

# OPTIMAL SOLUTIONS OF MULTIVARIATE COUPLING PROBLEMS

Ludger Rüschendorf  
University of Freiburg

## 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  $\ell_p$ -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  $\ell_1$ .

AMS subject classification: 62H20, 90C08

key words and phrases: optimal couplings,  $c$ -convex functions,  $\ell_p$ -metric, transportation problem

# 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  $|| \cdot ||$  a norm on  $\mathbb{R}^k$  the minimal  $\ell_p$ -metric (w.r.t. the distance  $|| \cdot ||$ )

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

all rv'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 to the second part of the problem. The multivariate coupling problem is a wellknown longtime 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 problem 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 [13]). 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 to determine optimal explicit coupling results in a series of interesting examples.

In the case  $p = 2$  and  $|| \cdot ||$  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  $X \stackrel{d}{=} P, Y \stackrel{d}{=} Q$  have finite second moments,  $|| \cdot ||$  the euclidean metric on  $\mathbb{R}^k$ .*

a)  *$(X, Y)$  is  $\ell_2$ -optimal if and only if*

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

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

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 (1987). A condition equivalent to (1.2) is that the support  $\Gamma$  of the distribution of  $(X, Y)$  is *cyclically monotone* i.e.  $\forall (x_1, y_1), \dots, (x_n, y_n) \in \Gamma, x_{n+1} := x_1$  holds:

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

(cf. Rockafellar (1970)). Theorem 1 allows to construct many examples of optimal transportation plans. The cases of normal distributions, radial transformations, spherically invariant distributions and others are considered in Cuesta, Rüschemdorf and Tuero (1992). If  $\phi$  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  $\ell_2$ -optimal pair for any r.v.  $X$  in the domain of  $\phi$ . This property proposes to call  $\phi$  an optimal coupling function. If  $\phi = (\phi_1, \dots, \phi_k)$  and  $\phi_i$  are continuously differentiable defined on a convex domain, then the  $\ell_2$ -optimality of  $\phi$  is equivalent to:

$$(1.4) \quad \begin{aligned} &\phi \text{ is cyclically monotone (i.e. the graph of } \phi, \\ &\Gamma = \{(x, \phi(x)); x \in \text{dom}\phi\} \text{ is cyclically monotone)} \end{aligned}$$

or, equivalently,

$$(1.5) \quad \begin{aligned} &\text{the matrix } \left( \frac{\partial \phi_i}{\partial x_j}(x) \right) \text{ is symmetric for all } x \in \text{dom}(\phi) \text{ and} \\ &\phi \text{ is monotone (i.e. } (y - x)(\phi(y) - \phi(x)) \geq 0 \text{ for all } x, y) \end{aligned}$$

or, equivalently,

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

(cf. [13] and [6]).

For radially continuous functions  $\phi$  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  $\int_{x_0 \rightarrow x} \phi(u) du$  is independent of the path of integration, i.e.  $\phi(u) du$  is closed (cf. Vainberg (1973), Th. 6.2) and in this case  $f(x) = f_0 + \int_0^1 \phi(x_0 + t(x - x_0))(x - x_0) dt = f_0 + \int_{x_0 \rightarrow x} \phi(u) du$ . Therefore, under closedness of  $\phi(u) du$ , convexity of  $f$  is equivalent to monotonicity of  $\phi$ .

**Remark 1** *Theorem 1 solves “one half” of the problem of construction of optimal  $\ell_2$ -transportation plans. It gives a characterization of all  $\ell_2$ -optimal transportation plans. The still open problem is to find to given  $P, Q$  an optimal coupling function  $\phi$ . If  $P, Q$  have densities  $f, g$  w.r.t.  $\lambda^k$  and if a regular invertible solution  $\phi$  exists, then by the transformation formula the problem to be solved is the Monge nonlinear partial differential equation:  
Find  $\phi$  (regular) cyclically monotone such that:*

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

$x$  in the support of  $Q$ .

