Research Article **Open Access**

Michael Preischl*

Bounds on integrals with respect to multivariate copulas

DOI 10.1515/demo-2016-0016
Received June 27, 2016; accepted November 2, 2016

Abstract: In this paper, we present a method to obtain upper and lower bounds on integrals with respect to copulas by solving the corresponding assignment problems (AP's). In their 2014 paper, Hofer and Iacó proposed this approach for two dimensions and stated the generalization to arbitrary dimensons as an open problem. We will clarify the connection between copulas and AP's and thus find an extension to the multidimensional case. Furthermore, we provide convergence statements and, as applications, we consider three dimensional dependence measures as well as an example from finance.

Keywords: Copulas, linear assignment problems, dependence measure, credit risk

MSC: 91G80, 90C08, 90C99

1 Introduction

A multidimensional distribution function with uniform marginals is called a *copula*. In contrast to the simplified approach of quantifying risk and dependence by single numbers like Spearman's ρ or Kendall's τ, modeling with copulas makes it possible to describe and encapsulate the entire dependence structure between random variables. On the other hand, an obvious downside of copulas is that, unlike simple concordance measures, they are often hard to treat analytically, especially in dimensions higher than two. Hence, instead of modeling the actual dependence structure, it is naturally interesting to ask for a "worst case" respectively a "best case" behaviour. In this paper, we propose a flexible method to approximate those extremal cases.

Assume that we are given a d-dimensional random vector $(X_1, \dots, X_d)$ and a function $f : \mathbb{R}^d \to \mathbb{R}$ that describes the quantity associated with $(X_1, \dots, X_d)$ which we wish to optimize. We further assume that the marginal distributions of $X_1, \dots, X_d$ are known and the dependence structure (i.e. the common distribution function) is completely unknown. This assumption is called *dependence uncertainty* and it is widely used in applications, mainly because compared to finding the dependence structure, information about the marginal laws can be relatively easily obtained.

By Sklar's Theorem it is always possible to reduce this maximization respectively minimization to a similar problem involving uniformly distributed random variables $X_1, \dots, X_d$ [14]. Therefore, we are justified in restricting our focus to finding copulas $C_{\min}$ and $C_{\max}$ such that

$$\int_{[0,1]^d} f(x_1, \dots, x_d) dC_{\min} \leq \int_{[0,1]^d} f(x_1, \dots, x_d) dC \tag{1}$$

and

$$\int_{[0,1]^d} f(x_1, \dots, x_d) dC \leq \int_{[0,1]^d} f(x_1, \dots, x_d) dC_{\max} \tag{2}$$

for all d-dimensional copulas C.

***Corresponding Author: Michael Preischl:** Graz University of Technology, Institute of Analysis and Number Theory, Kopernikusgasse 24/II, 8010 Graz, E-mail: preischl@math.tugraz.at

Equations (1) and (2) are special cases of the *Monge-Kantorovich* problem. This problem is very well studied in the case $d = 2$, however, due to its complexity, most analytic approaches to the Monge-Kantorovich problem in higher dimensions are restricted to particular situations. For example, one of Rüschendorf's many contributions to this field considered the case where f is a Δ-monotone function [18].
A more flexible, numerical take on this optimization problem that had a significant impact in recent years is the *rearrangement algorithm*, introduced by Puccetti and Rüschendorf [15]. This algorithm is impressively efficient in approximating the desired bounds even in high dimensions and thus suffices for most real world applications. The price for this is that it only works when f is a supermodular function and that convergence is not guaranteed. However, the cases where the algorithm does not converge are quite pathological and can be circumvented in practice. For a detailed description of how the rearrangement algorithm can be used to tackle optimization problems and also examples of non-convergence, see [16].
In two dimensions, Hofer and Iacó [7] combined the spirit of optimization theory with rigorous structural considerations and developed an algorithm that converges to the minimal respectively maximal values of equations (1) and (2) for any continuous function f. Their method connects optimality with a particular class of copulas, the Shuffles of *Min*. We state their main results in Theorem 2.2 in section 2.

2 Mathematical Foundations

As stated previously, a d-copula is a d-dimensional distribution function on $[0,1]^d$ with uniform marginals. Every d-copula C defines a measure μ_C on $([0,1]^d, \mathcal{B}([0,1]^d))$ which is d-fold stochastic, i.e., it fulfills

$$\mu_C(\underbrace{[0,1]\times[0,1]\times\cdots\times[0,1]}_{i-1 \text{ times}}\times A\times\underbrace{[0,1]\times[0,1]\times\cdots\times[0,1]}_{d-i \text{ times}}) = \lambda(A), \tag{3}$$

for all $i = 1,\dots,d$ and any Borel set $A \subset [0,1]$. Conversely, every d-fold stochastic measure on $([0,1]^d, \mathcal{B}([0,1]^d))$ defines a copula. We write $\mathcal{C}^d$ for the set of all d-copulas.

