APPLICATIONES MATHEMATICAE
23,3 (1995), pp. 325–338

L. RÜSCHENDORF (Freiburg)

OPTIMAL SOLUTIONS OF MULTIVARIATE COUPLING PROBLEMS

Abstract. Some necessary and some sufficient conditions are established for the explicit construction and characterization of optimal solutions of multivariate transportation (coupling) problems. The proofs are based on ideas from duality theory and nonconvex optimization theory. Applications are given to multivariate optimal coupling problems w.r.t. minimal ℓ_p-type metrics, where fairly explicit and complete characterizations of optimal transportation plans (couplings) are obtained. The results are of interest even in the one-dimensional case. For the first time an explicit criterion is given for the construction of optimal multivariate couplings for the Kantorovich metric ℓ_1.

1. Introduction. In this paper we deal with the following basic coupling problem. Let $P, Q \in M^1(\mathbb{R}^k, \mathbb{B}^k)$ be two probability measures on $(\mathbb{R}^k, \mathbb{B}^k)$ and define, for $p \geq 1$ and $|\ |$ a norm on $\mathbb{R}^k$, the minimal ℓ_p-metric (w.r.t. the distance $|\ |$)

$$\ell_p(P, Q) := \inf\{(E|X - Y|^p)^{1/p} : X \stackrel{d}{=} P,\ Y \stackrel{d}{=} Q\}, \tag{1.1}$$

all r.v.'s X, Y being defined on a rich enough probability space. The *transportation problem* (or *coupling problem*) is to determine the value of the optimal transportation $\ell_p(P, Q)$ and to construct an optimal pair (X, Y) of random variables. In this paper we restrict ourselves to the second part of the problem. The multivariate coupling problem is a well-known long-time open problem which has many applications in probability theory (cf. Rachev (1991)). The aim of the paper is to characterize optimal transportation plans (couplings), to describe the necessary notions and arguments from nonconvex optimization theory and consider extensions of the transportation prob-

1991 *Mathematics Subject Classification*: Primary 62H20.
Key words and phrases: optimal couplings, c-convex functions, ℓ_p-metric, transportation problem.

lem in (1.1) to general cost functions $c(x,y)$. We remark that several parts of this paper do not need the context of euclidean spaces (cf. also Rüschendorf (1991b)). In Sections 1 and 2 we review some basic notions and results which are up to now only available in some conference volumes. In the following Section 3 we develop some new criteria which allow us to determine optimal explicit coupling results in a series of interesting examples.

In the case $p = 2$ and $|\ |$ the euclidean metric, the following basic characterization of an optimal coupling (resp. an optimal solution of (1.1)) was given in Rüschendorf and Rachev (1990).

THEOREM 1. *Let* $X \stackrel{d}{=} P$ *and* $Y \stackrel{d}{=} Q$ *have finite second moments*, *and* $|\ |$ *be the euclidean metric on* $\mathbb{R}^k$.

(a) (X,Y) *is* ℓ_2*-optimal if and only if*

$$Y \in \partial f(X) \quad \textit{a.s. for some closed convex function } f, \tag{1.2}$$

where $\partial f(x)$ *denotes the subgradient of* f *at* x.

(b) *There exists an optimal pair in* (a).

Some previous versions of this result were developed in Knott and Smith (1984) and Smith and Knott (1992). A condition equivalent to (1.2) is that the support Γ of the distribution of (X,Y) is *cyclically monotone*, i.e. for all $(x_1,y_1),\ldots,(x_n,y_n) \in \Gamma$ and $x_{n+1} := x_1$ we have

$$\sum_{i=1}^{n}(x_{i+1}-x_i)y_i \le 0 \tag{1.3}$$

(cf. Rockafellar (1970)). Theorem 1 allows one to construct many examples of optimal transportation plans. The cases of normal distributions, radial transformations, spherically invariant distributions and others are considered in Cuesta, Rüschendorf and Tuero (1993). If ϕ is a function on $\mathbb{R}^k$ and $\phi = \nabla f$ is the gradient of a closed (= lsc) convex function f, then $(X, \phi(X))$ is an ℓ_2-optimal pair for any r.v. X in the domain of ϕ. This property suggests calling ϕ an optimal coupling function. If $\phi = (\phi_1,\ldots,\phi_k)$ and ϕ_i are continuously differentiable and defined on a convex domain, then the ℓ_2-optimality of ϕ is equivalent to

(1.4) ϕ is cyclically monotone (i.e. the graph of ϕ, $\Gamma = \{(x,\phi(x) : x \in \operatorname{dom}\phi\}$ is cyclically monotone)

or, equivalently,

(1.5) the matrix $(\partial\phi_i(x)/\partial x_j)$ is symmetric for all $x \in \operatorname{dom}\phi$ and ϕ is monotone (i.e. $(y-x)(\phi(y)-\phi(x)) \ge 0$ for all x, y)

or, equivalently,

$$\phi = \nabla f \quad \text{for some smooth convex function } f \tag{1.6}$$

(cf. Rüschendorf (1991b) and Levin (1992)).

