ON OPTIMAL MULTIVARIATE COUPLINGS

L. RÜSCHENDORF AND L. UCKELMANN
University of Freiburg
Institut für Mathematische Stochastik
Eckerstr. 1, 79104 Freiburg

Abstract. As consequence of a characterization of optimal multivariate coupling (transportation) problems we obtain the existence of optimal Monge solutions as well as an explicit construction method for optimal transportation plans in the case that one mass distribution is discrete. We also give a new characterization of an extension of the transportation problem with more than two mass distributions involved.

1. c-optimal couplings on $\mathbb{R}^k$

The multivariate coupling (transportation) problem on $\mathbb{R}^k$ for a transportation 'cost function' $c : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^1$ and given mass distributions $P, Q \in M^1(\mathbb{R}^k, \mathbb{B}^k)$ is defined as the problem

$$S_c(P,Q) \;=\; \sup\left\{\int c(x,y)d\mu(x,y); \mu \in M(P,Q)\right\} \tag{1.1}$$

where $M(P,Q)$ is the class of all probabilities on $\mathbb{R}^k \times \mathbb{R}^k$ with marginals P and Q. A pair of random variables $X \stackrel{d}{=} P, Y \stackrel{d}{=} Q$ is called *c-optimal* if

$$Ec(X,Y) \;=\; \sup\left\{Ec(U,V); U \stackrel{d}{=} P, V \stackrel{d}{=} Q\right\} \;=\; S_c(P,Q). \tag{1.2}$$

The calculation of the optimal value $S_c(P,Q)$ and the construction of c-optimal solutions (X,Y) is a basic problem in probability theory with many interesting applications (cf. Rachev (1991), Cuesta-Albertos, Matran, Rachev and Rüschendorf (1996)). (Note that the corresponding infimum problem can be reduced to the sup problem by switching from c to $-c$.)

The following characterization of optimal solutions given in Rüschendorf (1991) is basic. For its formulation we need some preliminary notions. Call a function f on $\mathbb{R}^k$ *c-convex* if for some index set I and $x_i \in \mathbb{R}^k, a_i \in \mathbb{R}^1, i \in I$

$$f(x) \;=\; \sup_{i\in I}(c(x,y_i)+a_i)\;. \tag{1.3}$$

The c-conjugate of a function f is defined by

$$f^*(y) \;=\; \sup_x(c(x,y)-f(x))\;, \tag{1.4}$$

the sup being over the domain of f. Define the doubly c-conjugate

$$f^{**}(x) \;=\; \sup(c(x,y)-f^*(y))\;. \tag{1.5}$$

Then f^* and 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. Also f^*, f^{**} are *admissible* in the sense that

$$f^*(y)+f^{**}(x) \;\geq\; c(x,y) \qquad \text{for all } x,y\;. \tag{1.6}$$

The (doubly) c-conjugate functions are basic for the theory of inequalities as in (1.6). The *c-subgradient* of a function f is defined by

$$\partial_c f(x) \;=\; \{y; f(z)-f(x)\geq c(z,y)-c(x,y) \quad \forall z\in \text{ dom } f\}\;. \tag{1.7}$$

Let $\mathcal{L}_m(P,Q)$ be 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 $f_1\in\mathcal{L}^1(P)$ and $f_2\in L^1(Q)$. The following characterization is then to be found in Rüschendorf (1991, 1995).

Theorem 1.1 *Let $c\in\mathcal{L}_m(P,Q)$ and assume that* $I(c)=\inf\left\{\int h_1dP+\int h_2dQ; c\leq h_1\oplus h_2, h_1\in L^1(P), h_2\in L^1(Q)\right\}<\infty$.

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

$$Y\in\partial_c f(X) \quad \textit{a.s. for some c-convex function } f\;; \tag{1.8}$$

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

The characterization in (1.8) 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 $x_{n+1}:=x_1$,

$$\sum_{i=1}^n(c(x_{i+1},y_i)-c(x_i,y_i)) \;\leq\; 0 \tag{1.9}$$

(cf. Knott and Smith (1993) and Rüschendorf (1996)) Some applications of (1.8) and (1.9) can be found in Cuesta-Albertos and Tuero-Diaz (1993), Rachev (1991), Rüschendorf (1995). If $c(\cdot, y)$ is locally Lipschitz and P has a Lebesgue-density, then for a c-optimal pair (X, Y) holds

$$\nabla f(X) \;=\; \nabla_x c(X,Y) \quad \text{a.s.} \tag{1.10}$$

(cf. Rüschendorf (1991), formula (73))

We next discuss two consequences of the characterization of c-optimal pairs (X, Y).