The usual boundary conditions of PDE's are replaced by the condition of cyclical monotonicity of  $\phi$ . For the (approximate) solution of (1.7) there

seem to be except in simple cases two strategies. Firstly to develop numerical solutions of (1.7) and secondly to give a “sufficiently” large list of examples  $\phi$  of optimal coupling functions and the resulting pairs of densities  $f, g$ . This second path has begun to be investigated in [1] but needs a lot of further extensions. The first path seems to be completely unsettled up to now.

## 2 $c$ -Convex Functions; the General Case

Theorem 1 has been extended to general cost functions  $c : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^1$  ( $c$  also might be defined on subsets) in [12], [13]. Call a pair  $X \stackrel{d}{=} P, Y \stackrel{d}{=} Q$   $c$ -optimal if

$$(2.1) \quad Ec(X, Y) = \sup\{Ec(U, V); U \stackrel{d}{=} P, V \stackrel{d}{=} Q\}.$$

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  $y_i \in \mathbb{R}^k, a_i \in \mathbb{R}^1, i \in I$

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

For  $c(x, y) = xy$  (on  $\mathbb{R}^k$ ) one obtains the closed convex functions. This case corresponds to the minimal  $\ell_2$ -metrics considered in Section 1.  $c$ -convex functions have been studied in several recent papers on non-convex optimization theory (cf. Elster and Nehse (1974) and Dietrich (1988) and references therein). Denote the  $c$ -conjugate of  $f$  by

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

the sup being over the domain of  $f$  resp.  $c$ , and the doubly  $c$ -conjugate

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

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

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

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

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

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  $\ell_2$ -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. [13])

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

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

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

For  $c(x, y) = |x - y|^p$ ,  $I(c) < \infty$  if  $c(\cdot, a) \in L^1(P)$ ,  $c(a, \cdot) \in L^1(Q)$ , i.e.  $P, Q$  have finite  $p$ -th 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  $\Gamma$  of  $(X, Y)$  is  $c$ -cyclically monotone, i.e.  $\forall (x_1, y_1), \dots, (x_n, y_n) \in \Gamma, x_{n+1} := x_1$  holds:

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

(cf. [2] and [14]). For a differentiable cost function  $c(x, y)$  and a function  $\phi$  on  $\mathbb{R}^k$  denote  $c_1(x, y) := \frac{\partial}{\partial x} c(x, y)$ . The differential form  $c_1(x, \phi(x))dx$  is called closed if its integral is path independent. For the case of regular  $\phi$  the following lemma was given in [13].

**Lemma 3** If  $c(\cdot, y)$  is differentiable for all  $y$  and  $c_1(u, \phi(x))dx$  is closed, then (2.7), (2.8) are equivalent to

$$(2.9) \quad \int_{y \rightarrow x} (c_1(u, \phi(y)) - c_1(u, \phi(x)))du \leq 0, \quad \forall x, y.$$

**Proof:** If (2.9) holds, then define for some  $x_0 \in \text{dom}\phi$

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

From (2.9) we conclude for all  $z$

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

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

**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*

$$(2.11) \quad \int_{y \rightarrow x} (h_y(u) - h_u(u))du \leq 0, \quad \forall x, y.$$

*This conclusion follows from the inequality*

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

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

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

### 3 $c$ -Optimal Coupling Functions

Call a function  $\phi$   *$c$ -cyclically monotone* if the graph of  $\phi$ ,  $\Gamma := \{(x, \phi(x)); x \in \text{dom}\phi\}$  is  $c$ -cyclically monotone. By Theorem 2 for any  $c$ -cyclically monotone function  $\phi$  and any r.v.  $X$  in the domain of  $\phi$  the pair  $(X, \phi(X))$  is a  $c$ -optimal pair. The idea of the following simple criterion for the construction of  $c$ -cyclically monotone functions  $\phi$  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 is 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  $\phi$*

$$(3.1) \quad c(x, \phi(x)) - c(x, \phi(y)) \leq h(x)(y - x),$$

*then  $\phi$  is cyclically monotone.*

**Proof:** For all  $x_1, \dots, x_n \in \text{dom}\phi$ ,  $x_{n+1} := x_1$  holds  $\sum_{i=1}^n (c(x_{i+1}, \phi(x_i)) - c(x_i, \phi(x_i))) \leq \sum_{i=1}^n h(x_i)(x_{i+1} - x_i) \leq 0$ , since  $h$  is cyclically monotone.  $\square$