We already mentioned that there is a class of copulas, which is closely related to assignment problems, the Shuffles of *Min* (or Shuffles of *M*). As the name suggests, these are obtained by a suitable rearrangement of the probability mass of the upper Fréchet-Hoeffding bound, or *Min* copula, $M(x_1,\dots,x_d) := \min(x_1,\dots,x_d)$. In two dimensions, C is a Shuffle of *Min* parametrized by $n \in \mathbb{N}$, a permutation π on $\{1,\dots,n\}$ and a function $\omega : \{1,\dots,n\} \to \{-1,1\}$ if C distributes the mass $\frac{1}{n}$ uniformly spread along the diagonal respectively antidiagonal of $[\frac{i-1}{n}, \frac{i}{n}]\times[\frac{\pi(i)-1}{n}, \frac{\pi(i)}{n}]$ whenever $\omega(i) = 1$ respectively $\omega(i) = -1$. The original, two dimensional definition is slightly more general and goes back to [12]. In higher dimensions, there are several versions of Shuffles of *Min* (see e.g. [5] for a discussion). A basic property of these Shuffles is that they are dense in the set of all copulas with respect to weak convergence. For more details and a survey of different metrics also see [5].

A concept which has proven very useful when solving two dimensional Monge-Kantorovich problems is that of *c-cyclical monotonicity*. It is a famous result in optimal transport theory that, under mild assumptions on c, a probability measure μ is optimal for the two dimensional Monge-Kantorovich problem if and only if it is concentrated on a c-cyclically monotone set. This optimality result follows from a dual formulation of the problem, for which we refer to the book of Villani [20].

Finding a similar statement for dimensions higher than two has been an open problem for many years. Griessler and Beiglböck ([1] and [6]) recently generalized c-cyclical monotonicity to arbitrary dimensions:

Definition 2.1 ([1] and [6]). *Let $X_1,\dots,X_d$ be Polish spaces and define $E := X_1\times\cdots\times X_d$. Let $c : E \to \mathbb{R}$ be Borel measurable. A set $\Gamma \subset E$ is called c-cyclically monotone if it fulfills one of the following conditions:*

(i) For any N and any points $(x_1^{(1)}, \dots, x_d^{(1)}), \dots, (x_1^{(N)}, \dots, x_d^{(N)}) \in \Gamma$ and permutations $\sigma_2, \dots, \sigma_d : \{1, \dots, N\} \to \{1, \dots, N\}$, one has

$$\sum_{i=1}^{N} c(x_1^{(i)}, \dots, x_d^{(i)}) \leq \sum_{i=1}^{N} c(x_1^{(i)}, x_2^{(\sigma_2(i))}, \dots, x_d^{(\sigma_d(i))}).$$

(ii) Any finite measure α concentrated on finitely many points in Γ is a (with respect to c) cost-minimizing transport plan between its marginals; i.e. if α' has the same marginals as α, then

$$\int c d\alpha \leq \int c d\alpha'.$$

They were also able to show that for any measurable cost function c, a measure μ which is optimal for the multidimensional Monge-Kantorovich problem is always concentrated on some c-cyclically monotone set. Griessler [6] recently showed the converse statement under more assumptions on c: If the cost function c is continuous and bounded by a sum of integrable functions, any measure which is concentrated on a c-cyclically monotone set is an optimal solution to the multidimensional Monge-Kantorovich problem.

Next we give a short overview of assignment problems. The mathematical formulation of a Linear Sum Assignment Problem is the following:

$$\min_{x \in \mathbb{R}^{n \times n}} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ij} \tag{4}$$

subject to

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \in \{1, \dots, n\}, \tag{5}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \forall j \in \{1, \dots, n\}, \tag{6}$$

$$x_{ij} \in \{0, 1\}. \tag{7}$$

The matrix (a_{ij}) is also called the *objective function* and the set of all $x \in \mathbb{R}^{n \times n}$ which fulfill all the constraints is called the *feasible region*.
It is not hard to see that a Linear Sum Assignment Problem can equivalently be written in the form

$$\min_{\pi \in S_n} \sum_{i=1}^{n} a_{i\pi(i)},$$

where S_n denotes the set of all permutations of $\{1, \dots, n\}$. Although the feasible region of this problem is actually n^2-dimensional, with n being the number of objects, we will refer to this version of the assignment problem as the "two dimensional assignment problem (2-AP)" since one can interpret this as matching two different kinds of objects. The assignment problem is, at least for the two dimensional case, very well-studied.

We are now ready to state the main result from [7] that connects Shuffles of *Min* and assignment problems to integrals with respect to copulas.