1.1. MONGE FUNCTIONS

The Monge problem is to find an optimal pair of the form $(X, \Phi(X)), X \stackrel{d}{=} P, \Phi(X) \stackrel{d}{=} Q$. Φ is then called an optimal Monge function. Since a Monge function Φ is determined by a nonlinear variational problem even its existence has been an open problem for a long time. Some sufficient conditions for the existence of (optimal) Monge functions have been given in the literature starting with Sudakov(1979) (cf. Cuesta-Albertos and Tuero-Diaz (1993), Gangbo and McCann (1996)). Gangbo and McCann (1996) have shown that an optimal Monge function exists for strictly convex cost functions of the form $c(x, y) = h(x - y)$ if P has a Lebesgue density. We remark that in this case formula (1.10) implies

$$Y \;=\; X - (\nabla h)^{-1}(\nabla f(X)) \quad \text{a.s.}\ ; \tag{1.11}$$

the right hand side defines a Monge function $\Phi(X)$. More generally this holds if (1.10) can be resolved uniquely in Y.

So the characterization formula (1.8) implies the existence of Monge solutions if P is Lebesgue-continuous and also gives an interesting relation between the gradient of c-convex functions f and their c-subgradients, the c-optimal functions Φ (cf. also Gangbo and McCann (1996) for an alternative derivation).

1.2. EXPLICIT SOLUTIONS FOR DISCRETE Q

Let $Q = \sum_{j=1}^n \alpha_j \varepsilon_{x_j}$ be a discrete distribution on $x_1, \ldots, x_n \in \mathbb{R}^k$ then by (1.3) we may restrict our discussion to c-convex functions of the form

$$f(x) \;=\; \sup_{1 \le j \le n} \left(c(x, x_j) + a_j\right) . \tag{1.12}$$

Consider the sets $A_j = \{x; f(x) = c(x, x_j) + a_j\}$, then

$$x_j \in \partial_c f(x) \text{ for } x \in A_j \ . \tag{1.13}$$

The subgradient is not unique only on the boundaries of A_j. The problem of finding c-optimal couplings therefore, is equivalent to finding suitable shifts a_j such that

$$P(A_j) \;=\; \alpha_j, 1 \le j \le n \;. \tag{1.14}$$

The optimal coupling then is given by $\phi = \sum_{j=1}^n x_j 1_{A_j}$.

If the boundaries of A_j have measure zero with respect to P then as a consequence one obtains the existence and uniqueness of an optimal Monge function for this problem. This application of the characterization in (1.8) also has been observed independently in Gangbo and McCann (1996).

In the case that $c(x,y) = -\|x-y\|^2$ the boundaries are linear and can be calculated explicitely for n not too large. The following example with P the uniform distribution on the unit square and

$$Q \;=\; \sum_{j=1}^{8} \alpha_j \varepsilon_{x_j},$$

has been solved approximatively in Abdellaoui (1994). The following solution is exact. For
$(\alpha_1, \cdots, \alpha_8) = (0.105, 0.2, 0.125, 0.125, 0.125, 0.12, 0.1, 0.1)$ and
$(x_1, \cdots, x_8) = ((0,1), (0.5, 0.5), (1,1), (1,0), (0,0), (1,4), (2,3), (1,2))$
one obtains
$(a_1, \cdots, a_8) = (0.18, -0.01, 0.43, 0.05, 0, 10.86, 7.22, 2.19)$.

