Numerical and analytical results for the transportation problem of Monge-Kantorovich

Ludger Rüschendorf* and Ludger Uckelmann

Institut für Mathematische Stochastik, University of Freiburg, Eckerstr. 1, 79104 Freiburg, Germany

Abstract

The Monge-Kantorovich transportation problem has a long and interesting history and has found a great variety of applications (see Rachev and Rüschendorf (1998)). Some interesting characterizations of optimal solutions to the transportation problem (resp. coupling problems) have been found recently. For the squared distance and discrete distributions they relate optimal solutions to generalized Voronoi diagrams. Numerically we investigate the dependence of optimal couplings on variations of the coupling function. Numerical results confirm also a conjecture on optimal couplings in the one-dimensional case for nonconvex coupling functions. A proof of this conjecture is given under some technical conditions.

1 Introduction

In the transportation problem of Monge-Kantorovich there are given two mass distributions P, Q on $(E, \mathcal{A})$ and a product measurable cost function $c : E \times E \to \mathbb{R}_+$. The aim is to determine an optimal transportation plan μ^* in $M(P, Q)$, the class of probability measures on $(E \times E, \mathcal{A} \otimes \mathcal{A})$ with marginals P, Q, such that the transportation cost is minimal. For technical reasons we consider the equivalent sup problem; for the inf problem just switch from 'c' to '$-c$'.

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

Note that any $\mu \in M(P, Q)$ has the interpretation as a transportation plan which transports the mass P to the mass Q. In terms of random variables an optimal

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

transportation plan μ^* corresponds to an optimal c-coupling (X, Y) of the distributions P, Q. If $Y = \Phi(X)$ is a function of X then we call Φ a Monge-function for the coupling problem.

The Monge-Kantorovich problem has a long history dating back to Monge (1781) and Kantorovich (1942). A detailed exposition of this problem and of its many applications is given in Rachev and Rüschendorf (1998). A basic result in this context is the duality theorem

$$\begin{aligned} &C(P,Q) \\ &= \inf\left\{\int f dP + \int g dQ; f \in L^1(P), g \in L^1(Q), f(x) + g(y) \geq c(x,y) \quad \forall x, y\right\}. \end{aligned} \tag{1.2}$$

This duality result holds true for all bounded (or uniformly integrable) measurable functions c if $(E, \mathcal{A}, P)$ is a perfect measure space. It has been found recently that in a somewhat stronger form it holds true in general only for perfect measure spaces (see Ramachandran and Rüschendorf (1998)).

A real function f on E is called c-convex if

$$f(x) = \sup_{(y,a) \in A \subset E \times \mathbb{R}} \Psi_{y,a}(x) \tag{1.3}$$

where A is any subset of $E \times \mathbb{R}$

$$\Psi_{y,a}(x) = c(x,y) + a \tag{1.4}$$

is determined by the shift a and the translation y. The c-subdifferential

$$\partial_c f(x) = \{y : f(z) - f(x) \geq c(z,y) - c(x,y) \quad \forall z \in \operatorname{dom} f\} \tag{1.5}$$

then is characterized in the following way (see Figure 1):
$y \in \partial_c f(x)$ if and only if there exists a shift a such that

$$\Psi_{y,a}(x) = f(x) \quad \text{and} \quad \Psi_{y,a}(z) \leq f(z), \quad \forall z \in \operatorname{dom} f. \tag{1.6}$$

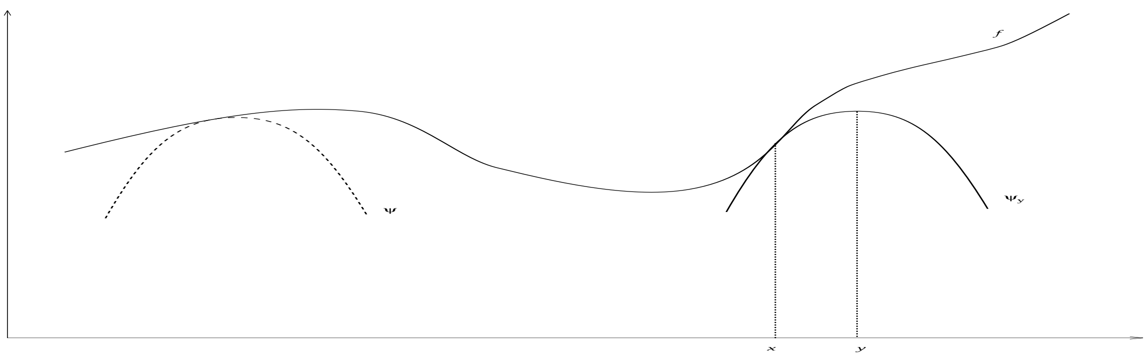


Figure 1: c-convexity and c-subdifferentiability

A basic characterization of optimal transportation plans resp. c-optimal couplings has been given in Rüschendorf (1991): A pair (X, Y) with $X \sim P$, $Y \sim Q$ is a c-optimal coupling if and only if

$$Y \in \partial_c f(X) \text{ a.s.} \tag{1.7}$$

for some c-convex function f.

