Generalized Hoeffding–Fréchet functionals and mass transportation

Ludger Rüschendorf, University of Freiburg

This note is concerned with some historical remarks on and a partial review of two interesting mathematical subjects, the generalized Hoeffding–Fréchet functionals and the Monge–Kantorovich mass transportation problem. Both topics have a different motivation and history and are often considered in the literature as different subjects. The main aim of this review is to point out the close connection of these topics and to indicate some possibly fruitful relationships. For the class of Hoeffding–Fréchet functionals risk bounds for a lot of well motivated additional model constraints have been worked out in recent years. These kind of constraints are motivated by risk applications. We indicate some interesting connections of these developments to mass transportation problems as, e.g., to the solution of non-linear mass transportation problems or to the use of stochastic ordering methods to the solution of mass transportation problems. We also briefly indicate the development of algorithms in the transportation problem by the regularization method (entropic optimal transport) and give corresponding references in the literature.

1 Mass transportation problem

The mass transportation problem was introduced by Kantorovich (1942, 1948) as a generalization of the Monge transportation problem to transport one mass-distribution P_1 with minimal cost to a second mass-distribution P_2. While Monge considered only transport maps T from P_1 to P_2 Kantorovich allowed splitting of the mass by means of more general transport kernels K such that $KP_1 = P_2$. In this formulation the set of all admissible transport plans is identified with the Fréchet class $M(P_1, P_2)$ of measures on the product space with marginals P_1, P_2. W.r.t. transport costs $c(x, y)$ the optimal transport (OT) problem then is given by

$$m(c) := \inf\left\{\int c(x,y)\mu(dx,dy);\ \ \mu \in M(P_1,P_2)\right\}. \tag{(OT), 1.1}$$

Kantorovich (1942) stated as main result the identity of the OT problem for a continuous cost function c with a dual problem (DT). The optimal solution of this dual problem is given by a 'potential translocation', i.e. a function T such that $|T(x)-T(y)| \le c(x, y)$. The stated dual problem and the characterization of optimal solutions by potentials however is correct only in the case that the cost c is a metric.

In more general form this duality theorem for the metric case was given in the paper of Kantorovich and Rubinstein (1957) and is known as Kantorovich–Rubinstein Theorem. It states for $c = d$ being a metric

$$m(d) = DT(d) := \sup\left\{\int f d(P_1 - P_2);\ \ f \in \mathrm{Lip}(1)\right\} \tag{1.2}$$

where Lip(1) is the class of Lipschitz functions with constant 1, i.e. $|f(x)-f(y)| \le d(x, y)$, $\forall\, x, y$.

General versions of the Kantorovich–Rubinstein theorem were given later in Levin (1975), Dudley (1976), Fernique (1981), de Acosta (1982), and Kellerer (1984a). Mass transportation problems were a main subject in the Russian probability literature and were developed in particular in the context of probability metrics (minimal metrics) and in connection with convergence results in central limit theorems, see Levin (1975), Levin and Milyutin (1979), Zolotarev (1976), Rachev (1985, 1991).

In particular also the related mass transshipment problem formulated in terms of masses μ with given difference of the marginals $\mu_2 - \mu_1 = P_2 - P_1$ was investigated in great detail. The main argument for their treatment is a reduction argument allowing for general cost functions c a reduction to an associated mass transportation problem with metric costs.

In recent time this problem has found a lot of interest in various mathematical fields like in analysis and PDEs, in Monge–Ampère–Boltzmann and evolution equations as well as for stochastic differential equations. It is an important tool in Riemannian geometry as for Ricci curvature bounds and gradient flows as well as for various classes of inequalities in probability and analysis, for isoperimetric inequalities and for transportation inequalities. The induced minimal metrics like the Kantorovich metric or the minimal L^2-metric (Wasserstein metric) are a main tool in the analysis of recursive stochastic equations and algorithms. They are of importance in statistics in risk theory and in mathematical finance for the construction of robust (neighbourhood) models (robust modelling) as well as in the area of image reconstruction and statistical clustering.

For some of these developments see the expositions in Rachev (1985, 1991), Rüschendorf (1991a,b), Cuesta-Albertos, Rüschendorf, and Tuero-Diaz (1993), Cuesta-Albertos, Matrán, Rachev, and Rüschendorf (1996), Rachev and Rüschendorf (1998a,b), Villani (2003), Ambrosio (2003), Ambrosio and Pratelli (2003), Ambrosio, Gigli, and Savaré (2005), Puccetti and Rüschendorf (2012a,b), Embrechts, Puccetti, and Rüschendorf (2013), Santambrogio (2015), Puccetti, Rüschendorf, Small, and Vanduffel (2017), Rüschendorf (2018), and Peyré and Cuturi (2018, 2019).

By all these developments the topic of mass transportation has become a most fruitful tool for various mathematical areas.

2 Generalized Hoeffding–Fréchet functionals

Independently from the motivations and development of the Monge–Kantorovich mass transportation problem there has been a parallel development of so called generalized Hoeffding–Fréchet functionals which aim to describe the range of possible influence of stochastic dependence on the expectation of a functional or more generally on a non-linear (convex) functional of the random vector.

The historical origin of this class of problems is to be found in early work of the Italian school of probability like Gini (1914), Salvemini (1949), and Dall'Aglio (1956) as well as by Fréchet (1940, 1951) and Hoeffding (1940, 1951), who established upper and lower bounds for a distribution function F of n variables, when marginals $F_1, \dots, F_n$ are prescribed as well as sharp upper and lower bounds in the real case $n = 2$ for EX_1X_2 with $X_i \sim F_i$, $i = 1, 2$. An extension to bounds for $E\varphi(X_1, X_2)$ for supermodular functions φ was given in Cambanis, Simons, and Stout (1976), Whitt (1976), Szulga (1978), and Tchen (1980). A general formulation of this class of topics was introduced in Rüschendorf (1979, 1980, 1983), and Gaffke and Rüschendorf (1981) in general (multimarginal) context as generalized Hoeffding–Fréchet functionals defined for probability measures P_i on $(\mathfrak{X}_i, \mathfrak{A}_i)$, $1 \leq i \leq n$ and functions φ on $\mathfrak{X} = \prod_{i=1}^n \mathfrak{X}_i$ by

$$M(\varphi) := \sup \left\{ \int \varphi \, \mathrm{d}P; \;\; P \in M(P_1, \dots, P_n) \right\}, \tag{2.1}$$

where $M(P_1,\dots,P_n)$ is the Fréchet class of measures P on the product

$$(\mathfrak{X},\mathfrak{A}) = \Big(\prod_{i=1}^n \mathfrak{X}_i, \otimes_{i=1}^n \mathfrak{A}_i\Big)$$

with marginals $P_1,\dots,P_n$. So $M(\varphi)$ describes the maximal value of the integral under all possible dependence structures and fixed marginals $P_1,\dots,P_n$. Similarly $m(\varphi)=\inf\{\dots\}$ gives the minimal value and the interval (open or closed) $(m(\varphi),M(\varphi))$ describes the range of values caused by dependence.

