Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

DEVELOPMENTS ON FRÉCHET-BOUNDS

BY LUDGER RÜSCHENDORF
University of Freiburg

This paper describes some developments in the field of Fréchet bounds since the last report given at the Rome conference in 1990. At first a review is given on product type representations of distributions with given marginals and their relations to the solution of Schrödinger type equations. The iterative proportional fitting procedure allows an approximate construction. Its convergence has been shown recently. We give a convergence proof of a modified algorithm under alternative assumptions. In the next part of the paper several sufficient and necessary conditions are given for the explicit construction of optimal multivariate couplings or, equivalently, tranportation plans. These results allow to calculate several multivariate examples, in particular examples for minimal ℓ_p-metrics. In the final part we consider some recent examples of optimal couplings under additional or relaxed restrictions. We discuss a problem involving order restrictions, the case of fixed difference of the marginals, the application of a duality principle for Monge functions and Fréchet bounds for marginal classes majorized by a finite measure.

1. Introduction. The paper is divided in three main parts. In the first part we review recent results on product-type representations of probability measures with fixed marginals. These representations are related to systems of integral equations introduced by Schrödinger (1931). They are of importance for the construction and analysis of Schrödinger bridges. Solutions may be approximated by the iterative proportional fitting procedure which was introduced in 1940 by Deming and Stephan. In the finite discrete case several convergence proofs were given in the sixties. However a general convergence proof was only found recently, and in this paper we establish convergence of a modified algorithm under somewhat different assumptions.

In the second part of the paper we consider recent extensions of the theory of optimal multivariate couplings, resp. tranportation problems, which we introduced in Section 3 of the report Rüschendorf (1991b) on Fréchet bounds given at the Rome conference. Some sufficient and necessary conditions for

AMS 1991 Subject Classification: Primary 60K30, 90C08.

Key words and phrases: Marginal, class, optimal coupling, Monge function, iterative proportional fitting.

optimality are established. In the case of minimal ℓ_p-metrics these results allow us to explicitly construct optimal couplings in several examples. The proofs exploit relations to nonconvex optimization problems.

In the final part we consider some examples of transportation problems (couplings) with additional or relaxed restrictions on the Fréchet class which were studied in recent papers of Olkin and Rachev (1990), Rachev and Rüschendorf (1994, 1993), and Levin (1992). The first problem, due to Rogers (1992), is concerned with optimal couplings subject to order restrictions. In the finite discrete case one can give a nice explicit solution. For the transportation problem with a fixed difference of the marginals, which is a generalization of the Kantorovich-Rubinstein problem, several explicit formulas and bounds for the optimal value were recently established in the univariate and also in the multivariate case. Next a simple duality principle for Monge functions is applied to two examples. Finally, a duality result for Fréchet classes majorized by a finite measure is established, and, as an application, a formula for the corresponding upper Fréchet bound is given.

2. Product Representation and Schrödinger Problem. The literature on probability representations in marginal problems has been concentrated to a large extent on the standard representation (copula representation). Any n-dimensional distribution function F with marginals $F_1, \dots, F_n$ has a representation of the form $F = C(F_1, \dots, F_n)$, where C is a df with uniform marginals. Extensions of this representation to general spaces are given in Scarsini (1989), Rachev and Rüschendorf (1990), and Rüschendorf (1991a). A review can be found in Dall'Aglio, Kotz, and Salinetti (1990), the conference volume of the 1990 conference in Rome.

In this section product-type representations of the form $\nu = \otimes_{i=1}^n f_i \mu$ are discussed. Here ν is a probability measure with marginals ν_i, μ is a probability measure with marginals μ_i, $\otimes_{i=1}^n f_i(x) = \prod_{i=1}^n f_i(x_i)$, $x \in E = \prod_{i=1}^n E_i$, $(E, \mathcal{A}) = \otimes_{i=1}^n (E_i, \mathcal{A}_i)$ the product space, and $\nu = f\mu$ denotes that ν has density f w.r.t. μ. Assume that $\mu = h \otimes_{i=1}^n \mu_i$ has a density h w.r.t. $\otimes \mu_i$ and that $\nu_i \ll \mu_i$ with density $r_i = \frac{d\nu_i}{d\mu_i}$. Then

$$\nu := \bigotimes_{i=1}^n f_i \mu \in M(\nu_1, \dots, \nu_n), \tag{2.1}$$

the class of distributions with marginals ν_i, iff the functions (f_i) satisfy the following system of integral equations

$$f_i(x_i) \int \prod_{j \neq i} f_j(x_j) h(x) \bigotimes_{j \neq i} \mu_j(\mathrm{d}x_j) = r_i(x_i) \; [\mu_i], \quad 1 \le i \le n. \tag{2.2}$$

In the case $n = 2$ these equations were introduced by Schrödinger (1931). A solution of the form (2.1) leads to the construction of Schrödinger bridges and permits the derivation of basic properties of these processes. For reference we refer to Föllmer (1988) and Wakolbinger (1992). The existence of solutions of (2.2) for $n = 2$ has been proved in many papers including Fortet (1940), Beurling (1960), Hobby and Pyke (1965) and Jamison (1974). A general existence result was recently obtained in Rüschendorf and Thomsen (1993b). The proof of the following extension to the case $n \geq 2$ is similar to the case $n = 2$ in Rüschendorf and Thomsen (1993a) and, therefore, is only sketched. Let $I(\nu \mid \mu) = \int \ln \left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right| \mathrm{d}\nu$ denote the Kullback-Leibler distance. Define for $M = M(\nu_1, \ldots, \nu_n)$

$$I(M \mid \mu) = \inf\{I(\nu \mid \mu);\ \nu \in \mathrm{M}\}. \tag{2.3}$$

THEOREM 2.1. *If $I(M \mid \mu) < \infty$, then there exists a solution $(f_1, \ldots, f_n)$ of the system (2.2) of integral equations. Moreover, $\otimes_{i=1}^n f_i$ is uniquely determined.*

PROOF: Since M is closed w.r.t. variation distance, there exists a unique I-projection ν^* of μ on M. Furthermore, by Theorem 3.1 of Csiszar (1975), $\ln \frac{\mathrm{d}\nu^*}{\mathrm{d}\mu}$ belongs to the closure of $\oplus_{i=1}^n L^1(\nu_i)$ in $L^1(\nu^*)$. From Theorem 3 resp. Remark 5 in Rüschendorf and Thomsen (1993a) any element in the closure of $\oplus_{i=1}^n L_+^1(\nu_i)$ is of the form $\oplus_{i=1}^n h_i$ for some measurable (but generally not integrable) functions h_i. Therefore, $\nu^* = \otimes_{i=1}^n f_i \mu \in M(\nu_1, \ldots, \nu_n)$, where $f_i := \exp(h_i)$ and $(f_1, \ldots, f_n)$ solves the Schrödinger equation (2.2) ∎

Consider a system $\mathcal{H} \subset \mathcal{P}(\{1, \ldots, n\})$ of subsets of $\{1, \ldots, n\}$ and let $\mu \in M^1(E, \mathcal{A})$, the class of probability measures on $(E, \mathcal{A})$, have multivariate marginals $\mu_H := \mu^{\pi_H}, H \in \mathcal{H}$. Let $\nu_H \ll \mu_H$, $H \in \mathcal{H}$, be a system of multivariate marginals with $r_H = \frac{\mathrm{d}\nu_H}{\mathrm{d}\mu_H}$ and assume that $M_{\mathcal{H}} := M(\nu_H, H \in \mathcal{H})$, the set of distributions with multivariate marginals ν_H is not empty. The following result describes a basic structural property of probabilites with given multivariate marginals.

