

The proposed greedy algorithm for this problem is as follows. Denote $H_i := \max\{F_i, G_i\}, 1 \le i \le n, \delta_1 := H_1, \delta_{i+1} := H_{i+1} - H_i, i \le n-1$, then (51) is equivalent to the standard transportation problem

$$\begin{array}{l} \text{minimize } \sum\sum c_{ij}x_{ij} \text{ subject to} \\ \sum_{j=1}^n x_{ij} = a_i, \sum_{i=1}^n x_{ij} = \delta_j, \qquad x_{ij} \ge 0, \end{array} \tag{53}$$

and the North-West corner rule applied to these new equality restrictions solves (51). For a detailed example and comparison cf. [75].

For a second example, let μ be a finite Borel measure on the plane and for any two probabilities P_1 and P_2 on $\Re^1$ and $A_i \times B_i \in \mathcal{B}^2, i \in I$ define

$$M^\mu(P_1, P_2) := \{P \in \mathcal{M}(P_1, P_2); P(A_i \times B_i) \le \mu(A_i \times B_i), i \in I\}, \tag{54}$$

the class of transportation plans with upper bounds on the capacity of sets $A_i \times B_i$.

By sharpness of the classical Fréchet bounds (cf. [83])

$$\min\{P(A_i \times B_i); P \in M(P_1, P_2)\} = \max\{P_1(A_i) + P_2(B_i) - 1, 0\} \tag{55}$$

we impose the necessary assumptions

$$\mu(A_i \times B_i) \ge \max(0, P_1(A_i) + P_2(B_i) - 1) \tag{56}$$

in order to avoid trivial cases.

Theorem 4.2 *(cf. [63], [75]) Define*

$$P^*(A \times B) = \min\left\{ \inf_{\substack{A_i \subset A \\ B_i \subset B}} \{\mu(A_i \times B_i) + P_1(A \backslash A_i) + P_2(B \backslash B_i)\}, \min\{P_1(A), P_2(B)\} \right\}, \tag{57}$$

then the generalized upper Fréchet bound

$$h_\mu(A \times B) := \sup\{P(A \times B); P \in M^\mu(P_1, P_2)\}$$

satisfies

a) $h_\mu(A \times B) \le P^*(A \times B)$,

b) If P^* *defines a measure, then* $P^* \in M^\mu(P_1, P_2)$ *and* $h_\mu(A \times B) = P^*(A \times B)$,

c) If $\{A_i \times B_i; i \in I\} = \{(-\infty, x] \times (-\infty, y]; x, y \in \Re^1\}$, *then* P^* *defines a measure and the bound in a) is sharp.*

Again, as in the first example, a greedy algorithm can be constructed for this problem (cf. [75]). The (generalized) transportation problem can be considered as a (generalized) moment problem with infinitely many moment type conditions specifying the marginal distributions. From this point of view some explicit moment type problems have been considered with moment type conditions on the marginal distribution functions. In [69] the problem of minimizing (maximizing) $E\|X_1-X_2\|^p$ is considered under the restrictions $E\|X_i\|^{q_j}=a_{ij}, i=1,2, j=1,\ldots,n$. This corresponds to a weakening of the marginal constraints.

An interesting problem related with moment type marginal constraints is considered in [49]. The problem arises in the context of computer tomography. Let Q_1, Q_2 be probability measures on $\Re^m$ with identical marginal distributions in a finite number n of directions $\vartheta_1,\ldots,\vartheta_n$.

We ask what can be said about the closeness of Q_1, Q_2 if they coincide in distribution in an increasing number n of directions? It is known that with respect to the supremum distance ρ, Q_1, Q_2 may differ considerably (this is known as "computer tomography paradox"). In a recent paper [49] it was shown that this paradox disappears when some weaker metrics like the λ-metric and the Lévy-Prohorov distance are used.

Consider the case $m=2$, and define

$$\lambda(P,Q) := \min_{T>0}\max\left\{\max_{\|t\|\le T}\left|\int e^{i\langle t,x\rangle}(P-Q)(dx)\right|, \frac{1}{T}\right\} \tag{58}$$

λ metrizes the topology of weak convergence [49].