For radially continuous functions ϕ on a convex set in $\mathbb{R}^k$ the condition that $\phi = \nabla f$ for some Gateaux differentiable function f is equivalent to the condition that the integral $\int_{x_0 \to x} \phi(u)\,du$ is independent of the path of integration, i.e. $\phi(u)\,du$ is closed (cf. Vainberg (1973), Th. 6.2) and in this case $f(x) = f_0 + \int_0^1 \phi(x_0 + t(x - x_0))(x - x_0)\,dt = f_0 + \int_{x_0 \to x} \phi(u)\,du$. Therefore, under closedness of $\phi(u)\,du$, convexity of f is equivalent to monotonicity of ϕ.

Remark 1. Theorem 1 solves "one half" of the problem of construction of optimal ℓ_2-transportation plans. It gives a characterization of all ℓ_2-optimal transportation plans. A still open problem is to find for given P, Q an optimal coupling function ϕ. If P, Q have densities f, g w.r.t. λ^k and if a regular invertible solution ϕ exists, then by the transformation formula the problem to be solved is the Monge nonlinear partial differential equation: Find ϕ (regular) cyclically monotone such that

$$g(x) = f(\phi^{-1}(x))|\det D_{\phi^{-1}}(x)| \tag{1.7}$$

for x in the support of Q.

The usual boundary conditions of PDE's are replaced by the condition of cyclical monotonicity of ϕ. For the (approximate) solution of (1.7) there seem to be two strategies, except in simple cases. Firstly, to develop numerical solutions of (1.7), and secondly, to give a "sufficiently" large list of examples ϕ of optimal coupling functions and the resulting pairs of densities f, g. This second path has begun to be investigated in Cuesta *et al.* (1993) but needs a lot of further extensions. Apparently, the first approach has not been taken up yet.

2. c-Convex functions; the general case. Theorem 1 has been extended to general cost functions $c : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^1$ (c also might be defined on subsets) in Rüschendorf (1991a, b). Call a pair $X \overset{d}{=} P, Y \overset{d}{=} Q$ *c-optimal* if

$$Ec(X, Y) = \sup\{Ec(U, V) : U \overset{d}{=} P,\ V \overset{d}{=} Q\}. \tag{2.1}$$

We consider the corresponding sup problem in order to avoid notational conflict with relevant notions from nonconvex optimization theory. For the inf problem just switch over from c to the cost function $-c$. A function f on $\mathbb{R}^k$ is called *c-convex* if for some index set I and $y_i \in \mathbb{R}^k$, $a_i \in \mathbb{R}^1$, $i \in I$,

$$f(x) = \sup_{i \in I}(c(x, y_i) + a_i). \tag{2.2}$$

For $c(x,y)=xy$ (on $\mathbb{R}^k$) one obtains the closed convex functions. This case corresponds to the minimal ℓ_2-metrics considered in Section 1. c-convex functions have been studied in several recent papers on nonconvex optimization theory (cf. Elster and Nehse (1974) and Dietrich (1988) and references therein). Denote the *c-conjugate* of f by

$$f^*(y) := \sup_x (c(x,y) - f(x)), \tag{2.3}$$

the sup being over the domain of f, and the *double c-conjugate* by

$$f^{**}(x) := \sup_y (c(x,y) - f^*(y)). \tag{2.4}$$

Then f^* and f^{**} are c-convex; f^{**} is the largest c-convex function majorized by f; $f = f^{**}$ if and only if f is c-convex (cf. Elster and Nehse (1974)); and f^*, f^{**} are "admissible" in the sense that

$$f^*(y) + f^{**}(x) \geq c(x,y), \qquad \forall x, y. \tag{2.5}$$

The (double) conjugate functions are basic for the theory of inequalities as in (2.5). The *c-subgradient* of a function f is defined by

$$\partial_c f(x) := \{y : f(z) - f(x) \geq c(z,y) - c(x,y),\ \forall z \in \operatorname{dom} f\} \tag{2.6}$$

The following result gives the basic characterization of c-optimal transportation plans (X,Y). It is the analogue to Theorem 1 for the case of ℓ_2-couplings.

Let $\mathcal{L}_m(P,Q)$ denote the set of all lower majorized measurable functions $c = c(x,y)$, i.e. $c(x,y) \geq f_1(x) + f_2(y)$ for some $f_1 \in L^1(P)$, $f_2 \in L^1(Q)$.