**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  $\phi$  is  $c$ -cyclically monotone.*

**Proof:** From concavity of  $c(\cdot, y)$  we obtain  $c(y, \phi(x)) - c(x, \phi(x)) \leq c_1(x, \phi(x))(y - x) = h(x)(y - x)$ ; i.e. condition (3.1) is satisfied.  $\square$

**Remark 3** *For the application to non-differentiable 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  $\phi$  is  $c$ -cyclically monotone.*

**Example 1** *a) (Optimal  $\ell_p$ -couplings)*

*Let  $c(x, y) := -|x - y|^p$ ,  $p > 1$ ,  $x, y \in \mathbb{R}^k$ ,  $|\cdot|$  the euclidean metric, i.e. we consider the problem to construct optimal transportation plans w.r.t. the minimal  $\ell_p$ -metric as defined in (1.1) (up to the  $p$ -th root). then  $c(\cdot, y)$  is (strictly) concave since the Minkowskis inequality*

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

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

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

*has the unique solution*

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

*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  $\phi$  the pair  $(X, \phi(X))$  is an optimal coupling w.r.t. minimal  $\ell_p$ -metrics. Some partial result in the case  $1 < p \leq 2$  was given in Smith and Knott (1992). For  $p = 2$  we obtain the condition that  $\phi(x) = h(x) + x$  is cyclically monotone. By Theorem 1 the necessary and sufficient condition is that  $\phi(x)$  is cyclically monotone.*

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

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

If  $h(x) = \alpha(|x|) \frac{x}{|x|}$ ,  $\alpha$  monotonically nondecreasing, is a radial transformation, then

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

where  $g(|x|) := (\alpha(|x|))^{\frac{1}{p-1}} + |x|$  is optimal.  $\phi$  again is a radial transformation. Optimality of radial transformations has been established before in [1].

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

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

Then as in example a) we obtain

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

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

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

$$(3.10) \quad = rh(x)$$

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  $q = \frac{p}{p-1}$  to  $p$ ,  $|h(x)|_q = \left( \sum |h_i(x)|^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} = |x - \phi(x)|_p^{r-1}$ . Therefore,

$$(3.11) \quad \begin{aligned} |x_i - \phi_i(x)|^{p-2} &= |h_i(x)|^{\frac{p-2}{p-1}} |x - \phi(x)|_p^{\frac{(p-r)(p-2)}{p-1}} \\ &= |h_i(x)|^{\frac{p-2}{p-1}} |h(x)|_q^{\frac{p-r}{r-1} \frac{p-2}{p-1}} \end{aligned}$$

and



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

From (3.9) and Corollary 5, therefore,

$$(3.12) \quad \phi(x) := \widehat{h}(x) + x \text{ is } c\text{-optimal},$$

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

d) (Kantorovich-metric  $\ell_1$ )

For  $c(x, y) = -|x - y|_p = -(\sum |x_i - y_i|^p)^{1/p}$ ,  $p \geq 1$ , a different situation occurs. Note that this cost function corresponds to the minimal  $\ell_1$ -metric (based on the  $p$ -norm in  $\mathbb{R}^k$ ). We consider at first the case  $p > 1$ . Then for  $f(x) = |x|_p$  the subdifferential is given by (cf. [10])

$$(3.13) \quad \partial f(x) = \begin{cases} \left( \frac{|x_i|^{p-2} x_i}{|x|_p^{p-1}} \right) = \nabla f(x) & \text{for } x \neq 0 \\ U_q & \text{for } x = 0 \end{cases}$$