Theorem 4.3 *(cf. [49])*
Let P,Q be probability measures on $\Re^2$ which have the same marginals in directions $\vartheta_1,\ldots,\vartheta_n$ no two of which are collinear. Suppose, that P has support in the unit disc, then

$$\lambda(P,Q) \le \left(\frac{2}{s!}\right)^{\frac{1}{s+1}} \quad \text{with } s := 2\left[\frac{n-1}{2}\right]. \tag{59}$$

Note that the right hand side satisfies $\left(\frac{2}{s!}\right)^{\frac{1}{s+1}} \sim \frac{e}{s}$ as $s\to\infty$. The assumption of coinciding marginals in directions $\vartheta_1,\ldots,\vartheta_n$ can be replaced by the assumption of coinciding moments up to order $n-1$ in these directions, and compactness of the support can be replaced by a Carleman type condition. Define

$$\begin{aligned} \mu_k &:= \sup_{\vartheta\in S^1}\int \langle x,\vartheta\rangle^k P(dx), k=0,1\ldots \\ \beta_s &:= \sum_{j=1}^{(s-2)/2} \mu_{2j}^{-\frac{1}{2j}} \end{aligned} \tag{60}$$

Theorem 4.4 *If $\vartheta_1,\ldots,\vartheta_n$ are not collinear directions in $\Re^2$, P has moments of any order and P,Qhave identical moments up to order $n-1$ in directions $\vartheta_1,\ldots,\vartheta_n$, then for some absolute constant C*

$$\lambda(P,Q) \le C\beta_s^{-1/4}\left(\mu_0+\mu_2^{1/2}\right)^{1/4} \tag{61}$$

For some extensions to higher dimensions and related results we refer to [50].

Finally we list some recent references of the Monge-Kantorovich problems in various areas of applied probability.

Acknowledgement: The authors would like to thank the editor and the reviewer for many remarks and suggestions concerning the style and the presentation of the paper.

References

[1] Abdellaoui, T. (1993). *Distances de deux lois dans les espaces de Banach.* Doctoral dissertation, Université de Rouen.

[2] Abdellaoui, T. (1994). Détermination d'un couple optimal du problème de Monge-Kantorovich. *C. R. Acad. Sci. Paris* 319, Série I, 981-984.

[3] Abdellaoui, T. and Heinich, H. (1994). Sur la distance de deux lois dans le cas vectoriel. *C. R. Acad. Sci. Paris* 319, Série I, 397-400.

[4] Aldous, D. and Steele, J.M. (1993). Introduction to the interface of probability and algorithms. *Statist. Science* 8, 3-9.

[5] Aurenhammer, F., Hoffmann, F. and Aronov, B. (1992). Minkowski-type theorems and least-square partitioning. *Reports of the Institute for Computer Science.* Dept. of Mathematics, Freie Universität Berlin.

[6] Balinski, M. and Rachev, S.T. (1993). Rounding proportions: rules of rounding. *Numer. Funct. Anal. Optimization* 14, 475-501.

[7] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press, New York.

[8] Barnes, E.R. and Hoffmann, A.J. (1985). On transportation problems with upper bounds on leading rectangles. *SIAM J. Alg. Disc. Math.* 6, 487-496.

[9] Barnsley, M. (1988). *Fractals Everywhere.* Academic Press, San Diego.

[10] Beneš, V. (1992). Extremal and optimal solutions in the transshipment problem. *Comment. Math. Univ. Carolinae.* 33, 97-112.

[11] Beneš, V. (1995). *Extremal and Optimal Solutions of the Marginal and Transshipment Problem.* Dr.Sci. Thesis, Dept. of Mathematics, FSI, Czech Technical University, Praha, Czech Republic.

[12] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Stat.* 9, 1196-1217.

[13] Blackwell, D. and Dubins, L.E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* 89, 691-692.

[14] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Maths.* XLIV, 375-417.

[15] Burkard, R.E., Klinz, B. and Rudolf, R. (1994). Perspectives of Monge properties in optimization. In *Optimierung und Kontrolle*, Spezialforschungsbereich F 003, Karl-Franzens-Universität Graz & Technische Universität Graz, Bericht Nr. 2.

[16] Cramer, M. (1995). Stochastische Analyse rekursiver Algorithmen mit idealen Metriken. Dissertation, Universität Freiburg.

[17] Cramer, M. and Rüschendorf, L. (1994). Convergence of a branching type recursion. Freiburger Zentrum für Datenanalyse und Modellbildung Preprint No. 3, 1994.