THEOREM 2 (cf. Rüschendorf (1991b)). *Let $c \in \mathcal{L}_m(P,Q)$ and assume that $I(c) := \inf\{\int h_1\, dP + \int h_2\, dQ : c \leq h_1 \oplus h_2,\ h_1 \in L^1(P),\ h_2 \in L^1(Q)\} < \infty$.*

(a) *$X \stackrel{d}{=} P$, $Y \stackrel{d}{=} Q$ is a c-optimal pair if and only if*

$$Y \in \partial_c f(X) \qquad \text{a.s. for some } c\text{-convex function } f. \tag{2.7}$$

(b) *If c is upper semicontinuous, then there exists an optimal pair (X,Y).*

For $c(x,y) = |x-y|^p$, we have $I(c) < \infty$ if $c(\cdot,a) \in L^1(P)$ and $c(a,\cdot) \in L^1(Q)$, i.e. if P and Q have finite pth moments. The same condition implies that $c(x,y) = -|x-y|^p$ is lower majorized, i.e. $c \in \mathcal{L}_m(P,Q)$. As in (1.4) and by a similar proof, condition (2.7) is equivalent to the condition that the support Γ of (X,Y) is *c-cyclically monotone*, i.e. for all $(x_1,y_1),\dots,(x_n,y_n) \in \Gamma$ and $x_{n+1} := x_1$ we have

$$\sum_{i=1}^n (c(x_{i+1},y_i) - c(x_i,y_i)) \leq 0 \tag{2.8}$$

(cf. Dietrich (1988) and Smith and Knott (1992)). For a differentiable cost function $c(x,y)$ and a function ϕ on $\mathbb{R}^k$ define $c_1(x,y) := \frac{\partial}{\partial x} c(x,y)$. The differential form $c_1(x, \phi(x))\, dx$ is called closed if its integral is path independent. For the case of regular ϕ the following lemma was given in Rüschendorf (1991b).

LEMMA 3. *If $c(\cdot, y)$ is differentiable for all y and $c_1(x, \phi(x))\, dx$ is closed, then* (2.7), (2,8) *are equivalent to*

$$\int_{y \to x} (c_1(u, \phi(y)) - c_1(u, \phi(u)))\, du \le 0, \qquad \forall x, y. \tag{2.9}$$

Proof. If (2.9) holds, then define, for some $x_0 \in \operatorname{dom} \phi$,

$$f(x) := \int_{x_0 \to x} c_1(u, \phi(u))\, du. \tag{2.10}$$

From (2.9) we conclude that for all z,

$$\begin{aligned} f(z) - f(x) = \int_{x \to z} c_1(u, \phi(u))\, du &\ge \int_{x \to z} c_1(u, \phi(x))\, du \\ &= c(z, \phi(x)) - c(x, \phi(x)). \end{aligned}$$

Therefore, $\phi(x) \in \partial_c f(x)$. The converse direction is similar. ∎

Remark 2. If $c(\cdot, y)$ is concave, then (without differentiability) the following version of Lemma 3 holds: If $-h_x(u)$ is in the subgradient $\partial(-c(\cdot, \phi(x)))$ at u and if $h_u(u)\, du$ is closed, then (2.7), (2.8) are equivalent to

$$\int_{y \to x} (h_y(u) - h_u(u))\, du \le 0, \qquad \forall x, y. \tag{2.11}$$

This conclusion follows from the inequality

$$c(z, \phi(x)) - c(x, \phi(x)) \le \int_{x \to z} h_x(u)\, du$$

and replacing (2.10) by $f(x) := \int_{x_0 \to x} h_u(u)\, du$.

For the application of the theory in this section one needs manageable criteria to determine c-subgradients resp. c-cyclically monotone functions ϕ. To establish these criteria is the main contribution of this paper.

3. c-Optimal coupling functions. Call a function ϕ *c-cyclically monotone* if the graph of ϕ, $\Gamma := \{(x, \phi(x)) : x \in \operatorname{dom} \phi\}$, is c-cyclically monotone. By Theorem 2 for any c-cyclically monotone function ϕ and any r.v. X in the domain of ϕ the pair $(X, \phi(X))$ is c-optimal. The idea of the following simple criterion for the construction of c-cyclically monotone functions ϕ is basically due to Smith and Knott (1992) who also consider the case of general pairs (X, Y). It is based on criterion (2.8) and relates c-cyclically monotone functions to cyclically monotone functions. This in an interesting relation since the cyclically monotone functions are studied in several papers.

THEOREM 4. *If for some cyclically monotone function h and all x, y in the domain of ϕ,*

$$c(x, \phi(x)) - c(x, \phi(x)) \le h(x)(y - x), \tag{3.1}$$

then ϕ is c-cyclically monotone.