A condition equivalent to (1.7) characterizes the c-optimal couplings by c-cyclical monotonicity of the support (see Smith and Knott (1992)). For the quadratic cost function $c(x, y) = -\|x - y\|^2$ a function f is c-convex if and only if $\widehat{f}(x) = f(x) + \frac{1}{2}\|x\|^2$ is convex and lower semicontinuous and $y \in \partial_c f(x)$ if and only if $y \in \partial \widehat{f}(x)$. The optimal coupling of multivariate normal distributions has been found in Olkin and Pukelsheim (1982) by direct methods. For the general case and several examples see Rüschendorf and Rachev (1990).

For the optimal coupling of P with a discrete measure $Q = \sum_{i=1}^n \alpha_i \delta_{y_i}$ one can restrict to c-convex functions of the form

$$f(x) = \sup_{1 \le i \le n} (c(x, y_i) + a_i). \tag{1.8}$$

as the c-subgradients are given by the translations (see (1.6)). Then with
$A_i := \{x : f(x) = c(x, y_i) + a_i\}$

$$y_i \in \partial_c f(x) \text{ if and only if } \quad x \in A_i. \tag{1.9}$$

Therefore, the problem to determine optimal couplings is reduced to determine shifts a_i, such that the 'Voronoi type' partitioning set A_i has the correct mass $P(A_i) = \alpha_i, 1 \le i \le n$. For coupling to discrete distributions and related examples see Rüschendorf and Uckelmann (1997). A connection to generalized Voronoi diagrams (power diagrams) in the case of squared distance $\|x - y\|^2$ has been given in Aurenhammer, Hoffmann, and Aronov (1998) where for P discrete supported by m points an algorithm is introduced which uses time of order $O(n^2 m \log m + nm \log^2 m)$.

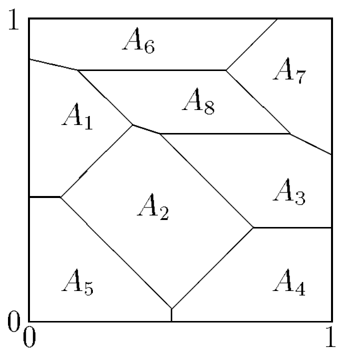


Figure 2: ℓ_2-partition of $[0, 1]^2$

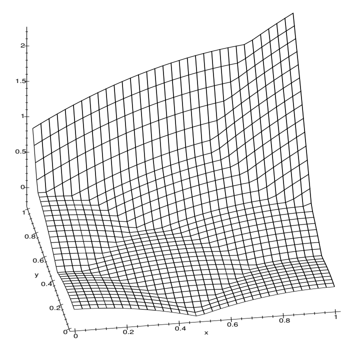


Figure 3: ℓ_2-convex function

In the following example based on the characterization in (1.9) (see Figures 2 and 3) the exact optimal solution is determined for the squared distance where $P = \mathcal{U}([0,1])^2$ is the uniform distribution on $[0,1]^2$, $n = 8$ with

$$y_i = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and masses $\alpha_i = 0.105, 0.2, 0.125, 0.125, 0.125, 0.125, 0.12, 0.1, 0.1$.

The next three dimensional example is for $P = \mathcal{U}([0,1]^3)$ and $n = 3$ with $y_i = \begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 0.2 \\ 0.2 \\ 0.8 \end{pmatrix}, \begin{pmatrix} 0.8 \\ 0.8 \\ 0.8 \end{pmatrix}$, (see Figure 4).

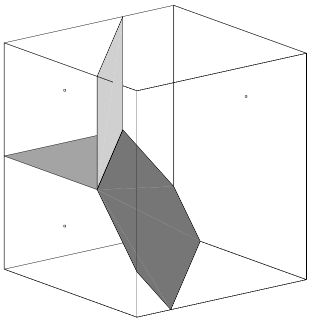


Figure 4:
ℓ_2-partition of unit cube

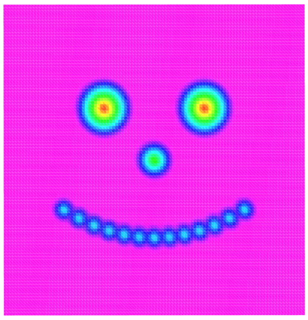


Figure 5:
optimal partition

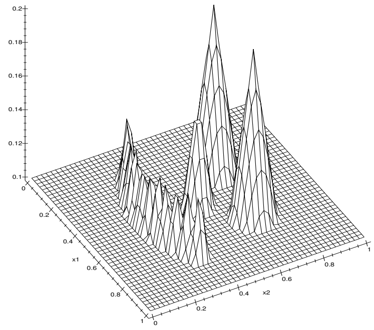


Figure 6:
c-convex function

The method also applies to nondifferentiable coupling functions. In the following example the coupling function c is nondifferentiable, $c(x,y) = -h(\|x-y\|)$, where $h(t) = \min(\beta, t)$, $P = \mathcal{U}([0,1]^2)$,

$$\begin{aligned} y_i &= \begin{pmatrix} 100 \\ 100 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}, \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \quad 1 \le i \le 4 \qquad \text{and} \\ y_k &= \begin{pmatrix} t_k \\ f(t_k) \end{pmatrix}, \quad t_k \in \left[\frac{1}{3}, \frac{2}{3}\right], \quad 4 \le k \le n, \quad f(t) = \left(t - \tfrac{1}{2}\right)^2 + \tfrac{1}{4}. \end{aligned}$$

So in this case the optimal coupling produces a 'smile effect' (see Figures 5 and 6).

