On the n-coupling problem

Ludger Rüschendorf* and Ludger Uckelmann

Institut für Mathematische Stochastik, University of Freiburg,

Eckerstr. 1, 79104 Freiburg, Germany

Abstract

In this paper we give a justification of the idea of Knott and Smith (1994) to solve three-coupling problems by using optimal couplings to the sum. In the normal case this leads to a complete solution. Under a technical condition this idea also works for general distributions and one obtains explicit results. We extend these results to the n-coupling problem and derive a characterization of optimal n-couplings by several 2-coupling problems. This leads to some constructive existence results for Monge solutions.

1 Introduction

Olkin and Rachev (1993) introduced and discussed the problem of simultaneous optimal coupling of three multivariate normal distributions. The problem is, given $P_i = N(0, \sum_i)$, $i = 1, 2, 3$ on $\mathbb{R}^d$, $\sum_i$ positive definite, to find random vectors $X \stackrel{d}{=} P_1$, $Y \stackrel{d}{=} P_2$, $Z \stackrel{d}{=} P_3$ such that

$$E\ ||X - Y||^2 + E\ ||Y - Z||^2 + E\ ||X - Z||^2 = \min; \tag{1.1}$$

the minimum over all random vectors X, Y, Z with distributions P_1, P_2, P_3. So in L^2-sense X, Y, Z are as close as possible in average given the marginal distributions P_i.

In comparison to the coupling of three or more distributions the coupling problem for two distributions is well investigated and an usable characterization of an optimal coupling is known (cf. Rüschendorf and Rachev (1990)). Obviously if it is possible to minimize each of the three summands in (1.1)separately by one triple (X, Y, Z), then one gets a solution of the three-coupling problem (1.1). But this assumption imposes severe symmetry conditions on $\sum_i$ such as commutativity $\sum_i \sum_j = \sum_j \sum_i$. So in general the three-coupling problem can not be reduced to the simpler two

*Corresponding author. E-mail: ruschen@pollux.mathematik.uni-freiburg.de

coupling problem directly.

Knott and Smith (1994) proposed an interesting idea to reduce the three-coupling problem to some related two-coupling problems. Note that problem (1.1) is equivalent to each of the following maximization problems

$$E\,(\langle X,Y\rangle + \langle Y,Z\rangle + \langle X,Z\rangle) = \max \tag{1.2}$$

or

$$E\,||X+Y+Z||^2 = \max \tag{1.3}$$

or

$$E\,(||X-T||^2 + ||Y-T||^2 + ||Z-T||^2) = \max, \tag{1.4}$$

where $T := X+Y+Z$ and the max is again over all random vectors X, Y, Z with distributions P_i. Therefore, Knott and Smith (1994) suggested that a triple (X, Y, Z) with the given marginal distributions 'should' be optimal if each of X, Y, Z is optimally coupled to its sum T. Using this idea they were able to construct in the normal case $P_i = N(0, \sum_i)$ an optimal triple (X, Y, Z) under the assumption that a positive definite solution $\sum_0$ of the matrix equation

$$\sum_{i=1}^{3} \left(\Sigma_0^{1/2}\ \Sigma_i\ \Sigma_0^{1/2}\right)^{1/2} = \Sigma_0 \tag{1.5}$$

can be found. This nonlinear matrix equation is a consequence of the 'coupling to the sum' idea.

For the construction of an optimal triple (X, Y, Z) let T be a random vector, $T \stackrel{d}{=} N(0, \sum_0)$ and define

$$S_i = \Sigma_i^{1/2} \left(\Sigma_i^{1/2}\ \Sigma_0\ \Sigma_i^{1/2}\right)^{-1/2} \Sigma_i^{1/2}. \tag{1.6}$$

S_i is the optimal coupling mapping between $N(0, \sum_0)$ and $N(0, \sum_i)$ (see Olkin and Pukelsheim (1982)). Then defining

$$X := S_1\, T, \qquad Y := S_2\, T, \qquad Z := S_3\, T \tag{1.7}$$

(1.5) implies that $X+Y+Z = T$ and Knott and Smith (1994) proved that this triple is optimal and

$$E\,||T||^2 = \operatorname{tr}(\Sigma_0)\,. \tag{1.8}$$

In this paper we prove existence of a positive definite solution of equation (1.5). Thus Knott and Smith's idea leads to a complete solution of the three-coupling problem in the normal case. We also show that up to some 'technical' assumption one can justify the idea of 'optimal coupling to the sum' for general distributions. We further derive a characterization of optimal solutions by some related two-coupling problems with respect to coupling functionals $F(x, y)$ different from $||x-y||^2$. This characterization can be used for explicit solutions in some concrete examples.

Finally we obtain a simple proof of a recent result of Gangbo and Swiech (1996) on the existence of Monge solutions and by means of our characterization result get more constructive results on the existence of Monge solutions. All results in this paper are extended to n-coupling problems involving n probability measures $P_1, \ldots, P_n$ on $\mathbb{R}^d$. The n-coupling problem is to find to given probability measures P_i on $\mathbb{R}^d, 1 \leq i \leq n$, random vectors X_i with distribution P_i such that

$$E \left\| \sum_{i=1}^{n} X_i \right\|^2 = \max, \tag{1.9}$$

equivalently

$$E \sum_{i<j} ||X_i - X_j||^2 = \min, \tag{1.10}$$

or

$$E \sum_{i<j} \langle X_i, X_j \rangle = \max \tag{1.11}$$

the max resp. min over all random vectors with distributions P_i. We use the notation $X_i \stackrel{d}{=} P_i$ for equality in distribution and assume throughout the paper that P_i have second moments.