[18] Cuesta-Albertos, J.A. and Matrán, C. (1987). Asymptotic behavior of p-predictions for vector valued random variables. *Proc. Amer. Math. Soc.* 100, 716-720.

[19] Cuesta-Albertos, J.A. and Matrán, C. (1988). Strong convergence of weighted sums of random elements through the equivalence of sequences of distributions. *J. Multivar. Anal.* 25, 311-322.

[20] Cuesta-Albertos, J.A. and Matrán, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* 17, 1264-1276.

[21] Cuesta-Albertos, J.A. and Matrán, C. (1991). Skorohod representation theorem and Wasserstein metrics. Preprint.

[22] Cuesta-Albertos, J.A. and Matrán, C. (1992). A review on strong convergence of weighted sums of random elements based on Wasserstein metrics. *J. Statist. Planning Infer.* 30, 359-370.

[23] Cuesta-Albertos, J.A. and Matrán, C. (1994). Stochastic convergence through Skorohod representation theorems and Wasserstein distances. *Suppl. Rendic. Circolo Matem. Palermo*, Serie II, 35, 89-113.

[24] Cuesta-Albertos, J.A., Matrán, C. and Tuero-Diaz, A. (1993). Optimal maps for the L^2-Wasserstein distance. Preprint.

[25] Cuesta-Albertos, J.A., Matrán, C. and Tuero-Diaz, A. (1993). On lower bounds for the L^2-Wasserstein metric in a Hilbert space. *J. Theoret. Probab.* (to appear).

[26] Cuesta-Albertos, J.A., Rüschendorf, L. and Tuero-Diaz, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. *J. Multivariate Anal.* 46, 355-361.

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

[28] Dall'Aglio, G., Kotz, S. and Salinetti, G. Eds. (1991). *Advances in Probability Distributions with Given Marginals.* Kluwer, Dordrecht.

[29] Dobrushin, R.L. (1970). Prescribing a system of random variables by conditional distributions. *Theor. Probab. Applns.* 15, 458-486.

[30] Dudley, R.M. (1989). *Real Analysis and Probability.* Wadsworth, Pacific Grove.

[31] Feldman, P., Rachev, S.T. and Rüschendorf, L. (1994). Limit theorems for recursive algorithms. *J. Comp. Appl. Maths.* 56, 169-182.

[32] Feldman, P., Rachev, S.T. and Rüschendorf, L. (1995). Limit distribution of the collision resolution interval. *Statistica Neerlandica* (to appear).

[33] Ferland, R. and Giroux, G. (1991). Le modèle Bose-Einstein de l'équation non-linéaire de Boltzmann: convergence vers l'équilibre. *Ann. Sci. Math. Québec.* 15, 23-33.

[34] Gani, J., Haynatzki, G., Rachev, S.T. and Yakowitz, S. (1995). Stochastic algorithms in AIDS spread modelling. Technical Report, UCSB, Santa Barbara.

[35] Gelbrich, M. (1990). On a formula for the L^2-Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math. Nachr.* 147, 185-203.

[36] Gelbrich, M. (1995). Simultaneous time and chance discretization for stochastic differential equations. *J. Comp. Appl. Maths.* (to appear).

[37] Gelbrich, M. and Rachev, S.T. (1995). Discretization for stochastic differential equations, L^2-Wasserstein metrics, and econometric models. Proceedings: *Distributions with Given Marginals*, IMS (to appear).

[38] Givens, C.R. and Shortt, R.M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* 31, 231-240.

[39] Gutmann, S., Kampermann, J.H.B. and Reeds J.A. (1991). Existence of probability measures with given marginals. *Ann. Probab.* 19, 1781-1791.

[40] Hanin, L.G. and Rachev, S.T. (1994). Mass transshipment problems and ideal metrics. *J. Comp. Appl. Maths.* 56, 183-196.

[41] Hanin, L.G. and Rachev, S.T. (1995). An extension of the Kantorovich-Rubinstein mass-transportation problem. *Numer. Funct. Anal. Optimization* (to appear).

[42] Heinich, H. and Lootgieter, J.C. (1993). Convergence des fonctions monotones. Preprint.

[43] Horowitz, J. and Karandikar, R.L. (1995). Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comp. Appl. Maths.* (to appear).