THEOREM 2.2. *(Rüschendorf and Thomsen (1993b).) If $I(M_{\mathcal{H}} \mid \mu) < \infty$, then there exist nonnegative functions $(f_H)_{H \in \mathcal{H}}$ on $E_H = \prod_{i \in H} E_i$ such that $\otimes_{H \in \mathcal{H}} f_H$ is measurable and $\nu := \otimes_{H \in \mathcal{H}} f_H \mu \in M_{\mathcal{H}}$.*

Theorem 2.2 implies the existence of solutions of the generalized system of Schrödinger equations

$$\int f_H(x_H) h(x) \bigotimes_{J \neq H} f_J(x_J) \bigotimes_{j \in H^c} \mu_j(\mathrm{d}x_j) = r_H(x_H)\, [\mu_H], \quad H \in \mathcal{H}. \tag{2.4}$$

It also implies that under an integrability condition there is a one-to-one relation between $M(\mu_H, H \in \mathcal{H})$ and $M(\nu_H, H \in \mathcal{H})$ given by densities of product-type. Moreover, ν is the I-projection of μ, and in this way the relation by product densities solves an optimization problem of a type related to Fréchet bounds.

3. Convergence of the Iterative Proportional Fitting Procedure. The iterative proportional fitting procedure (IPFP) is an algorithm for adjusting the marginal distributions of a probability measure to a-priori known marginals. It was introduced by Deming and Stephan (1940) in connection with adjusting estimates of cell probabilities in contingency tables subject to certain marginal constraints. In this section we restrict our attention to the case of two simple marginals. The aim of the IPFP is to find iteratively (approximatively) functions $a = a(x), b = b(y)$ such that

$$\begin{aligned} R(x,a,b) &:= a(x) \int h(x,y)b(y)\mu_2(\mathrm{d}y) = r_1(x)\ [\mu_1] \\ C(y,a,b) &:= b(y) \int h(x,y)a(x)\mu_1(\mathrm{d}\ x) = r_2(y)\ [\mu_2] \end{aligned} \tag{3.1}$$

for $x \in E_1, y \in E_2$, where h, r_i are defined as in Section 2. So the aim of the IPFP is to find solutions for the Schrödinger equation (2.2); the existence of solutions has been shown in Section 1.

The definition of the IPFP-algorithm is given recursively by:

$$\begin{aligned} b_0 := 1,\ a_0 := r_1,\ b_{n-1}(y) &:= \frac{r_2(y)}{\int h(x,y)a_n(x)\mu_1(\mathrm{d}x)} \quad \text{and} \\ a_{n-1}(x) &:= \frac{r_1(x)}{\int h(x,y)b_{n-1}(y)\mu_2(\mathrm{d}y)}. \end{aligned} \tag{3.2}$$

Convergence of the IPFP-algorithm in the finite discrete case was proved in papers by Brown (1959), Bishop and Fienberg (1969), Ireland and Kullback (1968), Fienberg (1970) and Csiszar (1975).

Defining the sequence of measures $(\mu^{(n)})$ by

$$\mu^{(2n)} \ := \ a_n \otimes b_n\mu, \quad \mu^{(2n-1)} \ := \ a_n \otimes b_{n-1}\mu \tag{3.3}$$

we find that

$$R(x, a_n, b_n) \ = \ r_1, \quad C(y, a_n, b_{n+1}) \ = \ r_2. \tag{3.4}$$

This means that $\mu^{(2n)}$ has the correct first marginal ν_1 while $\mu^{(2n+1)}$ has the correct second marginal ν_2, that is,

$$\mu^{(2n)} \in M(\nu_1), \quad \mu^{(2n+1)} \in M(\nu_2), \quad \text{for all } n \in \boldsymbol{N}, \tag{3.5}$$

and $(\mu^{(n)})$ is the sequence of alternating projections of $\mu = \mu^{(0)}$ on $M(\nu_2), M(\nu_1)$ w.r.t. the Kullback-Leibler distance. The idea is that $\mu^{(n)}$ converges to the projection ν^* of μ on $M(\nu_1, \nu_2) = M(\nu_1) \cap M(\nu_2)$.

In this sense the IPFP-algorithm is an analogue to the alternation algorithm (or backfitting algorithm) which was introduced in the case of Hilbert spaces by von Neumann (1950) and Aronszajn (1950), and extended to some further function spaces in many papers since then (cf. the survey in Light and Cheney (1985)). Both types of alternating projections have found important applications in fields such as tomography (cf. Hamaker and Solmon (1978)) in ridge type regression models, ACE (cf. Breiman and Friedman (1985), Stone (1985), Buja, Hastie and Tibshirani (1989)), in connection with Hoeffding's decomposition (cf. Rüschendorf (1985)), restricted least squares estimators (cf. Dykstra (1983), Gaffke and Mathar (1989)), projection pursuit density estimation (cf. Friedman, Stützle and Schroeder (1984)), and probabilistic expert systems (cf. Jirousek (1991)).

Recently a general convergence proof of the IPFP-algorithm was found under the following conditions.

THEOREM 3.1. *(Rüschendorf (1993c).) If for some positive constant* $c > 0, h/r_2 \geq c$ *a.s., then*

$$\begin{gathered} I\left(\mu^{(n)} \mid \mu\right) \to I(\nu^* \mid \mu) \qquad \text{and} \\ \mid \mu^{(n)} - \nu^* \mid \to 0, \quad I\left(\nu^* \mid \mu^{(n)}\right) \to 0, \end{gathered} \tag{3.6}$$

where $|\cdot|$ *denotes the total variation distance and* ν^* *is the* I*-projection of* μ *on* $M(\nu_1, \nu_2)$.

The proof of Theorem 3.1 is based on some geometric properties of the Kullback-Leibler distance. The essential step is to prove that the sequence $(a_n \otimes b_n) \subset L^1(\mu)$ is uniformly integrable. It is easy to see that the marginals $\pi_i\left(\mu^{(n)}\right)$ converge to the correct marginals ν_i. This, together with the lower semicontinuity of I w.r.t. the τ-topology (of setwise convergence), implies the convergence properties of $\mu^{(n)}$.

Motivated by the paper of Hobby and Pyke (1965), we next introduce a modification of the IPFP-algorithm which allows the use of monotonicity arguments to ensure pointwise convergence of a_n, b_n under alternative assumptions. Define the *modified IPFP-algorithm* recursively by:

1. $a_0(x) := \alpha > 0$.

2. If a_n is defined, then set

$$b_n(y) := \frac{r_2(y)}{\int h(x,y)a_n(x)\mu_1(\mathrm{d}x)}, \tag{3.7}$$

i.e. $C(y, a_n, b_n) = r_2(y)$.

3. If $R(x, a_n, b_n) < r_1(x)$, then define

$$a_{n+1}(x) := \frac{r_1(x)}{\int f(x,y)b_n(y)\mu_2(\mathrm{d}y)}, \tag{3.8}$$

i.e. $R(x, a_{n+1}, b_n) = r_1(x)$. If $R(x, a_n, b_n) \geq r_1(y)$, then define

$$a_{n+1}(x) := a_n(x). \tag{3.9}$$

THEOREM 3.2. *Assume that*

(A.1) $\quad h(x,y) \leq Cr_1(x)r_2(y)\ [\mu]$

(A.2) $\quad \int_A h(x,y)\mu_2(\mathrm{d}y) \geq u_1(x)\widetilde{\mu}_2(A), \quad A \in \mathcal{A}_2$ *for some finite measure* $\widetilde{\mu}_2 \approx \mu_2$ *and function* $u_1 \geq 0$ *with* $\int \frac{1}{u_1(x)}\nu_1(\mathrm{d}x) < \infty$*, then the modified IPFP-algorithm converges in total variation to the I-projection of* μ *on* $M(\nu_1, \nu_2)$ *and* a_n*,* b_n *converge pointwise.*

PROOF. Define

$$A_n := \{x \in E_1; \quad R(x, a_n, b_n) \leq r_1(x)\}. \tag{3.10}$$

By definition $a_n \leq a_{n+1}$, i.e., a_n converges isotonically to some function a and $b_{n+1} \leq b_n$, i.e., b_n converges antitonically to some function b. For $x \in A_n$ we have $R(x, a_n, b_n) \leq r_1(x)$ and, therefore, $R(x, a_{n+1}, b_{n+1}) = r_1(x)$. Since $b_{n+1} \leq b_n$, this implies that $R(x, a_{n+1}, b_{n+1}) \leq r_1(x)$, i.e. $x \in A_{n+1}$. Therefore, $A_n \subset A_{n+1}$ converges isotonically to some set A, i.e. $A_n \uparrow A$.