The corresponding optimal partition is given in the following Figure 1:

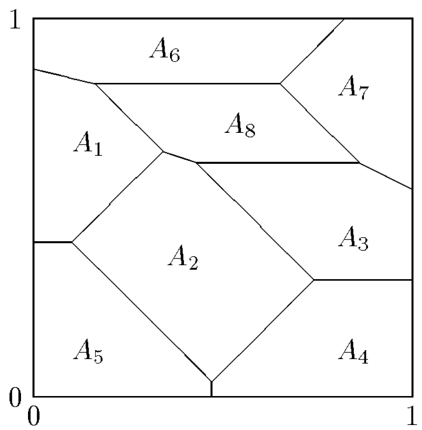

Figure 1

The corresponding c-convex function $f(x) = \sup_{1\le j\le n}(c(x,x_j)+a_j)$ is given in the following Figure 2:

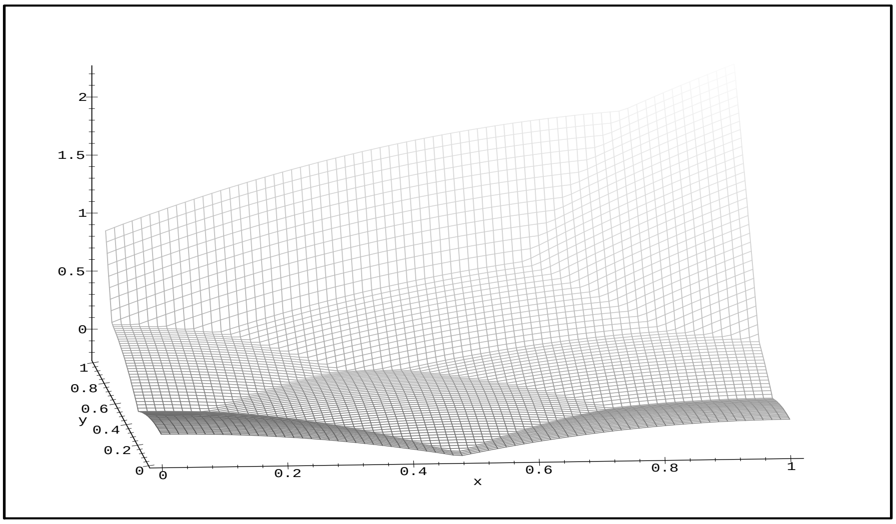

Figure 2

For the cost function $c(x,y) = -\|x-y\|^4$ it is more difficult to calculate the volume of the sets A_j. The following example is for P the uniform distribution on $[0,1]^2$ and $Q = \sum_{l=1}^4 \alpha_i \varepsilon_{x_i}$ where $(\alpha_1,\cdots,\alpha_4) = (0.34, 0.05, 0.22, 0.39)$, $(x_1,\cdots,x_4) = ((1,1),(0,0),(1,0),(0,1))$. One obtains $(a_1,\cdots,a_4) = (0,-0.5,-0.25,0)$

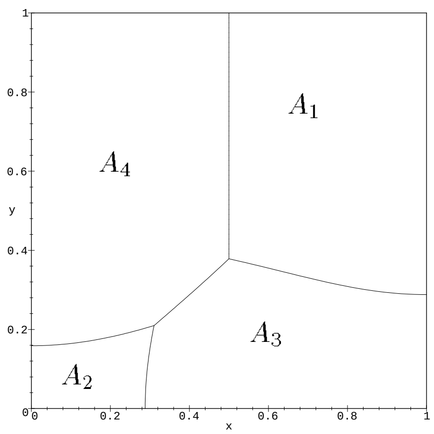

Figure 3

The c-convex function corresponding to the optimal solution is given in the following Figure 4:

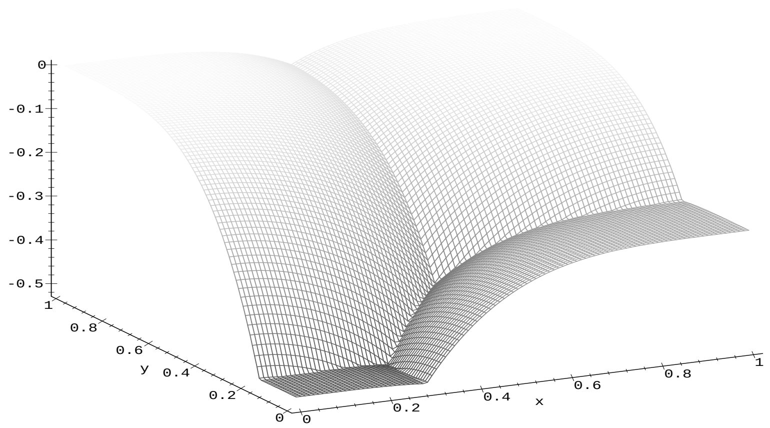

Figure 4

2. A generalized transportation problem

The general duality theorem also holds for cost functions of n-variables $c(x_1,\dots,x_n), x_i \in \mathbb{R}^k$ (cf. Rachev (1991), Rüschendorf (1981)) which are lower majorized (as in Theorem 1)