Theorem 2.2. *Let f be a continuous function on $[0, 1]^2$ and let the partition I^n for any n be given as*

$$I_{ij}^n := \left[\frac{i-1}{n}, \frac{i}{n}\right) \times \left[\frac{j-1}{n}, \frac{j}{n}\right) \quad \text{for } i, j = 1, \dots, n.$$

Then define

$$f_n^{max}(x_1, x_2) = a_{ij} := \max_{(x_1, x_2) \in I_{ij}^n} f(x, y) \qquad \forall (x_1, x_2) \in I_{ij}^n.$$

Now a copula C_n^{max} which fulfills

$$\int_{[0,1]^2} f_n^{max}(x_1, x_2) dC_n^{max} = \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) \tag{8}$$

is given as a shuffle of Min with parameters $(n, \pi^\star, 1)$ *where* $\pi^\star$ *is the permutation which solves the assignment problem*

$$\max_{\pi \in S_n} \sum_{i=1}^{n} a_{i\pi(i)}.$$

Moreover, the maximal value of (8) *is given as*

$$\max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} a_{i\pi^\star(i)}$$

and it holds

$$\lim_{n\to\infty} \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) = \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f(x_1, x_2) dC(x_1, x_2).$$

Furthermore, Iacó, Thonhauser and Tichy [8] showed that the sequence of maximizers C_n^{max} converges, at least along some subsequence, to a maximizer $C_{\max}$ of the problem

$$\max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f(x_1, x_2) dC(x_1, x_2).$$

3 Main Results

As we will see, some structural analogies are destroyed in the d-dimensional case, which is why a direct application of the method from [7] is not possible. For our main result, an extension of Theorem 2.2 to arbitrary dimensions along with a similar convergence result as in [8], we start by introducing the concept of a multidimensional assignment problem.

Define the index sets $\mathcal{I} := \{1, \dots, n\}^d$ and $\mathcal{I}_m^k := \{(i_1, \dots, i_d) \in \mathcal{I} : i_k = m\}$. The (axial) d-dimensional assignment problem (d-AP) on n items with objective function (a_i) is given as follows:

$$\min_{x \in \mathbb{R}^{n^d}} \sum_{i \in \mathcal{I}} a_i x_i \tag{9}$$

subject to

$$\sum_{i \in \mathcal{I}_m^k} x_i = 1, \qquad \forall m \in \{1, \dots, n\}, \quad \forall k \in \{1, \dots, d\}, \tag{10}$$

$$x_i \in \{0, 1\}. \tag{11}$$

Again, "d-dimensional" is meant with respect to the number of different object types. The feasible region in this case would actually be n^d-dimensional.

Theorem 3.1. *Let* n *be a positive integer and* $f : [0,1]^d \to \mathbb{R}$ *be constant on the cubes* I_i^n *with*

$$I_i^n := \left[\frac{i_1 - 1}{n}, \frac{i_1}{n}\right) \times \cdots \times \left[\frac{i_d - 1}{n}, \frac{i_d}{n}\right),$$

for $i = (i_1, \dots, i_d) \in \mathcal{I}$. *So* $f(x) = a_i$ *for all* $x \in I_i^n$.
Then a copula $C_{\min}$ *which fulfills*

$$\int_{[0,1]^d} f(x)\, dC_{\min} = \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} f(x)\, dC(x), \tag{12}$$

distributes uniformly on each square of type I_i^n the probability mass equal to $x_i^\star/n$, where $(x_i^\star)_{i\in\mathcal{I}}$ is an optimal solution to the relaxed d-AP with respect to the objective function $(a_i)_{i\in\mathcal{I}}$. By this construction, $C_{\min}$ is a so-called checkerboard copula ([4], [13]).

Here "relaxed" means that we are considering the continuous relaxation of the axial d-AP, i.e., we replace the integer constraint (11) by $x_i \geq 0$ for all i. Even though a seemingly subtle difference, this change yields an entirely different problem from the perspective of optimization theory.

Theorem 3.1 holds for any dimension d and indeed for the case $d = 2$ we get precisely the method proposed in [7]. This follows from Birkhoff's theorem, which states that the two dimensional assignment problem is identical to its continuous relaxation. However, Birkhoff's theorem does not hold for any dimension greater than two, which is why it is not possible to restrict the solution space to Shuffles of *Min* in higher dimensions. As a result, the optimizing copula that comes from the assignment problem will, in general, not be given as a Shuffle of *Min*. We refer to [3] for further details about assignment problems.

That means, for dimensions greater than two, the maximizer we find via this procedure will not have the nicely parametrized form that made Shuffles of *Min* so appealing. On the other hand, working with the relaxed assignment problem instead of the integer problem brings great advantages concerning computability. While the classical integer assignment problems are $\mathcal{NP}$ hard, their continuous relaxations lie in $\mathcal{P}$. Here the computational complexity is with respect to the number of objects that should be assigned, or in the context of copulas, the coarseness of the partition of $[0,1]^d$.