2 Approximative solutions and linear programs

For the approximative optimal c-couplings of $P = \mathcal{U}([0,1]^2)$ and $Q = \sum_{k=1}^m \alpha_k \varepsilon_{a_k}$, $a_k \in \mathbb{R}^2$ one can use a discretized version $\widetilde{P}$ of P of the form

$$\widetilde{P} = \frac{1}{(n+1)^2} \sum_{i=0}^{n} \sum_{j=0}^{n} \delta_{x_{i,j}}, \quad x_{ij} = \left(\frac{i}{n}, \frac{j}{n}\right) \tag{2.1}$$

and then solve the linear program

$$\sum_{k=1}^{m} \sum_{i,j=0}^{n} c_{ijk} x_{ijk} = min! \tag{2.2}$$

where $c_{ijk} = c(x_{ij}, a_k)$ and $\sum_{i,j=0}^{n} x_{ijk} = (n+1)^2 \alpha_k,\ \ k = 1, \ldots, m$, $\sum_{k=1}^{m} x_{ijk} = 1$.

The following examples were calculated using the program Soplex by Wunderling (1996). They show the dependence of optimal couplings on a variation of the distance. The qualitative form of dependence of optimal couplings on the coupling functions seen in the numerical examples is expected from the nature of the ℓ_p-metrics which for increasing p weight heavily misfits in the couplings of a larger magnitude. The first example is based on the same data as in Figure 2 for $m = 8$. It also shows that the numerical results are in good coincidence with the theoretical optimal coupling in Figure 2.

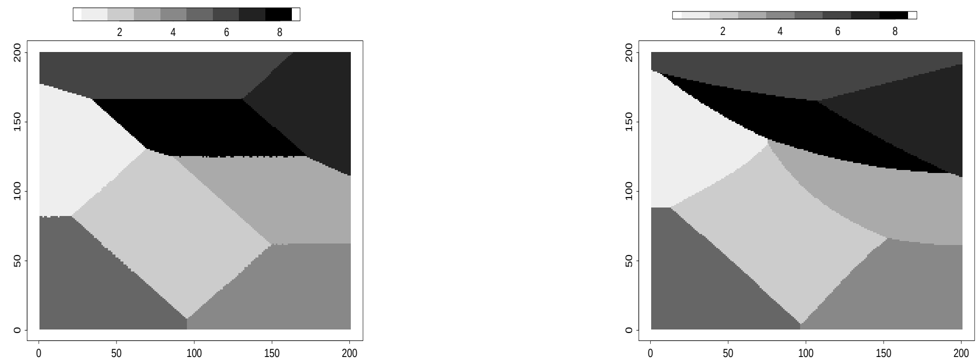


Figure 7: Approximative ℓ_2-, ℓ_4-partition, $m = 8$

The second example adds seven more points to the discrete distribution considered above, so $m = 15$ (see Figure 8).

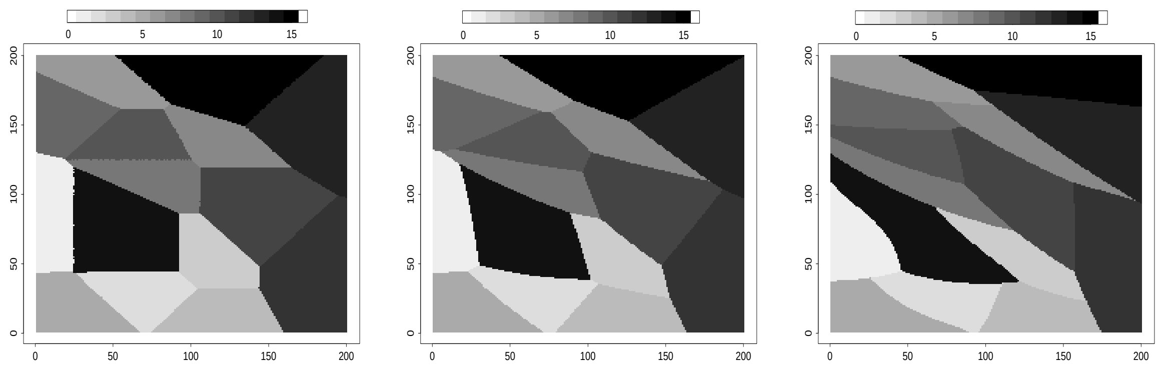


Figure 8: Approximative ℓ_2-, ℓ_3-, ℓ_{10}-partition, $m = 15$

Note that the linear program in (2.2) is not easy to solve. For 2 two-dimensional distributions with a discretized 50 x 50 grid one obtains a linear program with about 6 million variables. So one has to restrict to coarser grids or use carefully designed programs. The Soplex program can handle up to 2 million variables.

3 Optimal couplings of one-dimensional distributions

By a classical result of Dall'Aglio (1956) $E|X-Y|^\alpha$, $\alpha \geq 1$ is minimized in the class of random variables with one-dimensional distributions P, Q by $X = F_P^{-1}(U)$, $Y = F_Q^{-1}(U)$ where $U \stackrel{d}{=} \mathcal{U}([0,1])$ is uniformly distributed on $[0,1]$. This result has been generalized to quasi-monotone costs $c(x,y)$ such as $\Phi(|x-y|)$, or $\Phi(x+y), \Phi$ convex in Cambanis, Simons, and Stout (1976).