In modern risk theory, where X is a risk vector with known marginal distributions P_i of X_i this gives the range of best and worst case risk vector. In early papers on this topic there were given general duality results in Rüschendorf (1979, 1980, 1981), Gaffke and Rüschendorf (1981), and Kellerer (1984a,b), Ramachandran and Rüschendorf (1995, 2000) of the form

$$M(\varphi) = \inf\Big\{\sum\int \varphi_i \,\mathrm{d}P_i;\ \ \varphi \le \sum \varphi_i\Big\} \tag{2.2}$$

implying in particular sharpness of Fréchet bounds and improvements of classical inequalities like Hölder's, Cauchy–Schwarz's and Jensen's inequality, when the marginal distribution functions are known.

The generalized Hoeffding–Fréchet functional in the case $n=2$ is identical to the Kantorovich mass transportation problem, when φ represents the cost of transportation. As a consequence the above mentioned duality results for the Hoeffding–Fréchet functionals were the first valid duality results for the Kantorovich mass transportation problem with general (non-metric) costs.

They have a natural interpretation and were introduced from the beginning not only for two but for a general number of marginals (multimarginal transportation problems). From the perspective of mass transportation problems only later on the multi-marginal mass transportation problems were introduced, where the mass transport happens in several intermediate steps $x_1=x,x_2,\dots,x_n=y$ from x to y. Here not only the initial mass distribution $P=P_1$ and the final mass distribution $Q=P_n$ are prescribed but also the intermediate mass distributions P_i are prescribed and the cost of the transport depends on all intermediate steps, $c=c(x_1,x_2,\dots x_n)$, as, e.g., for $c(x_1,\dots,x_n)=\sum_{i=1}^{n-1} c(x_i,x_{i+1})$.

This connects to the much earlier development of the mass transshipment problem (as mentioned before) where only the difference between the final and the initial distributions $Q-P$ is prescribed, the number of intermediate steps varies over all natural numbers, and the intermediate masses P_i, $2\le i\le n-1$ can be chosen in a free way.

Hoeffding–Fréchet functionals are a main tool for establishing (sharp) risk bounds in risk theory under (complete) dependence uncertainty.

For the numerical solution of the problem of determining sharp risk bounds a description of the Hoeffding–Fréchet problem in terms of a rearrangement problem is fundamental. We state this result for the problem to determine the maximal tail risk of the sum, see Rüschendorf (1983).

Theorem 2.1 (Rearrangement = Dependence) *Let $\mathfrak{F}(F_1,\dots,F_d)$ be the set of all joint distribution functions on $\mathbb{R}^d$ with marginals $F_1,\dots F_d$.*
Let U be a random variable with uniform distribution $\mathcal{U}(0,1)$ on $(0,1)$. Then

$$\mathfrak{F}(F_1,\dots,F_d) = \big\{F_{(f_1(U),\dots,f_d(U))};\ \ f_i \sim_r F_i^{-1}, 1\le i\le d\big\}.$$

$$\begin{aligned} M(s) &= \sup\Big\{P\Big(\sum_{i=1}^d X_i \ge s\Big);\ \ X_i\sim F_i\Big\} \\ &= \inf\Big\{\alpha;\ \ \exists\, f_j^\alpha \sim_r F_j^{-1}\,|_{[\alpha,1]}, \sum_{j=1}^d f_j^\alpha \ge s\Big\}. \end{aligned}$$

Similarly,

$$M_{\leq}(s) = \sup\Big\{P\Big(\sum_{i=1}^{d} X_i \leq s\Big);\;\; X_i \sim F_i\Big\}$$

$$= \sup\Big\{\alpha \in [0,1];\;\; \exists f_j^\alpha \sim_r F_j^{-1}\,|_{[0,\alpha]}, \sum_{j=1}^{d} f_j^\alpha \leq s \text{ on } [0,\alpha]\Big\}.$$

Here $f \sim_r g$ denotes that g is a rearrangement of f, i.e. both functions have the same distribution function, while $f \sim g\,|_{[\alpha,1]}$ denotes that f is a rearrangement of g on $[\alpha, 1]$, i.e. $f\,|_{[\alpha,1]} \sim g\,|_{[\alpha,1]}$.

The structural result in Theorem 2.1 implies that for $d = 2$ the worst case distribution maximizing $M(s)$ resp. maximizing $\mathrm{Var}_\alpha(X_1 + X_2)$ is obtained by the countermonotonic coupling in the corresponding upper part of the distributions and the best case distribution minimizing $M(s)$ resp. minimizing $\mathrm{Var}_\alpha(X_1 + X_2)$ is obtained by the countermonotonic coupling in the lower part of the distribution.

Let $F_i^\alpha, (F_{i,\alpha})$ denote the distribution F_i restricted to the upper resp. lower α-part of F_i. Then for $d = 2$ for the worst resp. best case distributions the upper resp. lower parts of the distributions are flattened as much as possible. This principle also extends to $d \geq 2$, see Bernard, Rüschendorf, and Vanduffel (2017a, Theorem 2.5).

Theorem 2.2 (VaR-bounds and convex order) *Let F_i^α denote the upper α-part of F_i, then for $\mathrm{VaR}_\alpha^+(S) = \sup\{x \in \mathbb{R};\; F_S(x) \leq \alpha\}$ it holds*

a) $\overline{\mathrm{VaR}}_\alpha^+ = \sup\limits_{X_i \sim F_i} \mathrm{VaR}_\alpha^+\Big(\sum_{i=1}^{d} X_i\Big) = \sup\limits_{Y_i^\alpha \sim F_i^\alpha} \mathrm{VaR}_0^+\Big(\sum_{i=1}^{d} Y_i^\alpha\Big).$

b) If $X_i^\alpha, Y_i^\alpha \sim F_i^\alpha$ and $S^\alpha = \sum_{i=1}^{d} Y_i^\alpha \leq_{cx} \sum_{i=1}^{d} X_i^\alpha$, then

$$\mathrm{VaR}_0^+\Big(\sum_{i=1}^{d} X_i^\alpha\Big) \leq \mathrm{VaR}_0^+\Big(\sum_{i=1}^{d} Y_i^\alpha\Big).$$

As a result minimizing the sum in the upper part of the distribution in convex order implies, as in the case $d = 2$, minimizing the VaR. This minimization of the convex order is achieved in particular under mixing conditions on $F_1, \dots, F_d$. Here $F_1, \dots, F_d$ are called mixable if there exist $X_i \sim F_i$ and a constant c such that $\sum_{i=1}^{d} X_i = c$. Since by Theorem 2.2 for mixable distributions the mixing variables Y_i^α for F_i^α realize the convex minimum Theorem 2.2 implies the following corollary.