$$\begin{aligned} &\sup\left\{\int c(x_1,\dots,x_n)d\mu; \mu \in M(P_1,\dots,P_n)\right\} \\ &= \inf\left\{\sum_{i=1}^n \int f_i dP_i; f_i \in \mathcal{L}^1(P_i); c(x_1,\dots,x_n) \le \sum_{i=1}^n f_i(x_i)\right\} \quad (2.1) \end{aligned}$$

where $M(P_1,\dots,P_n)$ is the set of all probability measures on $\mathbb{R}^k \times \cdots \times \mathbb{R}^k$ with marginals $P_1,\dots,P_n$. Not many explicit results are known in this generalized case. Recently Olkin and Rachev (1993) and Knott and Smith (1993) have considered the case $n = 3$ and

$$c(x,y,z) := \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle \quad (2.2)$$

in particular for normal marginals $P_i = N(0,\Sigma_i)$.

Note that in this case the problem

$$\sup\{Ec(X,Y,Z); X \stackrel{d}{=} P_1, X \stackrel{d}{=} P_2, Z \stackrel{d}{=} P_3\} \quad (2.3)$$

is equivalent to

$$\inf\{E\|X-Y\|^2+\|Y-Z\|^2+\|Z-X\|^2; X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2, Z = P_3\}$$

and also to

$$\sup\{E\|X+Y+Z\|^2; X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2, Z \stackrel{d}{=} P_3\} \ .$$

In the duality theorem (2.1), for the solution of (2.3), we can restrict consideration to convex functions f_i (redefining $f_1(x_1)$ as $\sup\{c(x_1,x_2,x_3)-f_2(x_2)-f_3(x_3); x_2,x_3 \in \mathbb{R}^k\}$ and similarly f_2, f_3) and, therefore, for an optimal solution X,Y,Z holds:

$$Y+Z \in \partial f_1(X), X+Z \in \partial f_2(Y), X+Y \in \partial f_3(Z) \tag{2.4}$$

i.e. $Y+Z$ is optimally coupled to X, etc. Therefore, with $g_i(x) = f_i(x) + \frac{1}{2}\|x\|^2$ the sum $X+Y+Z$ is also optimally coupled with any of X,Y,Z

$$X+Y+Z \in \partial g_1(X) \cap \partial g_2(Y) \cap \partial g_3(Z). \tag{2.5}$$

For the corresponding minimum problem

$$\inf\{Ec(X,Y,Z); X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2, Z \stackrel{d}{=} P_3\} \tag{2.6}$$

one obtains similarly that

$$Y+Z \in \partial f_1(X), X+Z \in \partial f_2(Y), X+Y \in \partial f_3(Z) \tag{2.7}$$

for some concave functions f_i and so again $Y+Z$ is 'optimally' coupled to X, etc.

Remark 2.1

a) *Condition (2.7) is not sufficient for optimality as can be seen by the following one dimensional example where* $P_1 = P_2 = P_3 = \frac{1}{8}\sum_{i=1}^{8} \varepsilon_{\{i\}}$. *Consider random variables* X,Y,Z *given by permutations (with equal probabilities)* $X \simeq (1\ 4\ 2\ 3\ 8\ 7\ 6\ 5)$ $Y \simeq (3\ 5\ 7\ 8\ 4\ 1\ 2\ 6)$ $Z \simeq (8\ 5\ 4\ 2\ 1\ 7\ 6\ 3)$ *(i.e.* $P(X=1, Y=3, Z=8) = \frac{1}{8}$*), then (2.7) is fulfilled. Similarly the triple* $U \simeq (1\,3\,2\,4\,8\,7\,6\,5), V \simeq (4\,5\,7\,8\,3\,2\,1\,6)$ *and* $W \simeq (8\ 6\ 5\ 1\ 2\ 4\ 7\ 3)$ *satisfies (2.7) but* $E\|X+Y+Z\|^2 = E\|U+V+W\|^2 + 1/2$.