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

$$(3.14) \quad |h(x)|_q = 1$$

and consider the equation

$$(3.15) \quad \left( \frac{|x_i - \phi_i(x)|^{p-2}}{|x - \phi(x)|_p^{p-1}} (\phi_i(x) - x_i) \right) = h(x), \quad x \neq \phi(x).$$

Then from the definition of the subgradient we obtain

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

and, therefore, by Theorem 4 (or Remark 3)  $\phi$  is cyclically monotone.

Define for any nonnegative function  $\alpha(x) \geq 0$

$$(3.16) \quad \phi(x) := (\alpha(x))^{\frac{1}{p-1}} \left( \frac{h_i(x)}{|h_i(x)|^{\frac{p-2}{p-1}}} \right) + x.$$

Then

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

Furthermore,

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

and

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

This implies that  $\phi$  satisfies equation (3.14). So in the case of the  $\ell_1$ -metric the optimality equation has no longer a unique solution. In the case  $p = 2$ , (3.15) simplifies to

$$(3.17) \quad \phi(x) = \alpha(x)h(x) + x$$

and the optimality equation (3.14) reduces to the condition  $h(x) = \frac{\phi(x)-x}{|\phi(x)-x|}$  is cyclically monotone.

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

$$(3.18) \quad \begin{aligned} u \in \partial f(x) \text{ if and only if } u_i &= \frac{x_i}{|x_i|} \\ \text{for } x_i \neq 0 \text{ and } |u_i| \leq 1 &\text{ for } x_i = 0. \end{aligned}$$

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

$$(3.19) \quad \frac{\phi_i(x) - x_i}{|\phi_i(x) - x_i|} = h_i(x), x_i \neq \phi_i(x)$$

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

$$(3.20) \quad \phi_i(x) = \alpha_i(x)h_i(x) + x_i$$

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

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

$$(3.21) \quad \phi(x) = (\alpha(x))^{\frac{1}{p-1}} \left( \frac{h_i(x)}{|h_i(x)|^{\frac{p-2}{p-1}}} \right) + x \text{ for } p > 1$$

and

$$(3.22) \quad \phi(x) = (\alpha_i(x)h_i(x)) + x \text{ for } p = 1$$

are  $c$ -cyclically monotone.

So for all rv's  $X$  in the domain of  $\phi$  the pair  $(X, \phi(X))$  is an optimal coupling for the  $\ell_1$ -metric (Kantorovich metric) based on the  $p$ -norm distance  $|\cdot|_p$  on  $\mathbb{R}^k$ . It remains to be investigated how large the class of admissible  $h$  is.

From definition (2.6)  $c$ -subgradients of a function  $f$  have a characterization as solution of a nonconvex optimization problem:

$y^* \in \partial_c f(y)$  if and only if

$$(3.23) \quad \varphi^y(x) \geq \varphi^y(y) \text{ for all } x, y,$$

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

$$(3.24) \quad \frac{\partial}{\partial x} \varphi^y(y) = 0.$$

With the second derivatives

$$(3.25) \quad \begin{aligned} B(x, y) &:= -\frac{\partial^2}{\partial x \partial x'} \varphi^y(x) \\ &= \frac{\partial^2}{\partial x \partial x'} c(x, \phi(y)) - \frac{\partial^2}{\partial x \partial x'} c(x, \phi(x)) \\ &\quad - \frac{\partial^2}{\partial x \partial y} c(x, \phi(x)) D\phi(x) \end{aligned}$$

one obtains

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

a)  $B(x, y) \leq 0$  (in the sense of positive definiteness) implies that  $\phi$  is  $c$ -optimal.

b) if  $\phi$  is  $c$ -optimal, then

$$(3.26) \quad -B(y, y) = \frac{\partial^2}{\partial x \partial x'} c(y, \phi(y)) \phi(y) \geq 0$$

