Research Article

Oliver Grothe and Jonas Rieger*

Decomposition and graphical correspondence analysis of checkerboard copulas

https://doi.org/10.1515/demo-2024-0006
received November 17, 2023; accepted July 8, 2024

Abstract: We analyze optimal low-rank approximations and correspondence analysis of the dependence structure given by arbitrary bivariate checkerboard copulas. Methodologically, we make use of the truncation of singular value decompositions of doubly stochastic matrices representing the copulas. The resulting (truncated) representations of the dependence structures are sparse, in particular, compared to the number of squares on the checkerboard. The additive structure of the decomposition carries through to statistical functionals of the copula, such as Kendall's τ or Spearman's ρ, and also motivates similarity measures for checkerboard copulas. We link our analysis to continuous decompositions of copula densities and copula-generating algorithms and discuss further general properties of the decomposition and its truncation. For example, truncated series might lack nonnegativity, and approximation errors increase for monotonicity-like copulas. We provide algorithms and extensions that account for and counteract these properties. The low-rank representation is illustrated for various copula examples, and some analytical results are derived. The resulting correspondence analysis profile plots are analyzed, providing graphical insights into the dependence structure implied by the copula. An illustration is provided with an empirical data set on fuel injector spray characteristics in jet engines.

Keywords: correspondence analysis, checkerboard copulas, measures of concordance, singular value decomposition

MSC 2020: primary 62H05; secondary 62A09

1 Introduction

Copulas are a standard tool for modeling random vectors, as they separate marginal and dependence modeling. A copula contains information on the likelihood of joint occurrence of random variables on their intrinsic quantile scale. For two-dimensional vectors, the copula thus encodes a possibly large or infinite two-dimensional frequency table specifying the joint likelihood of the transformed random vector. If finite, square, and scaled appropriately, this table can be interpreted as a checkerboard copula [15,29]. The tables are generally large and may contain redundant information, and assessing the incorporated dependence information is not straightforward. We apply the well-known decomposition and dimensionality reduction techniques of high-dimensional data analysis to this table, thereby decomposing the copula. The decomposition opens a wide range of further analyses, for example, to compute and analyze copula characteristics, plot meaningful two-dimensional plots of the copula, or build simpler, reasonable approximations of complicated dependency

* **Corresponding author: Jonas Rieger**, Karlsruhe Institute of Technology, Institute for Operations Research, Analytics and Statistics, Kaiserstr. 12, 76137 Karlsruhe, Germany, e-mail: jonas.rieger@kit.edu
Oliver Grothe: Karlsruhe Institute of Technology, Institute for Operations Research, Analytics and Statistics, Kaiserstr. 12, 76137 Karlsruhe, Germany

structures. Through the low-rank approximation, one can drastically decrease the number of items to be stored compared to the full checkerboard matrix, i.e., the square of the lattice size.

Checkerboard copulas can be obtained from empirical data or, for example, by discretizing continuous copulas [16,27]. In either way, the copula frequency table is a doubly stochastic matrix. Taking the doubly stochastic matrix, we apply correspondence analysis methods that are mainly based on singular value decomposition (SVD).

Additive decompositions of copulas using variable-specific functions already exist in the literature, but only for continuous representations. Continuous decompositions are considered, for example, in Mesiar and Najjari [33] or Rodríguez-Lallena [40] for the generation of new copulas and in Cuadras [8] for the decomposition of copulas. The checkerboard case differs from existing approaches and yields different decompositions, as discussed in Section 3.3. Durrleman et al. [16] mentioned SVD of checkerboard copulas but did not go into detail, and Cuadras [9] considered discrete and continuous decompositions of general bivariate distributions. In contrast to these studies, we concentrate on the decomposition of doubly stochastic matrices that represent checkerboard copulas, allowing us to focus on the features of copulas. We provide formulas for important statistical functionals, including Spearman's ρ, Kendall's τ, and Pearson's ϕ^2. Through the Frobenius distance between the matrices, we express the similarity of two checkerboard copulas in terms of their ϕ^2.

Using the standard kit of correspondence analysis has obstacles for some copulas. Copulas such as the comonotonicity copula are costly to represent in standard SVD, as the corresponding frequency matrix is similar to an identity matrix, having full rank and many equally large singular vectors. Thus, approximations by truncating the SVD series have slowly decaying errors. Therefore, we propose to use a monotonicity-anchored representation (MAR) (adapted from [18] and [24]), taking into account the independence and comonotonicity-like characteristics. This representation does not change the singular vectors for symmetric copulas but can considerably reduce the approximation error. Also, the obtained truncations are not necessarily valid checkerboard copulas, as negative values can occur. We provide an algorithm that yields the nearest valid copula for the Frobenius norm by generalizing an algorithm by Zass and Shashua [47] and thus maps the obtained truncated (not doubly stochastic) matrix to the nearest doubly stochastic matrix. While this article is focused on the Frobenius error norm, we remark on using the Hellinger distance in Appendix B.

The frequency table decomposition corresponds to a decomposition of the discretized copula probability distribution function (PDF). Section 2.6 links our analysis to continuous decompositions, as in the literature on copula generation and continuous copula decomposition, and to cumulative distribution function (CDF) decompositions. Through the decomposition, we motivate a decomposition of the Gaussian copula into transformed Hermite polynomials.

Thus, this article makes several contributions. We define the decomposition of checkerboard copulas and give extensions of the approach for comonotonicity-like copulas and non-copula truncations. We link the approach to important existing copula concepts such as dependence measures, similarities of copulas, and continuous decompositions of copulas. We derive characteristics of the graphs obtained by the approach and thus provide a new method of graphical copula representations. Finally, we apply the approach to theoretical copula families of various complexities and an empirical data example from the engineering context.

This article is structured as follows. Section 2 describes the approach, including the extensions for comonotonicity-like copulas, non-copula truncations, and the computation of statistical functionals. We analyze the difference between decomposed copulas and draw the connection between discrete (checkerboard) and continuous decompositions. We provide the resulting decompositions for the well-known copulas of different complexities and for symmetric and asymmetric dependencies in Section 3. We use the graphical tools of correspondence analysis to interpret the two-dimensional graphs of copulas and apply the graphical tools to an empirical checkerboard from data on the fuel injection spray characteristics of jet engines in Section 4. Section 5 concludes this article.

2 Checkerboard copula decomposition and its characteristics

This section examines the SVD and its truncation for checkerboard copulas, i.e., doubly stochastic matrices. We introduce some notations in Section 2.1 and then define the truncated decomposition, including an MAR that accounts for dependencies similar to comonotonicity in Section 2.2. To correct negative matrix elements in the truncated representation, Section 2.3 formulates an algorithm to approximate the truncation by a doubly stochastic matrix. Sections 2.4 and 2.5 derive statistical functionals and similarity measures using the decomposition. Section 2.6 links the decompositions of continuous copulas and their discretized counterparts.

2.1 Doubly stochastic matrices from bivariate copulas

Let X and Y be random variables with CDF $F_{X,Y}$ and marginal CDFs F_X and F_Y, respectively. Through the well-known theorem of Sklar [45] the multivariate CDF $F_{X,Y}(x,y)$ can be decomposed as

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)),$$

whereby the copula C encodes the dependence structure of X and Y. The copula C can also be seen as CDF of $F_X(X)$ and $F_Y(Y)$, and thus, has the properties of a multivariate CDF with uniform margins, provided that X and Y are continuous. While the copula is unique for continuous random variables, it is only uniquely identified on the image of F_X and F_Y, respectively, in the discrete case.

A checkerboard copula [29] is a special type of copula that assumes a uniform mass within the squares of an evenly spaced lattice $I^n \times I^n$ ($I^n = \{0, 1/n, \ldots, 1\}$). Checkerboard copulas can be computed from empirical data or by the discretization of continuous copulas. The discretization facilitates the (asymptotic) comparison of discrete and continuous characteristics.

Any continuous copula C defines a doubly stochastic matrix $\mathbf{C}^n$ on the grid $I^n \times I^n$ by evaluating C on $I^n \times I^n$, i.e., $C \coloneqq C(u,v) \quad (u,v \in I^n)$ [27,31] and computing

$$\mathbf{C}^n_{i,j} = n\left[C\left(\frac{i}{n}, \frac{j}{n}\right) - C\left(\frac{i-1}{n}, \frac{j}{n}\right) - C\left(\frac{i}{n}, \frac{j-1}{n}\right) + C\left(\frac{i-1}{n}, \frac{j-1}{n}\right)\right] \quad \text{for } i,j = 1,\ldots,n. \tag{1}$$

The properties of $\mathbf{C}^n$ follow from the copula properties of C:

(1) $\mathbf{C}^n$ has nonnegative entries as the defining equation (1) coincides with a scaled version of the 2-volume of the copula, which is nonnegative.