Proof of Theorem 3.1. By definition, the value of (12) is given as

$$\int_{[0,1]^d} f(x)dC(x) = \sum_{i\in\mathcal{I}} a_i \mu_C(I_i^n).$$

Notice that the value of (12) does not depend on how the copula C distributes mass inside of each cube I_i^n, but only on how much mass is placed on each I_i^n. Hence, we can write $x_i := \mu_{C_{\min}}(I_i^n)$, with $i \in \mathcal{I}$ and are left with the following optimization problem:

$$\min \sum_{i\in\mathcal{I}} a_i x_i.$$

However, we still must encode constraints that ensure that the mass distribution x_i actually yields a copula. We recall that there is a one to one correspondence between d-copulas and d-fold stochastic measures, so it suffices to ensure that the measure $\mu_{C_{\min}}$ is d-fold stochastic. Since we already noted that the value of (12) is independent of the distribution inside the cubes I_i, we can assume that $\mu_{C_{\min}}$ distributes the mass inside of each cube I_i uniformly. The d-fold stochastic measures which distribute the mass x_i uniformly inside of the cube I_i for each $i \in \mathcal{I}$ are given by the equations

$$\sum_{i\in\mathcal{I}_m^k} x_i = \frac{1}{n}, \qquad \forall m \in \{1,\dots,n\}, \quad \forall k \in \{1,\dots,d\}. \tag{13}$$

This can be seen as follows: It is elementary that each d-fold stochastic measure satisfies the conditions (13). So let C fulfill (13) and let $0 \leq a < b \leq 1$. Now look for $1 \leq i_- \leq i^+ \leq n$ such that $a \in [\frac{i_- -1}{n}, \frac{i_-}{n}]$ and $b \in [\frac{i^+ -1}{n}, \frac{i^+}{n}]$. Without loss of generality let us consider the first coordinate, it holds

$$\begin{aligned}
\mu_C([a,b]\times[0,1]\times\cdots\times[0,1]) &= \mu_C\left(\left[a,\frac{i_-}{n}\right]\times[0,1]\times\cdots\times[0,1]\right) + \sum_{i_1=i_-+1}^{i^+-1} \mu_C\left(\left[\frac{i_1-1}{n},\frac{i_1}{n}\right]\times[0,1]\times\cdots\times[0,1]\right) \\
&\quad + \mu_C\left(\left[\frac{i^+-1}{n},b\right]\times[0,1]\times\cdots\times[0,1]\right) \\
&= \frac{i_-}{n} - a + \left(\sum_{i_1=i_-+1}^{i^+-1} \sum_{j\in\mathcal{I}_{i_1}^1} x_j\right) + b - \frac{i^+-1}{n} \\
&= b - a = \lambda([a,b]).
\end{aligned}$$

By standard arguments of measure theory, we can extend this result from intervals to arbitrary measurable sets A.
Hence the measure $C_{\min}$ is d-fold stochastic if and only if the constraints (13) are fulfilled. But those are exactly the constraints (10) from the d-AP with the right hand side $\frac{1}{n}$ instead of 1. Since scaling the right hand side of a linear optimization problem results in a similar scaling of the optimal solution, the optimal probability mass distribution (x_i) is given as $(x_i) = \frac{1}{n}(x_i^\star)$ with $(x_i^\star)$ being the optimal solution to the general d-AP with objective function (a_i). □

With the necessary adjustments, Theorem 3.1 is equally valid for a maximization instead of a minimization.

Also in the multidimensional case, it is possible to approximate integrals over continuous functions by a sequence of integrals over functions that are piecewise constant.

Theorem 3.2. *Let f be continuous on $[0,1]^d$ and bounded by a sum of integrable functions and let the sets I_i^n for $i \in \mathcal{I}$ be given as before. Then, set*

$$f_n^{max}(x) := \max_{y \in I_i^n} f(y) \qquad \forall x \in I_i^n,$$
$$f_n^{min}(x) := \min_{y \in I_i^n} f(y) \qquad \forall x \in I_i^n.$$

Now denote by C_n^{max} and C_n^{min} copulas which fulfill

$$\int_{[0,1]^d} f_n^{max}(x)\, dC_n^{max}(x) = \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} f_n^{max}(x)\, dC(x) \tag{14}$$

and

$$\int_{[0,1]^d} f_n^{min}(x)\, dC_n^{min}(x) = \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} f_n^{min}(x)\, dC(x).$$

Then

$$\lim_{n\to\infty} \int_{[0,1]^d} f_n^{min}(x) dC_n^{min}(x) = \lim_{n\to\infty} \int_{[0,1]^d} f_n^{max}(x) dC_n^{max}(x) = \inf_{C \in \mathcal{C}^d} \int_{[0,1]^d} f(x) dC(x).$$

Furthermore, the sequences of minimizers C_n^{max} and C_n^{min} converge, at least along some subsequence, to a minimizer $C_{\min}$ of the problem

$$\min_{C \in \mathcal{C}_d} \int_{[0,1]^d} f(x) dC(x).$$

Proof. We show directly that

$$\lim_{n\to\infty} \int_{[0,1]^d} f_n^{min}(x) dC_n^{min}(x) = \lim_{n\to\infty} \int_{[0,1]^d} f_n^{max}(x) dC_n^{max}(x).$$