[44] Kanagawa, S. (1986). The rate of convergence for approximate solutions of stochastic differential equations. *Tokyo J. Maths.* 12, 33-48.

[45] Kantorovich, L.V. (1942). On the transfer of masses. *Dokl. Akad. Nauk SSSR* 37, 7-8.

[46] Kellerer, H.G. (1984). Duality theorems for the marginal problems. *Z. Wahrsch. Verw. Geb.* 67, 399-432.

[47] Kellerer, H.G. (1993). Ambiguity in bounded moment problems. Proceedings: *Distributions with Given Marginals*, IMS (to appear).

[48] Kempermann, J.H.B. (1993). Moment problems for measures on $\mathbb{R}^n$ with given k-dimensional marginals. Proceedings: *Distributions with Given Marginals*, IMS (to appear).

[49] Klebanov, L.B. and Rachev, S.T. (1994). Proximity of probability measures with common marginals in a finite number of directions, Proceedings: *Distributions with Given Marginals*, IMS (to appear).

[50] Klebanov, L.B. and Rachev, S.T. (1995). The method of moments in computer tomography. *Math. Scientist* 20, 1-14.

[51] Knott, M. and Smith, C.S. (1984). On the optimal mapping of distributions. *J. Optimization Theory Applns.* 43, 39-49.

[52] Landers, D. and Rogge, L. (1980). Best approximants in L_ϕ-spaces. *Z. Wahrsch. verw. Gebiete* 51, 215-237.

[53] Levin, V.L. (1990). General Monge-Kantorovich problem and its applications in measure theory and mathematical economics. In: *Functional Analysis, Optimization and Mathematical Economics (A collection of papers dedicated to the Memory of L. V. Kantorovich)* (ed. L. J. Leifman), Oxford University Press, New York.

[54] Levin, V.L. (1990). Formula for the optimal value in the Monge-Kantorovich problem with a smooth cost function and a characterization by cyclic monotone mappings. *USSR Math. Sbornik.* 181,1694-1707.

[55] Levin, V.L. (1991). Some applications of set-valued mappings in mathematical exonomics. *J. Math. Economics* 20, 69-87.

[56] Lorentz, G.G. (1953). An inequality for rearrangements. *Amer. Math. Monthly* 60, 176-179.

[57] Maejima, M. (1988). Some limit theorems for stability methods of i.i.d. random variables. *Lecture Notes in Mathematics*, 1233, 57-68, Springer, Berlin.

[58] Maejima, M. and Rachev, S.T. (1995). Rates of convergence in the operator-stable limit theorem. *J. Theoret. Probab.* (to appear)

[59] Mallows, C.L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* 43, 508-515.

[60] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie des Sciences de Paris*, avec les Mémoires de Mathématiques et de Physique pour la même année, 666-704, avec 2 pl.

[61] Mosler, K. and Scarsini, M., eds. (1991). *Stochastic Order and Decision under Risk.* IMS Lecture Notes.

[62] Olkin, I. and Pukelsheim, F. (1982). The distance between two random vectors with given dispersion matrices. *Lin. Alg. Applns.* 48, 257-263.

[63] Olkin, I. and Rachev, S.T. (1990). Marginal problems with additional constraints. *Tech. Rep.* 270, Dept. of Statistics, Stanford Univ.

[64] Olkin, I. and Rachev, S.T. (1993). Maximum submatrix traces for positive definite matrices. *SIAM J. Matrix Anal. Applns.* 14, 390-397.

[65] Rachev, S.T. (1982). Minimal metrics in the random variable space. *Publ. Inst. Stat. Univ. Paris* XXVII, 27-47.

[66] Rachev, S.T. (1982). Minimal metrics in the random variables space. *Probability and Statistical Inference*, Proceedings of the 2nd Pannonian Symposium. W. Grossmann et al., eds., 319-327, D. Reidel Company, Dordrecht.

[67] Rachev, S.T. (1984). On a class of minimal functions on a space of probability measures. *Theory Probab. Applns.* 29, 41-49.

[68] Rachev, S.T. (1991). Mass transshipment problems and ideal metrics. *Num. Funct. Anal. Optimization* 12, 563-573.

[69] Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York.

[70] Rachev, S.T. and Rüschendorf, L. (1991). Recent results in the theory of probability metrics. *Statist. Decisions* 9, 327-373.