Corollary 2.3 *a) If $(F_i^\alpha)_{1\leq i\leq d}$ are mixable, then for mixing variables $Y_i^\alpha \sim F_i^\alpha$ with $S^\alpha = \sum_{i=1}^{d} Y_i^\alpha = c$ holds:*

$$\mathrm{VaR}_\alpha^+\Big(\sum_{i=1}^{d} X_i\Big) \leq \mathrm{VaR}_0^+(S^\alpha) = c.$$

b) If $(F_{i,\alpha})_{1\leq i\leq d}$ are mixable with mixing variables $(Y_{i,\alpha})$, then for any $X_{i,\alpha} \sim F_{i,\alpha}$, $S'_\alpha = \sum_{i=1}^{d} X_{i,\alpha}$, and $S_\alpha = \sum_{i=1}^{d} Y_{i,\alpha}$

$$\mathrm{VaR}_1(S_\alpha) \leq \mathrm{VaR}_1(S'_\alpha).$$

In consequence this implies: If S_α is a smallest sum in convex order w.r.t. $(F_{i,\alpha})$, then

$$\mathrm{VaR}_1(S_\alpha) = \inf\Big\{ \mathrm{VaR}_\alpha\Big(\sum_{i=1}^d X_i\Big),\ \ X_i \sim F_i \Big\}.$$

The mixability of distributions $F_1, \dots, F_d$ has been intensively investigated in the literature by R. Wang and coauthors, see e.g. Wang and Wang (2011) and Puccetti, Wang, and Wang (2013). By the above structural results these mixing results allow to determine worst and best case risks for several classes of distributions.

The reformulation of the tail risk problem as a rearrangement problem led to the introduction of the rearrangement algorithm (RA) which allows a precise determination of tail risk bounds or equivalently of VaR (= Value at risk) bounds, the best case VaR_α and the worst case VaR_α denoted by $\underline{\mathrm{VaR}}_\alpha$ in Puccetti and Rüschendorf (2012a).

The following example in Table 2.1 from Embrechts et al. (2013) for random vectors of Pareto(2) distributed risks, where the exact value of the upper bound $\overline{\mathrm{VaR}}_\alpha$ is known shows that the RA algorithm is precise. The example uses a discretization of the distributions of N steps. It is astonishing that the comonotonic VaR is closer to the best case VaR than to the worst case VaR in this example.

$\boldsymbol{d = 8}$	$N = 1.0e05$	*avg time:* 30 secs		
α	$\underline{\mathrm{VaR}}_\alpha$ (RA range)	VaR^+_α (exact)	$\overline{\mathrm{VaR}}_\alpha$ (exact)	$\overline{\mathrm{VaR}}_\alpha$ (RA range)
0.99	9.00 – 9.00	72.00	141.67	141.66 – 141.67
0.995	13.13 – 13.14	105.14	203.66	203.65 – 203.66
0.999	30.27 – 30.62	244.98	465.29	465.28 – 465.30
$\boldsymbol{d = 56}$	$N = 1.0e05$	*avg time:* 9 mins		
α	$\underline{\mathrm{VaR}}_\alpha$ (RA range)	VaR^+_α (exact)	$\overline{\mathrm{VaR}}_\alpha$ (exact)	$\overline{\mathrm{VaR}}_\alpha$ (RA range)
0.99	45.82 – 45.82	504.00	1053.96	1053.80 – 1054.11
0.995	48.60 – 48.61	735.96	1513.71	1513.49 – 1513.93
0.999	52.56 – 52.58	1714.88	3453.99	3453.49 – 3454.48
$\boldsymbol{d = 648}$	$N = 1.0e05$	*avg time:* 8 hrs		
α	$\underline{\mathrm{VaR}}_\alpha$ (RA range)	VaR^+_α (exact)	$\overline{\mathrm{VaR}}_\alpha$ (exact)	$\overline{\mathrm{VaR}}_\alpha$ (RA range)
0.99	530.12 – 530.24	5832.00	12302.00	12269.74 – 12354.00
0.995	532.33 – 562.50	8516.10	17666.06	17620.45 – 17739.60
0.999	608.08 – 608.47	19843.56	40303.48	40201.48 – 40467.92

Table 2.1 Estimates for $\overline{\mathrm{VaR}}_\alpha$ and $\underline{\mathrm{VaR}}_\alpha$ for random vectors of Pareto(2)-distributed risks. Comparision to the VaR_α of the comonotonic distribution denoted by VaR^+_α. Computation times on a standard laptop.

The example in Figure 2.1 shows based on the RA the wide VaR range for the classical Moscadelli (2004) data with $d = 8$ for operational risks with a generalized Pareto distribution (GPD).

This example shows that the unconstrained bounds corresponding to the unrestricted mass transportation problem are too wide to be usable in this application.

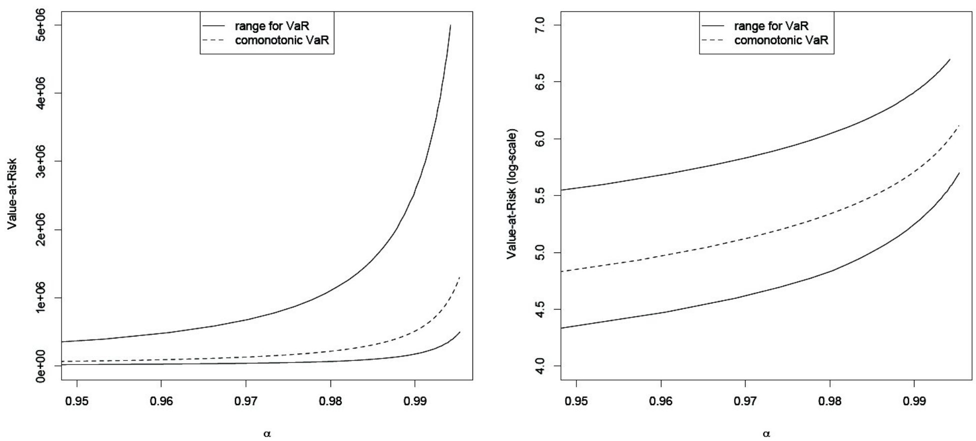


Figure 2.1 VaR range and comonotonic VaR(8) (in log-scale on the right) for the sum of $d = 8$ GPD risks with parameters following Moscadelli (2004), based on RA for $N = 1.0e05$.

3 Optimal multivariate couplings and transports

There is a difference in notation of researchers working in the more analytically oriented area of mass transportation problems and those working in the more probabilistic area of Hoeffding–Fréchet functionals. In the case $n = 2$ in the first group an optimal solution of the transportation problem is given by a measure μ in the Fréchet class $M(P_1, P_2)$ resp. a transport map T from P_1 to P_2. For the optimization problem in Hoeffding–Fréchet functionals a solution is typically given by an optimal coupling (X_1, X_2) of random variables with distributions $X_i \sim P_i$, $i = 1, 2$, i.e. $m(\varphi) = E\varphi(X_1, X_2)$.

This difference in notation may have led in several instances and presentations of the theory of mass transportation to neglect the corresponding developments from the area of Hoeffding–Fréchet functionals. An example of this kind is the basic transportation result in the case of optimal L^2-couplings with cost function $c(x, y) = \|x - y\|^2$. The basic result for this problem is the following theorem.

Theorem 3.1 (Optimal L^2-couplings)
Let $P_i \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$, $i = 1, 2$ with $\int \|x\|^2 \,\mathrm{d}P_i(x) < \infty$, then:

a) There exists an optimal L^2-coupling (X, Y) of P_1, P_2.

b) $X \sim P_1, Y \sim P_2$ is an optimal L^2-coupling of $P_1, P_2 \Leftrightarrow \exists$ convex, lsc $f \in L^2(P_1)$ such that $Y \in \partial f(X)$ a.s., where $\partial f(X)$ is the subgradient of f in X.

c) If $P_1 \ll \lambda\!\!\!\lambda^k$, then for f as in b) the subgradient $\partial f(X)$ of f in X is given by $\partial f(X) = \{\nabla f(X)\}$ a.s. and $(X, \nabla f(X))$ is an a.s. unique solution of the Monge transportation problem.