Let $\varepsilon > 0$. Since f is continuous, we may choose n large enough such that

$$|f_n^{min}(x) - f_n^{max}(x)| < \varepsilon \qquad \forall x \in [0,1]^d.$$

Furthermore, f_n^{min} is piecewise constant, so we have

$$\int_{[0,1]^d} f_n^{min}(x) dC_n^{min}(x) = \sum_{i \in \mathcal{I}} a_i \mu_{C_n^{min}}(I_i),$$

with $a_i := \min_{x \in I_i^n} f(x)$. Hence

$$\int_{[0,1]^d} f_n^{max}(x)\, dC_n^{max}(x) < \int_{[0,1]^d} f_n^{min}(x) + \varepsilon\, dC_n^{min}(x) = \sum_{i \in \mathcal{J}} \mu_{C_n^{min}}(I_i)(a_i + \varepsilon).$$

So we have

$$\left| \int_{[0,1]^d} f_n^{max}(x)\, dC_n^{max}(x) - \int_{[0,1]^d} f_n^{min}(x)\, dC_n^{min}(x) \right|$$
$$< \left| \sum_{i \in \mathcal{J}} \varepsilon\, \mu_{C_n^{min}}(I_i) \right| = \varepsilon.$$

For the proof that the sequence of optimizers converges to an optimizer for the continuous function, we start like in [8] by using Theorem 5.21 from [9] to deduce that C_n^{max} converges weakly along some subsequence to a copula $C_{\min}$.

Now according to Theorem 2.4 from [1], any measure which is an optimal solution to a transportation problem is necessarily concentrated on a c-cyclical monotone set. So since C_n^{max} is optimal for the transportation problem (14) it must be concentrated on a c-cyclical monotone set. Hence, for any $N \in \mathbb{N}$, the N-fold product measure $C_n^{max,\otimes N}$ is concentrated on the set $\mathcal{S}_n(N)$ of points $(x_1^{(1)}, \ldots, x_d^{(1)}), \ldots, (x_1^{(N)}, \ldots, x_d^{(N)})$ for which for any permutations $\sigma_2, \ldots, \sigma_d$

$$\sum_{i=1}^{N} f_n^{max}(x_1^{(i)}, \ldots, x_d^{(i)}) \le \sum_{i=1}^{N} f_n^{max}(x_1^{(i)}, x_2^{\sigma_2(i)}, \ldots, x_d^{\sigma_d(i)}).$$

Now fix $\varepsilon > 0$. Since f is continuous, we can choose n large enough such that $C_n^{max,\otimes N}$ is concentrated on the set $\mathcal{S}_\varepsilon(N)$ of points with

$$\sum_{i=1}^{N} f(x_1^{(i)}, \ldots, x_d^{(i)}) \le \sum_{i=1}^{N} f(x_1^{(i)}, x_2^{\sigma_2(i)}, \ldots, x_d^{\sigma_d(i)}) + \varepsilon.$$

Since f is continuous, the set $\mathcal{S}_\varepsilon(N)$ is closed. Therefore the limiting measure $C_{\min}^{\otimes N}$ is concentrated on $\mathcal{S}_\varepsilon(N)$ for all $\varepsilon > 0$. Now let $\varepsilon \to 0$ and we have that $C_{\min}^{\otimes N}$ is concentrated on a set of points with

$$\sum_{i=1}^{N} f(x_1^{(i)}, \ldots, x_d^{(i)}) \le \sum_{i=1}^{N} f(x_1^{(i)}, x_2^{\sigma_2(i)}, \ldots, x_d^{\sigma_d(i)}),$$

which means that $C_{\min}$ is concentrated on a c-cyclically monotone set. Since f is continuous and bounded by a sum of integrable functions, we can apply Theorem 1.2 from [6] to deduce that $C_{\min}$ is optimal. The proof for the sequence C_n^{min} is identical. □

In [7], the authors also give a convergence rate under the assumption of Lipschitz continuity. For completeness, we want to mention that this holds in very much the same way for the multidimensional setting.

Corollary 3.3. *Let the assumptions of Theorem 3.2 hold and, in addition, assume that f is Lipschitz continuous on $[0,1]^d$ with Lipschitz constant $L > 0$. Then*

$$\left| \int_{[0,1]^d} f_n^{max}(x) dC_n^{max}(x) - \int_{[0,1]^d} f_n^{min}(x) dC_n^{min}(x) \right| \le L \frac{\sqrt{d}}{n}.$$

Proof. by the Lipschitz continuity of f and the construction of f_n^{min} and f_n^{max}, we have that

$$|f_n^{min}(x) - f_n^{max}(x)| \le L \frac{\sqrt{d}}{n}.$$

Hence the proof follows by replacing ε with $L\frac{\sqrt{d}}{n}$ in the proof of Theorem 3.2. □

Remark 3.4. *The method presented generalizes the approach from [7] and furthermore facilitates the computation since the relaxed assignment problem is much easier to solve than the integer one. However, the solution vector is still n^d-dimensional. In practice this method can be applied on a standard laptop for $n^d \leq 10^7$ and since good approximations are already possible for $n \approx 25$, dimensions up to $d = 5$ can certainly be handled. So for most practical applications, the rearrangement algorithm by Puccetti and Rüschendorf [16] will, whenever applicable, still be the method of choice. The merit of our approach lies in the generality of the statement. It is not limited to supermodular functions but requires merely continuity and a notion of boundedness. Both of these assumptions might possibly be relaxed as research in the field of optimal transport progresses.*