If $\mu_1(A_n) = 1$, then $R(x, a_n, b_n) = r_1(x)$ a.s. and we are done. So assume w.l.g. that $\mu_1(A_n) < 1$, for all n, equivalently, $\mu_1(A_n^c) > 0$, for all n and $A_n^c \downarrow A^c$.

For $x \in A_n^c$ we have $R(x, a_n, b_n) > r_1(x)$ and, therefore, $a_n(x) = \alpha$. This implies that

$$\begin{aligned} r_1(x) < R(x, a_n, b_n) &= \alpha \int h(x,y)b_n(y)\mu_2(\mathrm{d}y) \\ &\leq \alpha C\ r_1(x) \int r_2(y)b_n(y)\mu_2(\mathrm{d}y). \end{aligned} \tag{3.11}$$

Therefore, $\int r_2(y)b_n(y)\mu_2(\mathrm{d}y) \geq \frac{1}{\alpha C} > 0$, for all n and in the limit $\int r_2(y)b(y)\mu_2(\mathrm{d}y) \geq \frac{1}{\alpha C}$. This implies the existence of positive constants $r, \sigma > 0$ with $\mu(\{b \geq r\}) \geq \sigma > 0$.

Next for any $x \in A_n \subset A$ and any $n \in \boldsymbol{N}$ we have

$$r_1(x) \geq R(x, a_n, b_n) = a_n(x) \int h(x,y)b_n(y)\mu_2(\mathrm{d}y) \tag{3.12}$$

and, therefore,

$$\begin{aligned} r_1(x) &\geq a(x) \int h(x,y)b(y)\mu_2(\mathrm{d}y) \\ &\geq a(x)r \int_{\{b \geq r\}} h(x,y)\mu_2(\mathrm{d}y) \\ &\geq r a(x) u_1(x) \widetilde{\mu}_2(\{b \geq r\}). \end{aligned}$$

Since $\mu_2 \approx \widetilde{\mu}_2$ we conclude that $\widetilde{\mu}_2(\{b \geq r\}) > 0$ and, therefore, we have

$$a(x) \leq \begin{cases} \overline{c}\frac{r_1(x)}{u_1(x)} & \text{for } x \in A \\ \alpha & \text{for } x \in A^c \end{cases} \tag{3.13}$$

Denote the r.h.s. of (3.13) by $\widetilde{a}(x)$; by assumption $\widetilde{a}$ is integrable. Furthermore, for each $n \in \boldsymbol{N}$ we have

$$\begin{aligned} r_2(y) &= \int h(x,y)a_n(x)b_n(y)\mu_1(\mathrm{d}x) \\ &\leq b_n(y) \int h(x,y)\widetilde{a}(x)\mu_1(\mathrm{d}x). \end{aligned} \tag{3.14}$$

Note that

$$\begin{aligned} \int \left(\int h(x,y) \frac{r_1(x)}{u_1(x)} \mu_1(\mathrm{d}x) \right) \mu_2(\mathrm{d}y) &= \int \frac{r_1(x)}{u_1(x)} \mu_2(\mathrm{d}x) \\ &= \int \frac{1}{u_1(x)} \nu_1(\mathrm{d}x) < \infty, \end{aligned}$$

so that the integrand in (3.14) is a.s. finite. Therefore, from (3.14) we obtain

$$b(y) \geq \frac{r_2(y)}{\widetilde{b}(y)}, \quad \widetilde{b}(y) := \int h(x,y)\widetilde{a}(x)\mu_1(\mathrm{d}x). \tag{3.15}$$

From the dominated convergence theorem it follows that

$$\begin{aligned} r_2(y) &= \lim_n C(y, a_n, b_n) \\ &= b(y) \int h(x,y)a(x)\mathrm{d}\mu_1(x). \end{aligned} \tag{3.16}$$

For $x \in A$ we have

$$\begin{aligned} r_1(x) &= \lim R(x, a_{n+1}, b_n) \\ &= \lim a_{n+1}(x) \int h(x,y) b_n(y) \mu_2(\mathrm{d}y) \\ &= a(x) \int h(x,y) b(y) \mu_2(\mathrm{d}y), \end{aligned} \tag{3.17}$$

the integral being positive and finite by (A.2). Since for $x \in A$, we have $r_1(x) \geq R(x, a_{n+1}, b_{n+1}) \geq R(x, a_{n+1}, b_n)$, it follows that $R(x, a_{n+1}, b_{n+1})$ converges to $r_1(x)$.

For $x \in A^c$ we have from (3.11)

$$\begin{aligned} r_1(x) &\leq \lim R(x, a_n, b_n) \\ &= a(x) \int h(x,y) b(y) \mu_2(\mathrm{d}y). \end{aligned} \tag{3.18}$$

Again by dominated convergence and from the definition we obtain

$$\int R(x, a_n, b_n) \mathrm{d}\mu_1(x) = \int C(y, a_n, b_n) \mathrm{d}\mu_2(y) = 1$$

and, therefore, equality holds in (3.18), i.e. a and b are solutions of the Schrödinger equation. Finally, from the pointwise dominated convergence of the densities we conclude that we have convergence in total variation. ∎

REMARKS. The assumptions (A.1), (A.2) can be modified in the proof given above. First note that (A.2) is satisfied if

(A.2′) $h(x,y) \geq c r_1(x) r_2(y)$ holds for some $c > 0$ (with $\widetilde{\mu}_2 := c\nu_2$).

Condition (A.1) is used in (3.11). Here we can also use the estimate

$$\int h(x,y) b_n(y) \mu_2(\mathrm{d}y) \leq \left(\int h^2(x,y) \mu_2(\mathrm{d}y) \right)^{1/2} \left(\int b_n^2(y) \mu_2(\mathrm{d}y) \right)^{1/2}.$$

So it is possible to replace (A.1) by the assumption

(A.1′) $\displaystyle\int \left(\frac{h(x,y)}{r_1(x)} \right)^2 \mathrm{d}\mu_2(y) \leq C, \quad \int r_2^2(y) \mu_2(\mathrm{d}y) < \infty.$ ∎

4. Optimal Multivariate Couplings, Minimal ℓ_p-Metrics. The aim of this section is to describe extensions of the theory of optimal multivariate couplings, or transportation problems, begun in Section 3 of the paper Rüschendorf (1991b) and continued in Rüschendorf (1993d). In particular sufficient criteria and examples for minimal ℓ_p-metrics are discussed.

Let $c = c(x,y)$ be a coupling function on $\boldsymbol{R}^k \times \boldsymbol{R}^k$ and for $P_i \in M^1(\boldsymbol{R}^k, \boldsymbol{B}^k)$, let

$$M(c) := \sup\left\{ \int c \mathrm{d}\mu; \quad \mu \in M(P_1, P_2) \right\}, \tag{4.1}$$

where $M(P_1,P_2)$ are the measures with marginals P_i. Call $\mu^* \in M(P_1,P_2)$ an *optimal c-coupling* if $\int c\mathrm{d}\mu^* = M(c)$. Assume that $c(x,y)$ is bounded below by functions of the form $f \oplus g(x,y) = f(x) + g(y)$, $f \in L^1(P_1)$, $g \in L^1(P_2)$, i.e. $f \oplus g \leq c$. Then by the duality theorem:

$$M(c) = \inf\left\{ \int f \mathrm{d}P_1 + \int g \mathrm{d}P_2;\ f \in L^1(P_1), g \in L^1(P_2), f \oplus g \geq c \right\}. \tag{4.2}$$

For an introduction to coupling problems and duality results of this kind we refer to Kellerer (1984).