Proof. For all $x_1, \ldots, x_n \in \operatorname{dom} \phi$ and $x_{n+1} := x_1$ we have

$$\sum_{i=1}^{n} (c(x_{i+1}, \phi(x_i)) - c(x_i, \phi(x_i))) \le \sum_{i=1}^{n} h(x_i)(x_{i+1} - x_i) \le 0,$$

since h is cyclically monotone. ∎

COROLLARY 5. *If $c(\cdot, y)$ is concave and differentiable for all y and if $h(u) := c_1(u, \phi(u))$ is cyclically monotone, then ϕ is c-cyclically monotone.*

Proof. From concavity of $c(\cdot, y)$ we obtain

$$c(y, \phi(x)) - c(x, \phi(x)) \le c_1(x, \phi(x))(y - x) = h(x)(y - x);$$

i.e. condition (3.1) is satisfied. ∎

Remark 3. For the application to nondifferentiable functions c like $c(x, y) = -|x - y|_p$ the following extension of Corollary 5 is of interest: If $c(\cdot, y)$ is concave and $-h(u)$ is in the subgradient of $-c(\cdot, \phi(u))$ at u, and if h is cyclically monotone, then ϕ is c-cyclically monotone.

EXAMPLE 1. (a) (Optimal ℓ_p-couplings) Let $c(x, y) := -|x - y|^p$, $p > 1$, $x, y \in \mathbb{R}^k$, where $|\ |$ is the euclidean metric, i.e. we consider the problem of constructing optimal transportation plans w.r.t. the minimal ℓ_p-metric as defined in (1.1) (up to the pth root). Then $c(\cdot, y)$ is (strictly) concave since the Minkowski inequality yields

$$\begin{aligned} c(\alpha x_1 + (1 - \alpha) x_2, y) &= -|\alpha(x_1 - y) + (1 - \alpha)(x_2 - y)|^p \\ &\ge -(\alpha |x_1 - y| + (1 - \alpha)|x_2 - y|)^p \\ &\ge -\alpha |x_1 - y|^p + (1 - \alpha)|x_2 - y|^p \\ &= \alpha c(x_1, y) + (1 - \alpha) c(x_2, y) \end{aligned} \tag{3.2}$$

by the convexity of $t \to t^p$ on the positive real line. (This argument holds true for any norm.) Let h be any cyclically monotone function. Then the equation

$$c_1(x, \phi(x)) = -|x - \phi(x)|^{p-2}(x - \phi(x)) = h(x) \tag{3.3}$$

has the unique solution

$$\phi(x) = |h(x)|^{-(p-2)/(p-1)} h(x) + x. \tag{3.4}$$

Therefore, from Corollary 5 for any cyclically monotone function h, the function $\phi = \phi_h$ from (3.4) is c-cyclically monotone and so for any r.v. X in the domain of ϕ the pair $(X, \phi(X))$ is an optimal coupling w.r.t. minimal ℓ_p-metrics. Some partial result in the case $1 < p \le 2$ was given in Smith and Knott (1992). For $p = 2$ we find that $\phi(x) = h(x) + x$ is cyclically monotone. By Theorem 1 cyclical monotonicity of h is necessary and sufficient in this case.

If in particular $h(x) := Ax$, A positive semidefinite, symmetric, linear (cf. (1.5)), then we obtain optimality of

$$\phi(x) = (x^T A^2 x)^{-(p-2)/(2(p-1))} Ax + x. \tag{3.5}$$

If $h(x) = \alpha(|x|)x/|x|$, with α nondecreasing, is a radial transformation, then

$$\phi(x) = g(|x|)\frac{x}{|x|}, \tag{3.6}$$

where $g(|x|) := (\alpha(|x|))^{1/(p-1)} + |x|$ is optimal. ϕ is again a radial transformation. Optimality of radial transformations has been established before in Cuesta *et al.* (1993).

(b) For $c(x, y) = -\sum_{i=1}^k |x_i - y_i|^p, p > 1$, and $h = (h_1, \dots, h_k)$ any cyclically monotone function define

$$\widehat{h}(x) := (|h_i(x)|^{-(p-2)/(p-1)}\ h_i(x)). \tag{3.7}$$

Then as in example (a) we conclude that

$$\phi(x) = \widehat{h}(x) + x \text{ is } c\text{-optimal.} \tag{3.8}$$

(c) For

$$c(x, y) = -\Big(\sum_{i=1}^k |x_i - y_i|^p\Big)^{r/p} = -|x - y|_p^r, \qquad p, r > 1,$$

consider the equation

$$c_1(x, \phi(x)) = -r|x - \phi(x)|_p^{r-p}\ (|x_i - \phi_i(x)|^{p-2}\ (x_i - \phi_i(x))) = rh(x) \tag{3.9}$$

for any cyclically monotone function h. From (3.9) we obtain

$$|h_i(x)| = |x - \phi(x)|_p^{r-p} |x_i - \phi_i(x)|^{p-1}$$

and with the conjugate index $q = p/(p-1)$ to p,

$$|h(x)|_q = \Big(\sum |h_i(x)|^{p/(p-1)}\Big)^{(p-1)/p} = |x - \phi(x)|_p^{r-1}.$$