Remark 3.5. *Considering only Shuffles of* Min *as minimizers respectively maximizers in equations* (1) *and* (2) *as Hofer and Iacó did in [7] is in some sense also the notion behind the rearrangement algorithm. In [17], Puccetti and Wang consider* rearrangements*, show how they are linked to copulas and illustrate that Shuffles of* Min *can be seen as a construction of particular rearrangements. As the authors of [17] mention, the fact that Shuffles of* Min *are dense in the set of copulas could then be used to approximate solutions to copula optimization problems arbitrarily well with Shuffles of* Min*. This would be a direct extension of the method from [7] to general dimensions, however, due to the complexity of the integer assignment problem, this approach is of very limited practical relevance. The rearrangement algorithm, on the other hand, cuts the solution space down to* oppositely ordered rearrangements*, resulting in the restriction to supermodular functions but also in an enormous increase of efficiency. For more details, we refer to [16].*

Remark 3.6. *Following the spirit of Bignozzi, Puccetti and Rüschendorf [2] or Lux and Papapantoleon [10], one might also consider including partial information about the distribution by simply adding suitable constraints to the linear program.*

4 Applications

4.1 Dependence Measures

A natural application for this technique is the approximation of upper and lower bounds on dependence measures. In the bivariate case, there are well-known and widely used measures such as *Spearman's ρ*, *Kendall's τ*, *Blomqvists β* and *Gini's γ*. See e.g. [11] for definitions.

We now focus on a multidimensional version of Spearman's ρ. Define

$$\rho(C) := \frac{d+1}{2^d - (d+1)} \left(2^d \int_{[0,1]^d} \Pi(u) dC(u) - 1 \right). \tag{15}$$

Here, Π denotes the independence copula, i.e. $\Pi(x_1, \dots, x_d) = \Pi_{i=1}^d x_i$. It is well-known that $\rho(C)$ is maximal when $C = M_d$, i.e. C is the *Min*-copula. It is also well-known that plugging in the lower Fréchet-Hoeffding bound (usually denoted by W_d) yields a lower bound on $\rho(C)$:

$$\rho(W_d) = \frac{2^d - (d+1)!}{d!(2^d - (d+1))} \leq \rho(C) \leq \rho(M_d), \qquad C \in \mathcal{C}^d, \quad d \geq 2.$$

However, since W_d is not a copula for $d \geq 2$, it is not a priori clear whether this lower bound is attained or not. Indeed, this has been stated as an open problem in [19]. In 2011, Wang and Wang [21] found an analytical solution to this long unresolved question. They give a formula to explicitly compute

$$\Lambda_d := \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} \Pi(u)\, dC(u)$$

for any $d \in \mathbb{N}$. Since the formula yields $\Lambda_3 = 5.4803 \times 10^{-2}$, it is straightforward that $\rho(W_3) = -\frac{2}{3}$ is not attained. We now want to give an alternative, numerical proof for this result. We chose this example because the fact that we know the exact analytical value will help us to validate the convergence of our method.

We use Theorem 3.2 to compute strict upper and lower bounds for Λ_3. In this case the monotonicity of Π facilitates finding the maximum respectively minimum functions in the algorithm as we simply have to evaluate Π at the vertices of the grid cubes that maximize respectively minimize the arguments. An approximated value is also obtained by evaluating Π in the center of each grid cube. Of course, this has to be adapted for other cost functions. Table 1 illustrates the results obtained using the method proposed in this paper with a grid of $n \in \{30, 40, 50, 60\}$ sections in each dimension[1] as well as the range the rearrangement algorithm (which uses the same discretization method as our algorithm) computes for a grid of 10^5 sections in each dimension [16].

Table 1: Approximation results for Λ_3. The values for $n = 30, \ldots, 60$ were calculated by the method proposed in this paper, whereas the values for $n = 10^5$ were calculated using the rearrangement algorithm [16].

n	lower bd.	approx. value	upper bd.
30	0.044414	0.054971	0.066341
40	0.046906	0.054897	0.063349
50	0.048433	0.054863	0.061587
60	0.049466	0.054844	0.060427
10^5 (RA)	0.054800	-	0.054807
analytic value	-	0.054803	-

Note that the range here is not a confidence interval but actually consists of deterministic upper and lower bounds on the true value. Already the lower bound for $n = 30$ suffices to prove that $\rho(C) > -\frac{2}{3} = \rho(W_3)$ for all copulas C. As can be seen, the rearrangement algorithm yields a more precise approximation for the same problem and that in considerably less time, even for higher dimensions d (more details can be found in [16]). However, since the lower bound computed by the rearrangement algorithm does not always have to be satisfied, we see the merit of our method in providing rigorous numerics for the fact that the lower bound $-\frac{2}{3}$ for $\rho(C)$ is not best-possible.