(2) From $C(u,1) = C(1,u) = u \quad (u \in [0,1])$ and $C(u,0) = C(0,u) = 0 \quad (u \in [0,1])$ follows for $j \in [n]$

$$\sum_{i=1}^{n} \mathbf{C}^n_{i,j} = \sum_{i=1}^{n} n\left[C\left(\frac{i}{n}, \frac{j}{n}\right) - C\left(\frac{i-1}{n}, \frac{j}{n}\right) - C\left(\frac{i}{n}, \frac{j-1}{n}\right) + C\left(\frac{i-1}{n}, \frac{j-1}{n}\right)\right] = 1.$$

An analogous computation shows $\sum_{j=1}^{n} \mathbf{C}^n_{i,j} = 1$ ($i \in [n]$), and thus, the row and column sums of $\mathbf{C}^n$ are 1.

The matrix $\mathbf{C}^n$ is by construction square, together with (1) and (2), a doubly stochastic matrix. The element $\mathbf{C}^n_{i,j}$ ($i,j \in [n]$, where $[n] \coloneqq \{1, 2, \ldots, n\}$) corresponds to the density of the checkerboard copula

$$\hat{c}(u,v) = n\mathbf{C}^n_{\lceil u\cdot n\rceil, \lceil v\cdot n\rceil} = n\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{C}^n_{i,j} \mathbb{1}_{u \in \left[\frac{i-1}{n}, \frac{i}{n}\right]} \mathbb{1}_{v \in \left[\frac{j-1}{n}, \frac{j}{n}\right]} \quad (u,v \in \mathbb{R}), \tag{2}$$

in the rectangle

$$R_{i,j} \coloneqq \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[\frac{j-1}{n}, \frac{j}{n}\right].$$

We denote by $\mathbb{1}_{\cdot}$ the indicator function. Thus, $\mathbf{C}^n_{i,j}$ can be interpreted naturally as a table of the likelihood of occurrence in the copula domain. Integration over equation (2) yields a checkerboard approximation of the

copula CDF $\hat{C}$ and the conditional CDF $\hat{C}_{u|V=v}$. The discretizations approximate the copula C with increasing n, and every copula C is the limit of its discretizations for $n \to \infty$ [see 27, Theorem 1].

2.2 SVD and MAR

Having the table of likelihood of occurrence, $\mathbf{C}^n$, correspondence analysis can be applied to the matrix $\mathbf{C}^n$ to analyze the structural properties of the matrix. Correspondence analysis uses the SVD to compute low-dimensional approximations of the matrix. To this end, the SVD is truncated, yielding the nearest matrix of the specified rank according to the Frobenius or the spectral norm [34]. In correspondence analysis, the matrix $\mathbf{C}^n$ is usually centered, and some scaling is applied to rows and columns to account for the sum differences of the rows or columns [18]. In the case of $\mathbf{C}^n$, the centering step is implemented by subtracting the matrix $\mathbf{\Pi}^n \coloneqq n^{-1}\mathbf{1}\mathbf{1}^\top$ from $\mathbf{C}^n$, where $\mathbf{1}$ is the vector of ones of suitable dimension. We denote this by

$$\mathbf{A}^n = G(\mathbf{C}^n) \coloneqq \mathbf{C}^n - \mathbf{\Pi}^n.$$

Note that $\frac{1}{\sqrt{n}}\mathbf{1}$ is a (left and right) singular vector of $\mathbf{C}^n$ with singular value 1, whereby 1 is the largest singular value for doubly stochastic matrices [37], and thus, the rank of $\mathbf{A}^n$ is at most $n-1$. We denote the SVD of $\mathbf{A}^n$ by

$$\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \quad \text{with } \mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n), \mathbf{S} = \operatorname{diag}(s_1, \dots, s_n), \mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n), \tag{3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the orthogonal matrices and the singular values s_k are in $[0, 1]$ and are sorted in descending order as usual.

The decomposition in equation (3) may be truncated by using only the $n^* \le n$ largest singular values of s, and the corresponding first n^* columns of $\mathbf{U}$ and $\mathbf{V}$:

$$T_{n^*}(\mathbf{A}^n) \coloneqq \mathbf{U}_{:,1:n^*}\mathbf{S}_{1:n^*,1:n^*}(\mathbf{V}_{:,1:n^*})^\top,$$

where we will use $T_{n^*}(\cdot)$ as a truncation operator of the argument's SVD in the following. The truncated $T_{n^*}(\mathbf{A}^n)$ yields an approximation of $\mathbf{C}^n$ by applying the inverse function of G, i.e.,

$$G^{-1}(T_{n^*}(\mathbf{A}^n)) = T_{n^*}(\mathbf{A}^n) + \mathbf{\Pi}^n.$$

The truncated SVD yields low-rank approximations with small errors for matrices with a few large and many small (or zero) singular values. We will show examples in Section 3. However, in the copula context, many copulas share characteristics with the comonotonicity copula, an identity matrix with singular value 1 with multiplicity n, and thus, high approximation errors for small-rank representations. To "remove" the comonotonicity copula characteristics before applying the SVD, we suggest transforming the matrix $\mathbf{C}^n$ so that we account for high frequencies on the diagonal of the matrix and thus the monotone dependence structures. We denote this transformation by $G_{\mathrm{MAR}}(\cdot)$ and call it MAR. As we argue in Lemma 1, through this representation, the singular vectors do not change for symmetric copulas, but the series of singular values s_k decreases faster, leading to better low-rank approximations. The MAR is given by

$$\tilde{\mathbf{A}}^n = G_{\mathrm{MAR}}(\mathbf{C}^n, \eta) \coloneqq c + \eta I_n - (1+\eta)\frac{1}{n}\mathbf{1}\mathbf{1}^\top, \tag{4}$$

with $\eta \in \mathbb{R}$ and I_n denoting the n-by-n identity matrix. The centering step is implemented by the last summand $-(1+\eta)\frac{1}{n}\mathbf{1}\mathbf{1}^\top$, i.e., for $\eta \in \mathbb{R}$,

$$(\tilde{\mathbf{A}}^n)\mathbf{1} = \mathbf{C}^n\mathbf{1} + \eta I_n\mathbf{1} - \frac{1+\eta}{n}\mathbf{1}\mathbf{1}^\top\mathbf{1} = \mathbf{1} + \eta\mathbf{1} - \frac{1+\eta}{n}\mathbf{1}\cdot n = 0\cdot\mathbf{1}$$

and analogously for $(\tilde{\mathbf{A}}^n)^\top\mathbf{1}$. The approach also suits strong negative dependence by rotating the copula first. A similar transformation to G_{MAR} can be found in Kazmierczak [24], in Greenacre [18, Section 8.6] formulated in the context of frequency tables. Unlike Greenacre [18] who used two parameters and demands them to be

chosen such that $(\tilde{\mathbf{A}}^n)_{i,j} \geq 0 \ \forall (i,j)$ for merely illustrative purposes, we do not require this additional restriction here. We scale the last summand by $1/n$ to preserve the margins shown earlier.

Note that $G(\mathbf{C}^n)$ is nested within the MAR by setting $\eta = 0$. Later, the parameter η is calculated such that the Frobenius distance between the inverse transformed version of $\tilde{\mathbf{A}}^n$, denoted by $G^{-1}_{\text{MAR}}(T_{n^*}(\tilde{\mathbf{A}}^n), \eta)$, and (the original) $\mathbf{A}^n$ is minimized.

Analogously to the aforementioned notation, we denote the SVD of $\tilde{\mathbf{A}}^n$ by

$$\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^\top, \quad \text{with } \tilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n), \tilde{\mathbf{S}} = \text{diag}(\tilde{s}_1, \ldots, \tilde{s}_n), \quad \text{and} \quad \tilde{\mathbf{V}} = (\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n).$$

The following lemma shows that singular values and vectors of $\tilde{\mathbf{A}}^n$ and $\mathbf{A}^n$ are closely connected, provided that $\mathbf{C}^n$ is symmetric.