More general nonconvex costs of the form $c(x,y) = \Phi(x-y)$ or $\Phi(|x-y|)$ have been dealt with in Uckelmann (1997) and McCann (1999). While Uckelmann used the duality based approach as described in section one to obtain explicit solutions in the uniform case, McCann gave an analysis for concave costs $\Phi(|x-y|)$ based on an elementary geometric no crossing rule already noticed by Monge. His detailed analysis reduces the optimal transportation to a (finite) parametric optimization problem for the coupling of general distributions.

Consider $c(x,y) = \Phi(x+y)$ where $\Phi : [0,2] \to \mathbb{R}$ is in C^2, Φ is convex on $[0,k_1] \cup [k_2,2]$ and concave on $[k_1,k_2]$ in other words Φ is convex, concave, convex. Let $P = Q$ be uniform on $[0,1]$ and let $0 \leq k_1 < k_2 \leq 2$, $0 < \alpha < \beta < 1$ be solutions of the equations

$$\begin{aligned} &\Phi(2\alpha) - \Phi(\alpha+\beta) + (\beta-\alpha)\Phi'(\alpha+\beta) = 0 \\ &\Phi(2\beta) - \Phi(\alpha+\beta) + (\alpha-\beta)\Phi'(\alpha+\beta) = 0 \end{aligned} \tag{3.1}$$

then an optimal coupling of P, Q w.r.t. $c(x,y) = \Phi(x+y)$ is given by $(U, \Gamma(U))$ where $U \stackrel{d}{=} \mathcal{U}([0,1])$ and

$$\Gamma(x) = \begin{cases} x & x \in [0,\alpha] \cup [\beta,1] \\ \alpha+\beta-x & x \in (\alpha,\beta) \end{cases} \tag{3.2}$$

(see Uckelmann (1997) (see Figure 9)).

Similar solutions are obtained for costs $\Phi(x-y), \Phi$ as above and for $\Phi(|x-y|)$ where Φ is concave, convex. These cases can be reduced directly to the case of costs of the form $\Phi(x+y)$ (see Figure 10).

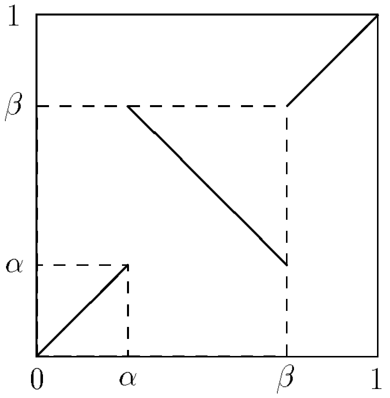


Figure 9:
Optimal measure for $c(x,y)=\Phi(x+y)$

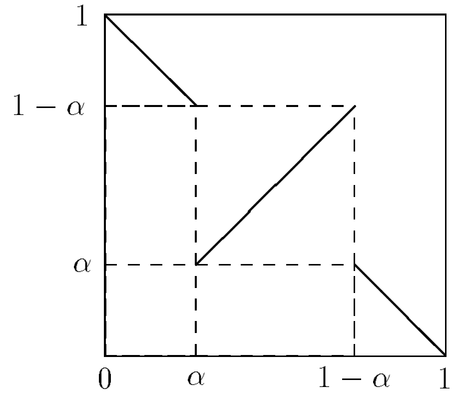


Figure 10:
$c(x,y)=\Phi(|x-y|)$, $\Phi(t)=at^4+bt^2+c$

The strategy to prove optimality is to define

$$f = f_1 1_{[0,\alpha]} + f_2 1_{[\alpha,\beta]} + f_3 1_{[\beta,1]} \tag{3.3}$$

where

$$\begin{aligned} f_1(x) &= \frac{1}{2}\Phi(2x), \\ f_2(x) &= \frac{1}{2}\Phi(2\alpha) + \Phi'(\alpha+\beta)(x-\alpha) \\ \text{and } f_3(x) &= \frac{1}{2}\Phi(2x) + \frac{1}{2}(\Phi(2\alpha)-\Phi(2\beta)) + (\beta-\alpha)\Phi'(\alpha+\beta) \end{aligned}$$

and, furthermore,

$$\begin{aligned} \Psi^1(\xi) &= \Phi(x+\xi) - \frac{1}{2}\Phi(2x) \qquad (3.4) \\ \Psi^2(\xi) &= \Phi(\alpha+\beta-x+\xi) + (x-\alpha)\Phi'(\alpha+\beta) + \frac{1}{2}\Phi(2\alpha) - \Phi(\alpha+\beta) \\ \Psi^3(\xi) &= \Phi(x+\xi) - \frac{1}{2}\Phi(2x) + \frac{1}{2}(\Phi(2\alpha)-\Phi(2\beta)) + (\beta-\alpha)\Phi'(\alpha+\beta). \end{aligned}$$

f is patched together by f_1, f_3 corresponding to the convex parts of Φ and f_2 corresponding to the concave part and $\Psi^1 = \Psi^1_x, x \in [0,\alpha]$, $\Psi^2 = \Psi^2_{\alpha+\beta-x}, x \in [\alpha,\beta]$, $\Psi^3 = \Psi^3_x, x \in [\beta,1]$ are as in (1.6) (see also Figure 11) for the convex and concave parts. By definition $f_i(x) = \Psi^i(x), \forall i, x$. Equation (3.1) allows to patch continuously together the Ψ-functions, $\Psi_{\Gamma(\alpha)}(\beta) = f(\beta)$, $\Psi_{\Gamma(\beta)}(\alpha) = f(\alpha)$.