A function f is called *c-convex* if for some index set I

$$f(x) = \sup_{i \in I}(c(x,y_i) + a_i).$$

This notation generalizes convexity which is a special case with $c(x,y) = x \cdot y$. It has been studied in several recent papers on non-convex optimization theory (cf. Elster and Nehse (1974), and Dietrich (1988). Denote the c-conjugate of f

$$f^*(y) := \sup_x (c(x,y) - f(x)). \tag{4.3}$$

Then f is c-convex if and only if $f = f^{**}$. The *c-subgradient* of f in x is defined by

$$\partial_c f(x) = \{ y : f(z) - f(x) \geq c(z,y) - c(x,y) \quad \text{for all } z \} \tag{4.4}$$

For the determination of optimal c-couplings the following result in Rüschendorf (1991b) is fundamental. Assume that $\int c(x,y)\mathrm{d}P_i(x) < \infty$, $i = 1,2$, and that c is bounded below as above. Then $\mu^* \in M(P_1,P_2)$ is an optimal c-coupling induced by random variables X, Y, if and only if

$$Y \in \partial_c f(X) \quad \text{a.s. for some } c\text{-convex function } f. \tag{4.5}$$

An interesting consequence of (4.5) is the existence of a pair X, Y with $X \sim P_1$, $Y \sim P_2$ and $Y \in \partial_c f(X)$ a.s. for some c-convex function f. A condition equivalent to (4.5) is that the support Γ of μ^* is *c-cyclically monotone* i.e. for all $(x_1,y_1), \ldots, (x_n,y_n) \in \Gamma$ we have

$$\sum_{i=1}^{n} (c(x_{i+1},y_i) - c(x_i,y_i)) \leq 0, \tag{4.6}$$

where $x_{n+1} := x_1$ (cf. Dietrich (1988) and Smith and Knott (1992).

The case in which $c(x,y) = \mid x-y \mid^p$, $x,y \in \boldsymbol{R}^k$, $p \geq 1$, has been discussed previously for the euclidean distance and $p = 2$ in Knott and Smith (1984) and Rüschendorf and Rachev (1990) and several examples have been constructed in Cuesta-Albertos, Rüschendorf, and Tuero-Diaz (1993), while the case of general minimal ℓ_p-metrics, $p > 1$ has been considered in Rüschendorf (1991a). In Rüschendorf (1991b) it was shown that for a continuous, differentiable, injective function ϕ, a pair of random variables $X \sim P_1$, $\phi(X) \sim P_2$ is an optimal c-coupling, if for all x, y in the support of P_1 we have

$$\int_{y \to x} (c_1(u, \phi(y)) - c_1(u, \phi(u)))\mathrm{d}u \leq 0, \tag{4.7}$$

where $c_1(u,v) = \frac{\partial}{\partial u} c(u,v)$ (assuming that $c_1(u,\phi(u))\mathrm{d}u$ is closed). It is easy to see that for $f(x) = \int_{0 \to x} c_1(u,\phi(u))\mathrm{d}u$ the condition that $\phi(y) \in \partial_c f(y)$ is equivalent to (4.7) for any (not necessarily differentiable) function ϕ (cf. Rüschendorf (1993d)). Therefore, (4.7) is a characterization of the c-optimality of any function ϕ.

In the following, we discuss some necessary and some sufficient conditions for the optimality of couplings induced by functions ϕ for the pairs $(X, \phi(X))$. For a more detailed discussion and proofs we refer to Rüschendorf (1993d). The idea of the following simple sufficient condition is from Smith and Knott (1992).

PROPOSITION 4.1. *If $c(\cdot, y)$ is concave and if $h(u) := c_1(u, \phi(u))$ is cyclically monotone on the support of P_1, then $(X, \phi(X))$ is an optimal c-coupling.*

PROOF. Let $x_1, \ldots, x_n$ be in the support of P_1, then by concavity of $c(\cdot, y)$:

$$\begin{aligned}\sum_{i=1}^n (c(x_{i+1}, \phi(x_i)) - c(x_i, \phi(x_i)) &\leq \sum_{i=1}^n c_1(x_i, \phi(x_i)) \cdot (x_{i+1} - x_i) \\ &= \sum_{i=1}^n h(x_i)(x_{i+1} - x_i) \leq 0. \qquad \blacksquare\end{aligned}$$

It is well known that cyclically monotone functions arise from gradients of convex functions. If ϕ is continuously differentiable, then ϕ is cyclically monotone if and only if $\left(\frac{\partial \phi_i}{\partial x_j}\right)$ is symmetric and ϕ is monotone, i.e. $(y - x)(\phi(y) - \phi(x)) \geq 0$ (cf. Rüschendorf (1991b) and Levin (1992)).

EXAMPLE 4.1. Consider the important example $c(x,y) = - \mid x - y \mid^p$, $x, y \in \boldsymbol{R}^k$, $p > 1$. Then $c_1(x,y) = -p \mid x - y \mid^{p-2} (x - y)$ and if $h(u)$ is cyclically monotone, then the equation

$$- \mid x - \phi(x) \mid^{p-2} (x - \phi(x)) \;=\; h(x) \tag{4.8}$$

is uniquely solved by

$$\phi(x) \;= \mid h(x) \mid^{-\frac{p-2}{p-1}} h(x) + x \;=\; \phi_h(x). \tag{4.9}$$

So for any monotone function h we obtain that $(X, \phi_h(X))$ is an optimal (minimal) ℓ_p-coupling. In particular, for $p = 2$ we recover the optimal coupling result for minimal ℓ_p-metrics. If $h(x) = Ax$ is a linear cylically monotone function, i.e. A is symmetric, positive semidefinite, then $\phi(x) = (x^T A^2 x)^{-\frac{p-2}{p-1}} Ax - x$. If $h(x) = \alpha(\mid x \mid)\frac{x}{|x|}$, where α is increasing, is a radial transformation, then

$$\begin{aligned}\phi(x) &= (\alpha(\mid x \mid))^{-\frac{p-2}{p-1}} \alpha(\mid x \mid)\frac{x}{\mid x \mid} + x = \left((\alpha(\mid x \mid))^{\frac{1}{p-1}} + \mid x \mid\right)\frac{x}{\mid x \mid} \\ &= h(\mid x \mid)\frac{x}{\mid x \mid}\end{aligned} \tag{4.10}$$

is again a radial transformation. For this case cf. also Cuesta-Albertos, Rüschendorf, and Tuero-Diaz (1993). A partial result in the case $1 < p \leq 2$ has been obtained previously by Smith and Knott (1992). Of course the assumption of concavity of c can be weakened to the condition that $\bar{c}(x, y) = c(x, y) + h(x)$ is concave in x for some function $h(x)$.