2 Optimal n-couplings

Optimal couplings for two probability measures P, Q on $\mathbb{R}^d$ w.r.t. the squared distance $c\,(x,y) = ||x-y||^2$ are characterized by the following result (see Rüschendorf and Rachev (1990)): $X \stackrel{d}{=} P, Y \stackrel{d}{=} Q$ are optimal if and only if there exists a convex lower semicontinuous function f such that

$$Y \in \partial\, f(X) \quad \text{a.s.}, \tag{2.1}$$

where $\partial f(x)$ is the subgradient of f in x. Equivalently, with the conjugate function $f^*(y) = \sup_x(\langle x, y\rangle - f(x))$ if and only if

$$X \in \partial f^*(Y) \quad \text{a.s.} \tag{2.2}$$

For general coupling functions $c\,(x, y)$ a corresponding result holds (see Rüschendorf (1991), (1995)), where convex functions are to be replaced by c-convex functions of the form

$$f(x) = \sup_I \{c\,(x, y_i) + a_i\} \tag{2.3}$$

for some index set I, $y_i \in \mathbb{R}^d$, $a_i \in \mathbb{R}$ and the subgradient set is to be replaced by the c-subgradient set

$$\partial_c f(x) = \{y \in \mathbb{R}^d;\ f(z) - f(x) \geq c(z,y) - c(x,y) \text{ for all } z \in \operatorname{dom} f\}. \quad (2.4)$$

Under an integrability condition on c a pair (X,Y), $X \stackrel{d}{=} P$, $Y \stackrel{d}{=} Y$ is c-optimal i.e.

$$Ec(X,Y) = \sup\{Ec(X_1,X_2);\quad X_1 \stackrel{d}{=} P,\ X_2 \stackrel{d}{=} Q\} \quad (2.5)$$

if and only if

$$Y \in \partial_c f(X) \ \text{ a.s.} \quad (2.6)$$

for some c-convex function f; equivalently,

$$X \in \partial_c f^c(Y) \ \text{ a.s.} \quad (2.7)$$

where

$$f^c(y) = \sup(c(x,y) - f(x)), \quad (2.8)$$

is the c-conjugate of f.

The following proposition states necessity of optimal coupling to the sum.

Proposition 2.1 (Necessity of optimal coupling to the sum)
Let $X_i \stackrel{d}{=} P_i$, $1 \leq i \leq n$, and let $X_1, \ldots, X_n$ be an optimal n-coupling for $P_1, \ldots, P_n$, then with $T_i := \sum_{j \neq i} X_j, T := \sum_{j=1}^n X_j$, X_i is optimally coupled to the sum T_i as well as to $T, 1 \leq i \leq n$.

Proof: Consider w.l.g. the case $i = 1$. The n-coupling problem (1.9) is equivalent to (1.11) i.e. to

$$E\left(\langle X_1, T_1\rangle + \sum_{i=2}^n \left\langle X_i, \sum_{j>i} X_j\right\rangle\right) = \max! \quad (2.9)$$

The second term depends only on $X_2, \ldots, X_n$. If X_1 were not optimally coupled to T_1, it would be possible to find a strict improvement of (2.9).
Furthermore, (2.9) is equivalent to

$$E\left(\langle X_1, T\rangle + \sum_{i=2}^n \left\langle X_i, \sum_{j>i} X_j\right\rangle\right) = \max!$$

(the difference depends only on the marginal distribution). Therefore by the same argument, X_1 has to be optimally coupled to the sum T as well. $\square$

We next prove that Knott and Smith's idea of optimal coupling to the sum leads to a complete characterization of solutions in the normal case $P_i = N(0, \sum_i)$, $1 \leq i \leq n$.

Theorem 2.2 (Coupling of multivariate normal distributions)
Let $X_i \stackrel{d}{=} P_i = N(0, \sum_i), 1 \le i \le n, \sum_i > 0$ positive definite. Then it holds: $X = (X_1, \dots, X_n)$ is an optimal n-coupling for $(P_1, \dots, P_n)$ if and only if $\sum_0 = \mathrm{Cov}\ T$, $T = \sum_{j=1}^n X_j$, is a positive definite solution of

$$\sum_{i=1}^n \left(\Sigma_0^{1/2}\ \Sigma_i\ \Sigma_0^{1/2} \right)^{1/2} = \Sigma_0 \tag{2.10}$$

Moreover in this case with $S_i = \sum_i^{-1/2} \left(\sum_i^{1/2}\ \sum_0\ \sum_i^{1/2} \right)^{1/2} \sum_i^{-1/2}$ one obtains that

$$X_i = S_i\ S_1^{-1}\ X_1 \text{a.s.}, 1 \le i \le n. \tag{2.11}$$

Proof: Let $X = (X_1, \dots, X_n)$ be an optimal n-coupling; then we may assume w.l.g. that (X_i) are jointly normal distributed. Otherwise replace X by a n-tuple with joint normal distribution and identical covariance matrix. This implies that also $T_n = \sum_{j<n} X_j$ and $T = \sum_{j=1}^n X_j$ are normal. By Proposition 2.1, T_n and X_n are optimally coupled i.e. (X_n, T_n) is an optimal pair for $N(0, \sum_n)$ and $Q := N(0, \sum_{T_n})$, where $\sum_{T_n} = \mathrm{Cov}(T_n)$. Note that it is not obvious that $\sum_{T_n}$ is positive definite. By Gelbrich (1990) (see also Olkin and Pukelsheim (1982)) an optimal coupling between $N(0, \sum_n)$ and Q is given by the pair $(X_n, A\ X_n)$ where $A = \sum_3^{-1/2} \left(\sum_3^{1/2}\ \sum_{T_n}\ \sum_3^{1/2} \right)^{1/2} \sum_3^{-1/2}$. Positive definiteness of $\sum_n$ implies uniqueness of optimal pairs then (see Cuesta-Albertos, Matran and Tuero-Diaz (1996) and Gangbo and McCann (1996)) and, therefore, $T_n = A\ X_n$ a.s. This implies that $T = \sum_{j=1}^n X_j = (A + I) X_n$ a.s. Since A is positive semidefinite and $\langle x, (A+I)x \rangle = \langle x, Ax \rangle + \langle x, x \rangle \ \ge\ \langle x, x \rangle\ > 0$ for $x \ne 0$, $A + I$ is positive definite and, therefore, $\sum_0 = \mathrm{Cov}(T_n + X_n) = (A + I)\ \sum_n\ (A+I)^T > 0$.