These equations are used to establish that $f(\xi) \geq \Psi^i(\xi), \xi \in [0,1], i = 1,2,3$. This inequality implies that f is c-convex and by definition of Ψ^i one obtains $\Gamma \in \partial_c f$, i.e. Γ is an optimal coupling function.

These results now seem to suggest a solution for costs of the form $\Phi(x+y)$ where Φ changes its convexity behavior more than two times. Assume e.g. that Φ is

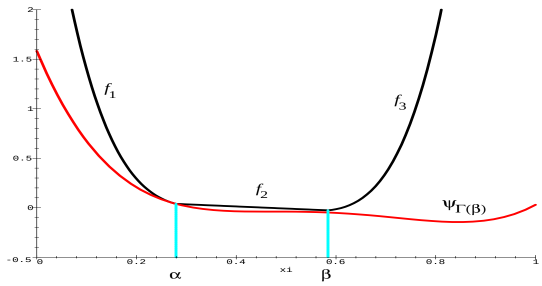


Figure 11: Patching of Ψ-functions

convex-concave-convex-concave-convex (i.e. four changes) and let $\alpha < \beta$ solve (3.1) and $0 < \alpha < \beta < \delta < \varepsilon < \gamma$ solve additionally

$$\begin{aligned} \Phi(2\delta) - \Phi(\delta+\varepsilon) + (\varepsilon-\delta)\Phi'(\delta+\varepsilon) &= 0 \\ \Phi(2\varepsilon) - \Phi(\delta+\varepsilon) + (\delta-\varepsilon)\Phi'(\delta+\varepsilon) &= 0 \end{aligned} \tag{3.5}$$

then

$$\Gamma(x) = \begin{cases} x, & x \in [0,\alpha] \cup [\beta,\delta] \cup [\varepsilon,1] \\ \alpha+\beta-x, & x \in (\alpha,\beta) \\ \delta+\varepsilon-x, & x \in (\delta,\varepsilon) \end{cases}$$

is conjectured to be optimal.

We do not have a formal proof of this statement (the admissibility part is not easy). One can see however analytically that Γ is optimal in the class of all distributions of this special form (i.e. with alternating linear increasing and decreasing parts); the equations for $\alpha, \beta, \delta, \varepsilon$ are the critical first order equations. Also examples calculated approximatively by the corresponding linear programs confirm this.

Consider e.g. $c(x,y) = \Phi(x+y)$ with $\Phi(t) = 0.092t^6 - 0.498t^5 + t^4 + 0.910t^3 + 0.406t^2$ then Φ has four convexity changes and one obtains $\alpha = 0.1, \beta = 0.3, \delta = 0.6, \varepsilon = 0.8$.

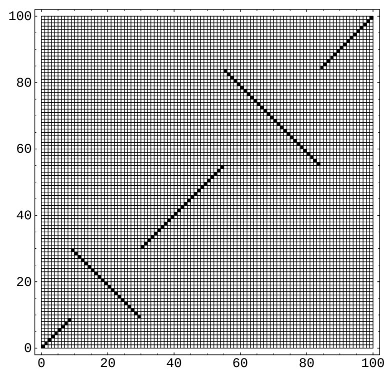


Figure 12: Solution for 4 convexity changes

The example is calculated by discrete approximation based on a 100×100 grid with help of Mathematica.

Changing the coupling function (by componentwise transformation) one can similarly produce coupling problems which have a shifted pattern of the linear terms as 'optimal' solutions; thus for any shuffling of min distribution (see Mikusinski, Sherwood, and Taylor (1991)) one can produce coupling functions c such that this distribution is a c-optimal coupling.

To deal with the optimal coupling for general distributions P, Q on $(\mathbb{R}^1, \mathcal{B}^1)$ with continuous distribution functions F, G and coupling function $c : \mathbb{R}^2 \to \mathbb{R}^1$ let $S = F^{-1}, T = G^{-1}$ and define the transformed costs

$$\tilde{c}(x,y) = c(S(x), T(y)) \text{ on } [0,1]^2$$

then for $X \stackrel{d}{=} P, Y \stackrel{d}{=} Q$

$$\begin{aligned} Ec(X,Y) &= Ec(S \circ F(X), T \circ G(Y)) \\ &= E\tilde{c}(F(X), G(Y)) \qquad \text{and, conversely, for} \\ U, V &\stackrel{d}{=} \mathcal{U}([0,1]) \\ E\tilde{c}(U,V) &= Ec(X,Y) \text{ with } X = S(U), Y = T(V). \end{aligned}$$

This implies

$$\sup_{X \stackrel{d}{=} P, Y \stackrel{d}{=} Q} Ec(X,Y) = \sup_{U \stackrel{d}{=} V \stackrel{d}{=} \mathcal{U}[0,1]} E\tilde{c}(U,V); \tag{3.6}$$

so the optimal coupling problem (P, Q, c) can be reduced to the optimal coupling of two uniform distributions w.r.t. $\tilde{c}$.