b) *Since* $E\|X+Y+Z\|^2 = E(\langle X,t\rangle + \langle Y,t\rangle + \langle Z,t\rangle)$ *with* $t = X+Y+Z$ *the following approach suggested in Knott and Smith (1994) is promising. Try to find* $X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2$ *and* $Z \stackrel{d}{=} P_3$ *which are optimal coupled to their sum* t*. In the normal case* $P_i = N(0,\Sigma_i)$ *and if* $t \stackrel{d}{=} N(0,\Sigma_0)$*;* Σ_0

nonsingular, it is well known that optimal coupling functions between t and X_i are given by

$$T_i = \Sigma_i^{1/2}\left(\Sigma_i^{1/2}\Sigma_0\Sigma_i^{1/2}\right)^{-1/2}\Sigma_i^{1/2}.$$

Then the condition $\sum_{i=1}^3 T_i = I$ is easily seen to be equivalent to

$$\sum_{i=1}^{3}\left(\Sigma_0^{1/2}\Sigma_i\Sigma_0^{1/2}\right)^{1/2} = \Sigma_0 \,. \tag{2.8}$$

Under the assumption that (2.8) has a positive definite solution Knott and Smith (1993) establish optimality of the triple defined via T_i. (2.5) gives a justification of this approach of Knott and Smith (1994).
If (X,Y,Z) is optimal then without loss in generality we may assume that $(X,Y,Z) \overset{d}{=} N(0,\Sigma)$ is jointly optimal. Assuming that $\Sigma_0^* =$ Cov $(X+Y+Z)$ is nonsingular we conclude by (2.5) and the uniqueness of optimal couplings in the normal case that equation (2.8) has a solution (namely Σ_0^*). Positive definiteness of Σ_0^* can be shown for special cases (e.g. the case of commutative $\Sigma_1,\Sigma_2,\Sigma_2$) but is an open problem in general.

The following result gives a necessary and sufficient characterization of optimal solutions of (2.2) in general.

Theorem 2.2 *Let $X \overset{d}{=} P_1, Y \overset{d}{=} P_2$ and $Z \overset{d}{=} P_3$, let P_i have finite covariance matrices, then (X,Y,Z) is optimal for problem (2.2) if and only if there exists a convex, lower semicontinuous function f and a F-convex function g with $F(y,z) := f^*(y+z) + \langle y,z\rangle$ such that*

$$\begin{aligned} &(1) \quad Y+Z \in \partial f(X) \quad a.s.\,, \\ &(2) \quad Z \in \partial_F g(Y) \quad a.s.\,. \end{aligned} \tag{2.9}$$

Proof: The duality theorem (2.1) is in the case (2.2)

$$\begin{aligned} &\sup_{\mu\in M(P_1,P_2,P_3)} \int (\langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle) d\mu(x,y,z) \\ &= \inf\left\{ \int f dP_1 + \int g dP_2 + \int h dP_3; f(x)+g(y)+h(z) \right. \\ &\qquad \left. \geq \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle \right\} \end{aligned} \tag{2.10}$$

Assume conditions (1) and (2) and define the F-conjugate of g by

$$h(z) := g^F(z) = \sup_y \{F(y,z) - g(y)\} . \tag{2.11}$$

Then the triple (f, g, h) is admissible, i.e.

$$\begin{aligned} & f(x) + g(y) + h(z) \\ = \; & f(x) + g(y) + \sup_{\overline{y}} \{ f^*(\overline{y} + z) + \langle \overline{y}, z\rangle - g(\overline{y})\} \\ = \; & f(x) + g(y) + \sup_{\overline{y}} \{ \sup_{\overline{x}} \{\langle \overline{x}, \overline{y} + z\rangle - f(\overline{x})\} + \langle \overline{y}, z\rangle - g(\overline{y})\} \\ = \; & f(x) + g(y) + \sup_{\overline{y}} \{ \sup_{\overline{x}} \{\langle \overline{x}, \overline{y}\rangle + \langle \overline{x}, z\rangle + \langle \overline{y}, z\rangle - f(\overline{x}) - g(\overline{y})\}\} \\ \geq \; & f(x) + g(y) + \langle x, y\rangle + \langle x, z\rangle + \langle y, z\rangle - f(x) g(y) \\ = \; & \langle x, y\rangle + \langle y, z\rangle + \langle x, z\rangle. \end{aligned} \tag{2.12}$$