Remark 3.2 (Optimal L^2-coupling theorem) *Part a) of this theorem follows from a standard existence result. The most important point of this theorem is part b) which was given in this form first in Rüschendorf and Rachev (1990). The proof there is based on the above stated duality theorem. Part c) is an immediate consequence of part b) since convex functions f are Lebesgue a.s. differentiable and thus $\partial f(x) = \{\nabla f(x)\}$ a.s. The sufficiency part of b) is contained already in Knott and Smith (1984, 1987).*

Brenier (1991) established the important particular case in b) for $P_1 \ll \lambda\!\!\!\lambda^k$ with bounded support, as well as the uniqueness of solutions in part c). In a large part of *the literature, in particular the analysis oriented literature, this theorem is denominated as 'Brenier's theorem'.*

By the above given history this terminology seems not justified and should be replaced by a more fair denomination as, e.g., the more neutral 'optimal L^2-coupling theorem' and mentioning all main contributions to this important result.

The optimal L^2-coupling theorem (Theorem 3.1) has been extended in Rüschendorf (1991a,b, 1996) to general cost functions c as follows.

Theorem 3.3 (Optimal c-coupling) *Let c be a lower majorized cost function (i.e. $c(x,y) \geq f_1(x) + f_2(y)$ for some $f_1 \in L^1(P_1)$, $f_2 \in L^1(P_2)$) and assume that $m(c) < \infty$.*

Then a pair (X,Y) with $X \sim P_1$, $Y \sim P_2$ is an optimal c-coupling of P_1, P_2, i.e., $m(c) = Ec(X,Y)$ if and only if

$$(X,Y) \in \partial_c f \text{ a.s.} \quad \text{for some } c\text{-convex funtion } f, \tag{3.1}$$

equivalently, $Y \in \partial_c f(X)$ a.s.

Here $\partial_c f(x)$ denotes the c-subgradient of f in x and $\partial_c f = \{(x,y);\ y \in \partial_c f(x)\}$.

For the notions of c-convexity and c-subgradient see Rüschendorf (1991a,b, 1996). This theorem was reformulated in Knott and Smith (1994) in terms of c-cyclical monotonicity. For further extensions see Gangbo and McCann (1996), Ambrosio and Pratelli (2003), and Schachermayer and Teichmann (2009).

A more detailed review of this result is given in Rüschendorf (2007).

It is interesting to remind that the optimal coupling results in Theorem 3.1 and 3.3 for the case $n = 2$ are also basic for the solution of a class of non-linear generalized Hoeffding–Fréchet problems in risk theory namely to determine worst case risks for a specified law invariant, convex risk measure ϱ on $\mathbb{R}^k$, i.e., to solve for given $P_i \in M^1(\mathbb{R}^k, \mathfrak{B}^k)$, $k = 1, \dots, n$,

$$\varrho(X) = \sup\{\varrho(Y);\ Y_i \sim P_i, 1 \leq i \leq n\}. \tag{3.2}$$

A representation result in Rüschendorf (2006) gives a representation of law invariant convex risk measures for risks X with multidimensional components $X_i \sim P_i$ in terms of 'max-correlation risk measures' Ψ_μ. These are defined for distributions μ of scenario densities $Y = (Y_1, \dots, Y_k)$, $Y \sim \mu$ with $Y_i \geq 0$, $Y_i \in L^q$ and $EY_i = 1$, $1 \leq i \leq n$, by

$$\Psi_\mu(X) = \sup_{\widetilde{X} \sim X} E\widetilde{X} \cdot Y = \sup_{\widetilde{Y} \sim Y} EX \cdot \widetilde{Y} \tag{3.3}$$

as the solution of a L^2-transportation problem. A convex risk measure Ψ on L_k^p with components in L^p is shown to be law invariant if and only if

$$\Psi(X) = \sup_{\mu \in A} (\Psi_\mu(X) - \alpha(\mu)) \tag{3.4}$$

for some class A of scenario measures μ and with a penalty function α (see Rüschendorf (2006). This representation result allows to determine the solution of the non-linear Hoeffding–Fréchet problem in two steps. In step 1 a worst case scenario measure $\mu_0 \in A$ has to be determined maximizing the average risk functional $F(\mu) = \frac{1}{n} \sum_{i=1}^n \Psi_\mu(X_i) - \alpha(\mu)$. In the second step n μ_0-comonotone solutions $X_1, \dots, X_n$ of the usual L^2-optimal coupling problems of P_i, μ_0 have to be determined. Thus the non-linear optimal mass transportation resp. Hoeffding–Fréchet functional problem can be reduced to the solution of a variational problem and a class of n optimal L^2-coupling problems. For more details see Burgert and Rüschendorf (2006) and Rüschendorf (2006, 2012).

4 Generalized Hoeffding-Frechet functionals and – mass transportation under additional constraints

The transportation problems and problems of Hoeffding–Fréchet functionals considered in the first part of this review concern the unconstrained case, i.e., when all possible dependence structures resp. transportation plans are allowed. The corresponding optimal risk bounds resp. optimal transportation results however are often not acceptable in applications. This led in recent years to considerable effort to deal with Hoeffding–Fréchet functionals with additional constraints on the dependence structure as well as of structural kind.

The following figure describes some of the modifications considered in the literature. Various questions and extensions of the standard mass transportation problems resp. the

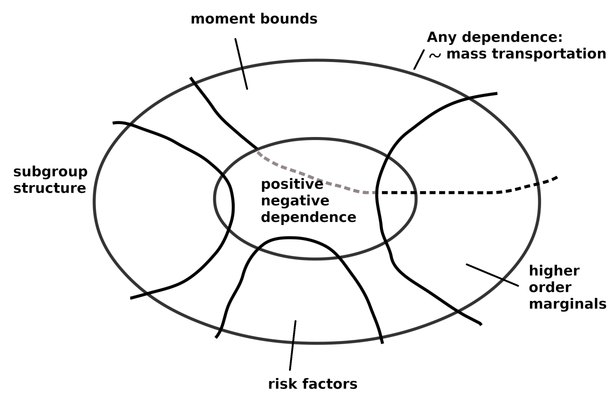


Figure 4.1

corresponding Hoeffding–Fréchet functional problems concern the following points:

a) Hoeffding–Fréchet functionals and mass transportation with additional constraints. Generalized moment constraints, multivariate marginals, positive negative dependence information, additional structural constraints as partial information on risk factors (partially specified risk factor models) or models with inherent subgroup structure and transports with local or global capacity or flow rate constraints.

b) additional martingale constraints lead to improved price bounds

c) new ordering methods within subclasses

d) worst case risks w.r.t. risk measures correspond to non-linear mass transportation, case of higher dimensional risks

e) entropic optimal transport

The general intuition is that positive dependence information allows to increase lower risk bounds (but not to decrease upper bounds) while negative dependence information allows to decrease upper risk bounds (but not to increase lower bounds).