Lemma 1. *For the SVD of* $\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^\top$ *and* $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ *of symmetric* $\mathbf{C}^n$*, i.e., for* $\mathbf{U} = \mathbf{V}$*, it holds that*

$$\mathbf{u}_k = \tilde{\mathbf{u}}_k = \mathbf{v}_k = \tilde{\mathbf{v}}_k \quad k = 1, 2, \ldots \quad \textit{and} \quad \tilde{s}_k = (s_k + \eta) \quad k = 1, 2, \ldots, n.$$

Proof. From $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \perp \mathbf{1}$ follows for $k = 1, \ldots, n-1$

$$\begin{aligned}
\tilde{\mathbf{A}}^n \cdot \mathbf{v}_k &= \left[\mathbf{C}^n + \eta I_n + \frac{1+\eta}{n}\mathbf{1}\mathbf{1}^\top\right] \cdot \mathbf{v}_k \\
&= \mathbf{C}^n\mathbf{v}_k + \eta I_n \mathbf{v}_k + \frac{1+\eta}{n}\mathbf{1}\underbrace{\mathbf{1}^\top\mathbf{v}_k}_{=0} \\
&= s_k\mathbf{u}_k + \eta\mathbf{v}_k.
\end{aligned}$$

For symmetric matrices, thus,

$$\tilde{\mathbf{A}}^n \cdot \mathbf{v}_k = (s_k + \eta)\mathbf{v}_k. \qquad \square$$

For asymmetric $\mathbf{C}^n$, the singular values and vectors of $\mathbf{A}^n$ and $\tilde{\mathbf{A}}^n$ differ. Lemma 1 yields the n^*-truncated representation of $\tilde{\mathbf{A}}^n$

$$\begin{aligned}
T_{n^*}(\tilde{\mathbf{A}}^n) &= \sum_{k=1}^{n^*} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top \\
&\overset{(*)}{=} \sum_{k=1}^{n^*} (s_k + \eta)\mathbf{u}_k\mathbf{u}_k^\top \quad |(*): \quad \text{for symmetric } \mathbf{C}^n
\end{aligned}$$

and thus after backtransformation of equation (4)

$$G^{-1}_{\text{MAR}}(T_{n^*}(\tilde{\mathbf{A}}^n), \eta) = \left(\sum_{k=1}^{n^*} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top - \eta I_n + (1+\eta)\frac{1}{n}\mathbf{1}\mathbf{1}^\top\right) \tag{5}$$

$$\overset{(*)}{=} \left(\sum_{k=1}^{n^*} \mathbf{u}_k\mathbf{u}_k^\top(s_k + \eta) - \eta I_n + (1+\eta)\frac{1}{n}\mathbf{1}\mathbf{1}^\top\right) \tag{6}$$

$$= \sum_{k=1}^{n^*} \mathbf{u}_k\mathbf{u}_k^\top(s_k + \eta) - \eta I_n + (1+\eta)\mathbf{\Pi}^n \tag{7}$$

and equations (6) and (7) are, again, only valid for symmetric copulas. PDF and CDF can be computed using $G^{-1}_{\text{MAR}}(T_{n^*}(\tilde{\mathbf{A}}^n), \eta)$ analogously to equation (2). The parameter η of equation (4) can be determined by minimizing some error norm of interest. For example, we calculate the fraction η that minimizes the residual inertia (thus, Frobenius error) for a specified approximation of rank n^* by

$$\eta^*(\mathbf{C}^n, n^*) = \arg\min_{\eta \in \mathbb{R}} \left\| \sum_{k=1}^{n^*} \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^\top(\tilde{s}_k) - \eta I_n + (1+\eta)\mathbf{\Pi}^n - \mathbf{C}^n \right\|_F^2. \tag{8}$$

For a symmetric matrix $\mathbf{C}^n$ and an approximation of dimension n^*, this yields

$$\eta^*(\mathbf{C}^n, n^*) = \arg\min_{\eta \in \mathbb{R}} \sum_{k=n^*+1}^{n} (s_k + \eta)^2 = -\sum_{k=n^*+1}^{n} \frac{s_k}{n - n^*}. \tag{9}$$

For asymmetric matrices $\mathbf{C}^n$, the problem in equation (8) can be solved numerically. The simulations in Section 3 examine the choices of η and the resulting matrices $\tilde{\mathbf{A}}^n$.

2.3 Ensuring double stochasticity of truncations

As noted earlier, truncations of the SVD can yield low error approximations with considerably lower rank matrices. In general, truncations of the SVD are not necessarily doubly stochastic matrices. Truncations keep the property of having row and column sums of one as the singular vectors $\mathbf{u}_k$ and $\mathbf{v}_k$, or $\tilde{\mathbf{u}}_k$ and $\tilde{\mathbf{v}}_k$, respectively, are perpendicular to $\mathbf{1}$ for $k \in [n-1]$, but the truncations do not necessarily have nonnegative elements. One can approximate the truncation by the nearest, doubly stochastic matrix to ensure nonnegativity. This step does not increase the complexity of the representation, as it does not include any information other than the truncated matrix. We give a general idea of the algorithms for symmetric and asymmetric matrices here; they are more specifically described in Appendix A.

Zass and Shashua [47] proposed an algorithm to find the nearest doubly stochastic matrix for any symmetric matrix $\mathbf{A}^{\text{sym}} \in \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} = \mathbf{A}^\top\}$ according to the Frobenius norm, i.e., a solution to the problem $P(\mathbf{A})$, with

$$\begin{aligned} P(\mathbf{A}) = \arg\min_{\mathbf{B} \in \mathbb{R}^{n \times n}} \ & \|\mathbf{A} - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}\mathbf{1} = \mathbf{1} \\ & \mathbf{B}^\top \mathbf{1} = \mathbf{1} \\ & \mathbf{B}_{i,j} \geq 0, \quad \forall i, j \in [n]. \end{aligned}$$

According to Zass and Shashua [47], $P(\mathbf{A})$ can be solved for symmetric $\mathbf{A}$ iteratively by solving two problems, i.e., $P(\mathbf{A}) = P_2 P_1 P_2 \ldots P_1(\mathbf{A})$, with

$$\begin{aligned} P_1(\mathbf{A}_1) = \arg\min_{\mathbf{B} \in \mathbb{R}^{n \times n}} \ & \|\mathbf{A}_1 - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}\mathbf{1} = \mathbf{1} \\ & \mathbf{B}^\top \mathbf{1} = \mathbf{1} \end{aligned}$$

and

$$\begin{aligned} P_2(\mathbf{A}_2) = \arg\min_{\mathbf{B} \in \mathbb{R}^{n \times n}} \ & \|\mathbf{A}_2 - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}_{i,j} \geq 0, \quad \forall i, j \in [n]. \end{aligned}$$

$\mathbf{A}_1$ and $\mathbf{A}_2$ refer to the iterative solutions of P_2 and P_1, respectively. Algorithm 1 formulates the algorithm explicitly. P_1 and P_2 have closed-form solutions and calculations for the solution of P_1 and P_2 are carried out in A. In the case of asymmetric matrices, Algorithm 1 retains its overall structure while incorporating a deflection component to maintain validity [17]. The resulting Algorithm 2 is shown in A. Note that there are algorithms for approximations with a particular interest in keeping the sparsity structure of $\mathbf{A}$. Rontsis and Goulart [41] formulated an algorithm for a slightly modified problem that accounts for the sparsity of the matrix $\mathbf{A}$ based on the alternate direction method of multipliers and applied to symmetric and asymmetric matrices $\mathbf{A}$. Sparsity thereby refers to zero entries in matrix $\mathbf{A}$. In general, the SVD approximations typically contain many small, nonzero values, and thus, the approximation does not benefit from exploiting the sparsity structure.

2.4 Statistical functionals of decompositions and truncations

Various statistical properties can be computed using the decomposition, including dependence measures such as Kendall's τ, Spearman's ρ_S, and Pearson's ϕ^2. We start by expressing the well-known dependency measures Kendall's τ and Spearman's ρ_S through the decomposition. The structure of both measures inherits the SVD structure of the checkerboard copula, and thus, copulas with many high singular values tend to have a measure representation with many terms, subject to the direction of the singular vectors. The empirical computation of dependence measures using the checkerboard copula itself might be inefficient, and the use of a low-rank approximation might be more robust. We leave the asymptotics of the decomposed measures for further research.

Durrleman et al. [16] showed that for checkerboard copulas, Kendall's τ and Spearman's ρ_S can be computed by

$$\tau(\mathbf{C}^n) \coloneqq 1 - \frac{1}{n^2}\operatorname{trace}(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top) \quad \text{and} \quad \rho_S(\mathbf{C}^n) \coloneqq \frac{3}{n}\operatorname{trace}(\mathbf{\Omega}\mathbf{C}^n) - 3,$$

with $\mathbf{E} \in \mathbb{R}^{n\times n}$ and

$$\mathbf{E}_{i,j} = \begin{cases} 1, & \text{if} \quad i = j, \\ 2, & \text{if} \quad i > j, \\ 0, & \text{if} \quad i < j, \end{cases}$$

and $\mathbf{\Omega} \in \mathbb{R}^{n\times n}$, where

$$\mathbf{\Omega} \coloneqq \check{\boldsymbol{\omega}}\check{\boldsymbol{\omega}}^\top, \quad \text{with } \check{\boldsymbol{\omega}} = \frac{1}{n}(2n - 2\cdot 1 + 1, 2n - 2\cdot 2 + 1, \dots)^\top.$$