Since $N(0, \sum_0)$ and $N(0, \sum_i)$ are optimally coupled by the mappings S_i (cp. (1.6)) i.e. $(T,\ S_i\ T)$ is an optimal pair for $(N(0, \sum_0),\ N(0, \sum_i))$, and since optimal coupling to the sum is a necessary condition by Proposition 2.1 we obtain from the same uniqueness result, that $X_i = S_i\ T$ a.s.

This implies that

$$T = \sum_{j=1}^n\ X_j = \left(\sum_{j=1}^n S_j \right) T \qquad \text{i.e.} \qquad \sum_{j=1}^n S_j = I.$$

By some simple algebra this shows that $\sum_0$ is a solution of equation (2.10).

For the converse direction of Theorem 2.2 the proof of Knott and Smith (1994) for the case $n = 3$ can easily be extended to general n. □

Remark: Theorem 2.2 in particular implies existence of a positive solution $\sum_0$ of (1.5) resp. (2.10). In order to find a positive definite solution $\sum_0$ of (2.10)

$$\sum_{i=1}^n (K_0\ K_i^2\ K_0)^{1/2} = K_0^2, \quad K_i := \Sigma_i^{1/2} \tag{2.12}$$

Knott and Smith (1994) suggest for ($n = 3$) to use the iterative procedure

$$K_0^{(k+1)} = \left(\sum_{i=1}^{n} \left(K_0^{(k)}\ K_i^2\ K_0^{(k)} \right)^{1/2} \right)^{1/2}. \tag{2.13}$$

It turns out by extensive simulations (for $n = 3$) with random initial matrices that the iteration converges in dimension $\mathrm{d} = 2$ (typically one needs about 100 iteration steps for exactness up to 8 digits). But for dimension $\mathrm{d} = 3$ only for favourable initial matrices convergence is observed. □

Without some technical assumption optimal coupling of random vectors (X_i) to the sum $T = \sum_{j=1}^n X_j$ is not sufficient to optimality for the n-coupling problem. Let e.g. X_j, $1 \le j \le n$, be random vectors with $\sum_{j=1}^n X_j = 0$. Constructions of (X_j) with this property exist for several nontrivial distributions (P_i). Obviously, any X_i is optimally coupled to the (trivial) sum T but also $X = (X_j)$ is not optimal for the n-coupling (P_j).

Nevertheless the following theorem justifies the coupling to the sum idea of Knott and Smith (1994) under the assumption that the distribution of T is Lebesgue continuous.

Theorem 2.3 (Coupling to the sum principle)
Let P_i be distributions on $\mathbb{R}^{\mathrm{d}}$, $1 \le i \le n$ with finite second moments and let $X_i \stackrel{d}{=} P_i$, $1 \le i \le n$, be such that X_i are optimally coupled to the sum $T = \sum_{j=1}^n X_j$. If P^T is Lebesgue-continuous, then $X = (X_1, \dots, X_n)$ is an optimal n-coupling of $(P_1, \dots, P_n)$.

Proof: By the characterization of optimal couplings in (2.1) there exist convex functions f_i such that $X_i \in \partial f_i(T)$ a.s.

The convex functions f_i are locally Lipschitz continuous and, therefore, by Rademachers theorem λ^{d} a.s. differentiable. Since $P^T \ll \lambda^{\mathrm{d}}$ this implies that

$$X_i = \bigtriangledown f_i(T) =: \Phi_i(T) \quad \text{a.s.} \tag{2.14}$$

Since $T = \sum_{i=1}^n X_i$, we obtain $\sum_{i=1}^n \bigtriangledown f_i = I$ or, equivalently, $\sum_{i=1}^n f_i(t) = \frac{1}{2}\, ||t||^2$.

Furthermore, by definition of convex conjugate functions f_i^* for $x_i \in \mathbb{R}^{\mathrm{d}}$, $t = \sum_{i=1}^n x_i$ holds

$$\langle x_i, t\rangle \ \le\ f_i(t) + f_i^*(x_i) \tag{2.15}$$

Therefore,

$$\begin{aligned} ||t||^2 = \langle t, t\rangle &= \left\langle \sum x_i, t \right\rangle \\ &\le \sum_i f_i(t) + \sum_i f_i^*(x_i) \\ &= \frac{1}{2}||t||^2 + \sum_i f_i^*(x_i), \ \text{i.e.} \\ \frac{1}{2}||t||^2 &\le \sum_i f_i^*(x_i). \end{aligned} \tag{2.16}$$

The condition $X_i \in \partial f_i(T)$ a.s. implies that

$$\langle X_i, T\rangle = f_i(T) + f_i^*(X_i) \text{ a.s.} \tag{2.17}$$

i.e. equality holds in (2.15) on the support of the distribution of (X_i, T) and, therefore, $\frac{1}{2}\,||T||^2 = \sum_i f_i^*(X_i)$ a.s.