REMARK. Proposition 4.1 also holds more generally for the support of optimal couplings as formulated in Smith and Knott (1992). It can also be derived directly from criterion (4.7). Let $g_y(u) := c_1(u, \phi(y)) - c_1(u, \phi(u))$. If $c(\cdot, z)$ is concave, then $-c_1(\cdot, \phi(y))$ is cyclically monotone. This implies $(g_y(v) - g_y(u), v - u) = (c_1(v, \phi(y)) - c_1(u, \phi(y)), v - u) \leq 0$ and, therefore, the path-integral

$$F_y(x) \;:=\; \int_{x_o \to x} g_y(u) \cdot \mathrm{d}u \tag{4.11}$$

is concave. From (4.11) we conclude that

$$F_y(x) - F_y(y) \;\leq\; g_y(y)(x - y) \;=\; 0, \tag{4.12}$$

i.e., that condition (4.7) holds. ∎

Note that by definition $y^* \in \partial_c f(y)$ iff

$$f(x) - c(x, y^*) \;\geq\; f(y) - c(y, y^*) \tag{4.13}$$

for all x, i.e. $\varphi^y(x) \geq \varphi^y(y)$, for all x, where $\varphi^y(y) := f(x) - c(x, y^*)$. For $y^* = \phi(y)$, $\varphi^y(x) = \int_{y \to x}(c_1(u, \phi(u)) - c_1(u, \phi(y))\mathrm{d}u = -F_y(x)$ if $c_1(u, \phi(u))\mathrm{d}u$

is closed. So (4.13) is equivalent to the condition that $x \to F_y(x)$ has its maximum in $x = y$. Since

$$\frac{\partial}{\partial x} F_y(y) \;=\; -\frac{\partial}{\partial x}\varphi^y(y) \;=\; 0, \tag{4.14}$$

and with $y^* = \phi(y)$,

$$\begin{aligned} B(x,y) &:= -\frac{\partial^2}{\partial x \partial x'}\varphi^y(x) \\ &= \frac{\partial^2}{\partial x \partial x'} c(x, \phi(y)) - \frac{\partial^2}{\partial x \partial x'} c(x, \phi(x)) \\ &\quad - \frac{\partial^2}{\partial x \partial y} c(x, \phi(x)) \frac{\partial \phi}{\partial x}(x), \end{aligned} \tag{4.15}$$

we obtain (cf. Rüschendorf (1993d)).

Proposition 4.2. *Let c be differentiable and $c_1(u, \phi(u))\mathrm{d}u$ be closed.*

a) A sufficient condition for the c-optimality of ϕ is given by

$$B(x,y) \;\leq\; 0. \tag{4.16}$$

b) A necessary condition for the c-optimality of ϕ is given by

$$-B(y,y) \;=\; \frac{\partial^2}{\partial x \partial y} c(y, \phi(y)) D\phi(y) \;\geq\; 0. \tag{4.17}$$

For $c(x,y) = -p \mid x - y \mid^p$, $1 < p$, $\mid \cdot \mid$ the euclidean distance, (4.17) leads to the necessary condition $\left((p-2)\frac{(y-\phi(y))\cdot(y-\phi(y))^T}{|y-\phi(y)|^2} - I\right) D\phi(y) \leq 0$, which in the case $p = 2$ is equivalent to the necessary and sufficient conditon $D\phi(y) \geq 0$.

The next sufficient condition does not assume that $c(\cdot, y)$ is concave.

Proposition 4.3. *If $c_1(u, \phi(u))\mathrm{d}u$ is closed and if for all (x, y) in the support of the distribution of (X, Y)*

$$\langle x - y, c_1(x, \phi(x)) - c_1(x, \phi(y))\rangle \;\geq\; 0, \tag{4.18}$$

then ϕ is a c-optimal coupling for $P_1 = P^X$, $P_2 = P^Y$.

Proof. For the c-optimality of ϕ it is sufficient by (4.7) to prove that

$$F_y(x) = \int_{y \to x} (c_1(u, \phi(y)) - c_1(u, \phi(u)))\mathrm{d}u \leq 0 = F_y(y).$$

Let $x_t := y + t(x-y)$, $t \geq 0$ and $H(t) := F_y(x_t)$, then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}H(t) &= \left\langle \frac{\partial}{\partial x}F_y(x_t), x-y \right\rangle \\ &= \langle c_1(x_t,\phi(y)) - c_1(x_t,\phi(x_t)), x-y\rangle \qquad (4.19)\\ &= \frac{1}{t}\langle c_1(x_t,\phi(y)) - c_1(x_t,\phi(x_t)), x_t-y\rangle \ \leq\ 0.\end{aligned}$$

This implies that

$$F_y(x) = F_y(y) + \int_0^1 \frac{\mathrm{d}H}{\mathrm{d}t}(t)\mathrm{d}t \ \leq\ 0. \qquad \blacksquare$$

REMARKS.

a) In the case $c(x,y) = -\mid x-y\mid^p$, define $g_x(y) :=\mid x-\phi(y)\mid^{p-2}(x-\phi(y))$. Then condition (4.18) reads

$$\langle x-y, \mid x-\phi(x)\mid^{p-2}(x-\phi(x)) - \mid x-\phi(y)\mid^{p-2}(x-\phi(y))\rangle \leq 0;$$

for $p=2$ this is equivalent to the monotonicity of ϕ. As a consequence we obtain that all functions of the form $\phi(x) =\mid h(x)\mid^{\frac{2-p}{p-1}} h(x)+x$, where h is a cyclically monotone function, are c-optimal coupling functions. For this and several further applications to ℓ_1-metrics we refer to Rüschendorf (1993d). Even in the one-dimensional case Proposition 4.3 leads to new results.

b) Some ideas related to this section can be found in the recent paper of Levin (1992) in the context of the transshipment problem (with fixed difference of the marginals). For differentiable cost functions Levin obtains an explicit formula for the optimal value of the problem, while we obtain some characterizations of the optimal transportation plans (cf. also Section 5.2).

c) The sufficient condition for optimality, $B(x,y) \leq 0$, implies that $F_y(x)$ is concave. If we can assure the weaker condition that F_y is quasi-concave, i.e.

$$F_y(\alpha x_1 + (1-\alpha)x_2) \ \geq\ \min(F_y(x_1), F_y(x_2)),$$

then a local minimum is either situated in a domain where F_y is constant or it is already a global minimum (cf. Roberts and Varberg (1973)). Therefore, the sharpened necessary condition that $B(y,y) < 0$, is already a sufficient condition for c-optimality of ϕ.

5. Optimal Couplings Under Additional and Relaxed Restrictions. Recently, some modifications of the usual coupling (transportation) problem in which additional or relaxed restrictions on the underlying Fréchet class of distributions with given marginals are introduced have been concidered. The aim is to obtain Fréchet bounds for these modified Fréchet classes. Some classes of restrictions of this type have been investigated in Olkin and Rachev (1990), Rachev and Rüschendorf (1994, 1993), and Levin (1992). In the final part of this paper we review some of the recent developments.

5.1. Order Restrictions. The following coupling (transportation) problem was posed by Rogers (1992). Let F, G be 1-dimensional distribution functions, with $F \leq_{\mathrm{st}} G$, i.e. F stochastically smaller than G, let $C := \{(x,y) \in \boldsymbol{R}^2;\ x \leq y\}$, and let

$$M_C(F,G) := M(F,G) \cap \{\mu \in M^1(\boldsymbol{R}^2, \boldsymbol{B}^2);\quad \mu(C) = 1\}$$

be the set of all measures with marginals F, G which are concentrated on the order cone C. We consider the following problem: For a strict convex function φ, determine

$$\sup\left\{\int \varphi(y-x)\mu(\mathrm{d}x, \mathrm{d}y);\quad \mu \in M_C(F,G)\right\} \tag{5.2}$$

Note that the corresponding inf problem is well-known and independent of the order restriction. The motivation for problem (5.1) is to get a good monotone coupling of random walks (S_n), (S'_n) with $S'_0 = x \geq S_0 = 0$, $S'_n \geq S_n$ for all n and $S'_n = S_n$ for all large enough n. Without the order restriction a solution of (5.2) is given by the random variables $X = F^{-1}(U)$, $Y = G^{-1}(1-U)$ where U is a rv uniformly distributed on $(0,1)$. It is intuitively clear that a solution of (5.2) should concentrate as much mass as possible on the diagonal. This is indeed true, as was shown by Rogers (1992):

Each solution (X,Y) of (5.2) has the property that

$$P(X=Y) = |\, F \wedge G \,| = \int f \wedge g \mathrm{d}m \tag{5.3}$$

if $F = fm$, $G = gm$. Moreover a solution of (5.2) exists.

It is possible to characterize optimal solutions by an ordering property (cf. Rachev and Rüschendorf (1993)).