Let, as in Section 2.2, the SVD of the centered $\mathbf{C}^n$ be denoted by $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, and additionally, $\mathbf{u}_0 = \mathbf{v}_0 = \frac{1}{\sqrt{n}}\mathbf{1}^\top$ and $s_0 = 1$, such that $\mathbf{C}^n = \sum_{k=0}^{n-1}\mathbf{u}_k s_k \mathbf{v}_k$. Then, follows Spearman's ρ_S with $\boldsymbol{\omega} \coloneqq \|\check{\boldsymbol{\omega}}\|^{-1}\check{\boldsymbol{\omega}}$ as

$$\begin{aligned} \rho_S(\mathbf{C}^n) &= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\sum_{k=0}^{n-1}\mathbf{u}_k s_k \mathbf{v}_k^\top\right) - 3 \\ &= (4 - 1/n^2)\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle - 3 \\ &= (4 - 1/n^2)\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle \end{aligned} \tag{10}$$

and for Kendall's τ

$$\begin{aligned} \tau(\mathbf{C}^n) &\coloneqq 1 - \frac{1}{n^2}\operatorname{trace}(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top) \\ &= 1 - \frac{1}{n^2}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} s_k s_l \langle \mathbf{u}_l, \mathbf{E}\mathbf{u}_k\rangle\langle \mathbf{v}_k, \mathbf{E}\mathbf{v}_l\rangle. \end{aligned} \tag{11}$$

Details of the calculations are provided in Appendix C. Both dependence measures can also be put in terms of the MAR, for example,

$$\rho_S(\mathbf{C}^n) = \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1}\tilde{s}_k\langle\tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle\tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - \left(4 - \frac{1}{n^2}\right)\eta + \frac{3}{n}(1+\eta)\right] - 3 \tag{12}$$

$$= \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1}(s_k + \eta)\langle\tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle^2 - \left(4 - \frac{1}{n^2}\right)\eta + \frac{3}{n}(1+\eta)\right] - 3 \quad |(\text{for symmetric } \mathbf{C}^n). \tag{13}$$

Note that η refers to the MAR coefficient of equation (4). The calculations are performed in Appendix C. The decompositions of ρ_S and τ are both based on the singular-value-weighted sum of scalar products containing

the singular vectors. Thus, they account for the importance and the direction of the component. For τ, the projection vector contains transformations of the other singular vectors, reflecting the integral's measure being the copula CDF. The representations in equations (10) and (11) yield approximations for Spearman's ρ_S and Kendall's τ in terms of the truncated representations, i.e.,

$$\hat{\rho}_{S_{n^*}}(\mathbf{C}^n) \coloneqq (4 - 1/n^2) \sum_{k=1}^{n^*} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k \rangle \langle \mathbf{v}_k, \boldsymbol{\omega} \rangle \quad \text{and} \tag{14}$$

$$\hat{\tau}_{n^*}(\mathbf{C}^n) \coloneqq 1 - \frac{1}{n^2} \sum_{k=0}^{n^*} \sum_{l=0}^{n^*} s_k s_l \langle \mathbf{u}_l, E\mathbf{u}_k \rangle \langle \mathbf{v}_k, E\mathbf{v}_l \rangle. \tag{15}$$

In the SVD representation, Pearson's ϕ^2 boils down to the total inertia of the copula from independence [44, p. 223]

$$\begin{aligned} \phi^2(\mathbf{C}^n) &= \int_0^1 \int_0^1 \hat{c}^2(u, v) \mathrm{d}u \mathrm{d}v - 1 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} n^2 (\mathbf{C}^n_{ij})^2 \frac{1}{n^2} - 1 \\ &= \|\mathbf{A}^n\|_F^2 \\ &= \sum_{k=1}^{n} s_k^2, \end{aligned}$$

where $s_1, s_2, \ldots$ are the singular values of the centered $\mathbf{A}^n$. Note that this is proportional to Pearson's χ^2 statistic for testing independence in an empirical contingency table (for Pearson's χ^2 statistic in the copula context, see, e.g., [43]). The truncated representation ϕ^2 is

$$\hat{\phi}^2{}_{n^*}(\mathbf{C}^n) = \sum_{k=1}^{n^*} s_k^2 = \phi^2(G^{-1}(T_{n^*}(\mathbf{A}^n))).$$

In correspondence analysis, the ratio of the total inertia of approximation and the original matrix is a standard measure for the approximation's goodness of fit, i.e.,

$$\frac{\sum_{k=1}^{n^*} s_k^2}{\sum_{k=1}^{n} s_k^2} = \frac{\hat{\phi}^2{}_{n^*}(\mathbf{C}^n)}{\phi^2(\mathbf{C}^n)}.$$

Counting the number of nonzero singular values yields an estimate of the dimensionality of the representation, i.e.,

$$\phi_g(\mathbf{C}^n) = |\{s_k : s_k > 0, k \in [n]\}| = \operatorname{rank}(\mathbf{A}^n).$$

It counts the dimensions needed to model all information in $\mathbf{C}^n$ and does not consider the strength of the information, in contrast to, for example, Pearson's ϕ^2. Cuadras and Díaz [12] called this the geometric dimension of a copula. For discretizations of a continuous copula, the values of ϕ^2 and ϕ_g depend on the grid resolution n and are, therefore, the discretized copula's properties and not of the continuous counterpart. The following lemma formulates this explicitly, and the example in Section 2.6 shows that the geometric dimension can decrease with increasing grid size when the grid sizes are not nested.

Lemma 2. *Let $n_1 > n_2$, with $n_1 = n_2 m$ ($m \in \mathbb{N}$), be the grid resolutions of the discretizations $\mathbf{C}^n[n_1]$ and $\mathbf{C}^n[n_2]$ of a copula C. Then,*

$$\phi_g(\mathbf{C}^n[n_1]) \geq \phi_g(\mathbf{C}^n[n_2]).$$

Proof. Let M_1 and M_2 be the discretized copula C with grid sizes n_1 and n_2, respectively. Then,

$$M_2 = A M_1 A^\top,$$

with $A \in \mathbb{R}^{n_2 \times n_1}$ and $A_{ij} = \frac{1}{\sqrt{m}} \mathbb{1}_{j \in ((i-1)\cdot m, i\cdot m]}$ and

$$\begin{aligned}\operatorname{rank}(M_2) &= \operatorname{rank}(AM_1A^\top) \le \min(\operatorname{rank}(A), \operatorname{rank}(M_1), \operatorname{rank}(A^\top)) \\ &= \min(\operatorname{rank}(A), \operatorname{rank}(M_1)) \le \operatorname{rank}(M_1).\end{aligned}$$ □

2.5 Similarity of copulas

Using the decomposition makes it easy to compute the similarity of copulas if they have a shared grid size. We show that this similarity in terms of the Frobenius distance is mainly driven by Pearson's ϕ^2 of the product of the two copulas. The Frobenius distance is highly dependent on the grid size n; thus, we propose two normalizations. Let $\mathbf{A}^A = \mathbf{U}^A\mathbf{S}^A(\mathbf{V}^A)^\top$ and $\mathbf{A}^B = \mathbf{U}^B\mathbf{S}^B(\mathbf{V}^B)^\top$ be the two matrices after centering the bistochastic matrices $\mathbf{C}^A$ and $\mathbf{C}^B$, respectively. For ease of notation, we omit the common grid size n. Then,

$$\begin{aligned}\|\mathbf{A}^A - \mathbf{A}^B\|_F^2 &= \|\mathbf{C}^A - \mathbf{C}^B\|_F^2 \\ &= \operatorname{trace}\mathbf{C}^A(\mathbf{C}^A)^\top - 2\operatorname{trace}\mathbf{C}^B(\mathbf{C}^A)^\top + \operatorname{trace}\mathbf{C}^B(\mathbf{C}^B)^\top \\ &= \phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B) - 2\phi^2(\mathbf{C}^P) \\ &= \sum_{k=1}^{n-1}(s_k^A)^2 + \sum_{k=1}^{n-1}(s_k^B)^2 - 2\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top + 2,\end{aligned} \tag{16}$$

with the product copula $\mathbf{C}^P = \mathbf{C}^B(\mathbf{C}^A)^\top$ (see [27], p. 700). Whereas the terms $\phi^2(\mathbf{C}^A)$ and $\phi^2(\mathbf{C}^B)$ depend on the individual copulas solely, $\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top$ depends on the relative orientation of the singular vectors, i.e.,

$$\begin{aligned}\operatorname{trace}\mathbf{A}^B(\mathbf{A}^A)^\top &= \sum_{k=1}^{n-1}\sum_{l=1}^{n-1} s_k^B s_l^A \langle \mathbf{v}_k^B, \mathbf{v}_l^A\rangle\langle \mathbf{u}_l^A, \mathbf{u}_k^B\rangle \\ &= \sum_{k=1}^{n-1}\sum_{l=1}^{n-1} s_k^B s_l^A \cos\alpha(\mathbf{v}_k^B, \mathbf{v}_l^A)\cos\alpha(\mathbf{u}_l^A, \mathbf{u}_k^B),\end{aligned}$$

where $\alpha(\cdot,\cdot)$ is the angle between the two vectors. Thus, the copula similarity is driven by the similarity of the singular vectors weighted by the singular values.