This implies that $X = (X_i)$ is an optimal n-coupling, since for any $Y_i \stackrel{d}{=} P_i$ holds

$$E\frac{1}{2}\,||\sum_{i=1}^n Y_i||^2 \le E\,\sum_i f_i^*(Y_i) = E\,\sum_i f_i^*(X_i) = \frac{1}{2}E||T||^2.$$

This implies optimality of $(X_1, \dots, X_n)$. □

In a recent paper Gangbo and Swiech (1997) have proven existence of Monge solutions for the n-coupling problem, i.e. of solutions of the form $(X_1, \Phi_2(X_1), \dots, \Phi_n(X_1))$ if all P_i are Lebesgue-continuous. We obtain a simple proof of this result based on the necessity of the coupling to the sum principle in Proposition 2.1.

Theorem 2.4 (Monge solutions)
Let $P_i \ll \lambda^{\mathrm{d}}, 1 \le i \le n$, with finite second moments. Then there exists a Monge solution of the form $(X_1, \Phi_2(X_1), \dots, \Phi_n(X_1)), X_1 \stackrel{d}{=} P_1$, of the n-coupling problem.

Proof: Let $X = (X_1, \dots, X_n)$ be a solution of the n-coupling problem $(P_i), X_i \stackrel{d}{=} P_i$. Then by Proposition 2.1 X_i are optimally coupled to the sums $T_i = \sum_{j\neq i} X_j$ and so by 2.1 $T_i \in \partial f_i(X_i)$ a.s. for some convex functions f_i. By Rademachers theorem f_i are λ^{d} a.s. differentiable so that by assumption $P_i \ll \lambda^{\mathrm{d}}$, $\partial f_i(X_i) = \{\bigtriangledown f_i(X_i)\}$ a.s., i.e. $T_i = \bigtriangledown f_i(X_i)$ a.s.
Therefore, defining $\overline{f_i}(x) := f_i(x) + \frac{1}{2}||x||^2$ we obtain $\overline{f_i}$ is strictly convex, $\bigtriangledown\overline{f_i}$ exists, λ^{d} a.s. and is invertible and $T = T_i + X_i = \bigtriangledown\overline{f_i}(X_i)$ a.s. This implies that $X_i = (\bigtriangledown\overline{f_i})^{-1}\,(T)$ a.s. for all i and, therefore,

$$X_i = \left(\bigtriangledown\overline{f_i}\right)^{-1}\,(\bigtriangledown f_1(X_1)) = \Phi_i(X_i) \text{ a.s.} \tag{2.18}$$

which is the wished Monge solution. □

Remark: Note that the proof of Theorem 2.4 is not constructive. If we take convex functions f_i and define X_i by (2.18) we do generally not obtain optimal n-couplings. An improved constructive version of Monge solutions is given in the next section. □

3 Characterization of optimal n-couplings and examples

Based on (2.6), (2.7) the following reduction result for the three coupling problem has been proved in Rüschendorf and Uckelmann (1997) for the case $n = 3$.

Theorem 3.1 *Let $X \stackrel{d}{=} P_1, Y \stackrel{d}{=} P_2, Z \stackrel{d}{=} P_3$ and let P_i have finite second moments. Then (X, Y, Z) is a solution of the three coupling problem (1.1) if and only if there exists a convex lsc function f and an F-convex function g, where $F(y, z) := f^*(y + z) + \langle y, z \rangle$, such that*

$$\begin{aligned} &(1) \qquad Y + Z \ \in \partial f(X) \quad a.s. \\ &(2) \qquad\quad\ \ Z \ \in \partial_F g(Y) \quad a.s. \end{aligned} \tag{3.1}$$

Remarks:

a) From the characterization in Theorem 3.1 one obtains in the case $n = 3$ a more concrete coupling to the sum result in comparison to Theorem 2.4. If X, Y, Z is an optimal three coupling for P_1, P_2, P_3 then

$$Y + Z \in \partial f(X),\ \ X + Z \in \partial g(Y) \text{ and } X + Y \in \partial g^F(Z) \text{ a.s.} \tag{3.2}$$

with f, g as in Theorem 3.1, g^F the $F-$conjugate of g. By (2.1) this implies that any of X, Y, Z is optimally coupled to the sum of the two others. Further with $f_1(x) = f(x) + \frac{1}{2}||x||^2$, $f_2(x) = g(x) + \frac{1}{2}||x||^2$, $f_3(x) = g^F(x) + \frac{1}{2}||x||^2$ holds:

$$T = X + Y + Z \in \partial f_1(X) \cap \partial f_2(Y) \cap \partial f_3(Z) \text{ a.s.} \tag{3.3}$$

i.e. X, Y, Z are optimally coupled to the sum. Note that in the form (3.2) these conditions are by Theorem 3.1 also sufficient for optimal three coupling.

b) If $\Phi : \mathrm{IR}^d \to \mathrm{IR}^d$ is a cyclically monotone function and $X = \Phi(Y + Z)$, then X is optimally coupled to $Y + Z$. If we solve the two-coupling problem

$$E \langle \Phi(Y + Z), Y + Z \rangle + \langle Y, Z \rangle \ = \ \max! \tag{3.4}$$

over $Y \stackrel{d}{=} P_2, Z \stackrel{d}{=} P_3$, then (X, Y, Z) is an optimal solution for the three-coupling problem (P_1, P_2, P_3) where P_1 is the distribution of X. This simple method allows to construct some explicit examples.

If e.g. $\Phi = A$ is a linear, positive semidefinite, symmetric function then (3.4) amounts to a linear problem which can be solved explicitely (as in the normal case). Theorem 3.1 essentially implies that up to some technicals any three-coupling problem can be solved in this simple way. The problem is however to find to given (P_i) the correct Φ (or f).