The points a)–c) concern mainly modifications of the generalized Hoeffding–Fréchet functional problem. A general treatment of some topics as in a) is given in Rüschendorf (2013) and a far extended exposition in the recent textbook Rüschendorf, Vanduffel, and Bernard (2024). The results to point a) are coming and motivated to a great extent from the generalized Hoeffding–Fréchet functionals. For mass transportation problems there have been investigated several classes of constraints on the transportation plans like capacity constraints,locally and globally, describing f.e. allowed or forbidden regions of the transport plan or posing constraints on flow rates in a dynamical formulation of the transport problem. For some references on this see Barnes and Hoffman (1985), Rachev and Rüschendorf (1994) , Cuesta-Albertos et al. (1996), Rachev and Olkin (1999), Korman and McCann (2015), Ekren and Soner (2018) and Dong et al. (2024).

A particular interesting class of constraints for the mass transportation problem are the quite recently intensively studied martingale mass transportation problems in point b) where the motivation of the additional martingale restriction comes from the martingale structure induced by the financial martingale pricing measures (see e.g., Beiglböck et al. (2013)). Point c) of the list indicates that the solution of such kind of problems is often based on the extension and use of several stochastic ordering results in these subclasses.

The problem of non-linear mass transportation problems in point d) of the list was already described in brief form at the end of Section 3.

The topic of entropic optimal transport in e) is concerned with so called entropic regularization which plays a key role in establishing efficient algorithms with provable convergence. Entropic regularization is a popular choice since it allows for the application of the iterative proportion fitting procedure (IPFP) also called Sinkhorn's algorithm that can be implemented in large scale applications and is simultaneously analytically tractable.

The basic entropically regularized transport problem is of the form

$$m_r(c,\varepsilon) = \inf\left\{ \int c(x,y)\mu(dx,dy) + \varepsilon H(\mu \mid P_1 \otimes P_2);\ \ \mu \in M(P_1,P_2) \right\},$$

where $H(\mu \mid P_1 \otimes P_2)$ denotes the relative entropy w.r.t. the product measure resp. the Kullback–Leibler distance of μ and $P_1 \otimes P_2$, and $\varepsilon > 0$ is the regularization parameter. Under general conditions it can be shown that solutions of the regularized transport problems for approximating sequences $P_1^n \to P_1$, $P_2^n \to P_2$ of marginals, in particular for discrete approximating sequences, converge to solutions of the transportation problem for P_1, P_2. For a presentation and review of the ample literature on this topic we refer to Nenna (2016), Benamou and Brenier (2000), Nutz (2021), Bernton, Ghosal, and Nutz (2022).

In the following examples we give a small impression of the quality and range of improvement of the VaR bounds which can be achieved by additional constraints as in point a).

4.1 Risk bounds under moment constraints

Let $X_i \sim F_i$, $1 \le i \le n$, and assume we are given an upper bound on the variance of $S_n = \sum_{i=1}^n X_i$

$$\mathrm{Var}(S_n) \le s^2. \tag{4.1}$$

This is a simple additional information often available for risk models. We denote for given α

$$\begin{aligned} M &= \sup\{\mathrm{VaR}_\alpha(S_n);\ \ X_i \sim F_i, 1 \le i \le n, \mathrm{Var}(S_n) \le s^2\} \\ m &= \inf\{\mathrm{VaR}_\alpha(S_n);\ \ X_i \sim F_i, 1 \le i \le n, \mathrm{Var}(S_n) \le s^2\}. \end{aligned} \tag{4.2}$$

Then by means of a Cantelli type bound it holds (see Bernard et al. (2017a, Theorem 3.2)).

Theorem 4.1 *$\alpha \in (0,1)$ and $\mathrm{Var}(S_n) \le s^2$, then*

$$a := \max\left(\mu - s\sqrt{\frac{\alpha}{1-\alpha}}, A\right) \le m \le \mathrm{VaR}_\alpha(S_n) \le M$$
$$\le b := \min\left(\mu + s\sqrt{\frac{\alpha}{1-\alpha}}, B\right), \quad \mu = ES_n.$$

Here $A = \sum_{i=1}^n \mathrm{LTVaR}_\alpha(X_i)$ and $B = \sum_{i=1}^n \mathrm{TVaR}_\alpha(X_i)$ are defined in terms of the left tail resp. right tail value at risk.

An extension of the RA, the ERA as introduced in Bernard et al. (2017b), allows to solve these kind of problems precisely. For this problem with additional bounds on the variance of the joint portfolio S_n the example of Pareto(3) distributed risks in Table 4.1 and for various levels of variance bounds induced by the correlation ϱ between the variables shows that the value of the ERA (which corresponds to a real dependence structure) is quite close to the upper Cantelli bound. This implies that both – the ERA and the Cantelli bounds – are precise.

For more details we refer to Bernard, Rüschendorf, Vanduffel, and Wang (2017b) including higher order moment bounds like on the first 3 or 4 moments of S_n which allow to improve the upper risk bounds, see Bernard et al. (2017a).

The upper moment bounds on S_n are negative dependence constraints and thus indicate improvements of the upper risk bounds.

4.2 Partially specified risk factor models

A practically most relevant additional model assumption are the partially specified risk factor models.

Let $X = (X_1, \dots, X_n)$ be a risk vector; Z a risk factor variable such that

$$X_j = f_j(Z, \varepsilon_j), \quad 1 \le j \le n, \tag{4.3}$$

where Z is a systematic risk factor and ε_j are individual risk factors. We assume that the distributions of $(X_j, Z) \sim H_j$, $1 \le j \le n$ are known, but the joint distribution of X, Z is unknown. This implies that the marginal distributed functions F_j of X_j and the conditional distribution functions $F_{j|z}$ of X_j given $Z = z$ are known. Then the model

$$A(H) = \{(X, Z); \ (X_j, Z) \sim H_j, 1 \le j \le n\} \tag{4.4}$$

is called partially specified risk factor model. Precise but quite involved improved upper and lower bounds for the partially specified risk factor model can be formulated in terms of the conditional models. The risk factor Z may introduce positive or negative dependence information and thus leads to reduction of the risk bounds. This model was introduced and investigated in Bernard et al. (2017a) and in Bernard et al. (2017b). Based on a mixing representation of the model a simple to use representation of the sharp upper VaR_α bound is given there which is based on the conditionally comonotonic vector. This model is shown to lead to potentially considerable reduction of the upper risk bounds. An insightful example showing this effect is the following.