PROPOSITON 5.1. *Let X, Y be rv's with df's F, G and $X \leq Y$ a.s. Then (X,Y) defines a solution of (5.2) iff*

$$X(\omega) < X(\omega') \leq Y(\omega) \leq Y(\omega') \quad \textit{implies} \quad Y(\omega') = Y(\omega) \tag{5.4}$$

a.s. for the pairs (ω, ω') w.r.t. the product measure.

For finite discrete distributions one can explicitly construct optimal pairs with the ordering property given in (5.4). First consider the case of equiprobable atoms. Let $\mu_1 = \frac{1}{n}\sum_{i=1}^n \varepsilon_{a_i}$, $\mu_2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_{b_i}$ be the measures corresponding to F, G, where $a_1 \leq \ldots \leq a_n$, $b_1 \leq \ldots \leq b_n$ and $a_i \leq b_i$, for all i. Problem (5.2) is equivalent to the following problem:

Find a permutation $\pi \in \Upsilon_n$ such that

$$\sum_{i=1}^n \varphi(b_i - a_{\pi(i)}) \;=\; \max! \tag{5.5}$$

over all permutations $\pi \in \Upsilon_n$ such that $a_{\pi(i)} \leq b_i$, $1 \leq i \leq n$, which we call admissible permutations. An optimal admissible permutation is essentially unique (up to indices with equal values of a_i) and given by the following proposition (cf. Rachev and Rüschendorf (1993)).

PROPOSITION 5.2. *Define $\pi^* \in \Upsilon_n$ inductively by:*

$$\begin{aligned} &\pi^*(1) := \max\{k \leq n;\ a_k \leq b_1\} \\ &\pi^*(k) := \max\{\ell \leq n;\ \ell \notin \{\pi^*(1), \ldots, \pi^*(k-1)\}, a_\ell \leq b_k\}, \quad 2 \leq k \leq n. \end{aligned} \tag{5.6}$$

Then $\pi^* \in \Upsilon_n$ is the optimal admissible permutation.

So up to a simultaneous permuation of the probability space an optimal pair of rv's is essentially unique.

REMARK. Proposition 5.2 extends to the case in which $\mu_1 = \sum_{i=1}^n p_i \varepsilon_{a_i}$, $\mu_2 = \sum_{i=1}^n q_i \varepsilon_{b_i}$ with rational p_i, q_i, by representing p_i, q_i in the formal equiprobable case. By an approximation argument as given in Rogers (1992) this allows one to approximate optimal couplings for F, G with compact support. The general case then can be approximated via ordering criterion (5.4) using truncation.

5.2. Marginals with a Given Difference. Consider the class $\mathcal{F}_H$ of all distributions on $\boldsymbol{R}^2$ with fixed difference of the marginal df's $F_1 - F_2 = H$ and consider the transportation problem with the following relaxed condition on the marginals:

$$\text{Minimize } \int c(x,y)\mathrm{d}F(x,y) \quad \text{subject to} \quad F \in \mathcal{F}_H, \tag{5.7}$$

where $c(x,y)$ is a symmetric, nonnegative and continuous cost function. Problem (5.7) is a generalization of the Kantorovich-Rubinstein problem. The following result was proved by Rachev and Rüschendorf in 1991. (The corresponding paper is Rachev and Rüschendorf (1993).)

THEOREM 5.3. *(cf. Rachev and Rüschendorf (1994)).). Assume that* $c(x,y) = \mid x - y \mid \zeta(x,y) \geq 0$, *where* $\zeta(x,y)$ *is symmetric and continuous on the diagonal,* $t \to \zeta(t,t)$ *is locally bounded and* $\zeta(t,t) \leq \zeta(x,y)$ *holds for* $x < t < y$, *then*

$$\inf\left\{\int c(x,y)\mathrm{d}F(x,y);\ F \in \mathcal{F}_H\right\} = \int \zeta(t,t) \mid H \mid (t)\mathrm{d}t. \tag{5.8}$$

From (5.8) one sees that only the behaviour of the cost function on the diagonal enters the formula.

In the multivariate case the following upper bound was established in Rachev and Rüschendorf (1994) for the case of $c_p(x,y) = \mid x - y \mid_p = \left(\sum \mid x_i - y_i \mid^p\right)^{1/p}, 1 \leq p$. Let $\mathcal{F}_H$ be the class of distributions on $\boldsymbol{R}^k \times \boldsymbol{R}^k$ with fixed difference H of the marginals.

THEOREM 5.4. *(cf. Rachev and Rüschendorf (1994)). If* H *has Lebesgue density* h, *then*

a) $\inf\left\{\int_{\boldsymbol{R}^{2k}} \mid x - y \mid_p \mathrm{d}F(x,y);\ F \in \mathcal{F}_H\right\} \leq \int_{\boldsymbol{R}^k} \mid y \mid_p J_H(y) \mid \mathrm{d}y$, *where* $J_H(y) = \int_0^1 t^{-(k+1)} h\left(\frac{y}{t}\right) \mathrm{d}t.$

b) If there exists a continuous function $g : \boldsymbol{R}^k \to \boldsymbol{R}^1$, a.e. *differentiable and such that for* $p = 1$,

$$\nabla g(y) = (\ sg(y_i\ J_H(y)))\ \text{a.e.} \tag{5.9}$$

and for $p > 1$,

$$\nabla g(y) = \left(sg(y_i\ J_H(y))\left(\frac{\mid y_i \mid}{\mid y \mid_q}\right)^{q/p}\right), \tag{5.10}$$

where $\frac{1}{q} + \frac{1}{p} = 1$ *and* sg *denotes the sign function, then equality in* a) *holds.*

The derivatives in (5.9), (5.10) may be considered in the weak sense.

For the case of bounded, differentiable cost functions on $\boldsymbol{R}^k \times \boldsymbol{R}^k$ Levin (1992) obtained the following interesting result:

THEOREM 5.5. *Let* c *be bounded and defined on* $X \times X$, *where* X *is a domain in* $\boldsymbol{R}^k$, $c(x,x) = 0$, $\{c \leq \alpha\}$ *analytic for all* α, c *continuously differentiable in some open neighbourhood of the diagonal and assume the existence* *of some function h on X with $h(x) - h(y) \leq c(x, y)$, then*

$$\begin{aligned} \inf\left\{\int c(x,y)\mathrm{d}F(x,y);\ F \in \mathcal{F}_H\right\} &= \int_X u_0(x)\mathrm{d}H(x), \quad \textit{where} \\ u_0(x) &= \int_{x_0 \to x} c_1(u,u)\mathrm{d}u, \quad c_1(u,v) := \frac{\partial}{\partial u}c(u,v). \end{aligned} \tag{5.11}$$

The proof of (5.11) shows that differentiability on the diagonal is crucial since it reduces the dual problem to a trivial situation. In the real (and euclidean) case the most natural cost function $\mid x - y \mid$ is excluded while for the differentiable functions $\mid x - y \mid^\alpha$, $\alpha > 1$, the infimum is trivially zero (cf. (5.8)). It would, therefore, be of interest to have a version of (5.11) in the nondifferentiable case. From the proofs it is not clear how an optimal plan (if it exists) can be constructed.

5.3. A Duality Principle. Let $c = c(u, v)$, $u, v \in \boldsymbol{R}^1$ be a Monge function, i.e., $\Delta_x^y c \leq 0$ for all $x, y \in \boldsymbol{R}^2$, $x \leq y$ where Δ_x^y is the multivariate difference operator. It is well known that the Kantorovich-functional

$$\widehat{A}_c(P_1, P_2) = \inf\{Ec(X,Y);\ X \sim P_1,\ Y \sim P_2\} \tag{5.12}$$

is given by

$$\widehat{A}_c(P_1, P_2) = Ec(F_1^{-1}(U), F_2^{-1}(U)), \tag{5.13}$$

here each F_i is the df of P_i and U is a rv uniformly distributed on $(0,1)$. The following simple property of Monge functions leads to an interesting duality principle (cf. Rachev and Rüschendorf (1993)).

$$\begin{aligned} &\text{If } c \text{ is a Monge function, then} \\ &\widetilde{c}(u,v) := -c(-u,v) \text{ is a Monge function.} \end{aligned} \tag{5.14}$$

As a first application consider a Monge function c and define the sup-problem:

$$\widetilde{A}_c(P_1, P_2) := \sup\{Ec(X,Y);\ X \sim P_2,\ Y \sim P_2\}. \tag{5.15}$$

From the duality result in (5.14) we obtain the following (well-known) dual result for the sup-problem which in this case is also obvious from the lower Fréchet bounds.