□

The following result based on Theorem 3.1 is a constructive version on the existence of Monge solutions in Theorem 2.4. Also the continuity assumption on the P_i in Theorem 2.4 is weakened.

Theorem 3.2
Let P_i be probability measures on $\mathbb{R}^d$, $1 \le i \le 3$, and let $P_1, P_2 \ll \lambda^d$.

a) There exists a Monge solution $(X, \Psi_1(X), \Psi_2(X))$ of the three-coupling problem.

b) There exist convex, a.s. differentiable functions f, g such that with $\Phi_1(y) = (\nabla h)^{-1}(\nabla g(y) + y), h(t) := f^*(t) + \frac{1}{2}||t||^2$, the functions Ψ_1, Ψ_2

$$\begin{aligned} \Psi_1(X) &= \Phi_1^{-1}(\nabla f(X)) \\ \Psi_2(X) &= \nabla f(X) - \Psi_1(X) \end{aligned} \tag{3.5}$$

define Monge solutions.

Proof: Consider the functions f, g, F as stated in Theorem 3.1. Then f, g are λ^d a.s. differentiable, $F(\cdot, z)$ is convex and λ^d a.s. differentiable for all $z \in \mathbb{R}^d$. For $z \in \partial_F\ g(y)$ the function $h(t) := g(t) - F(t, z)$ has a local minimum in $t = y$ and, therefore, z solves the equation

$$0 = \nabla u(y) = \nabla g(y) - \nabla_1 F(y, z) \tag{3.6}$$

if g is differentiable in y.

Further, ∇h exists λ^d a.s. and for λ^d a.a. y holds

$$\begin{aligned} \nabla g(y) &= \nabla_1 F(y, z) &&= \nabla(f^*(y+z) + \langle y, z\rangle) \\ &= \nabla f^*(y+z) + z &&= \nabla h(y+z) - y. \end{aligned}$$

h is strictly convex and, therefore, ∇h is strictly monotone and invertible. This implies

$$z = (\nabla h)^{-1}(\nabla g(y) + y) - y =: \Phi_2(y). \tag{3.7}$$

Since by Theorem 3.1, $Z \in \partial_F\ g(Y)$ a.s. and $P^Y \ll \lambda^d$ we obtain $Z = \Phi_2(Y)$ a.s. Furthermore, $P^X = P_1 \ll \lambda^d$ implies that $\partial f(X) = \{\nabla f(X)\}$ a.s. and by Theorem 3.1 $Y + Z = \nabla f(X)$ a.s.

This implies that

$$\begin{aligned} Z &= \Phi_2(Y) = (\nabla h)^{-1}(\nabla g(Y) + Y) - Y \\ &= (\nabla h)^{-1} \nabla \overline{g}(Y) - Y \\ &=: \Phi_1(Y) - Y \end{aligned}$$

where $\overline{g}(y) := g(y) + \frac{1}{2}||y||^2$. $\overline{g}$ is strictly convex and, therefore, Φ_1 is invertible. This finally implies that

$$\begin{aligned} \nabla f(X) &= Y + Z = \Phi_1(Y) \text{ i.e.} \\ Y &= \Phi_1^{-1}\ (\nabla f(X)) = \Psi_1(X). \end{aligned}$$

Then we also obtain a representation of Z,

$$\begin{aligned} Z &= \Phi_1(Y) - Y = \bigtriangledown f(X) - \Phi_1^{-1}\left(\bigtriangledown f(X)\right) \\ &=: \Psi_2(X). \end{aligned}$$

□

Remark: Note that (3.5) is also a sufficient condition for optimality if g is even F-convex. If f is strict convex one obtains an alternative representation for an optimal coupling by $(X,\ \Phi_1(X),\ \Phi_2(X))$ with

$$\begin{aligned} \Phi_1(X) &= (\bigtriangledown f)^{-1}((\bigtriangledown h)^{-1}(\bigtriangledown g(X) + X)) \\ \Phi_2(X) &= (\bigtriangledown h)^{-1}(\bigtriangledown g(X) + X) - X, \end{aligned} \tag{3.8}$$

where $h(t) = f^*(t) + \frac{1}{2}||t||^2$.

This representation is again sufficient for optimality if g is F-convex (cp. Theorem 3.1). We will use this sufficient condition in the following to construct some examples. □

An extension of the characterization of optimal solutions in Theorem 3.1 to n-coupling problems is given in the following theorem.

Theorem 3.3 (Characterization of optimal n-couplings)
Let P_i be probability measures on $\mathbb{R}^d$, $1 \le i \le n$ with finite second moments. Then $X = (X_1, \dots, X_n)$ is n-optimal if and only if functions $f_1, \dots, f_{n-1}$ exist such that f_1 is convex, lsc, f_k is F_{k-1}-convex where

$$F_0\,(s,t) = \langle s,t \rangle\,, \quad F_k(s,t) = f_k^{F_{k-1}}\,(s+t)\, + \langle s,t \rangle\,, \tag{3.9}$$

$f^{F_{k-1}}$ is the F_{k-1} conjugate of f_k, $\quad 1 \le k \le n-1$, and

$$\sum_{i=2}^n X_i \in \partial\, f_1(X_1), \qquad \sum_{i=k+1}^n X_i \in \partial_{F_{k-1}}\, f_k(X_k) \tag{3.10}$$

$2 \le k \le n-1$, a.s.