Example 4.2 (Pareto distributions with dependence parameter p)
Let $\varepsilon_i^j \sim$ Pareto(4), $U \sim \mathcal{U}(0,1)$ and $I \sim \mathfrak{B}(1,p)$. Assume that we have two groups of risks X_j^1, X_j^2, $1 \le j \le n/2$ where X_j^i are given by

$$X_i^1 = (1-U)^{-1/3} - 1 + \varepsilon_i^1$$
$$X_i^2 = I((1-U)^{-1/3} - 1) + (1-I)(Z^{-1/3} - 1) + \varepsilon_i^2.$$

Panel A: Approximate sharp bounds obtained by the ERA

(m, M)		$n = 10$		
		$\varrho = 0$	$\varrho = 0.15$	$\varrho = 0.3$
	$\mathrm{VaR}_{95\%}$	(4.401; 15.72)	(4.091; 21.85)	(3.863; 26.19)
$d = 10{,}000$	$\mathrm{VaR}_{99\%}$	(5.486; 28.69)	(4.591; 43.45)	(4.492; 53.22)
	$\mathrm{VaR}_{99.5\%}$	(6.820; 39.48)	(5.471; 59.60)	(4.850; 73.11)

(m, M)		$n = 100$		
		$\varrho = 0$	$\varrho = 0.15$	$\varrho = 0.3$
	$\mathrm{VaR}_{95\%}$	(47.96; 84.72)	(42.48; 188.9)	(39.61; 243.3)
$d = 10{,}000$	$\mathrm{VaR}_{99\%}$	(48.99; 129.5)	(46.61; 366.0)	(45.36; 489.5)
	$\mathrm{VaR}_{99.5\%}$	(49.23; 162.8)	(47.54; 499.1)	(46.68; 671.5)

Panel B: Variance-constrained bounds

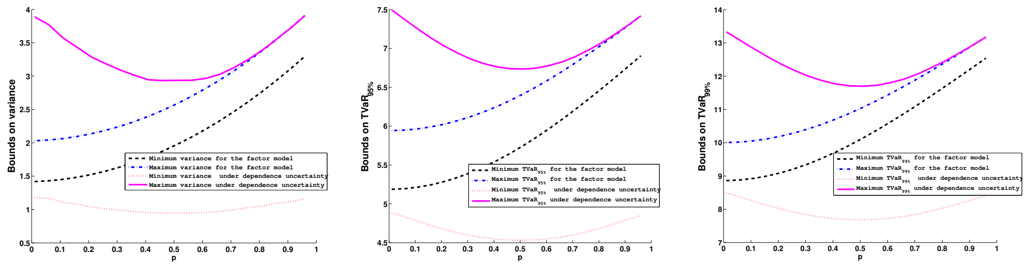
(a, b)		$n = 10$		
		$\varrho = 0$	$\varrho = 0.15$	$\varrho = 0.3$
	$\mathrm{VaR}_{95\%}$	(4.398; 16.03)	(4.089; 21.92)	(3.861; 26.23)
$d = 10{,}000$	$\mathrm{VaR}_{99\%}$	(4.725; 30.20)	(4.589; 43.64)	(4.490; 53.50)
	$\mathrm{VaR}_{99.5\%}$	(4.800; 40.74)	(4.705; 59.80)	(4.634; 73.77)
	$\mathrm{VaR}_{95\%}$	(4.372; 16.94)	(4.037; 23.30)	(3.791; 27.96)
$d = +\infty$	$\mathrm{VaR}_{99\%}$	(4.725; 32.25)	(4.578; 46.77)	(4.470; 57.41)
	$\mathrm{VaR}_{99.5\%}$	(4.806; 43.63)	(4.702; 64.22)	(4.634; 77.72)

(a, b)		$n = 100$		
		$\varrho = 0$	$\varrho = 0.15$	$\varrho = 0.3$
	$\mathrm{VaR}_{95\%}$	(47.96; 84.74)	(42.48; 188.9)	(39.61; 243.4)
$d = 10{,}000$	$\mathrm{VaR}_{99\%}$	(48.99; 129.6)	(46.59; 367.3)	(45.33; 491.7)
	$\mathrm{VaR}_{99.5\%}$	(49.23; 162.9)	(47.54; 500.0)	(46.65; 676.3)
	$\mathrm{VaR}_{95\%}$	(48.01; 87.75)	(42.09; 200.3)	(38.99; 259.2)
$d = +\infty$	$\mathrm{VaR}_{99\%}$	(49.13; 136.2)	(46.53; 393.1)	(45.18; 527.4)
	$\mathrm{VaR}_{99.5\%}$	(49.39; 172.2)	(47.56; 536.4)	(46.60; 726.9)

Panel C: Unconstrained bounds independent of ϱ

(A, B)		$n = 10$	$n = 100$
	$\mathrm{VaR}_{95\%}$,	(3.646; 30.33)	(36.46; 303.3)
$d = 10{,}000$	$\mathrm{VaR}_{99\%}$	(4.447; 57.76)	(44.47; 577.6)
	$\mathrm{VaR}_{99.5\%}$	(4.633; 74.11)	(46.33; 741.1)
	$\mathrm{VaR}_{95\%}$	(3.647; 30.72)	(36.47; 307.2)
$d = +\infty$	$\mathrm{VaR}_{99\%}$	(4.448; 59.62)	(44.48; 596.2)
	$\mathrm{VaR}_{99.5\%}$	(4.635; 77.72)	(46.35; 777.2)

Table 4.1 Bounds on Value-at-Risk of sums of Pareto distributed risks ($\theta = 3$), d denotes the discretization level



(a) bounds for the variance, TVaR at 95 % and TVaR at 99 %, p dependence parameter; $p = 0$ $\sim$ strong negative dependence; $p = 1 \sim$ strong positive dependence

$n = 50$	VaR_α	$\mathrm{TVaR}_\alpha(S^c)$	$\mathrm{VaR}_\alpha\left(T_Z^+\right)$	$\mathrm{LTVaR}_\alpha(S^c)$	$\mathrm{VaR}_\alpha\left(T_Z^-\right)$	Δ
$p = 0.0$	157	378	266	68	149	62 %
$p = 0.2$	158	354	267	69	151	59 %
$p = 0.4$	164	340	271	70	157	58 %
$p = 0.5$	169	338	274	70	161	58 %
$p = 0.6$	175	340	278	70	167	59 %
$p = 0.8$	189	354	289	69	181	62 %
$p = 1.0$	205	378	300	68	198	67 %

(b) upper and lower VaR bounds, $\theta_2 = 4$, VaR_α in dependence on p

Figure 4.2 $p \approx 0 \Rightarrow$ strong negative dependence, $p \approx 1 \Rightarrow$ strong positive dependence

Then the common risk factor $Z = (1-U)^{-1/3} - 1$ *is Pareto(3) distributed and thus dominates the individual risk factors. For* p *small close to* 0 *the common risk factor introduces a strong form of negative dependence, for* p *large close to* 1 *it induces a strong form of positive dependence in the factor model. As consequence we expect for* p *small a strong improvement of the upper risk bound and for* p *large a strong improvement of the lower risk bounds. This effect is clearly shown in the following figure (Figure 4.2).*

Several examples of a similar kind show a similar improvement effect which in the above example is of about 60 % of the range. This indicates that this kind of structural information can be very useful in real examples.

4.3 Positive and negative dependence information, subgroup structure models and stochastic ordering

Under one- or two-sided positive or negative dependence information improvements of the standard bounds for tail risks can be given. In Bignozzi, Puccetti, and Rüschendorf (2015) and Puccetti et al. (2017) this method has been applied to models having a subgroup structure with independent subgroups leading in concrete applications in insurance and in concrete risk portfolios in a banking context to considerable reduction of the risk bounds. This comparison has been extended in Rüschendorf and Witting (2017) to a systematic study of a combination of ordering within the subgroups with ordering conditions of the copulas between the subgroups without assuming necessarily the independence of the subgroups.