Although the distance (squared) in equation (16) is straightforward to compute, it depends on the grid size n, as the range of values increases with n. Clearly, $\|\mathbf{A}^A - \mathbf{A}^B\|_F^2 \ge 0$ and $\|\mathbf{A}^A - \mathbf{A}^B\|_F^2 = 0$ for $\mathbf{A}^A = \mathbf{A}^B$. The maximum

$$\|\mathbf{A}^A - \mathbf{A}^B\|_F^2 = \phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B) - 2\phi^2(C^P) \le n + n - 2\cdot 0 = 2n \tag{17}$$

is attained, for example, for $\mathbf{C}^A = I_n$ and $\mathbf{C}^B$ any doubly stochastic matrix with ones on off-diagonal elements, for example,

$$\mathbf{C}^B = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & & & & \\ 1 & 0 & \dots & & \end{pmatrix}.$$

Thus, the use of the Frobenius distance suffers from a high dependence on the grid size n, and we propose two simple rescalings of the distance that account for the increase in n. The first one uses the maximal distance from equation (17), yielding

$$\delta_1(\mathbf{C}^A, \mathbf{C}^B) = \frac{\|\mathbf{C}^A - \mathbf{C}^B\|_F}{\sqrt{2n}},$$

so that the values lie within $[0, 1]$. The examples in Section 3.4 indicate that this normalization overcorrects, resulting in decreasing δ_1 for checkerboard approximations of the same copulas with increasing n.

Another approach is to standardize the distance by the sum of Pearson's ϕ^2 of the copulas $\mathbf{C}^A$ and $\mathbf{C}^B$, i.e.,

$$\delta_2(\mathbf{C}^A, \mathbf{C}^B) = \frac{||\mathbf{C}^A - \mathbf{C}^B||_F}{\sqrt{\phi^2(\mathbf{C}^A) + \phi^2(\mathbf{C}^B)}}.$$

As $||\mathbf{C}^A - \mathbf{C}^B||_F \geq 0$, $\delta_2(\cdot,\cdot) \geq 0$ and from $\frac{||\mathbf{C}^A - \mathbf{C}^B||_F^2}{\phi^2(\mathbf{A}^A) + \phi^2(\mathbf{A}^B) - 2\phi^2(C^P)} = 1$ follows $\delta_2(\cdot,\cdot) \leq 1$. This standardization yields values that exhibit less variation with n. The similarity measures are applied to copulas in Section 3.4.

2.6 Some considerations on the link to continuous decompositions

Cuadras and Díaz [12] and Cuadras [8] defined continuous PDF decompositions for continuous copulas. In the following, we briefly expand on the connection between the continuous decomposition and the decomposition of the corresponding checkerboard copulas. Let again C denote the copula CDF, c the copula PDF and

$$c(u, v) = 1 + \sum_{k\geq 1} \lambda_k a_k(u) b_k(v), \tag{18}$$

with complete orthonormal sets $\{a_k\}$ and $\{b_k\}$. Cuadras and Díaz [12] call the cardinality of the set $\{\lambda_k : \lambda_k \geq 0\}$ geometric dimensionality, γ, of the copula, provided that γ is finite, analogously to the discretized case. The decomposition exists if the copula's ϕ^2, i.e., $\phi^2(C) = \sum_k \lambda_k^2$, is finite and induces a decomposition of the copula CDF

$$\begin{aligned} C(u, v) &= \int_0^u \int_0^v \left(1 + \sum_{k=1}^{\gamma} \lambda_k a_k(\bar{u}) b_k(\bar{v})\right) \mathrm{d}\bar{u}\mathrm{d}\bar{v} \\ &= uv + \sum_{k=1}^{\gamma} \int_0^u \int_0^v \lambda_k a_k(\bar{u}) b_k(\bar{v}) \mathrm{d}\bar{u}\mathrm{d}\bar{v}. \end{aligned}$$

The discretized copula of grid size n yields

$$\begin{aligned} C_{ij}^n &= C\left(\frac{i}{n}, \frac{j}{n}\right) - C\left(\frac{i-1}{n}, \frac{j}{n}\right) - C\left(\frac{i}{n}, \frac{j-1}{n}\right) + C\left(\frac{i-1}{n}, \frac{j-1}{n}\right) \\ &= \sum_{k=1}^{\gamma} \lambda_k \int_{\frac{i-1}{n}}^{\frac{i}{n}} a_k(u)\mathrm{d}u \int_{\frac{j-1}{n}}^{\frac{j}{n}} b_k(v)\mathrm{d}v, \end{aligned}$$

and with the additional vectors

$$\mathfrak{a}_k = \left(\int_0^{\frac{1}{n}} a_k(u)\mathrm{d}u, \int_{\frac{1}{n}}^{\frac{2}{n}} a_k(u)\mathrm{d}u, \ldots\right)^\top, \quad (k \in [\gamma]), \tag{19}$$

$$\mathfrak{b}_k = \left(\int_0^{\frac{1}{n}} b_k(v)\mathrm{d}v, \int_{\frac{1}{n}}^{\frac{2}{n}} b_k(v)\mathrm{d}v, \ldots\right)^\top, \quad (k \in [\gamma]), \tag{20}$$

$$\mathbf{C}^n = \sum_{k=1}^{\gamma} \lambda_k \mathfrak{a}_k \mathfrak{b}_k^\top. \tag{21}$$

Note that equation (21) denotes an exact decomposition of $\mathbf{C}^n$, but not necessarily the SVD decomposition. The difference becomes particularly evident if $n < \gamma$ and the summation in (21) has more summands than the dimensionality of the $n \times n$ matrix $\mathbf{C}^n$. To be the SVD, the vectors $\mathfrak{a}_k$ and $\mathfrak{b}_k$ must be left and right singular vectors. Take $\mathfrak{a}_l$ with $l \in [\gamma]$,

$$(\mathbf{C}^n)^\top \mathfrak{a}_l = \left(\sum_{k=1}^{\gamma} \lambda_k \mathfrak{a}_k \mathfrak{b}_k^\top\right)^\top \mathfrak{a}_l \tag{22}$$

$$= \left(\sum_{k=1}^{\gamma} \lambda_k \mathfrak{b}_k \mathfrak{a}_k^\top\right) \mathfrak{a}_l \tag{23}$$

$$= \sum_{k=1}^{\gamma} \lambda_k \mathfrak{b}_k (\mathfrak{a}_k^\top \mathfrak{a}_l). \tag{24}$$

Thus, $\mathfrak{a}_l$ is the left singular vector if it is orthogonal to the other $\{\mathfrak{a}_k : k \in [\gamma], k \neq l\}$ and if $\mathfrak{b}_l$ is a right singular vector. The corresponding singular value is λ_k. The orthogonality condition is trivially fulfilled for $\gamma = 1$ and generally depends on the grid size n. The orthogonal $\{a_k\}$ and $\{b_k\}$ do not induce the orthogonality of the vectors $\{\mathfrak{a}_k : k \in [n]\}$ and $\{\mathfrak{b}_k : k \in [n]\}$.

In addition, the decomposition in equation (21) bounds the geometric dimension of the discretized decomposition by the geometric dimension of the continuous decomposition. The trivial matrix-order bound is $n - 1$. Example 1 shows that this is, indeed, an upper bound and not an equality. A representation with fewer summands could be possible with fewer orthogonal vectors.

Example 1. Let C_1 be a continuous copula with uniform support on the rectangles shown in Figure 1(a). The continuous copula's decomposition has geometric dimension $\gamma(C_1) = 3$. Figure 1(c) shows the geometric dimensions of discretizations of C_1 with various grid sizes. For $n = 4$, the geometric dimension of the discretization is 1, and thus strictly smaller than the continuous geometric dimension and $n - 1$ (Figure 1(b)).

Similar to the decompositions of the continuous copula, the decompositions of the copula CDF do not directly yield decompositions of the PDF. A continuous decomposition of the CDF with d summands is in general form

$$C(u, v) = uv + \sum_{k=1}^{d} \lambda_k F_k(u) G_k(v), \tag{25}$$

with orthogonal $F_1, \ldots, F_d$ and $G_1, \ldots, G_d$. It implies a decomposition of the PDF for differentiable F_k and G_k ($k = 1, \ldots, d$),

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v) = 1 + \sum_{k=1}^{d} \lambda_k \frac{\partial F_k}{\partial u}(u) \frac{\partial G_k}{\partial v}(v),$$

that generally lacks the orthogonality of the function $\partial F_1(u)/\partial u, \partial F_2(u)/\partial u, \ldots$. However, the aforementioned calculation shows that the number of summands for a representation of PDF is, at most, the number of summands of CDF, such that d is an upper bound for the geometric dimension, γ, of a PDF decomposition.