Proof: The proof is based in both directions on the duality theorem (see Kellerer (1984), Rachev (1991), Rüschendorf (1981, 1991)). With $M(P_1, \dots, P_n)$ the set of measures with marginals $P_1, \dots, P_n$ holds

$$\begin{aligned} &\sup\left\{\int c\,(x_1, \dots, x_n)\,\mathrm{d}\mu; \quad \mu \in M(P_1, \dots, P_n)\right\} \\ &= \inf\left\{\sum_{i=1}^n \int f_i\,\mathrm{d}P_i; \quad f_i \in \mathcal{L}^1(P_i), \sum_{i=1}^n f_i(x_i) \ge c\,(x_1, \dots, x_n)\right\}. \end{aligned} \tag{3.11}$$

This duality theorem is applied to $c\,(x) = \sum_{i<j} \langle x_i, x_j \rangle$. Given condition (3.10) define

$$f_n(t) = f_{n-1}^{F_n-2}(t) = \sup_s\{F_{n-2}(s,t) - f_{n-1}(s)\};$$

then $(f_1, \dots, f_n)$ is admissible for the dual problem in (3.11) as can be seen by recursive insertion of the definition of f_i (for details see Uckelmann (1998)).

Since $\sum_{k=2}^n X_k \in \partial\, f_1(X_1)$, it holds a.s. that

$$f_1(X_1) + f_1^*\left(\sum_{k=2}^n X_k\right) = \left\langle X_1, \sum_{k=2}^n X_k \right\rangle. \tag{3.12}$$

Further from $\sum_{k=3}^n X_k \in \partial_{F_1}\, f_2(X_2)$ one obtains

$$\begin{aligned} f_2(X_2) + f_2^{F_1}(\sum_{k=3}^n X_k) &= F_1(X_2, \sum_{k=3}^n X_k) \\ &= f_1^*(\sum_{k=2}^n X_k) + \left\langle X_2, \sum_{k=3}^n X_k \right\rangle \text{ a.s.} \end{aligned}$$

and generally

$$f_m(X_m) + f_m^{F_{m-1}}\left(\sum_{k=m+1}^n X_k\right) = f_{m-1}^{F_{m-2}}\left(\sum_{k=m}^n X_k\right) + \left\langle X_m, \sum_{k=m+1}^n X_k \right\rangle \text{ a.s.} \tag{3.13}$$

Finally, $X_n \in \partial_{F_{n-2}}\, f_{n-1}\,(X_{n-1})$ a.s. implies
$f_{n-1}(X_{n-1}) + f_{n-1}^{F_{n-2}}(X_n) = F_{n-2}\,(X_{n-1}, X_n)$ a.s., equivalently, by definition of f_n

$$f_{n-1}(X_{n-1}) + f_n(X_n) = f_{n-2}^{F_{n-3}}\,(X_{n-1} + X_n) + \langle X_{n-1}, X_n \rangle. \tag{3.14}$$

Summing over these equations yields

$$\sum_{k=1}^n f_k(X_k) = c\,(X_1, \dots, X_n) \text{ a.s.} \tag{3.15}$$

This implies by the duality theorem (3.11) optimality of $X = (X_1, \dots, X_n)$.

For the converse direction let $\mu \in M(P_1, \dots, P_n)$ be an optimal measure and $(f_1, \dots, f_n)$ be a solution of the dual problem (3.11) (see Kellerer (1984), Rüschendorf (1981)). Define $f = f_1^{**}$ i.e. $f(x) = \sup_y\{\langle x, y\rangle - f_1^*(y)\}$ then $f \le f_1$ and $f^* = f_1^*$. With

$$\begin{aligned} G(x_2, \dots, x_n) &= \sup_{x_1}\{c\,(x_1, \dots, x_n) - f(x_1)\} \\ &= \sup_{x_1}\{\left\langle x_1, \sum_{i=2}^k x_i \right\rangle - f(x_1)\} + \sum_{i<j} \langle x_i, x_j\rangle \\ &= f^*\left(\sum_{i=2}^n x_i\right) + \sum_{i<j}\langle x_i, x_j\rangle \end{aligned}$$

holds

$$\sum_{i=1}^n f_i(x_i) \ge f(x_1) + \sum_{i=2}^n f_i(x_i) \ge f(x_1) + G(x_2, \dots, x_n) \ge c(x_1, \dots, x_n).$$

As $(f_1, \dots, f_n)$ is a solution of the dual problem, this inequality implies

$$\int (f(x_1) + G(x_2, \dots, x_n) - c\,(x_1, \dots, x_n))\,\mathrm{d}\mu = 0$$

and, therefore, $f(x_1) \ + \ G(x_2, \dots, x_n) = c\,(x_1, \dots, x_n)\,\mu$ a.s.

This implies that μ a.s.,

$$f(x_1) + f^*\left(\sum_{i=2}^n x_i\right) = \left\langle x_1, \sum_{i=2}^n x_i \right\rangle$$

and, therefore, $\sum_{i=2}^n x_i \in \partial f(x_1) \quad \mu$ a.s., the first condition of (3.10) holds.

Similarly, one constructs successively improvements of the admissible solution $f_1, \dots, f_n$ which result in the conditions of the theorem (for details see Uckelmann (1998)).

$$\sum_{i=k}^n X_i \in \partial_{F_{k-2}} f_{k-1}(X_{k-1}) \text{ a.s.} \qquad \square$$

Remark: Note that the functions $f_k, f_k^{F_{k-1}}$ in Theorem 3.3 are convex and lsc. As consequence of Theorem 3.3 one obtains as in Theorem 3.2 an improved constructive version of the existence of Monge solutions. The proof can be given along the lines of the proof of Theorem 3.2. $\square$

For the application to concrete examples the following calculation of $F(t, z) = f^*(t + z) + \langle t, z\rangle$ is of interest. We use the notation of Theorem 3.1.