The constructions in the following convex resp. concave case are used in the further development of the problem.

Example 3.1

a) *convex costs*

Let F, G be continuous strictly monotone and $c(x,y) = \Phi(x+y)$, Φ convex, then

$$(F^{-1}(U), G^{-1}(U)) \text{ is a } c\text{-optimal coupling.} \tag{3.7}$$

We prove this wellknown result based on the characterization of optimal couplings as given in the introduction.

Let

$$\begin{aligned} f(x) &:= \Phi(S(x) + T(x)) + \int_x^1 T'(t)\Phi'(S(t) + T(t))\, dt \\ \Psi_x(\xi) &:= \Phi(S(\xi) + T(x)) + \int_x^1 T'(t)\Phi'(S(t) + T(t))\, dt \end{aligned}$$

then

$$f(x) = \Psi_x(x)$$

and

$$\begin{aligned} f(\xi) - \Psi_x(\xi) &= \Phi(S(\xi)+T(\xi)) - \Phi(S(\xi)+T(x)) + \int_\xi^x T'(t)\Phi'(S(t)+T(t))\,dt \\ &= \int_\xi^x T'(t)\;[\Phi'(S(t)+T(t)) - \Phi'(S(\xi)+T(t))]\;dt. \end{aligned}$$

Convexity of Φ *implies*

$$\Phi'(S(t)+T(t)) - \Phi'(S(\xi)+T(t)) \quad \begin{cases} \geq 0 & for \quad t \geq \xi \\ \leq 0 & for \quad t \leq \xi. \end{cases}$$

Using $T'(t) \geq 0$ *this implies*

$$f(\xi) \geq \Psi_x(\xi) \qquad \text{for all } x, \xi \in [0,1]$$

and the pair (U, U) *is an optimal* $\tilde{c}$ *coupling,* $\tilde{c}(x,y) = \Phi(S(x)+T(y))$. *Therefore, by* (3.6) $(F^{-1}(U), G^{-1}(U))$ *is an optimal* $c-$*coupling of* P, Q.

b) <u>*concave case*</u>
If Φ *is concave,* $c(x,y) = \Phi(x+y)$, *then*

$$(F^{-1}(U), G^{-1}(1-U) \tag{3.8}$$

is c-optimal.

The proof is similar to that of (3.7) *using now*

$$g(x) = \Phi(S(x)+T(1-x)) - \int_x^1 T'(1-t)\Phi'(S(t)+T(1-t))dt$$

as a $\tilde{c}$*-convex function and showing that* $1-x \in \partial_{\tilde{c}} g(x)$ *for all* $x \in [0,1]$.

Consider next the case where the convexity behavior of Φ changes once. Let $\Phi \in C^2$, be strictly convex on $(-\infty, \lambda]$ and strictly concave on $[\lambda, \infty)$, i.e.

$$\Phi''(t) \begin{cases} > 0 & for \quad t < \lambda \\ < 0 & for \quad t > \lambda. \end{cases}$$

Define

$$\begin{aligned} \eta(t) = \eta_\alpha(t) &= F^{-1}(t) + G^{-1}(1+\alpha-t) \\ &= S(t) + T(1+\alpha-t), \\ H_x(t) &= \Phi'(S(x)+T(1+\alpha-t)) - \Phi'(\eta(t)), \\ G_y(t) &= \Phi'(S(t)+T(1+\alpha-y)) - \Phi'(\eta(t)), \end{aligned} \tag{3.9}$$

then the following explicit optimal coupling result holds.

Theorem 1 *Let P, Q have bounded support and strictly increasing differentiable distribution functions F, G. Let $\alpha \in (0,1)$ be a solution of the equation*

$$\Phi(S(\alpha)+T(1)) - \Phi(S(\alpha)+T(\alpha)) \;=\; \int_\alpha^1 T'(1+\alpha-t)\,\Phi'(\eta(t))\,dt. \tag{3.10}$$

and assume that

a) $\eta(t) \geq \lambda$ *for* $t \in [\alpha, 1]$

b) $\Phi'(S(\alpha)+T(\alpha)) \leq \min\{\Phi'(\eta(\alpha)), \Phi'(\eta(1))\}$

c) For $y \in [\alpha, 1]$ *the functions* H_α, G_y *have at most two zeros in* $[\alpha, 1]$.

Then $(F^{-1}(U), G^{-1}(\Gamma U))$ *is a c-optimal coupling with*

$$\Gamma(x) = \begin{cases} x & \text{for} \quad x \in [0,\alpha] \\ 1+\alpha-x & \text{for} \quad x \in (\alpha, 1] \end{cases}$$

i.e. $G^{-1} \circ \Gamma \circ F$ *is a c-optimal coupling function.*

Proof: For the proof we construct a $\tilde{c}$-convex function f, with $\tilde{c}(x,y) = \Phi(S(x)+T(y))$ such that $\Gamma(u) \in \partial_{\tilde{c}} f(u), u \in [0,1]$. Then the result follows from (3.6).