Equations (25) and (18) enable constructing copulas from appropriate $\{\lambda_k\}$, $\{f_k\}$, and $\{g_k\}$. Rodríguez-Lallena [40] formulated conditions on the components to ensure the validity of the resulting copula. Mesiar

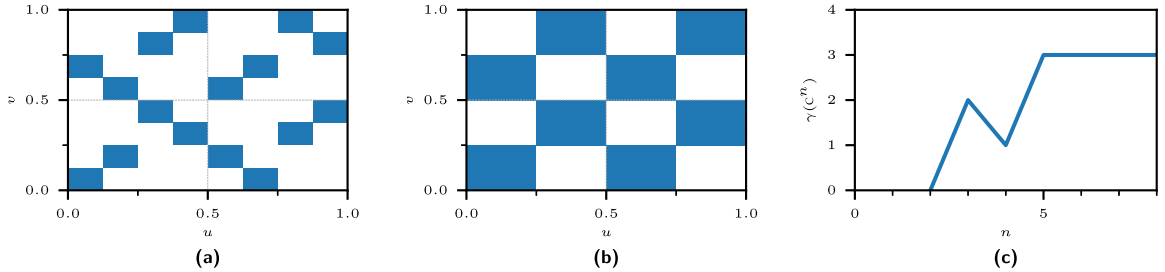


Figure 1: Example for a copula C_1 with corresponding doubly stochastic matrix $\mathbf{C}^n$ that has strictly smaller geometric dimension $\gamma(\mathbf{C}^n)$ than $\min(\gamma(C), n–1)$ for $n = 4$: (a) the continuous copula's mass is uniformly distributed over the blue rectangles, (b) the checkerboard copula's mass is uniformly distributed over the blue rectangles for $n = 4$, and (c) the geometric dimension $\gamma(\mathbf{C}^n)$ for increasing n.

and Najjari [33] extended this construction to higher (finite) dimensions. The construction only yields copulas without tail dependence and thus excludes, for example, the Cuadras-Augé (CA) or Gumbel copula. Instead of estimating all components, some parts in equation (25) can be fixed. Bakam and Pommeret [2], for example, used a Legendre polynomial basis and only fitted the remaining coefficients. In Section 3.3, we show that for (transformed) Hermite polynomials and certain $\{\lambda_k\}$, the Gaussian copula arises. Allowing not only pairs in equation (18) but also the cross products for $a_k(u)b_l(v)$ $(k \neq l)$ in the summation leads to further copula decomposition methods (see, for example, called generalized partition of unity copulas) [30,38].

We give further examples of the difference between continuous and discretized decomposition for the Farlie-Gumbel-Morgenstern (FGM) copula in Section 3.1 and for the Gaussian copula in Section 3.3.

3 Illustrative SVDs of copulas

This section provides the resulting decompositions for some checkerboard approximations of parametric copula families. Section 3.1 focuses on symmetric copulas, whereas Section 3.2 analyzes asymmetric copulas. These sections give examples of the resulting singular values and singular vectors, and we expand on the Frobenius norm-minimizing choice of η in the MAR. At the end of Section 3.1, we provide examples of invalid, i.e., non-copula, truncations, and the use of Algorithm 1. Section 3.3 compares the checkerboard and continuous decomposition, as introduced in Section 2.6, for the Gaussian copula. Section 3.4 applies the similarity measures of Section 2.5 to various checkerboard copulas.

In this section, we will denote the rank of the truncation by $n^* \in \mathbb{N}_0$ and refer to the non-MAR model by *raw* model.

3.1 Decompositions of symmetric copulas

We start with simple copulas with low geometric dimensions and obtain up to high geometric-dimensional copulas with tail dependence in the later examples in this section. The independence copula

$$C^{\Pi}(u, v) = uv$$

yields the checkerboard copula $\mathbf{C}^n = \Pi$ of geometric dimension 0. The comonotonicity copula

$$C^{M}(u, v) = \min(u, v)$$

yields the checkerboard copula $\mathbf{C}^n = I_n$ with geometric dimension $n - 1$. The MAR with $\eta = -1$ fully recovers the matrix for $n^* \geq 0$. Thus, the geometric dimensionality is significantly reduced in the MAR for the comonotonicity copula.

The FGM copula family with CDF

$$C^{\theta,FGM}(u, v) = uv + \theta uv(1 - u)(1 - v)$$

for $\theta \in [-1, 1]$ is of geometric dimensionality 1. Figure 2(a) depicts the first singular vector with respect to n and Figure 2(b) the first singular value, $s_1 = |\theta|/3$ for the continuous representation with respect to θ. The first singular vector is $u_1 = a(1, 1 - 2/(n - 1), \ldots, -1)^\top$ $(a \in \mathbb{R}$ such that $||u_2|| = 1)$, being the checkerboard analog of $a_1(x) = \sqrt{3}(1 - 2x)$ according to equation (19). The singular vector is the piecewise integrated a_1 since the geometric dimension is one. For $n^* = 1$, the MAR following the optimization in (9) has parameters $\eta = s_1/(n - 1) = |\theta|/(3(n - 1))$. A numerical optimization in MATLAB yields numerically equivalent values, as shown in Figure 2(c). The approximation is improved with the MAR, but the gain is smaller than for the comonotonicity copula (Figure 2(d)). The matrix can be fully recovered for any $n^* \geq 1$. The calculation of Spearman's ρ_S according to the representation in equation (10) yields the result for the FGM copula $(n^* \geq 1)$

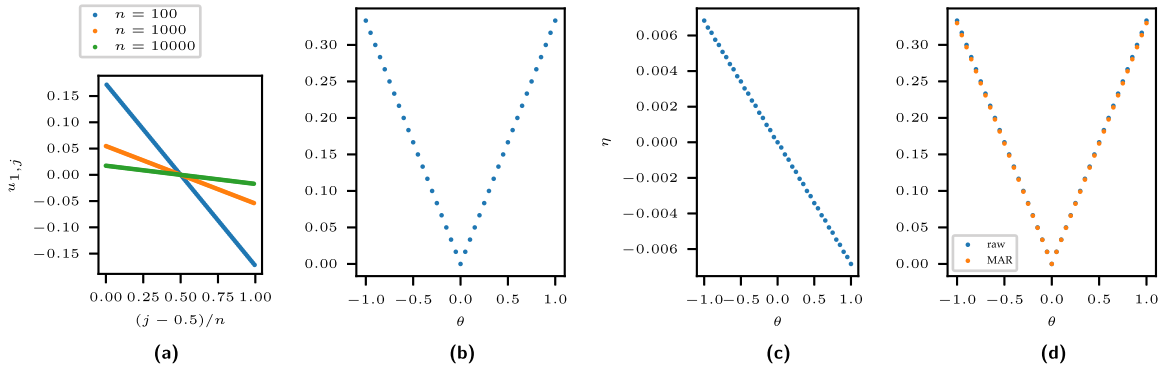


Figure 2: Analysis of the FGM checkerboard copula decompositions using the raw and MAR model: (a) elements $u_{1,j}$ ($j \in [n]$) of the first singular vector u_1 for $\theta = 0.8$ and various grid resolutions n. The same plots arise for other values of $\theta \neq 0$. The different slopes result from the normalization of the singular vector, (b) the first singular value s_1 for various values of θ. The value is, by definition, positive, (c) the values of η in the MAR minimizing the Frobenius error for a 0-truncation, which is only the MAR. The values for η are obtained by numerical minimization using MATLAB's `fminsearch`. The resulting values of η coincide with their theoretical counterparts (see equation (9), and (d) Frobenius error of the MAR and the standard representation for 0-truncations. The values of η are in plot c). The MAR reduces the error slightly.

$$\begin{aligned}\hat{\rho}_S(\hat{C}^{\theta,FGM}) &= (4 - 1/n^2)\sum_{k=1}^{n^*} s_k\langle\omega, u_k\rangle\langle v_k, \omega\rangle \\ &= (4 - 1/n^2)\frac{\theta}{3}\left(\alpha\frac{1}{\|\tilde{\omega}\|}\langle\tilde{\omega}, (1, 1 - 2/n, \ldots, -1)^\top\rangle\right)^2 \\ &= \frac{\theta}{3}(1 - 1/n^2).\end{aligned}$$

The approximated $\hat{\rho}_S$ yields the FGM copula's analytical Spearman's ρ_S of $\theta/3$ for $n \to \infty$.

The CA family of copulas [10] with CDF

$$C^{\theta,\mathrm{CA}}(u, v) = \begin{cases} uv^{1-\theta} & , u \le v, \\ u^{1-\theta}v & , u \ge v, \end{cases}$$

for $\theta \in [0, 1]$ has an upper tail dependency of $\lambda_u = \theta$. The corresponding centered, doubly stochastic matrix is of rank $n - 1$. For $\theta = 0$, $C^{0,\mathrm{CA}} = C^{\Pi}$, while $C^{1,\mathrm{CA}} = C^{M}$ with the decompositions argued earlier. Figure 3 shows the computed singular vectors and values for $\theta \in (0, 1)$. The singular vectors in Figure 3(a) drop near $u = 1$. The decay of singular values starting from s_1 is similar for the different values of θ, but it is shifted upward for higher values of θ, as shown in Figure 3(b). Figure 3(d) shows that the reconstruction is significantly improved when the MAR is used, especially for large θ. For large absolute values of θ, larger absolute values of η in the MAR are chosen (Figure 3(c)).