In some recent papers of Ansari and Rüschendorf (2021a,b, 2024) worst case portfolios have been identified for several classes of elliptical models, of partially specified risk factor models and for various classes of general factor models. Thus in these classes of models the corresponding generalized Hoeffding–Fréchet functionals resp. mass transportation problems are solved by means of newly developed stochastic ordering methods. Of particular importance in this context is the construction of corresponding mass-transfers and the use of mass-transfer theory as in Müller (2013).

For more details and concrete classes of examples we refer to Rüschendorf et al. (2024).

5 Conclusion

As described in the review above the mass transportation problem and the problem of generalized Hoeffding–Fréchet functionals have from the beginning on a different motivation and arise from different historical sources. For the class of generalized Hoeffding–Fréchet functionals in recent years a great amount of modifications of the underlying models induced by various forms of dependence constraints or of structural information have been worked out. An extended presentation of this kind of results is given in the recent book by Rüschendorf et al. (2024). From the point of view and motivation by mass transportation also some classes of constraints have been dealt with in the literature from the beginnings on.

A class of examples given by martingale optimal mass transportation lying somewhere between these two kinds of motivations has found a lot of interest in recent publications. There are some fruitful connections as shown above and related developments in these two areas like the basic characterization of L^2 – or more general of c-optimal couplings. The determination of worst case dependence structures w.r.t. convex law invariant risk measures can also be seen as solution to a class of interesting non-linear mass transportation problems. It has been treated and solved in connection with generalized Hoeffding–Fréchet functionals.

There is also a common development concerning the consideration of practically relevant statistical robustness models as given by constraints on the class of dependence structures or given by additional structural information. These kind of constraints lead to new classes of relevant mass transportation problems, many of them still waiting for being solved. For several of the considered classes of dependence or structural constraints new stochastic ordering methods, as based on the development of suitable mass transfer theory, play a key role for the solution of the generalized Hoeffding–Fréchet functionals resp. for the related mass transportation problems.

References

L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces. Lectures given at the C. I. M. -C. I. M. E. joint Euro-summer school, Madeira, Funchal, Portugal, July 3–9, 2000*, pages 1–52. Berlin: Springer, 2003.

L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications. Lectures given at the C. I. M. E. summer school, Martina Franca, Italy, September 2–8, 2001*, pages 123–160. Berlin: Springer, 2003.

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Basel: Birkhäuser, 2005.

J. Ansari and L. Rüschendorf. Ordering results for elliptical distributions with applications to risk bounds. *Journal of Multivariate Analysis*, 182:14, 2021a. Id/No 104709.

J. Ansari and L. Rüschendorf. Sklar's theorem, copula products, and ordering results in factor models. *Depend. Model.*, 9:267–306, 2021b.

J. Ansari and L. Rüschendorf. Supermodular and directionally convex comparison results for general factor models. *Journal of Multivariate Analysis*, 201:20, 2024. Id/No 105264.

E. R. Barnes and A. J. Hoffman. On transportation problems with upper bounds on leading rectangles. *Siam Journal on Algebraic and Discrete Methods*, 6:487–496, 1985. URL `https://api.semanticscholar.org/CorpusID:120871163`.

M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices – a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

C. Bernard, L. Rüschendorf, and S. Vanduffel. Value-at-Risk bounds with variance constraints. *Journal of Risk and Insurance*, 84(3):923–959, 2017a.

C. Bernard, L. Rüschendorf, S. Vanduffel, and R. Wang. Risk bounds for factor models. *Finance and Stochastics*, 21(3):631–659, 2017b.

C. Bernard, L. Rüschendorf, S. Vanduffel, and J. Yao. How robust is the value-at-risk of credit risk portfolios? *The European Journal of Finance*, 23(6):507–534, 2017c.

E. Bernton, P. Ghosal, and M. Nutz. Entropic optimal transport: geometry and large deviations. *Duke Math. J.*, 171(16):3363–3400, 2022.

V. Bignozzi, G. Puccetti, and L. Rüschendorf. Reducing model risk via positive and negative dependence assumptions. *Insur. Math. Econ.*, 61:17–26, 2015.

Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

C. Burgert and L. Rüschendorf. Consistent risk measures for portfolio vectors. *Insur. Math. Econ.*, 38:289–297, 2006.

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

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

J. A. Cuesta-Albertos, C. Matrán, S. T. Rachev, and L. Rüschendorf. Mass transportation problems in probability theory. *Mathematical Scientist*, 21:37–72, 1996.

G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppie. *Ann. Scuola Normale Superiore Di Pisa, Cl. Sci.*, 3:33–74, 1956.

A. de Acosta. Invariance principles in probability for triangular arrays of B-valued random vectors and some applications. *Annals of Probability*, 10:346–373, 1982.

A. Dong, A. Stephanovitch, and T. T. Georgiou. Monge–kantorovich optimal transport through constrictions and flow-rate constraints. *Automatica*, 160:111448, 2024. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2023.111448. URL `https://www.sciencedirect.com/science/article/pii/S0005109823006155`.

R. M. Dudley. *Probabilities and metrics. Convergence of laws on metric spaces, with a view to statistical testing*, volume 45 of *Lect. Notes Ser., Aarhus Univ.* Matematisk Institut, Aarhus Universitet, Aarhus, 1976.

I. Ekren and H. M. Soner. Constrained optimal transport. *Archive for Rational Mechanics and Analysis*, 227(3):929–965, 2018.

P. Embrechts, G. Puccetti, and L. Rüschendorf. Model uncertainty and VaR aggregation. *Journal of Banking Finance*, 37(8):2750–2764, 2013.

X. Fernique. Sur le théorème de Kantorovich–Rubinstein dans les espaces polonais. Seminaire de probabilites XV, Univ. Strasbourg 1979/80, 1981.

M. Fréchet. *Les probabilités associées à un système d'événements compatibles et dépendants; I. Événements en nombre fini fixe*, volume 859 of *Actual. Sci. Industr.* Paris: Hermann & Cie., 1940.

M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Series 3*, 14:53–77, 1951.

N. Gaffke and L. Rüschendorf. On a class of extremal problems in statistics. *Math. Operationsforschung Statist.*, 12:123–135, 1981.

W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

C. Gini. Di una misura della dissomiglianza tra due gruppi di quantita e delle sue applicazioni allo studio della relazioni statistiche. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, A.A.*, 74:185–213, 1914.

W. Hoeffding. *Maßstabinvariante Korrelationstheorie.* Ph.D.Thesis, Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin 5, 181–233 (1940) and Berlin: Dissertation, 1940.

W. Hoeffding. A combinatorial central limit theorem. *Annals of Mathematical Statistics*, 22:558–566, 1951.

L. V. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS, n. Ser.*, 37:199–201, 1942.

L. V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk*, 3:225–226, 1948. In Russian.

L. V. Kantorovich and G. S. Rubinstein. On a function space in certain extremal problems. *Dokl. Akad. Nauk USSR*, 115:1058–1061, 1957.

H. G. Kellerer. Duality theorems and probability metrics. In *Proc. 7th Brasov Conf. (Bucuresti, Romania, 1982)*, pages 211–220, 1984a.