Define

$$f(x) \;=\; f_1(x)\mathbb{1}_{[0,\alpha]}(x) + f_2(x)\mathbb{1}_{[\alpha,1]}(x) \tag{3.11}$$

with

$$\begin{aligned} f_1(x) &:= \Phi(S(x)+T(x)) - \int_\alpha^x T'(t)\,\Phi'(S(t)+T(t))dt \\ f_2(x) &:= \Phi(S(x)+T(1+\alpha-x)) - \int_x^1 T'(1+\alpha-t)\,\Phi'(\eta(t))dt. \end{aligned}$$

Define for $x \in [0,\alpha]$

$$\begin{aligned} \Psi_1(\xi) &:= \tilde{c}(\xi, x) + a_1 \\ &= \Phi(S(\xi)+T(x)) - \int_\alpha^x T'(t)\,\Phi'(S(t)+T(t))dt \end{aligned}$$

and for $x \in (\alpha, 1]$

$$\begin{aligned} \Psi_2(\xi) &:= \tilde{c}(\xi, 1+\alpha-x) + a_2 \\ &= \Phi(S(\xi)+T(1+\alpha-x)) - \int_x^1 T'(1+\alpha-t)\,\Phi'(\eta(t))dt. \end{aligned}$$

Note that f_i, Ψ_i correspond to the convex resp. concave case in Example 3.1. Since

$$f(x) = \Psi_1(x)\mathbb{1}_{[0,\alpha]}(x) + \Psi_2(x)\mathbb{1}_{[\alpha,1]}(x)$$

it remains to prove that

$$f(\xi) \geq \psi_i(\xi) \text{ for all } \xi \in [0,1], \quad i = 1,2 \tag{3.12}$$

which implies that $\Gamma(u) \in \partial_{\tilde{c}} f(u)$. Note that from (3.10)

$$0 = \int_\alpha^1 T'(1+\alpha-t)\, [\Phi'(\eta(t)) - \Phi'(H_\alpha(t))]\, dt.$$

Therefore, there exists $\tau \in (\alpha, 1)$ such that $\Phi'(\eta(\tau)) = \Phi'(H_\alpha(\tau))$. This implies $S(\alpha) + T(1+\alpha-\tau) \leq \lambda \leq \eta(\tau)$ and using monotonicity of T, Φ' on $[(-\infty, \lambda]$ it follows that $S(\alpha) + T(\alpha) \leq \lambda$ and so $\Phi(S(t) + T(t))$ is convex on $[0, \alpha]$.

To prove (3.12) one has to consider several cases. For $x, \xi \in [0, \alpha]$ is

$$\begin{aligned} D_1(\xi, x) &:= f_1(\xi) - \Psi_1(\xi) \\ &= \Phi(S(\xi) + T(\xi)) - \Phi(S(\xi) + T(x)) - \int_x^\xi T'(t)\, \Phi'(S(t) + T(t))\, dt \\ &\geq 0 \end{aligned}$$

as in Example 3.1 a).

Similarly, using the assumptions of the Theorem one obtains (after a lot of calculations and case distinctions) for $(x, \xi) \in [0,\alpha] \times [\alpha, 1] \cup [\alpha, 1] \times [0, \alpha] \cup [\alpha, 1] \times [\alpha, 1]$ that $f_i(\xi) \geq \Psi_j(\xi)$ (with i, j corresponding to the combination of intervals). The proof of these inequalities is technically somewhat involved; for details we refer to the dissertation of Uckelmann (1998). □

Example 3.2

a) Consider $c(x,y) = -(x+y-1)^3$ *and* P, Q *with densities* $f(x) \equiv 1$, $g(x) = 12(x-\frac{1}{2})^2$. *Then* $G(x) = 4\left(x - \frac{1}{2}\right)^3 + \frac{1}{2}$, $F(x) = x$ *and* $G^{-1}(u) = \left(\frac{u}{4} - \frac{1}{8}\right)^{1/3} + \frac{1}{2}$. *The assumptions of the theorem are fulfilled with* $\Phi(t) = -(t-1)^3$, $\lambda = 1$. *This can be seen from Figure 13.*

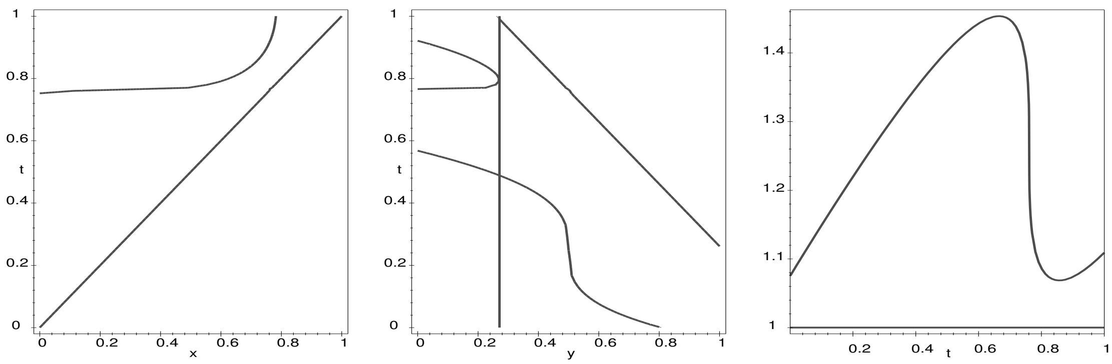
Therefore,

$$\Pi_1(u) := G^{-1}(u) \mathbb{1}_{[0,\alpha]}(u) \;+\; G^{-1}(1-u) \mathbb{1}_{(\alpha,1]} \qquad \textit{with} \qquad \alpha = 0.261$$

is a c-optimal coupling of P, Q.