Therefore,

$$|x_i-\phi_i(x)|^{p-2} = |h_i(x)|^{\frac{p-2}{p-1}}\ |x-\phi(x)|_p^{\frac{(p-r)(p-2)}{p-1}} = |h_i(x)|^{\frac{p-2}{p-1}}\ |h(x)|_q^{\frac{p-r}{r-1}\frac{p-2}{p-1}} \tag{3.10}$$

and

$$\begin{aligned}\phi_i(x)-x_i &= h_i(x)|x-\phi(x)|_p^{p-r}|x_i-\phi_i(x)|^{2-p}\\ &= h_i(x)|h(x)|_{\frac{p}{p-1}}^{\frac{p-r}{r-1}}|h_i(x)|^{\frac{2-p}{p-1}}|h(x)|_q^{\frac{r-p}{r-1}\frac{p-2}{p-1}}\\ &= h_i(x)|h_i(x)|^{\frac{2-p}{p-1}}|h(x)|_q^{\frac{p-r}{r-1}}.\end{aligned}$$

From (3.9) and Corollary 5, therefore,

$$\phi(x) := \widehat{h}(x)+x \text{ is } c\text{-optimal,} \tag{3.11}$$

where $\widehat{h}(x) := |h(x)|_q^{(p-r)/(r-1)}(h_i(x)/|h_i(x)|^{(p-2)/(p-1)})$.

(d) (Kantorovich metric ℓ_1) For

$$c(x,y) = -|x-y|_p = -\Big(\sum |x_i-y_i|^p\Big)^{1/p}, \qquad p\ge 1,$$

a different situation occurs. Note that this cost function corresponds to the minimal ℓ_1-metric (based on the p-norm in $\mathbb{R}^k$).

We first consider the case $p>1$. Then for $f(x)=|x|_p$ the subdifferential is given by (cf. Rockafellar (1970))

$$\partial f(x) = \begin{cases} (|x_i|^{p-2}x_i/|x|_p^{p-1}) = \nabla f(x) & \text{for } x\ne 0,\\ U_q & \text{for } x=0,\end{cases} \tag{3.12}$$

where U_q is the unit ball w.r.t. q-norm. Therefore, for any $y\in\partial f(x)$ we have $|y|_q=1$ for $x\ne 0$, while $|y|_q\le 1$ for $x=0$. Let h be any cyclically monotone function with

$$|h(x)|_q = 1 \tag{3.13}$$

and consider the equation

$$\left(\frac{|x_i-\phi_i(x)|^{p-2}}{|x-\phi(x)|_p^{p-1}}(\phi_i(x)-x_i)\right) = h(x), \qquad x\ne\phi(x). \tag{3.14}$$

Then from the definition of the subgradient we obtain

$$c(y,\phi(x)) - c(x,\phi(x)) \le h(x)(y-x)$$

and, therefore, by Theorem 4 (or Remark 3), ϕ is cyclically monotone.

For any nonnegative function $\alpha(x)\ge 0$ define

$$\phi(x) := (\alpha(x))^{1/(p-1)}\left(\frac{h_i(x)}{|h_i(x)|^{(p-2)/(p-1)}}\right)+x. \tag{3.15}$$

Then

$$|\phi(x)-x|_p^{p-1} = \alpha(x)|h(x)|_q = \alpha(x).$$

Furthermore,

$$|\phi_i(x)-x_i| = (\alpha(x)|h_i(x)|)^{1/(p-1)}$$

and

$$\begin{aligned}\phi_i(x)-x_i &= (\alpha(x))^{1/(p-1)}\frac{h_i(x)}{|h_i(x)|^{(p-2)/(p-1)}}\\ &= (\alpha(x)h_i(x))\frac{\alpha(x)^{(2-p)/(p-1)}}{|h_i(x)|^{(p-2)/(p-1)}} = \alpha(x)h_i(x)/|\phi_i(x)-x_i|^{p-2}.\end{aligned}$$

This implies that ϕ satisfies equation (3.14). So in the case of the ℓ_1-metric the optimality equation has no longer a unique solution. In the case $p=2$, (3.15) simplifies to

$$\phi(x) = \alpha(x)h(x)+x \tag{3.16}$$

and the optimality equation (3.14) reduces to the condition of $h(x) = (\phi(x)-x)/(|\phi(x)-x|)$ being cyclically monotone.