For a Monge function c we have

$$\widetilde{A}_c(P_1, P_2) = Ec(F_1^{-1}(1 - U), F_2^{-1}(U)), \tag{5.16}$$

where U is a rv uniformly distributed on $(0,1)$.

For an application with additional restrictions let $\Gamma(x,y)$ be a measure generating function with

$$\Gamma(x,y) \geq (F_1(x)+F_2(y)-1)_+ =: \underline{F}(x,y). \tag{5.17}$$

Define $\mathcal{F}_\Gamma = \mathcal{F}_\Gamma(P_1,P_2)$ to be the set of all distribution functions F with marginals P_1, P_2 and such that F is bounded above by Γ

$$F(x,y) \leq \Gamma(x,y), \quad \text{for all } x,y. \tag{5.18}$$

Define

$$\begin{aligned} F^*(x,y) := &\inf_{\substack{u\leq x\\ v\leq y}} \{\Gamma(u,v)+(F_1(x)-F_1(u))+(F_2(y)-F_2(v))\} \\ &\wedge \min\{F_1(x),F_2(y)\}, \end{aligned} \tag{5.19}$$

then Barnes and Hoffmann (1985) in the discrete case and Olkin and Rachev (1990) in the general case proved:

$$F^* \in \mathcal{F}_\Gamma \text{ and } \inf\left\{\int c(x,y)\mathrm{d}F(x,y);\ F\in\mathcal{F}_\Gamma\right\} = \int c(x,y)\mathrm{d}F^*(x,y). \tag{5.20}$$

In particular F^* is the upper Fréchet bound corresponding to $\mathcal{F}_\Gamma$.

From the duality principle we can infer the dual result to (5.20). For a measure μ let G_μ be defined by

$$G_\mu(x,y) := P(X\leq x, Y\geq y), \tag{5.21}$$

where $(X,Y)\sim\mu$. Define

$$\mathcal{G}_\triangle := \{G_\mu;\ \mu\in M(P_1,P_2),\ G_\mu(x,y) \leq \triangle(x,y) \quad \text{for all } x,y\}, \tag{5.22}$$

where $\triangle$ is a measure defining function defining a measure δ by $\triangle(x,y) = \delta((-\infty,x]\times[y,\infty))$. Define

$$\widetilde{A}_\triangle(c) := \sup\left\{\int c\mathrm{d}\mu;\ G_\mu \leq \triangle,\ \mu\in M(P_1,P_2)\right\} \tag{5.23}$$

and

$$\widetilde{F}_2(y) := 1-F_2((-y)-) \tag{5.24}$$

$$\begin{aligned} \widetilde{F}^*(x,y) := &\inf_{\substack{u\leq x\\ v\geq -y}} \{\triangle(u,v)+(F_1(x)-F_1(u)) \\ &+F_2((-v)-)-F_2((-y)-)\} \wedge \min\{F_1(x),\widetilde{F}_2(y)\}. \end{aligned} \tag{5.25}$$

The simple duality argument of (5.14) allows one to obtain the solution of the transportation problem (5.23) without additional calculation. A direct proof of the following result was given in Olkin and Rachev (1990).

PROPOSITION 5.6. *$\widetilde{F}^*$ is a df with marginals F_1, F_2. The corresponding measure μ^* satisfies: $G_{\mu^*} \in \mathcal{G}_\triangle$ and*

$$\widetilde{A}_\triangle(c) \;=\; \int c\mathrm{d}\mu^*. \tag{5.26}$$

5.4. Majorized Fréchet Bounds. On a measure space $(X, \mathcal{B})$, let $\mathcal{B}_i \subset \mathcal{B}$ be sub-σ-algebras, $1 \le i \le n$, and $P_i \in M^1(X, \mathcal{B}_i)$, let μ be a finite measure on $(X, \mathcal{B})$, and define

$$M_\mu \;:=\; \{P \in M^1(X, \mathcal{B});\; P/\mathcal{B}_i = P_i,\; 1 \le i \le n,\; P \le \mu\}. \tag{5.27}$$

Assume that $M_\mu \ne \emptyset$ and define

$$U_\mu(\varphi) \;:=\; \inf\left\{U(\varphi_0) + \int h\mathrm{d}\mu;\; h \ge 0,\; \varphi_0 + h \ge \varphi\right\}, \tag{5.28}$$

where

$$U(\varphi_0) \;:=\; \inf\left\{\sum_{i=1}^n \int f_i\mathrm{d}P_i;\; f_i \in L^1(\mathcal{B}_i, P_i),\; \varphi_0 \le \sum_{i=1}^n f_i\right\}.$$

U is the dual operator for the pure marginal problem and for all lower majorized measurable functions φ_0

$$\sup\left\{\int \varphi_0\mathrm{d}P;\; P \in M(P_1, \ldots, P_n)\right\} \;=\; U(\varphi_0) \tag{5.29}$$

holds (cf. e.g. Kellerer (1984), Rüschendorf (1991b)). Similarly, U_μ is the dual operator for the majorized marginal problem. A linear operator S is majorized by U_μ,

$$S \le U_\mu \text{ iff } S \ge 0,\; S/\mathcal{B}_i = P_i,\; 1 \le i \le n \quad\text{and}\quad S \le \mu. \tag{5.30}$$

Therefore, the approach to duality theorems as developed in Rüschendorf (1991b), Section 2.1, yields the duality theorem

$$U_\mu(\varphi) \;=\; \sup\left\{\int \varphi\mathrm{d}P;\; P \in M_\mu\right\} \;=:\; M_\mu(\varphi) \tag{5.31}$$

for upper semicontinuous or uniformly approximable integrable functions φ in the situation of compactly approximable measure spaces $(X, \mathcal{B}_i, P_i)$ with countable topological basis. In some sense (5.31) gives the duality result for the general case of order restrictions as considered e.g. in Sections 5.1, 5.2.

We want to consider the question of more explicit evaluations of the dual operator U_μ for the case $\varphi = 1_B, B \in \mathcal{B}$. In this sense we try to establish the sharp upper Fréchet bounds in the class M_μ. Define $M_\mu(B) := M_\mu(1_B)$ and assume the duality (5.31).

PROPOSITION 5.7. *(cf. Rachev and Rüschendorf (1993)).*

$$M_\mu(B) = \sup_{P \in M(P_1,\dots,P_n)} P \wedge \mu(B), \tag{5.32}$$

where $P \wedge \mu$ is the infimum in the lattice of measures.