Furthermore, from (1) $Y + Z \in \partial f(X)$ a.s. and, therefore, $f(X) + f^*(Y + Z) = \langle X, Y + Z\rangle$ a.s. which implies

$$f(X) + f^*(Y+Z) + \langle Y, Z\rangle = f(X) + F(Y,Z) = \langle X, Y\rangle + \langle X, Z\rangle + \langle Y, Z\rangle \text{ a.s.}$$

From (2) $Z \in \partial_F g(Y)$ a.s. and so $g(Y) + g^F(Z) = F(Y,Z)$ a.s. This implies

$$f(X) + g(Y) + g^F(Z) = \langle X, Y\rangle + \langle Y, Z\rangle + \langle X, Z\rangle \quad a.s. \tag{2.13}$$

The inequality (2.12) and (2.13) imply optimality, since for any random variables $\widetilde{X} \stackrel{d}{=} P_1, \widetilde{Y} \stackrel{d}{=} P_2$ and $\widetilde{Z} \stackrel{d}{=} P_3$,

$$\begin{aligned} & E(\langle \widetilde{X}, \widetilde{Y}\rangle + \langle \widetilde{Y}, \widetilde{Z}\rangle + \langle \widetilde{X}, \widetilde{Z}\rangle) \\ \leq \; & E(t(\widetilde{X}) + g(\widetilde{Y}) + h(\widetilde{Z})) \\ = \; & E(t(X) + g(Y) + h(Z)) \\ = \; & E(\langle X, Y\rangle + \langle Y, Z\rangle + \langle X, Z\rangle). \end{aligned} \tag{2.14}$$

For the opposite direction there exist optimal solutions (f_1, f_2, f_3) of the dual problem. So for an optimal measure $\mu \in M(P_1, P_2, P_3)$

$$\int (\langle x, y\rangle + \langle y, z\rangle + \langle x, z\rangle) d\mu = \int f_1 dP_1 + \int f_2 dP_2 + \int f_3 dP_3 , \tag{2.15}$$

where $f_1(x) + f_2(y) + f_3(z) \geq c(x, y, z)$, and equality holds on the support of μ.

Define $f(x) = f_1^{**}(x) = \sup_y\{\langle x, y\rangle - f_1^*(y)\}$ where $f_1^*(y) = \sup_x\{\langle x, y\rangle - f_1(x)\}$ is the conjugate. Then f is convex and lower semicontinuous and $f(x) \le f_1(x)$. Define

$$\begin{aligned} F(y,z) &:= \sup_x\{\langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle - f(x)\} \\ &= \sup_x\{\langle x, y+z\rangle - f(x)\} + \langle y,z\rangle \\ &= f^*(y+z) + \langle y,z\rangle \;. \end{aligned}$$

Then it holds

$$\begin{aligned} f_1(x) + f_2(y) + f_3(z) &\ge f(x) + f_2(y) + f_3(z) \\ \ge f(x) + F(y,z) &\ge \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle. \end{aligned}$$

As (f_1, f_2, f_3) is a solution of the dual problem we obtain

$$0 = \int (f(x) + F(y,z) - (\langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle))d\mu \tag{2.16}$$

and, therefore,

$$f(x) + F(y,z) = \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle \quad \mu \text{ a.s.} \tag{2.17}$$

This implies that

$$f(x) + f^*(y+z) = \langle x, y+z\rangle \quad \mu \text{ a.s.}$$

and, therefore,

$$Y + Z \in \partial f(X) \qquad \text{a.s..}$$

For condition (2) define the double F-conjugate

$$g(y) = f_2^{FF}(y) = \sup_z\{F(y,z) - f_2^F(z)\}$$

where

$$f_2^F(z) := \sup_y\{F(y,z) - f_2(y)\}.$$

We have that g is F-convex; with

$$\begin{aligned} &h(z) := g^F(z)(= f_2^F(z)) \quad \text{holds (as above)} \\ &f(x) + f_2(y) + f_3(z) \ge f(x) + g(y) + h(z) \\ &\ge \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle. \end{aligned}$$