H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, 67:399–432, 1984b.

M. Knott and C. S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43:39–49, 1984.

M. Knott and C. S. Smith. Note on the optimal transportation of distributions. *Journal of Optimization Theory and Applications*, 52:323–329, 1987.

M. Knott and C. S. Smith. On a generalization of cyclic monotonicity and distances among random vectors. *Linear Algebra and its Applications*, 199:363–371, 1994.

J. Korman and R. J. McCann. Optimal transportation with capacity constraints. *Transactions of the American Mathematical Society*, 367:1501–1521, 2015. URL `https://api.semanticscholar.org/CorpusID:7233885`.

V. L. Levin. On the problem of mass transfer. *Soviet Math. Doklady*, 16:1349–1353, 1975.

V. L. Levin and A. A. Milyutin. The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems. *Trans Russian Math. Surveys*, 34:1–78, 1979.

M. Moscadelli. The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee. Available at `http://ideas.repec.org/p/bdi/wptemi/td_517_04.html`, July 2004.

A. Müller. Duality theory and transfers for stochastic order relations. In *Stochastic Orders in Reliability and Risk. In honor of Professor Moshe Shaked. Based on the talks presented at the international workshop on stochastic orders in reliability and risk management, SORR2011, Xiamen, China, June 27–29, 2011*, pages 41–57. New York, NY: Springer, 2013.

L. Nenna. *Numerical Methods for Multi-Marginal Optimal Transportation.* Theses, Université Paris Sciences et Lettres, 223 pg., 2016.

M. Nutz. Introduction to Entropic Optimal Transport. *Lecture Notes Columbia Univ., 76 pg.*, 2021.

G. Peyré and M. Cuturi. Computational optimal transport. With applications to data sciences. *Found. Trends Mach. Learn.*, 11(5-6):1–262, 2018.

G. Peyré and M. Cuturi. Editorial. IMA IAI – Information and Inference special issue on optimal transport in data sciences. *Inf. Inference*, 8(4):655–656, 2019.

G. Puccetti and L. Rüschendorf. Bounds for joint portfolios of dependent risks. *Statistics & Risk Modeling*, 29:107–132, 2012a.

G. Puccetti and L. Rüschendorf. Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236: 1833–1840, 2012b.

G. Puccetti, B. Wang, and R. Wang. Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insur. Math. Econ.*, 53(3):821–828, 2013.

G. Puccetti, L. Rüschendorf, D. Small, and S. Vanduffel. Reduction of Value-at-Risk bounds via independence and variance information. *Scand. Actuar. J.*, 2017(3):245–266, 2017.

S. T. Rachev. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and Its Applications*, 29:647–676, 1985.

S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, 1991.

S. T. Rachev and I. Olkin. Mass transportation problems with capacity constraints. *Journal of Applied Probability*, 36:433–445, 1999.

S. T. Rachev and L. Rüschendorf. Solution of some transportation problems with relaxed or additional constraints. *SIAM Control and Optimization*, 32:673–689, 1994.

S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol. I: Theorey.* Probability and its Applications. Springer, 1998a.

S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol. II: Applications.* Probability and its Applications. Springer, 1998b.

D. Ramachandran and L. Rüschendorf. A general duality theorem for marginal problems. *Probability Theory and Related Fields*, 101:311–319, 1995.

D. Ramachandran and L. Rüschendorf. On the validity of the Monge–Kantorovich duality theorem. *Theory of Probability and Its Applications*, 45:350–356, 2000.

L. Rüschendorf. *Vergleich von Zufallsvariablen bzgl. integralinduzierter Halbordnungen.* Habilitationsschrift, Mathematisch-Naturwissenschaftliche Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen. 133 S., 1979.

L. Rüschendorf. Inequalities for the expectation of Δ-monotone functions. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, 54:341–354, 1980.

L. Rüschendorf. Sharpness of Fréchet bounds. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, 57:293–302, 1981.

L. Rüschendorf. Solution of a statistical optimization problem by rearrangement methods. *Metrika*, 30:55–62, 1983.

L. Rüschendorf. Bounds for distributions with multivariate marginals. In K. Mosler and M. Scarsini, editors, *Stochastic Order and Decision under Risk*, volume 19, pages 285–310. IMS Lecture Notes, 1991a.

L. Rüschendorf. Fréchet-bounds and their applications. In G. Dall'Aglio, S. Kotz, and G. Salinetti, editors, *Advances in Probability Measure with Given Marginals (Rome, Italy, 1990)*, volume 67 of *Math. Appl.*, pages 151–187. Kluwer, 1991b.

L. Rüschendorf. Developments on Fréchet bounds. In *Proceedings of Distributions with Fixed Marginals and Related Topics*, volume 28, pages 273–296. IMS Lecture Notes Monograph Series, 1996.

L. Rüschendorf. Law invariant convex risk measures for portfolio vectors. *Statistics & Decisions*, 24:97–108, 2006.

L. Rüschendorf. Monge–Kantorovich transportation problem and optimal couplings. *Jahresbericht der DMV*, 109:113–137, 2007.

L. Rüschendorf. Risk bounds, worst case dependence, and optimal claims and contracts. In *Proceedings of AFMathConf 2012: Actuarial and Financial Mathematics Conference*, pages 23–36, 2012.

L. Rüschendorf. *Mathematical Risk Analysis. Dependence, risk bounds, optimal allocations and portfolios.* Springer Ser. Oper. Res. Financ. Eng. Berlin: Springer, 2013.

L. Rüschendorf. Risk bounds with additional information on functionals of the risk vector. *Depend. Model.*, 6:102–113, 2018.

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

L. Rüschendorf and J. Witting. VaR bounds in models with partial dependence information on subgroups. *Depend. Model.*, 5:59–74, 2017.

L. Rüschendorf, S. Vanduffel, and C. Bernard. *Model Risk Management. Risk bounds under uncertainty.* Cambridge: Cambridge University Press, 2024.

T. Salvemini. Nuovi procedimenti di calcolo degli indici di dissomiglianza e di connessione. *Statistica, Milano*, 9:3–26, 1949.

F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, volume 87 of *Prog. Nonlinear Differ. Equ. Appl.* Cham: Birkhäuser/Springer, 2015.

W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge–Kantorovich problem. *Proc. Am. Math. Soc.*, 137(2):519–529, 2009.

A. Szulga. On the Wasserstein metric. Information theory, statistical decision functions, random processes; Trans. 8th Prague Conf., Vol. B, Prague 1978, 1978.

A. H. Tchen. Inequalities for distributions with given marginals. *Annals of Probability*, 8: 814–827, 1980.

C. Villani. *Topics in Optimal Transportation.* Number 58 in *Graduate Studies in Mathematics.* American Mathematical Society, Providence, 2003.

B. Wang and R. Wang. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102:1344–1360, 2011.

M. Whitt. Bivariate distributions with given marginals. *Annals of Statistics*, 4:1280–1289, 1976.

V. M. Zolotarev. Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory of Probability and Its Applications*, 21:721–737, 1976.

Department of Mathematical Stochastics
University of Freiburg
Ernst-Zermelo-Str. 1
D – 79104 Freiburg (Germany)
ruschen@stochastik.uni-freiburg.de