[71] Rachev, S.T. and Rüschendorf, L. (1992). A new ideal metric with applications to multivariate stable limit theorems. *Probab. Theory Related Fields* 94, 163-187.

[72] Rachev, S.T. and Rüschendorf, L. (1992). Probability metrics and recursive algorithms. *Adv. Appl. Probab.* (to appear).

[73] Rachev, S.T. and Rüschendorf, L. (1993). On the rate of convergence in the CLT with respect to the Kantorovich metric. *9th Conference in Probability on Banach Spaces.* Eds. J. Kuelbs, M. Marcus and J. Hoffmann-Jorgensen.

[74] Rachev, S.T. and Rüschendorf, L. (1995). Propagation of chaos and contraction of stochastic mappings. *Siberian Adv. Maths.* 4, 114-150.

[75] Rachev, S.T. and Rüschendorf, L. (1994). Solution of some transportation problems with relaxed or additional constraints. *SIAM J. Control Optimization* 32, 673-689.

[76] Rachev, S.T. and Samorodnitsky, G. (1995). Limit laws for a stochastic process and random recursion arising in probabilistic modelling. *J. Appl. Probab.* (to appear).

[77] Rachev, S.T. and Shortt, R.M. (1990). Duality theorems for Kantorovich-Rubinstein and Wasserstein functionals. *Dissert. Mathematicae* CCXCIX, 647-676.

[78] Ramachandran, D. and Rüschendorf, L. (1995). A general duality theorem for marginal problems. *Probab. Theory Related Fields* 101, 311-319.

[79] Rockafeller, R.T. (1970). *Convex Analysis.* Princeton Univ. Press, Princeton, NJ.

[80] Römisch, W. and Schultz, R. (1991). Stability analysis of stochastic programs. *Ann. Operat. Res.* 30, 241-266.

[81] Römisch, W. and Schultz, R. (1993). Stability of solutions for stochastic programs with complete recourse. *Mat. Operat. Res.* 18, 590-609.

[82] Rösler, U. (1991). A limit theorem for Quicksort. *Informatique Théor. Applicns.* 25, 85-100.

[83] Rüschendorf, L. (1981). Sharpness of Fréchet bounds. *Z. Wahrsch. Verw. Geb.* 57, 293-302.

[84] Rüschendorf, L (1991). Bounds for distributions with multivariate marginals. Proceedings: *Stochastic Order and Decision under Risk.* Eds. K. Mosler and M. Scarsini. IMS Lecture Notes, 19, 285-310.

[85] Rüschendorf, L. (1991). Fréchet bounds and their applications. *Advances in Probability Measures with Given Marginals.* Eds. G. Dall'Aglio, S. Kotz and G. Salinetti, 151-188.

[86] Rüschendorf, L. (1993). Optimal solutions of multivariate coupling problems. *Appl. Mathematicae* (to appear).

[87] Rüschendorf, L. and Rachev, S.T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* 32, 48-54.

[88] Smith, C.S. and Knott, M. (1987). Note on the optimal transportation of distributions. *J. Optimization Theory Applns.* 52, 323-329.

[89] Smith, C.S. and Knott, M. (1992). On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.* 40, 328-334.

[90] Sudakov, V.N. (1976). *Geometric problems in the theory of infinite dimensional probability distributions.* Tr. Math. Inst. V.A. Steklov, Akad. Nauk, SSSR, 141. (English translation *Proc. Steklov Inst. Math.* (1979) no 2, 141)

[91] Tanaka, H. (1973). An inequality for a functional of probability distributions and its applications to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrsch. Verw. Gebiete.* 27, 47-52.

[92] Tuero-Diaz, A. (1991). *Aplicaciones crecientes: Relaciones con las métricas de Wasserstein.* Doctoral dissertation, Universidad de Cantabria.

[93] Tuero-Diaz, A. (1993). On the stochastic convergence of representations based on Wasserstein metrics. *Ann. Probab.* 21, 72-85.

[94] Wang, Y.H. (1986). Coupling methods in approximations. *Canad. J. Statist.* 14, 69-74.

[95] Zolotarev, V.M. (1975). On the continuity of stochastic sequences generated by recurrent procedures. *Theory Probab. Appl.* 20, 819-832.