**Proof:**

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

b) is a well-known necessary condition for local optimality of  $\varphi^y$  in  $x = y$ . □

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

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

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

$$(3.27) \quad -B(y, y) = -p|y - \phi(y)|^{p-2} \left( (p-2) \frac{(y - \phi(y))(y - \phi(y))^T}{|y - \phi(y)|^2} - I \right) D\phi(y) \geq 0.$$

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

The following sufficient condition for  $c$ -optimality of  $\phi$  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  $\phi$

$$(3.28) \quad (x - y)(c_1(x, \phi(x)) - c_1(x, \phi(y))) \geq 0,$$

then  $\phi$  is a  $c$ -optimal function.

**Proof:** For  $c$ -optimality of  $\phi$  it is by (3.22) and (2.9) sufficient to prove that

$$\begin{aligned} F_y(x) &:= \int_{y \rightarrow x} (c_1(u, \phi(y)) - c_1(u, \phi(u))) du \\ &\leq 0 = F_y(y). \end{aligned}$$

Let for  $t \geq 0$ ,  $x_t := y + t(x - y)$  and  $H(t) := F_y(x_t)$ . Then by (3.27)

$$\begin{aligned} (3.29) \quad \frac{d}{dt}H(t) &= \frac{\partial}{\partial x} F_y(x_t)(x - y) \\ &= (c_1(x_t, \phi(y)) - c_1(x_t, \phi(x_t)))(x - y) \\ &= \frac{1}{t}(c_1(x_t, \phi(y)) - c_1(x_t, \phi(x_t)))(x_t - y) \leq 0 \end{aligned}$$

This implies that

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

□

**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  $h_x(u)du$  is closed and if for all  $x, y \in \text{dom}(\phi)$

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

then  $\phi$  is  $c$ -cyclically monotone. This gives an alternative to the criterion in Corollary 5 and Remark 3, which is of advantage in some examples.

**Example 2** a) In the case  $c(x, y) = -|x - y|^p$ ,  $p > 1$ ,  $||$  the euclidean metric, condition (3.27) amounts to

$$(3.30) \quad \begin{aligned} &(g_x(x) - g_x(y))(x - y) \geq 0 \text{ for all } x, y, \\ &\text{with } g_x(y) := |x - \phi(y)|^{p-2}(x - \phi(y)). \end{aligned}$$

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

$$(3.31) \quad g_x(y) = h_x(y), \quad \forall y$$

has the unique solution

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

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

$$(3.32) \quad h_x(y) = (\psi(y) - x)|\psi(y) - x|^{\frac{r}{1-r}}$$

and

$$(3.33) \quad \phi(y) = |h(y)|^{-r} h(y).$$

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  $c(x, y) = -|x - y|_p = -(\sum |x_i - y_i|^p)^{1/p}$ ,  $p \geq 1$ ,  $c_1(x, y) := -\frac{1}{|x-y|_p^{p-1}}(|x_i - y_i|^{p-2}(x_i - y_i))$ ,  $x \neq y$  defines a subgradient and condition (3.27) resp. Remark 5 amounts to:

$$(3.34) \quad \sum_{i=1}^k (x_i - y_i) \left( \frac{|x_i - \phi_i(x)|^{p-2}}{|x - \phi(x)|_p^{p-1}} (\phi_i(x) - x_i) - \frac{|x_i - \phi_i(y)|^{p-2}}{|x - \phi(y)|_p^{p-1}} (\phi_i(y) - x_i) \right) \geq 0.$$

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

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

So Theorem 8 has the following interesting consequence for the  $\ell_1$ -metric w.r.t. the euclidean distance  $|\cdot| = |\cdot|_2$ .