Again from the optimality equation (2.15)

$$f(x) + g(y) + h(z) = \langle x,y\rangle + \langle y,z\rangle + \langle x,z\rangle \quad \mu \text{ a.s.}$$

This implies that

$$g(Y) + h(Z) \; = \; F(Y, Z) \quad \text{a.s.}$$

and, as $h = g^F$ it follows that

$$Z \in \partial_F g(Y) \ .$$

□

Remark 2.3

a) If g is F-convex, then g is convex as supremum of the convex functions of the type $f^*(\cdot + z) + \langle \cdot, z\rangle + a$. Condition (2) can be reformulated using (1.10) by

$$\nabla g(Y) \; = \; \nabla f^*(Y + Z) + Z \quad a.s.$$

using continuity of the involved distributions.
Define $h(y) := f^*(y) + \frac{1}{2}\|y\|^2$, then we obtain

$$(\nabla h)^{-1}(\nabla g(Y) + Y) - Y \; = \; Z \tag{2.18}$$

b) An analog result holds true for the inf-problem

$$\inf_{X,Y,Z} E(\langle X, Y\rangle + \langle Y, Z\rangle + \langle X, Z\rangle) \tag{2.19}$$

The corresponding characterizations are: (X, Y, Z) is optimal for (2.17)

$$\begin{aligned} \Leftrightarrow \quad & (1) \quad Y + Z \in -\partial(-f)(X) \\ & (2) \quad Z \in \partial_{-F} - g(Y) \end{aligned} \tag{2.20}$$

where f is concave, f^* the concave conjugate and g is F-concave.

Example 2.4 *Consider the univariate example $P_1 = P_2 = P_3 = U([0,1])$ the uniform distribution on $[0,1]$. Then the* inf *problem (2.17) is solved by $(X, \Phi_1(X), \Phi_2(X))$, where*

$$\begin{aligned} X \overset{d}{=} U([0,1]), \quad \Phi_1(x) &:= \begin{cases} 1 - 2x, & x \le \frac{1}{2} \\ 2 - 2x, & x > \frac{1}{2} \end{cases} \\ \Phi_2(x) &:= \begin{cases} x + \frac{1}{2}, & x \le \frac{1}{2} \\ x - \frac{1}{2}, & x > \frac{1}{2} \end{cases} \end{aligned} \tag{2.21}$$

For the proof define $f(x) := \frac{3}{2}x - \frac{1}{2}x^2 = g(x)$. Then f is concave and

$$-\partial(-f(x)) \; = \; f'(x) \; = \; \frac{3}{2} - x \; = \; \Phi_1(x) + \Phi_2(x)$$

and, therefore,

$$\Phi_1(x) + \Phi_2(x) \in -\partial(-f(x)) \tag{2.22}$$

i.e. condition (1).

Further,

$$\begin{aligned} f^*(x) &= \inf_y \{xy - f(y)\} \\ &= \inf_y \{xy - \frac{3}{2}y + \frac{1}{2}y^2\} \;=\; -\frac{1}{2}\left(\frac{3}{2} - x\right)^2. \end{aligned}$$

Therefore,

$$\begin{aligned} F(y,z) &= f^*(y+z) + yz \\ &= -\frac{9}{8} + \frac{3}{2}y + \frac{3}{2}z - \frac{1}{2}y^2 - \frac{1}{2}z^2 - yz + yz \\ &= g(y) + g(z) - \frac{9}{8} \end{aligned}$$

This implies

$$F(y',z) - F(y,z) \;=\; g(y') - g(y). \tag{2.23}$$

Since $g(y) = F(y,z) - g(z) + \frac{9}{8}$, *g is F concave and* $z \in \partial_F g(y)$. *Therefore, the second condition* $\Phi_2(x) \in \partial_{-F} - g(\Phi_1(x))$ *is fulfilled and the optimality is established by (2.18).*

As consequence of the characterization in Theorem 2.2 one obtains the following more specific coupling property of an optimal pair in problem (2.3) to the sum.