In the case $p=1$ we have, analogously, with $f(x)=|x|_1$, the following characterization of $\partial f(x)$:

$$u\in\partial f(x) \quad \text{if and only if} \quad \begin{array}{l} u_i = x_i/(|x_i|) \quad \text{for } x_i\neq 0 \text{ and}\\ |u_i|\le 1 \text{ for } x_i = 0.\end{array} \tag{3.17}$$

Similar to the preceding calculations or by Remark 3 we obtain the optimality equation

$$\frac{\phi_i(x)-x_i}{|\phi_i(x)-x_i|} = h_i(x), \qquad x_i\neq\phi_i(x), \tag{3.18}$$

and $|h_i(x)|\le 1, h$ cyclically monotone. The solutions of (3.18) are given by

$$\phi_i(x) = \alpha_i(x)h_i(x)+x_i \tag{3.19}$$

for some nonnegative functions $\alpha_i(x)$.

Corollary 6. *Let $c(x,y) = -|x-y|_p$, $p\ge 1$, let $\alpha(x)$ with $\alpha_i(x)\ge 0$ be measurable, and let $h(x)$ be cyclically monotone with $|h(x)|_q = 1$ for all x, where $q=p/(p-1)$ is the conjugate index to p. Then*

$$\phi(x) = (\alpha(x))^{1/(p-1)}\left(\frac{h_i(x)}{|h_i(x)|^{(p-2)/(p-1)}}\right)+x \qquad \textit{for } p>1 \tag{3.20}$$

and

$$\phi(x) = (\alpha_i(x)h_i(x))+x \qquad \textit{for } p=1 \tag{3.21}$$

are c-cyclically monotone.

So for all r.v.'s X in the domain of ϕ the pair $(X,\phi(X))$ is an optimal coupling for the ℓ_1-metric (Kantorovich metric) based on the p-norm distance $|\ |_p$ on $\mathbb{R}^k$. It remains to be investigated how large the class of admissible h is.

From definition (2.6) the c-subgradient of a function f has a characterization as solution of a nonconvex optimization problem: $y^*\in\partial_c f(y)$ if and only if

$$\varphi^y(x)\ge\varphi^y(y)\qquad\text{for all } x,y,\tag{3.22}$$

where $\varphi^y(x):=f(x)-c(x,y^*)$; i.e. φ^y has its minimum at $x=y$. If $y^*=\phi(y)$ and $c_1(u,\phi(u))\,du$ is closed, then by (2.10),

$$\frac{\partial}{\partial x}\varphi^y(y)=0.\tag{3.23}$$

With the second derivatives

$$\begin{aligned}B(x,y)&:=-\frac{\partial^2}{\partial x\partial x'}\varphi^y(x)\\&=\frac{\partial^2}{\partial x\partial x'}c(x,\phi(y))-\frac{\partial^2}{\partial x\partial x'}c(x,\phi(x))\\&\quad-\frac{\partial^2}{\partial x\partial y}c(x,\phi(x))D\phi(x)\end{aligned}\tag{3.24}$$

one obtains

Proposition 7. *If c is differentiable in the first component and $c_1(u,\phi(u))\,du$ is closed*, *then*:

(a) $B(x,y)\le 0$ (*in the sense of negative definiteness*) *implies that ϕ is c-optimal.*

(b) *If ϕ is c-optimal*, *then*

$$-B(y,y)=\frac{\partial^2}{\partial x\partial x'}c(y,\phi(y))\phi(y)\ge 0.\tag{3.25}$$

Proof. (a) If $B(x,y)\le 0$, then φ^y is convex and $\frac{\partial}{\partial x}\varphi^y(y)=0$. This implies that y is a global minimum of φ^y and, therefore, by (3.22), $\phi(y)\in\partial_c f(y)$, i.e. ϕ is a c-optimal function.

(b) is a well-known necessary condition for local optimality of φ^y at $x=y$. ■

Remark 4. In the case $c(x,y)=-|x-y|^p$, $1<p$, with $|\ |$ the euclidean metric,

$$\frac{\partial^2}{\partial y_j\partial x_i}c(x,y)=\begin{cases}-p(p-2)|x-y|^{p-3}s(y_j-x_j)(x_i-y_i) & \text{for } i\ne j,\\ -p(p-2)|x-y|^{p-3}s(y_i-x_i)(x_i-y_i)+p|x-y|^{p-2} & \text{for } i=j,\end{cases}$$

where s is the sign function and, therefore, the necessary condition (3.25) reads

$$\begin{aligned} (3.26)\qquad & -B(y,y) \\ &= -p|y-\phi(y)|^{p-2}\biggl((p-2)\frac{(y-\phi(y))(y-\phi(y))^T}{|y-\phi(y)|^2}-I\biggr)D\phi(y)\ge 0. \end{aligned}$$

For $p=2$, (3.26) is equivalent to the necessary and sufficient condition $D\phi\ge 0$.