**Corollary 9** (Optimal couplings w.r.t. Kantorovich-metric)

If  $\frac{\phi(u)-u}{|\phi(u)-u|} du$  is closed and  $\phi$  satisfies the normalized angle monotonicity condition

$$(3.36) \quad (x - y) \left( \frac{\phi(x) - x}{|\phi(x) - x|} - \frac{\phi(y) - x}{|\phi(y) - x|} \right) \geq 0 \text{ for all } x, y.$$

then  $(X, \phi(X))$  is an optimal coupling for the  $\ell_1$ -metric w.r.t. euclidean distance on  $\mathbb{R}^k$  for any r.v.  $X$  in the domain of  $\phi$ .

**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  $\ell_1$ , as monotonicity is related to optimality w.r.t. the  $\ell_2$ -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 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:

$$(3.37) \quad \begin{array}{ccc} & \geq & \\ x & & y \text{ implies } c_1(x, \phi(x)) & \geq & c_1(x, \phi(y)), \\ & \leq & & \leq & \end{array}$$

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

$$(3.38) \quad h(x) := c_1(x, \phi(x)) \text{ is monotonically nondecreasing.}$$

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

$$(3.39) \quad \phi(x) - x \text{ is nondecreasing,}$$

while condition (3.36) amounts to

$$(3.40) \quad \phi \text{ is nondecreasing.}$$

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  $\varphi^y$  (cf. (3.18)) is convex. If we can assure the weaker condition that  $\varphi^y$  is quasi-convex, i.e.

$$(3.41) \quad \varphi^y(\alpha x + (1 - \alpha)y) \leq \min(\varphi^y(x), \varphi^y(y)),$$

then a local minimum of  $\varphi^y$  is either situated in a domain where  $\varphi^y$  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  $\phi$ .

- 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

- [1] Cuesta-Albertos, J. A.; Rüschemdorf, L. and Tüero-Díaz, A. (1993): Optimal coupling of multivariate distributions and stochastic processes. *J. Mult. Anal.* 46, 335-361
- [2] Dietrich, H. (1988): Zur  $c$ -Konvexität und  $c$ -Subdifferenzierbarkeit von Funktionalen. *Optimization* 19, 355-371
- [3] Elster, K.H. and Nehse, R. (1974): Zur Theorie der Polarfunktionalen. *Optimization* 5, 3-21
- [4] Kellerer, H. (1984): Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete* 67, 399-432
- [5] Knott, M. and Smith, C. (1984): On the optimal mapping of distributions. *J. Opt. Th. Appl.* 43, 39-49
- [6] 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
- [7] Rachev, S. T. (1991): *Probability Metrics and the Stability of Stochastic Models*. Wiley
- [8] Rachev, S. T. and Rüschemdorf, L. (1991): Solution of some transportation problems with relaxed or additional constraints. To appear in: *SIAM J. Control Optimiz.*, 1994
- [9] Roberts, A. W. and Varberg, D. E. (1973): *Convex Functions*. Academic Press
- [10] Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press
- [11] Rüschemdorf, L. and Rachev, S. T. (1990): A characterization of random variables with minimum  $L^2$ -distance. *J. Mult. Analysis* 32, 48-54

- [12] Rüschendorf, L. (1991a): Bounds for distributions with multivariate marginals. In: Proceedings: Stochastic Order and Decision under Risk. Eds. K. Mosler and M. Scarsini. IMS Lecture Notes, Vol. 19, 285-310
  
- [13] Rüschendorf, L. (1991b): Fréchet-bounds and their applications. In: Advances in Probability Measures with given Marginals. Eds. Dall'Aglio, Kotz and Salinetti, 151-188
  
- [14] Smith, C. and Knott, M. (1992): On Hoeffding Fréchet bounds and cyclic monotone relations. *J. Mult. Anal.* 40, 328-334
  
- [15] Vainberg, M. M. (1973): Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations. Wiley.

Ludger Rüschendorf  
Institut für Mathematische Stochastik  
Hebelstr. 27  
79104 Freiburg  
Germany

e-mail: [ruschen@buffon.mathematik.uni-freiburg.de](mailto:ruschen@buffon.mathematik.uni-freiburg.de)