The Gumbel family of copulas with CDF

$$C^{\theta,\mathrm{Gu}}(u, v) = \exp\left(-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}\right)$$

for $\theta \in [1, \infty)$ is an Archimedean copula and exhibits upper tail dependence like the CA copula. The checkerboard copula contains high values in the upper right part (Figures 4(a) and 5(a)). The Gumbel copula is the independence copula for $\theta = 1$ and approaches the comonotonicity copula for $\theta \to \infty$. The singular vectors in Figure 6(a) contain jumps next to $u = 1$ like the singular vectors for the CA copula. Again, the approximation improves considerably when using the MAR, particularly for higher values of θ, as shown in Figure 6(c) for approximations of rank one or in Figure 6(d) for approximations of rank five. The difference in the MAR and the raw representation approximation reduces when the approximation order increases (Figure 6(b)).

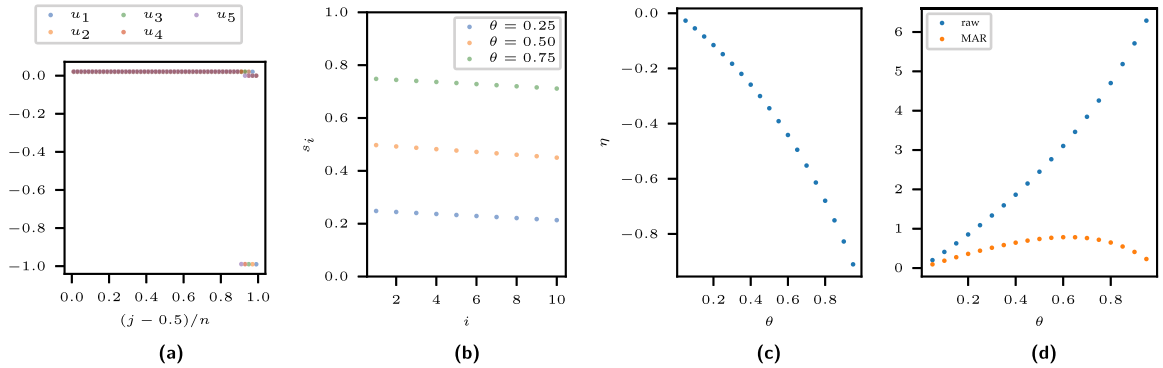


Figure 3: Analysis of the CA checkerboard copula decompositions using the raw and MAR model for various values of θ and $n = 50$: (a) elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the first five singular vectors for $\theta = 0.5$. The singular vectors have a similar course for other values of $\theta \in (0, 1)$, (b) the singular values for $\theta \in \{0.25, 0.5, 0.75\}$, (c) Frobenius-norm minimizing choice of η in the MAR for approximations of rank one, and (d) Frobenius error of the MAR and raw representation for approximations of rank one.

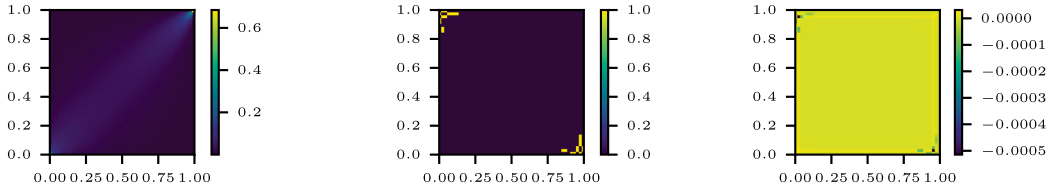


Figure 4: Analysis of the truncation of order 10 of a Gumbel checkerboard copula with $\theta = 2.5$ and $n = 50$: (a) the checkerboard PDF, (b) the yellow squares indicate negative matrix elements in $T_{10}(\mathbf{A}^n[50])$, and (c) difference of approximation and corrected approximation, $G^{-1}(T_{10}(\mathbf{A}^n[50])) - P(G^{-1}(T_{10}(\mathbf{A}^n[50])))$, using Algorithm 1. Note the different scaling compared to (a).

For higher parameter values θ, the truncated representations of the Gumbel copula contain negative entries. Figures 4 and 5 show the discretized PDF, its approximation, and an indicator plot for the invalidity of the elements. The negative elements in Figures 4(b) and 5(b) have a waveform. The Gumbel copula contains higher peaks for higher parameters θ, and the approximation tends to have more negative elements. After applying the correction algorithm, all elements are nonnegative, and the Frobenius distance between the (corrected) approximation and the discretized PDF is smaller (Table 1). The corrections are smaller for small values of θ (Figure 4(c)) than for larger values of θ (Figure 5(c)).

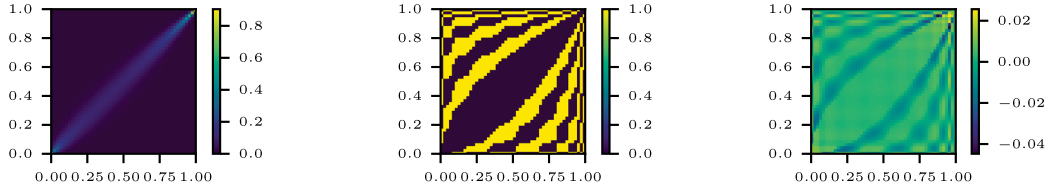


Figure 5: Analysis of the truncation of order 10 of a Gumbel checkerboard copula with $\theta = 7.5$ and $n = 50$: (a) the checkerboard PDF, (b) the yellow squares indicate the negative matrix elements in $T_{10}(\mathbf{A}^n[50])$. They occur more frequently than for $\theta = 2.5$, and (c) difference of approximation and corrected approximation, $G^{-1}(T_{10}(\mathbf{A}^n[50])) - P(G^{-1}(T_{10}(\mathbf{A}^n[50])))$, using Algorithm 1. Note the different scaling compared to (a).

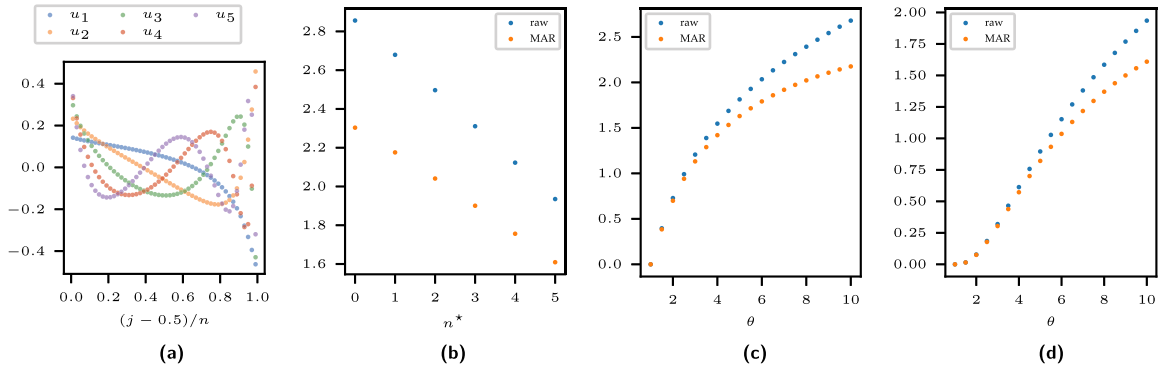


Figure 6: Analysis of the Gumbel checkerboard copula decompositions using the raw and MAR model for $\theta = 10$ and $n = 50$: (a) elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the first singular vectors for $\theta = 10$. The continuous Gumbel copula has an upper tail dependence, (b) the Frobenius error of the approximation for a Gumbel copula with $\theta = 10$ and increasing approximation order n^*. The MAR reduces the error considerably, (c) Frobenius error for approximations of rank one, and (d) Frobenius error for approximations of rank five.

Table 1: Frobenius distances for the approximation of a Gumbel checkerboard copula with parameter θ and $n = 50$

	$\theta = 2.5$	$\theta = 7.5$
$\|\|\mathbf{C}^n[50] - G^{-1}(T_{10}(\mathbf{A}^n[50]))\|\|_F$	0.0084	0.6449
$\|\|\mathbf{C}^n[50] - P(G^{-1}(T_{10}(\mathbf{A}^n[50])))\|\|_F$	0.0084	0.5476
$\|\|G^{-1}(T_{10}(\mathbf{A}^n[50])) - P(G^{-1}(T_{10}(\mathbf{A}^n[50])))\|\|_F$	0.0008	0.3099
$\|\|G^{-1}(T_{10}(\mathbf{A}^n[50])) - P(G^{-1}(T_{10}(\mathbf{A}^n[50])))\|\|_F / \|\|\mathbf{C}^n[50]\|\|_F$	0.05%	11.47%

$G^{-1}(T_{10}(\mathbf{A}^n[50]))$ denotes the truncation, and $P(G^{-1}(T_{10}(\mathbf{A}^n[50])))$ the result of Algorithm 1. The distance between the original and the approximation decreases with the application of Algorithm 1. The last row displays the relative change through Algorithm 1 with respect to the Frobenius norm of the raw matrix $\mathbf{C}^n[50]$.