The following sufficient condition for c-optimality of ϕ does not assume that the cost function is concave.

THEOREM 8. *If $c(\cdot,y)$ is differentiable for all y and $c_1(u,\phi(u))\,du$ is closed and if for all x,y in the domain of ϕ,*

$$(3.27)\qquad (x-y)(c_1(x,\phi(x))-c_1(x,\phi(y)))\ge 0,$$

then ϕ is a c-optimal function.

Proof. By (3.22) and (2.9) it is sufficient to prove that

$$F_y(x):=\int\limits_{y\to x}(c_1(u,\phi(y))-c_1(u,\phi(u)))\,du\le 0=F_y(y).$$

For $t\ge 0$, let $x_t:=y+t(x-y)$ and $H(t):=F_y(x_t)$. Then by (3.27),

$$\begin{aligned} (3.28)\qquad \frac{d}{dt}H(t) &= \frac{\partial}{\partial x}F_y(x_t)(x-y)=(c_1(x_t,\phi(y))-c_1(x_t,\phi(x_t)))(x-y) \\ &= \frac{1}{t}(c_1(x_t,\phi(y))-c_1(x_t,\phi(x_t)))(x_t-y)\le 0. \end{aligned}$$

This implies that

$$F_y(x)=F_y(y)+\int\limits_0^1\frac{d}{dt}H(t)\,dt\le 0. \ \blacksquare$$

Remark 5. If $c(\cdot,y)$ is concave, then as in Remark 2 we have (without differentiability) the following modified version of Theorem 8: If $-h_x(u)\in\partial(-x(\cdot,\phi(x)))(u)$, if $h_u(u)\,du$ is closed and if for all $x,y\in\operatorname{dom}\phi$,

$$(x-y)(h_x(x)-h_y(x))\ge 0,$$

then ϕ is c-cyclically monotone. This gives an alternative to the criterion in Corollary 5 and Remark 3, which is advantageous in some examples.

EXAMPLE 2. (a) In the case $c(x,y)=-|x-y|^p$, $p>1$, where $|\ |$ is the euclidean metric, condition (3.27) amounts to

$$(3.29)\qquad (g_x(x)-g_x(y)(x-y)\ge 0\qquad\text{for all } x,y,$$

with $g_x(y):=|x-\phi(y)|^{p-2}(x-\phi(y))$. A rough sufficient condition for (3.29) is that for all x, g_x is cyclically monotone. For a cyclically monotone function h_x the equation

$$g_x(y) = h_x(y), \qquad \forall y, \tag{3.30}$$

has the unique solution

$$\phi_x(y) = |h_x(y)|^{-(p-2)/(p-1)} h_x(y) + x.$$

The assumption that $\phi_x = \phi$ does not depend on x leads to the following equation with $h := h_0$, $r := (p-2)/(p-1)$, $\psi(y) := h(y)/|h(y)|^r$:

$$h_x(y) = (\psi(y) - x)|\psi(y) - x|^{r/(1-r)} \tag{3.31}$$

and

$$\phi(y) = |h(y)|^{-r} h(y). \tag{3.32}$$

But now it has to be checked whether h_x is cyclically monotone for all x. For $p = 2$, $\phi(y) = h(y)$, this is trivially true.

(b) For $c(x, y) = -|x - y|_p = -(\sum |x_i - y_i|^p)^{1/p}$, $p \geq 1$, the function

$$c_1(x, y) := -\frac{1}{|x - y|_p^{p-1}} (|x_i - y_i|^{p-2}(x_i - y_i)), \qquad x \neq y,$$

defines a subgradient and conditions (3.27) and Remark 5 amount to

$$\sum_{i=1}^{k} (x_i - y_i) \left(\frac{|x_i - \phi_i(x)|^{p-2}}{|x - \phi(x)|_p^{p-1}} (\phi_i(x) - x_i) - \frac{|x_i - \phi_i(y)|^{p-2}}{|x - \phi(y)|_p^{p-1}} (\phi_i(y) - x_i) \right) \geq 0. \tag{3.33}$$

For $p = 2$, (3.33) reduces to

$$(x - y) \left(\frac{\phi(x) - x}{|\phi(x) - x|} - \frac{\phi(y) - x}{|\phi(y) - x|} \right) \geq 0 \qquad \text{for all } x, y. \tag{3.34}$$

So Theorem 8 has the following interesting consequence for the ℓ_1-metric w.r.t. the euclidean distance $|\ | = |\ |_2$.