Proposition 2.5 *If* (X,Y,Z) *is an optimal solution for* $\sup\{E\langle X,Y\rangle + \langle Y,Z\rangle + \langle X,Z\rangle\}$ *then for the following convex functions* f_1, f_2, f_3 *given (in the notation of Theorem 2.2) by* $f_1(x) = f(x) + \frac{1}{2}\|x\|^2$, $\quad f_2(x) = g(x) + \frac{1}{2}\|x\|^2$ *and* $f_3(x) = g^F(x) + \frac{1}{2}\|x\|^2$.

(1) $X + Y + Z \in \partial f_1(X)$ *a.s.*

(2) $X + Y + Z \in \partial f_2(Y)$ *a.s.*

(3) $X + Y + Z \in \partial f_3(Z)$ *a.s.*

i.e. X, Y, Z *are optimally coupled to the sum.*

Proof: From (2.9) in Theorem 2.2 $Y + Z \in \partial f(X)$ and, therefore, for any x'

$$\begin{aligned} \langle X+Y+Z, X-x'\rangle &= \langle Y+Z, X-x'\rangle + \langle X, X-x'\rangle \\ &\geq f(X) - f(x') + \frac{1}{2}\|X\|^2 - \frac{1}{2}\|x'\|^2. \end{aligned}$$

This implies $X + Y + Z \in \partial f_1(X)$.

Since $X \in \partial f^*(Y + Z)$

$$\langle X, Y + Z - \xi\rangle \;\geq\; f^*(Y + Z) - f^*(\xi), \forall\xi.$$

Therefore, from (2.9) relation (2):

$$F(Y,Z) - F(y',Z) = f^*(Y+Z) - f^*(y'+Z) + \langle Y - y', Z\rangle \geq g(Y) - g(y').$$

g is convex by Remark 2.2a) and as $\langle X+Z, Y-y'\rangle = \langle X, Y+Z-(y'+Z)\rangle + \langle Y-y', Z\rangle \geq f^*(Y+Z) - f^*(y'+Z) + \langle Y-y', Z\rangle$ we obtain $X+Z \in \partial g(Y)$ and, therefore, $X+Y+Z \in \partial f_2(Y)$ a.s.

Finally, from (2) in (2.9) $Y \in \partial_F g^F(Z)$. The F conjugate of g is convex. Then we can argue as a above to obtain $X+Y+Z \in \partial f_3(Z)$. □

References

1. Abdellaoui, T. (1994): Optimal solution of the Monge-Kantorovich transference problem, *Preprint.*
2. Cuesta-Albertos, J.A. and Tuero-Diaz, A. (1993): A characterization for the solution of the Monge-Kantorovich mass transference problem, *Stat. Prob. Letters* **16**, 147-152.
3. Cuesta-Albertos, J.A., Matran, C., Rachev, S.T. and Rüschendorf, L. (1996): Mass transportation problems in probability theory, *Math. Scientist* **21**, 34-72.
4. Gangbo, W. and McCann, R.J. (1996): The geometry of optimal transportation, *Preprint.*
5. Knott, M. and Smith, C.S. (1993): On a generalization of cyclic-monotonicity and distances among random vectors, *Preprint.*
6. Olkin, I. and Rachev, S.T. (1993): Maximum submatrix traces for positive definite matrices, *SIAM J. Matrix Anal. Appl.* **14**, 390-397.
7. Rachev, S.T. (1991): Probability Matrices and the stability of stochastic models, Wiley.
8. Rüschendorf, L. (1981): Sharpness of Fréchet-bounds, *Z. für Wahrscheinlichkeitstheorie verw. Geb.* **57**, 293-302.
9. Rüschendorf, L. (1991): Fréchet-bounds and their applications. In: Advances in Probability Measures with Given Marginals. Eds.: G. Dall'Aglio, S. Kotz and G. Salinetti, Kluwer Acad. Publ., 151-188.
10. Rüschendorf, L. (1995): Optimal solution of multivariate coupling problems, *Appl. Mathematicae* **23**, 325-338.
11. Rüschendorf, L. (1996): On c-optimal random variables, *Stat. Prob. Letter* **27**, 267-270.
12. Smith, C. and Knott, M. (1992): On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Mult. Anal.* **40**, 328-334.
13. Sudakov, V.N. (1979): Geometric problems in the theory of infinite dimensional probability distributions, *Proc. Steklov Inst. Math.* **141**, 1-178.