Lemma 3.4 *If* $f : \mathbb{R}^d \to \mathbb{R}$ *is strictly convex, then*

$$F(y, z) = \left\langle y + z, (\nabla f)^{-1}(y + z)\right\rangle - f\left((\nabla f)^{-1}(y + z)\right) + \langle y, z\rangle\,. \tag{3.16}$$

Proof: Strict convexity of f implies existence of $(\nabla f)^{-1}$. Define $\eta_s(t) := \langle s, t\rangle - f(t)$, then $\nabla\eta_s(t) = s - \nabla f(t) = 0$ if and only if $t_0 = (\nabla f)^{-1}(s)$. Since $D^2\eta_s(t) = -D^2 f(t) < 0$ for all t, η_s is strictly concave and t_0 is a global supremum of η_s. Therefore, $f^*(s) = \sup_t\{\langle s, t\rangle - f(t)\} = \langle s, (\nabla f)^{-1}(s)\rangle - f((\nabla f)^{-1}(s))$ which implies (3.16). $\square$

Lemma 3.5 *For* $\alpha > 0$ *and* $p \geq 2$ *consider* $f(t) := \alpha||t||^p$, *then* $F(y, z) = \frac{p-1}{p}\ (\alpha p)^{\frac{1}{1-p}}\ ||y + z||^{\frac{p}{p-1}} + \langle y, z\rangle$.

Proof: Note that

$$\begin{aligned} \nabla f(t) &= \alpha p||t||^{p-2}t \\ &\text{and} \\ (\nabla f)^{-1}(s) &= (\alpha p)^{1/1-p}||s||^{-\frac{p-2}{p-1}}. \end{aligned} \tag{3.17}$$

This implies

$\left\langle s, (\bigtriangledown f)^{-1}(s)\right\rangle = (\alpha p)^{\frac{1}{1-p}}||s||^{\frac{p}{p-1}}$ and $f\left((\bigtriangledown f)^{-1}(s)\right) = \alpha(\alpha p)^{\frac{p}{1-p}}||s||^{\frac{p}{p-1}}$.

This applied to (3.16) yields (3.17). □

In the case $p = 2$ we obtain the following consequence for F-subdifferentials.

Lemma 3.6 *Let $f(t) := \alpha||t||^2$ and let $g \in C^2(\mathrm{I\!R}^{\mathrm{d}})$ satisfy $D^2 g(t) - \frac{1}{2\alpha} I_{\mathrm{d}\times\mathrm{d}} > 0$.*
Then g is F-convex and

$$z = \frac{2\alpha}{1+2\alpha} \bigtriangledown g(y) - \frac{1}{1+2\alpha} y \in \partial_F\, g(y). \tag{3.18}$$

For $g(t) = G(t) + \beta||t||^2, G \in C^2(\mathrm{I\!R}^{\mathrm{d}})$ convex, $\alpha\beta > \frac{1}{4}$ the above assumption is fulfilled.

Proof: By Lemma 3.5

$$F(t,z) = \frac{1}{4\alpha}\,||t+z||^2 + \langle t, z\rangle\,.$$

Define $\Psi_z(t) := g(t) - \frac{1}{4\alpha}\,||t+z||^2 - \langle t, z\rangle$ then $\bigtriangledown\Psi_z(t) = \bigtriangledown g(t) - \frac{1+2\alpha}{2\alpha} z - \frac{1}{2\alpha} t$ and $D^2\Psi_z(t) = D^2 g(t) - \frac{1}{2\alpha} I_{\mathrm{d}\times\mathrm{d}}$. For $z = \frac{2\alpha}{1+2\alpha} \bigtriangledown g(t) - \frac{1}{1+2\alpha} t$ holds $\bigtriangledown\Psi_z(t) = 0$ and $D^2\Psi_z(t) > 0$ implies that $t = t_z$ is a global minimum of Ψ_z.

To prove F-convexity of g consider

$$\begin{aligned} \mathcal{E} &= \{(z, g(t_z) - F(t_z, z));\quad z \in \mathrm{I\!R}^{\mathrm{d}}\}. \\ \text{Since } g(t) &\geq F(t,z) + g(t_z) - F(t_z, z) \\ \text{and } g(t_z) &= F(t_z, z) + g(t_z) - F(t_z, z) \text{ it holds that} \\ g(t) &= \sup\nolimits_{(z,a)\in\mathcal{E}}\, \{F(t,z) + a\}. \end{aligned}$$

Therefore, g is F-convex and $z \in \partial_F\, g(y)$.

If $g(t) = G(t) + \beta||t||^2$, then

$$\begin{aligned} D^2 g(t) - \frac{1}{2\alpha} I_{\mathrm{d}\times\mathrm{d}} &= D^2 G(t) + (2\beta - \frac{1}{2\alpha}) I_{\mathrm{d}\times\mathrm{d}} \\ &= D^2 G(t) + \frac{4\alpha\beta - 1}{2\alpha} I_{\mathrm{d}\times\mathrm{d}}. \end{aligned}$$

From our assumptions the condition of the first part of Lemma 3.5 is fulfilled. □

Example 3.7 Let $\alpha\beta \geq \frac{1}{4}, \alpha, \beta > 0$ and let $G \in C^2(\mathrm{I\!R}^{\mathrm{d}})$ be convex and define $f(t) = \alpha||t||^2, g(t) = G(t) + \beta||t||^2$. Define $h(t) = f^*(t) + \frac{1}{2}||t||^2 = \frac{1}{4\alpha}||t||^2 + \frac{1}{2}||t||^2$. Then $\bigtriangledown h(t) = \frac{1+2\alpha}{2\alpha} t, \quad (\bigtriangledown h)^{-1}(s) = \frac{2\alpha}{1+2\alpha} s$ and $\bigtriangledown g(t) + t = \bigtriangledown G(t) + (2\beta + 1)t$. By