An interesting extension to the minimization of Λ_d is considering non-uniform marginal distributions. While the result of Wang and Wang [21] is restricted to identical marginal distributions having an increasing density, Sklar's Theorem allows us to treat any marginal laws by inserting the quantile functions of the desired distributions. Table 2 contains the approximated value as well as upper and lower bounds of

$$\Theta(\mu_1, \mu_2, \mu_3) = \min_{\mu \in \mathcal{M}(\mu_1, \mu_2, \mu_3)} \int_{[0,1]^3} \Pi(x) d\mu(x)$$

for different choices of μ_1, μ_2, μ_3. Here $\mathcal{M}(\mu_1, \ldots, \mu_d)$ denotes the set of probability measures on $[0, 1]^d$ which have marginal distributions $\mu_1, \ldots, \mu_d$. Distributions with unbounded support require some minor adjustments, however the method is still applicable. Again, the rearrangement algorithm as proposed by Puccetti and Rüschendorf [16] will produce more accurate results in shorter calculation time and is thus probably the preferable choice in applications.

1 This result was obtained using the "lpSolve" package for the open source program *R*. This package is built on the free Mixed Integer Linear Program solver lp_solve, which utilizes the revised simplex method and the Branch-and-bound method. No presolve routines or any other kind of advanced techniques were used.

Table 2: Approximation results for inhomogenous marginals, obtained for a discretization of $n = 60$ sections in each dimension.

μ_1	μ_2	μ_3	lower bd.	approx. value	upper bd.
Par(2, 0.2)	$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$	0.043123	0.046767	0.054226
$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$	$Beta(2,5)$	0.030489	0.033876	0.037232
$exp(1)$	$\mathcal{U}([0,1])$	$Beta(2,5)$	0.044647	0.051213	0.058365

4.2 First-to-default Swaps

In the examples so far, we always minimized the expectation of the product of random variables. The product function is *supermodular* in the following sense.

Definition 4.1. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called supermodular, if it fulfills*

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y) \qquad \forall x, y \in \mathbb{R}^d.$$

Here, $x \wedge y$ resp. $x \vee y$ means the componentwise minimum resp. maximum of x and y.

Since the rearrangement algorithm can only be applied to approximate the expectation of supermodular functions, it is interesting to look at an example involving a non-supermodular function. We want to consider so called *First-to-default* Swaps. They can be thought of as an insurance contract for portfolios of risky assets. The protection seller (PS) compensates the losses if one of the assets in the portfolio of the protection buyer (PB) defaults. In turn, the PB pays fixed premiums at fixed points in time (e.g. quarterly or annually) until the first default occurs or the maturity of the contract is reached. There are no payments for any event after the first default or after maturity. We consider a portfolio consisting of three risky assets and use the following assumptions and notations.

The default times τ_1, τ_2, and τ_3 of the assets follow an exponential distribution with parameters λ_1, λ_2 and λ_3 respectively. The notionals of all three assets are 1 and the recovery rates R_1, R_2 and R_3 describe the amount of money that can be liquidized if the corresponding asset defaults. So the total loss for a default of asset i is $(1 - R_i)$. The times of premium payments are denoted by $0 = t_0 < t_1 < \cdots < t_k = T$ with T denoting the time of maturity. Note in particular that the first payment is due at time $t_0 = 0$. Finally, we assume there is a constant, risk free interest rate $r \geq 0$. Now the premiums p are given as

$$p = \int_{[0,1)^3} \frac{e^{-r \min(F_1^{-1}(x_1), F_2^{-1}(x_2), F_3^{-1}(x_3))}}{\sum_{i=0}^{k} e^{-r t_i} \mathbb{1}_{\{t_i < \min(F_1^{-1}(x_1), F_2^{-1}(x_2), F_3^{-1}(x_3))\}}} \cdot \Big(\mathbb{1}_{\{F_1^{-1}(x_1) \leq \min(F_2^{-1}(x_2), F_3^{-1}(x_3), T)\}} (1 - R_1)$$
$$+ \mathbb{1}_{\{F_2^{-1}(x_2) \leq \min(F_1^{-1}(x_1), F_3^{-1}(x_3), T)\}} (1 - R_2) + \mathbb{1}_{\{F_3^{-1}(x_3) \leq \min(F_1^{-1}(x_1), F_2^{-1}(x_2), T)\}} (1 - R_3) \Big) \, dC(x),$$

with C denoting the copula of the distribution function of (τ_1, τ_2, τ_3) and F_i^{-1} being the quantile function corresponding to the distribution of τ_i. Since our assumptions and the valuation method we want to use are basically the same, we refer to [7] for the precise deduction of the premium heights. We calculate bounds for the minimal as well as for the maximal premium for three payment times $t_0 = 0$, $t_1 = 1$ and $t_2 = T = 2$. The results are listed in Table 3.