PROOF. From (5.31)

$$\begin{aligned} M_\mu(B) &= \inf\{\mu(h) + U(\varphi);\ h \geq 0,\ \varphi + h \geq 1_B\} \\ &= \inf\{\mu(h) + U(1_B - h);\ 0 \leq h \leq 1_B\}. \end{aligned} \tag{5.33}$$

The last step follows by taking $\varphi = (1_B - h)_+$, so $0 \leq \varphi$ and w.l.g. $h \leq 1_B$. Next using the integration trick of Strassen (1965)

$$\begin{aligned} M_\mu(B) &= \inf_{0 \leq h \leq 1_B} \sup_{P \in M(P_1,\dots,P_n)} \{\mu(h) + P(1_B - h)\} \\ &= \inf_{0 \leq h \leq 1_B} \sup_P \left\{ \int_0^1 \mu(h > t) \mathrm{d}t + \int_0^1 P(1_B - h \geq 1 - t) \mathrm{d}t \right\}. \end{aligned} \tag{5.34}$$

With $C_t := \{h > t\} \subset B$ we have

$$\{x : h(x) \leq 1_B(x) - 1 + t\} = \{x \in B;\ h(x) \leq t\} = B \setminus C_t.$$

Therefore,

$$\begin{aligned} M_\mu(B) &= \inf_{0 \leq h \leq 1_B} \sup_P \int_0^1 (\mu(C_t) + P(B \setminus C_t)) \mathrm{d}t \\ &\geq \sup_P \inf_{C \subset B} \{\mu(C) + P(B \setminus C)\} \\ &= \sup_P \mu \wedge P(B). \end{aligned}$$

On the other hand obviously

$$\begin{aligned} M_\mu(B) &= \sup\{P(B);\ P \in M(P_1,\dots,P_n),\ P \leq \mu\} \\ &= \sup\{P \wedge \mu(B);\ P \in M(P_1,\dots,P_n),\ P \leq \mu\} \\ &\leq \sup_{P \in M(P_1,\dots,P_n)} P \wedge \mu(B). \end{aligned}$$

∎

Proposition 5.7 allows one to reduce the problem of the majorized Fréchet bounds to the problem of "usual" Fréchet bounds but for a more complicated functional. It remains an open problem to determine more explicit formulas for $M_\mu(B)$. Some special instances of explicit results for this case of local upper bounds have been solved in Rachev and Rüschendorf (1994).

References

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68**, 377–404.

BARNES, E. R. and HOFFMANN, A. J. (1985). On transportation problems with upper bounds on leading rectangles. *SIAM J. Alg. Discr. Meth.* **6**, 487–496.

BEURLING, A. (1960). An automorphism of product measures. *Ann. Math.* **72**, 189–200.

BISHOP, Y. M. and FIENBERG, S. E. (1969). Incomplete two dimensional contingency tables. *Biometrics* **55**, 119–128.

BREIMAN, L. and FRIEDMAN, J. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80**, 580–619.

BROWN, D. T. (1959). A note on approximations to discrete probability distributions. *Inform. and Control* **2**, 386–392.

BUJA, A., HASTIE, T. J. and TIBSHIRANI, R. J. (1989). Linear smoothers and additive models. *Ann. Statist.* **17**, 453–555.

CSISZAR, I. (1975): I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3**, 146–158.

CUESTA-ALBERTOS, J. A., RÜSCHENDORF, L. and TUERO-DIAZ, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. *J. Multivariate Anal.* **46**, 335–361.

DALL'AGLIO, G., KOTZ, S. and SALINETTI, G. (Eds.) (1991). Advances in Probability Distributions with Given Marginals. Kluwer Academic Publishers.

DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11**, 427–444.

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

DYKSTRA, R. L. (1983). An algorithm for restricted least squares regression. *J. Amer. Statist. Assoc.* **78**, 837–842.

ELSTER, K. H. and NEHSE, R. (1974). Zur Theorie der Polarfunktionale. *Optimization* **5**, 3–21.

FIENBERG, S. E. (1970). An iterative procedure for estimation in contingency tables. *Ann. Math. Statist.* **41**, 907–917.

FÖLLMER, H. (1988). Random fields and diffusion processes. In: LNM 1362, 101–203, Springer.

FORTET, P. R. (1940). Résolution d'un système d'équations de M. Schrödinger. *J. Math. Pures et Appl.* **9**, 83–105.

FRIEDMAN, J. H., STÜTZLE, W. and SCHRÖDER, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79**, 599–608.

GAFFKE, N. and MATHAR, R. (1989). A cyclic projection algorithm via duality. *Metrika* **36**, 29–54.

HAMAKER, C. and SOLMON, D. C. (1978). The angles between the null spaces of x-rays. *J. Math. Anal. Appl.* **62**, 1–23.

HOBBY, C. and PYKE, R. (1965). Doubly stochastic operators obtained from positive operators. *Pac. J. Math.* **15**, 153–157.

IRELAND, C. T. and KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika* **5**, 179–188.

JAMISON, B. (1974). Reciprocal processes. *Z. Wahrsch. verw. Gebiete* **30**, 65–86.

JIROUSEK, R. (1991). Solution of the marginal problem and decomposable distributions. *Kybernetika* **27**, 1–9.

KELLERER, H. (1984). Duality theorems for marginal problems. *Z. Wahrsch. verw. Gebiete* **67**, 399–432.

KNOTT, M. and SMITH, C. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43**, 39–49.

LEVIN, V. I. (1992): A formula for the optimal value in the Monge-Kantorovich problem with a smooth cost function and a characterization of cyclically monotone mappings. *Math. USSR Sbornik* **71**, 533–548.

LIGHT, W. A. and CHENEY, E. W. (1985). Approximation Theory in Tensor Product Spaces. LNM 1169. Springer.

OLKIN, I. and RACHEV, S. T. (1990). Marginal problems with additional constraints. Technical Report 270, Stanford University.

RACHEV, S. T. and RÜSCHENDORF, L. (1990). A transformation property of minimal metrics. *Theory Probab. Appl.* **35**, 131–137.

RACHEV, S. T. and RÜSCHENDORF, L. (1994). Solution of some transportation problems with relaxed or additional contraints. SIAM J. Control Optimiz. 32, 673-689

RACHEV, S. T. and RÜSCHENDORF, L. (1993). On constrained transportation problems. *Proceedings of 32nd Conference on Decision and Control*, San Antonio, IEEE, 2896–2900.

ROBERTS, A. W. and VARBERG, D. E. (1973). *Convex Functions.* Academic Press.

ROGERS, C. (1992). Coupling of random walks. Private communication.

RÜSCHENDORF, L. (1985). Projections and iterative procedures. In: *Proceedings Sixth Intern. Symp. Multivariate Anal.*, Pittsburgh, Ed. P.R. Krishnaiah, 485–593.

RÜSCHENDORF, L. and RACHEV, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal*, **32**, 48–54.

RÜSCHENDORF, L. (1991a). Bounds for distributions with multivariate marginals. In: *Proceedings: Stochastic Order and Decision under Risk.* Ed. K. Mosler and M. Scarsini. IMS Lecture Notes, Vol. 19, 285–310.

RÜSCHENDORF, L. (1991b). Fréchet bounds and their applications. In: *Advances in probability Measures with given Marginals.* Eds. Dall'Aglio, Kotz and Salinetti, 151–188.

RÜSCHENDORF, L. and THOMSEN, W. (1993a). Note on the Schrödinger equation and I-projections. *Statist. Probab. Lett.* **17**, 369–375.

RÜSCHENDORF, L. and THOMSEN, W. (1993b). Closedness of sum spaces and the generalized Schrödinger problem. To appear in: *Theory Probab. Appl.*

RÜSCHENDORF, L. (1993c). Convergence of the iterative proportional fitting procedure. To appear in: *Ann. Statist.*

RÜSCHENDORF, L. (1993d). Optimal solution of multivariate transportation problems. To appear in: *Applicationes Mathematicae.*

SCARSINI, M. (1989). Copulae of probability measures on product spaces. *J. Multivariate Anal.* **31**, 201–213.

SCHRÖDINGER, E. (1931). Über die Umkehrung der Naturgesetze. Sitzungsberichte, *Preuss. Akad. Wiss. Berlin, Phys. Math.* **144**, 144–153.

SMITH, C. and KNOTT, M. (1992). On Hoeffding Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.* **40**, 328–334.

STONE, C. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **2**, 689–705.

VON NEUMANN, J. (1950). *Functional Operators*, Volume 2. Princeton Univ. Press

WAKOLBINGER, A. (1992). Schrödinger bridges from 1931 to 1991. In: *Proc. 4th Latin America Congress in Probab. Math. Statist.* Ed: E. Cabana, 61–79.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT FREIBURG
HEBELSTRASSE 27
D-79104 FREIBURG, GERMANY
ruschen@mathematik.uni.freiburg.de