b) If $g(x) = 2x$, *then similarly as in a)* $\Pi_2(u) = \sqrt{u} \mathbb{1}_{[0,\alpha]}(u) + \sqrt{1-u}\, \mathbb{1}_{[0,\alpha]}(u)$, *with* $\alpha = 0.121$ *is c-optimal.*

c) <u>approximative solutions</u>
If we discretize the transportation problem on the grid $\frac{i}{n+1}$, $0 \leq i \leq n$, *with marginal densities* $f\left(\frac{i}{n+1}\right)$, $g\left(\frac{i}{n+1}\right)$ *then we obtain with help of Mathematica (for* $n = 50$*) the following solutions in examples a) resp. b) which confirm the result of Theorem 1 (see Figures 14 and 15).*



$\{(x,t) : H_x(t) = 0\}$ $\{(y,t) : G_y(t) = 0\}$ $\eta(t)$

Figure 13: Validity of assumptions in Theorem 1

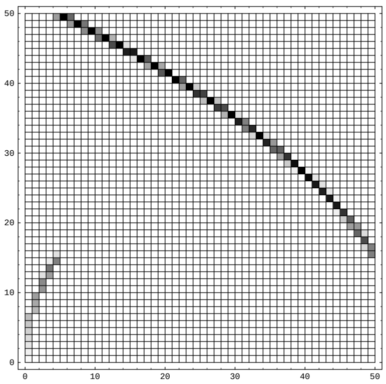


Figure 14:
solution in the linear case $g(x) = 2x$

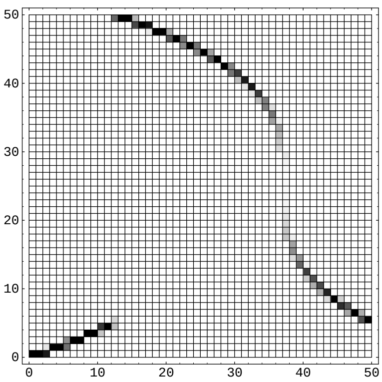


Figure 15:
quadratic case $g(x) = 12(x - \frac{1}{2})^2$

Open Question. It is not known whether the conditions in Theorem 1 are necessary to prove optimality of the pair $(F^{-1}(U), G^{-1}(\Gamma U))$. From several numerical results it seems that this coupling is optimal in general for the considered type of coupling function.

References

Aurenhammer, F., F. Hoffmann, and B. Aronov (1998). Minkowski-type theorems and least-squares clustering. *Algorithmica 20*, 61 ff.

Cambanis, S., G. Simons, and W. Stout (1976). Inequalities for $Ek(X,Y)$ when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie verw. Gebiete 36*, 285–294.

Dall'Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di repartitione doppia. *Ann. Scuola Norm. Sup. Pisa 10*, 35–74.

Kantorovich, L. (1942). On the transfer of masses. *Dokl. Akad. Nauk USSR 37*, 7–8.

McCann, R. (1999). Exact solutions to the transportation problem on the line. *Proc. R. Soc. London A 455*, 1341–1380.

Mikusinski, P., H. Sherwood, and M. Taylor (1991). Probabilistic interpretations of copulas and their convex sum. In *Advances in Probability distributions with given marginals*. G. Dall' Aglio, S. Kotz, G. Salinetti.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématiques et de Physique, pour la même Année, Tirés des Registres de cette Académie, Paris*, pp. 666–704. Mémoires de l'Académie Royale.

Olkin, I. and F. Pukelsheim (1982). The distance between two random vectors with given dispersion matrices. *Journal of Linear Algebra and its Applications 48*, 257–263.

Rachev, S. and L. Rüschendorf (1998). *Mass Transportation Problems, Vol. I: Theory, Vol. II: Applications*. Springer.

Ramachandran, D. and L. Rüschendorf (1998). On the Monge Kantorovich duality theorem. To appear in: Theory Probab. Appl.

Rüschendorf, L. (1991). Fréchet-bounds and their applications. In G. Dall'Aglio, S. Kotz, and G. Salinetti (Eds.), *Advances in Probability Measure with Given Marginals*, pp. 151–188. Amsterdam: Kluwer.

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

Rüschendorf, L. and L. Uckelmann (1997). On optimal multivariate couplings. In V. Beneš and J. Štěpán (Eds.), *Distributions with Given Marginals and Moment Problems*, pp. 261–273. Kluwer Acad. Publ.

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

Uckelmann, L. (1997). Optimal couplings between onedimensional distributions. In V. Beneš and J. Štěpán (Eds.), *Distributions with Given Marginals and Moment Problems*, pp. 275–282. Kluwer Acad. Publ.

Uckelmann, L. (1998). *Über das Monge Kantorovich Transportproblem und dessen Verallgemeinerungen*. Ph. D. thesis, Universität Freiburg.

Wunderling, R. (1996). Paralleler und objektorientierter Simplex-Algorithmus. Technical Report TR 96-09, ZIB (Berlin) technical report.