Remark 4.2. *Note that the integrand in the last example is not continuous. Our method is not restricted to continuous functions but can be applied as long as the integrand f can, with respect to the L^1 norm, be approximated by a (subsequence of a) sequence of functions $(f_n)_n$ that are constant on the cubes I_i^n with $i \in \mathcal{I}$ (as defined in Theorem 3.1). Hence, by a simple denseness argument, the algorithm actually works for any function f in $L^1([0,1]^d)$, which is why we still have valid bounds in our last example. However, the speed of convergence can be very slow for functions with many discontinuities. For example, it can happen that for $n_1 < n_2$, the bounds for a discretization of n_2 sections are worse than those for a discretization with n_1 sections. Also the convergence of the sequence of optimizers is not guaranteed if we choose an integrand f which is not continuous.*

Table 3: Upper and lower bounds for the minimal and the maximal premium. The interest rate was set to $r = 0.05$.

$\lambda_1 = \frac{1}{3}, \quad R_1 = 0.5,$	$\lambda_2 = \frac{1}{2}, \quad R_2 = 0.7,$	$\lambda_3 = \frac{2}{5}, \quad R_3 = 0.6$
n	range min. premium	range max. premium
30	0.14093 – 0.16104	0.37090 – 0.40287
40	0.14566 – 0.16072	0.37693 – 0.40086
50	0.14572 – 0.15775	0.37777 – 0.39690
70	0.14777 – 0.15424	0.37889 – 0.39102

Acknowledgement: The author gratefully acknowledges the support of the Austrian Science Fund (FWF) Project F5510 (part of the Special Research Program (SFB) "Quasi-Monte Carlo Methods: Theory and Applications"). Furthermore, the author would like to thank Robert Tichy and Stefan Thonhauser from TU Graz for helpful discussions and comments as well as Woden Kusner (also from TU Graz) for corrections and language editing. Finally, thanks go to two anonymous referees whose detailed reports helped to improve the manuscript.

References

[1] Beiglböck, M. and C. Griessler (2014). An optimality principle with applications in optimal transport. Available at arXiv:1404.7054.
[2] Bignozzi, V., G. Puccetti, and L. Rüschendorf (2015). Reducing model risk via positive and negative dependence assumptions. *Insurance Math. Econom. 61*(1), 17–26.
[3] Burkard, R. E., M. Dell'Amico, and S. Martello (2009). *Assignment Problems. Revised Reprint.* Siam, Philadelphia, PA.
[4] Durante, F., J. Fernández-Sánchez, J. J. Quesada-Molina, and M. Úbeda-Flores (2015). Convergence results for patchwork copulas. *European J. Oper. Res. 247*(2), 525–531.
[5] Durante, F. and J. F. Sánchez (2012). On the approximation of copulas via shuffles of min. *Statist. Probab. Lett. 82*(10), 1761–1767.
[6] Griessler, C. (2016). c-cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge-Kantorovich problem. Available at arXiv:1601.05608.
[7] Hofer, M. and M. R. Iacò (2014). Optimal bounds for integrals with respect to copulas and applications. *J. Optim. Theory Appl. 161*(3), 999–1011.
[8] Iacò, M. R., S. Thonhauser, and R. F. Tichy (2015). Distribution functions, extremal limits and optimal transport. *Indag. Math. 26*(5), 823–841.
[9] Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer Verlag, New York.
[10] Lux, T. and A. Papapantoleon (2016). Improved Fréchet–Hoeffding bounds on d-copulas and applications in model-free finance. Available at arXiv:1602.08894.
[11] Mai, J.-F. and M. Scherer (2014). *Financial Engineering with Copulas Explained*. Palgrave Macmillan, Hampshire.
[12] Mikusinski, P., H. Sherwood, and M. D. Taylor (1992). Shuffles of Min. *Stochastica 13*(1), 61–74.
[13] Mikusiński, P. and M. D. Taylor (2010). Some approximations of n-copulas. *Metrika 72*(3), 385–414.
[14] Nelsen, R. B. (2006). *An Introduction to Copulas*. Second Edition. Springer, New York.
[15] Puccetti, G. and L. Rüschendorf (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comput. Appl. Math. 236*(7), 1833–1840.
[16] Puccetti, G. and L. Rüschendorf (2015). Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals. *Comm. Statist. Simulation Comput. 44*(3), 705–718.
[17] Puccetti, G. and R. Wang (2015). Extremal dependence concepts. *Statist. Sci. 30*(4), 485–517.
[18] Rüschendorf, L. (1980). Inequalities for the expectation of δ-monotone functions. *Probab. Theory Related Fields 54*(3), 341–349.
[19] Schmid, F., R. Schmidt, T. Blumentritt, S. Gaißer, and M. Ruppert (2010). Copula-based measures of multivariate association. In *Copula Theory and its Applications*, pp. 209–236. Springer, Berlin Heidelberg.
[20] Villani, C. (2009). *Optimal Transport: Old and New.* Springer-Verlag, Berlin.
[21] Wang, B. and R. Wang (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal. 102*(10), 1344–1360.