Lemma 3.6 g is F-convex and by (3.8) an optimal Monge pair is given by (Φ_1, Φ_2) with

$$\begin{array}{rclcl} \Phi_1(t) & = & (\bigtriangledown f)^{-1}((\bigtriangledown h)^{-1}(\bigtriangledown g(t)+t)) & = & \frac{1}{1+2\alpha} \bigtriangledown G(t) + \frac{1+2\beta}{1+2\alpha} t \\ \Phi_2(t) & = & (\bigtriangledown h)^{-1}(\bigtriangledown g(t)+t) - t & = & \frac{2\alpha}{1+2\alpha} \bigtriangledown G(t) + \frac{4\alpha\beta-1}{1+2\alpha} t. \end{array} \tag{3.19}$$

Example 3.8 Let $A, B > 0$ be positive definite matrices and consider $f(x) = \frac{1}{2}\langle x, Ax\rangle, g(y) = \frac{1}{2}\langle y, By\rangle$. Then g is F-convex and an optimal Monge tuple (Φ_1,Φ_2) is given by

$$\begin{array}{rcl} \Phi_1(x) & = & A^{-1}(A^{-1}+I_{\mathrm{d}\times\mathrm{d}})^{-1}(B+I_{\mathrm{d}\times\mathrm{d}})x \\ \Phi_2(x) & = & ((A^{-1}+I_{\mathrm{d}\times\mathrm{d}})^{-1}(B+I_{\mathrm{d}\times\mathrm{d}}) - I_{\mathrm{d}\times\mathrm{d}})x. \end{array} \tag{3.20}$$

For the proof note that $f^*(x) = \langle A^{-1}x, x\rangle$ and $F(y,z) = \langle A^{-1}(y+z), y+z\rangle + \langle y, z\rangle$. Consider $\Psi_z(t) := g(t) - F(t,z)$, then $\bigtriangledown\Psi_z(t) = Bt - A^{-1}(t+z) - z = 0$ if and only if $z = (A^{-1}+I_{\mathrm{d}\times\mathrm{d}})^{-1}(B - I_{\mathrm{d}\times\mathrm{d}})t$. By assumption $D^2\Psi_z(t) = B - A^{-1} > 0$ i.e. Ψ_z is convex. Then by Lemma 3.5 g is F-convex and

$$(A^{-1}+I_{\mathrm{d}\times\mathrm{d}})^{-1}(B - I_{\mathrm{d}\times\mathrm{d}})t \in \partial_F\, g(t).$$

Next apply formula (3.8) to obtain (3.20).

Remark: It is easy to rederive from our general characterization results that in dimension $\mathrm{d} = 1$ optimal couplings are given by $X_i = F_i^{-1}(U)$, where F_i are the distribution functions of P_i and U is uniformly distributed on $[0,1]$. The normal examples for optimal three-couplings in Olkin and Rachev (1993) resp. Knott and Smith (1994) were so far the only cases treated in the literature. Based on our characterization results and the description of F-convex functions we have added some more examples. One can construct several further examples using the two-coupling results in Rüschendorf (1995) where general coupling functionals are considered. □

References

[1] Cuesta-Albertos, J.A.; Matran, C.; Tuero-Diaz, A. (1996): Properties of the optimal maps for the L^2-Monge-Kantorovich transportation problem, *Preprint.*

[2] Gangbo, W.; McCann, R.J. (1996): The geometry of optimal transportation, *Acta Math.* **177**, 113–161.

[3] Gangbo, W.; Swiech, A. (1996): Optimal maps for the multidimensional Monge-Kantorovich Problem, *Preprint.*

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

[5] Kellerer, H.G. (1984): Duality theorems for marginal problems, *Z. für Wahrscheinlichkeitstheorie verw. Geb.* **67**, 399–432.

[6] Knott, M.; Smith, C.S. (1994): On a generalization of cyclic-monotonicity and distances among random vectors, *Linear Algebra and its Appl.* **199**, 363–371.

[7] Olkin, I.; Pukelsheim, F. (1982): The distance between two random vectors with given dispersion matrices, *Linear Algebra and its Appl.* **48**, 257–263.

[8] Olkin, I.; Rachev, S.T. (1993): Maximum submatrix traces for positive definite matrices, *SIAM J. Matrix Anal. Appl.* **14**, 390–397.

[9] Rachev, S.T. (1991): Probability Metrics and the Stability of Stochastic Models, Wiley.

[10] Rachev, S.T.; Rüschendorf, L. (1998): Mass Transportation Problems, Vol. I: Theory, Springer.

[11] Rüschendorf, L. (1981): Sharpness of Fréchet-bounds, *Z. für Wahrscheinlichkeitstheorie verw. Geb.* **57**, 293–302.

[12] 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.

[13] Rüschendorf, L. (1995): Optimal solution of multivariate coupling problems, *Appl. Mathematicae* **23**, 325–338.

[14] Rüschendorf, L.; Rachev, S.T. (1990): A characterization of random variables with minimum L^2-distance, *J. Mult. Anal.* **32**, 48–54.

[15] Rüschendorf, L.; Uckelmann, L. (1997): On optimal multivariate couplings, In: Distributions with given marginals and moment problems, Beneš, V; Štěpán, J. (Eds.), Kluwer Acad. Publ., 613–273.

[16] Uckelmann, L. (1998): Über das Monge-Kantorovich Problem und dessen Verallgemeinerungen, *Dissertation, Universität Freiburg.*