COROLLARY 9 (Optimal couplings w.r.t. Kantorovich metric). *If $\frac{\phi(u) - u}{|\phi(u) - u|} du$ is closed and ϕ satisfies the normalized angle monotonicity condition*

$$(x - y) \left(\frac{\phi(x) - x}{|\phi(x) - x|} - \frac{\phi(y) - x}{|\phi(y) - x|} \right) \geq 0 \qquad \textit{for all } x, y. \tag{3.35}$$

then $(X, \phi(X))$ is an optimal coupling for the ℓ_1-metric w.r.t. euclidean distance on $\mathbb{R}^k$ for any r.v. X in the domain of ϕ.

Remark 6. (a) Condition (3.35) has an obvious geometric interpretation. If we consider $\phi(x)$ and $\phi(y)$ in the (by x) translated coordinate system and normalized to norm 1, then this difference has an angle with the difference of x and y (or y in the translated system) of less than 90 degrees.

Without translation and normalization this is just the usual monotonicity. Corollary 9 suggests to develop a theory of (3.35) since this notion is related to optimality w.r.t. the Kantorovich metric ℓ_1, as monotonicity is related to optimality w.r.t. the ℓ_2-metric (cf. Theorem 1).

(b) (One-dimensional case) Even in the one-dimensional case the conclusions of this section are of interest and new. They allow one to construct optimal couplings in some cases of cost functions c which are not of Monge type. On the real line condition (3.27) is equivalent to

$$x \left\{ \begin{matrix} \geq \\ \geq \end{matrix} \right\} y \quad \text{implies} \quad c_1(x, \phi(x)) \left\{ \begin{matrix} \leq \\ \leq \end{matrix} \right\} c_1(x, \phi(y)), \tag{3.36}$$

while in the concave case Corollary 5 gives the sufficient condition for optimality:

$$h(x) := c_1(x, \phi(x)) \text{ is nondecreasing.} \tag{3.37}$$

To compare these criteria consider the case $c(x,y) = (x-y)^2$, $x, y \in \mathbb{R}^1$. Then condition (3.37) is equivalent to

$$\phi(x) - x \text{ is nondecreasing,} \tag{3.38}$$

while condition (3.36) amounts to

$$\phi \text{ is nondecreasing.} \tag{3.39}$$

So in this case Theorem 8 gives the best possible answer while Corollary 5 has a stronger sufficient condition.

(c) The sufficient condition $B(x,y) \leq 0$ for c-optimality in Proposition 7 implies that φ^y (cf. (3.18)) is convex. If we can assure the weaker condition that φ^y is quasi-convex, i.e.

$$\varphi^y(\alpha x + (1-\alpha)y) \leq \min(\varphi^y(x), \varphi^y(y)), \tag{3.40}$$

then a local minimum of φ^y is either situated in a domain where φ^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.

(d) Similar ideas to those in this section appear in a recent paper of Levin (1992) on the Kantorovich–Rubinstein problem. Levin obtains for this problem (with fixed difference of the marginals) an explicit formula for the value of the optimal transshipment problem in the case of differentiable cost functions but no characterization of optimal plans. In contrast we obtain in this paper explicit results for the form of optimal transportation plans but no explicit formula for the optimal value. For the proof of the optimal value formula in the transshipment problem the differentiability of the cost function at the diagonal is a crucial assumption. Note that this assumption excludes the natural cost functions $c_p(x,y) = |x-y|_p$, $p \geq 1$; the differentiable powers

$|x-y|_p^\alpha$, $\alpha > 1$, lead to trivial results in the transshipment problem. Some explicit results and bounds in the nondifferentiable case of the transshipment problem were established in Rachev and Rüschendorf (1991).

Since the transshipment and transportation problems coincide for cost functions satisfying the triangle inequality, the results of this paper can also be seen in this context.

References

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

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

K. H. Elster und R. Nehse (1974), *Zur Theorie der Polarfunktionale*, ibid. 5, 3–21.

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

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

V. I. Levin (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-Sb. 71, 533–548.

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

S. T. Rachev and L. Rüschendorf (1991), *Solution of some transportation problems with relaxed or additional constraints*, SIAM J. Control Optim., to appear.

A. W. Roberts and D. E. Varberg (1973), *Convex Functions*, Academic Press.

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

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

L. Rüschendorf (1991b), *Fréchet-bounds and their applications*, in: *Advances in Probability Measures with Given Marginals*, G. Dall'Aglio, S. Kotz and G. Salinetti (eds.), Kluwer Acad. Publ., 151–188.

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

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

M. M. Vainberg (1973), *Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations*, Wiley.

LUDGER RÜSCHENDORF
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
HEBELSTR. 27
FREIBURG, GERMANY
E-mail: RUSCHEN@BUFFON.MATHEMATIK.UNI-FREIBURG.DE

Received on 2.12.1994