3.2 Decompositions of asymmetric copulas

For asymmetric copulas, the left and right singular vectors do not coincide. We use an asymmetric construction method from Nelsen [35, p. 84], which yields copulas with cubic sections. The copula CDF is

$$C^{a,b,\,\mathrm{asym}}(u,v) = uv + uv(1-u)(1-v)[(a-b)v(1-u) + b], \tag{26}$$

where $|b| \le 1$, $[b - 3 - (9 + 6b - 3b^2)^{1/2}]/2 \le a \le 1$, and $a \ne b$. The conditions on a and b ensure the validity of the resulting copula. For $a = b \in [-1, 1]$ the FGM copula with parameter b arises. Figure 7 shows the resulting SVD for two configurations of a and b. For $a = 0.5$ and $b = -0.5$ the resulting copula CDF is

$$C^{0.5,-0.5,\,\mathrm{asym}}(u,v) = uv + u(1-u)^2v^2(1-v) - 0.5uv(1-u)(1-v),$$

and for $a = -1.5$ and $b = 0.5$

$$C^{-1.5,0.5,\,\mathrm{asym}}(u,v) = uv - 2uv^2(1-u)^2(1-v) + 0.5uv(1-u)(1-v).$$

In both cases, the left and right singular vectors are the polynomials of degree two. The geometric dimension of the discretized copula is 2. Thus, the singular values in Figure 7(e) drop at 3 to zero. The singular values are larger for the first singular value combination than for the second. The left singular vectors in Figure 7(a) and (c) have similar courses but change order. The right singular vectors (Figure 7(b) and (d)) exhibit a greater variation between the combinations of parameters than the left singular values. They show y-axis mirroring but also change slope and are shifted.

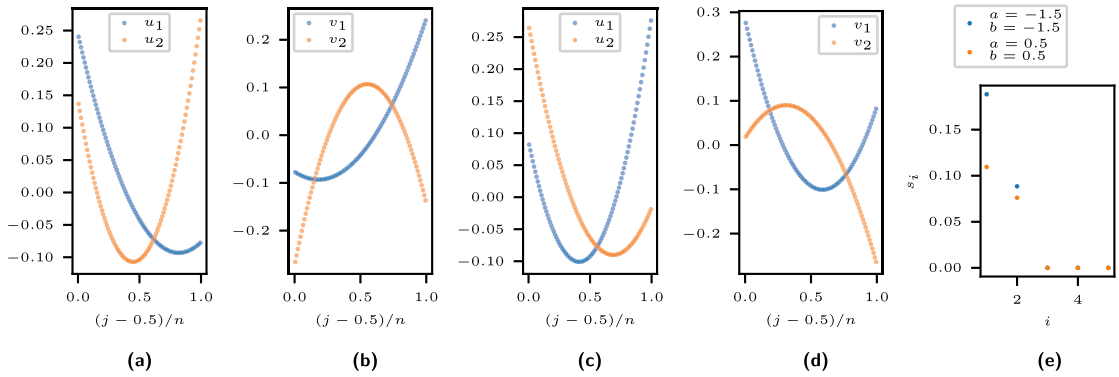


Figure 7: Analysis of the asymmetric checkerboard copula decomposition of the copula following equation (26) with $n = 50$ and two configurations of a and b. The left singular vectors are similar between the two parameter configurations, whereas the right singular values exhibit strong differences: (a) elements $u_{i,j}$ ($i \in [2], j \in [n]$) of the *left* singular vectors u_i with $a = 0.5$ and $b = -0.5$, (b) elements $v_{i,j}$ ($i \in [2], j \in [n]$) of the *right* singular vectors v_i with $a = 0.5$ and $b = -0.5$, (c) elements $u_{i,j}$ ($i \in [2], j \in [n]$) of the *left* singular vectors u_i with $a = -1.5$ and $b = 0.5$, (d) elements $v_{i,j}$ ($i \in [2], j \in [n]$) of the *right* singular vectors v_i with $a = -1.5$ and $b = 0.5$, and (e) the singular values s_i drop to zero after s_2 as the geometric dimension is 2.

3.3 Gaussian copula

We end the section with the Gaussian copula and apply the notions of Section 2.6. The Gaussian copula models the dependence structure of multivariate Gaussian distributions. Let F_ρ denote a bivariate Gaussian CDF with correlation ρ, variance $(1, 1)^\top$ and mean $(0, 0)^\top$, the PDF by f_ρ and the standard univariate Gaussian counterparts by Φ and φ, respectively. Then, the CDF of a Gaussian copula with correlation $\rho \in [-1, 1]$ is given by

$$C^{\rho,Ga} = F_\rho(\Phi^{-1}(u), \Phi^{-1}(v)).$$

Figure 8(a) and (b) shows the resulting (PDF) decompositions for the checkerboard copula. As proven in the following, the singular vectors are identical for different $|\rho| \in (0, 1)$ in the continuous decomposition. No noticeable differences can be observed for the singular vectors of the checkerboard approximations for different values of ρ.

For a bivariate Gaussian distribution, Hill [19] showed a PDF decomposition using Hermite polynomials. The following theorem extends its results to the Gaussian copula, yielding a representation in terms of

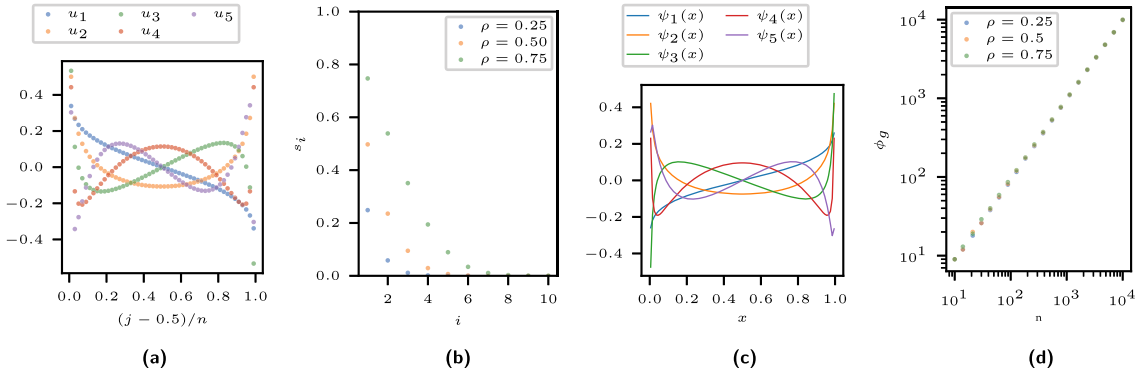


Figure 8: Checkerboard decomposition of the Gaussian family of copulas for $n = 50$, the transformed probabilist's Hermite polynomials, and numerical estimates for the geometric dimension: (a) elements $u_{i,j}$ ($i \in [5], j \in [n]$) of the singular vectors for $\rho = 0.5$. No discernible difference is evident in the plots for the other ρ, (b) the singular values s_i increase for larger values of ρ, (c) the first five (normalized) transformed probabilist's Hermite polynomials ψ_i, and (d) the numerical estimations of the geometric dimensions increase with the grid size and are comparable for the different values of ρ.

transformed Hermite polynomials. We use the representation of the probabilist's Hermite polynomial ψ_i of order i by

$$\psi_i(x) = (-1)^i \exp(x^2/2)\frac{\mathrm{d}^i}{\mathrm{d}x^i}\exp(-x^2/2).$$

Theorem 1. *Let c_ρ be a Gaussian copula density with parameter $-1 < \rho < 1$, Φ the standard Gaussian CDF, and ψ_i the probabilist's Hermite polynomial of order i. Then,*

$$c_\rho(u, v) = 1 + \sum_{i=1}^{\infty} \frac{\rho^i}{i!}\psi_i(\Phi^{-1}(u))\psi_i(\Phi^{-1}(v)), \quad u, v \in (0, 1). \tag{27}$$

Proof.

$$\begin{aligned}
c_\rho(u, v) &= \frac{f_\rho(\Phi^{-1}(u), \Phi^{-1}(v))}{\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v))} && | x \coloneqq \Phi^{-1}(u), y \coloneqq \Phi^{-1}(v) \\
&= \frac{f_\rho(x, y)}{\varphi(x)\cdot\varphi(y)} && | \text{ Hill [26]} \\
&= \frac{\frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2 + y^2)\right]\left\{1 + \sum_{i=1}^{\infty}\frac{\rho^i}{i!}\psi_i(x)\psi_i(y)\right\}}{\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)} \\
&= \sum_{i=0}^{\infty}\frac{\rho^i}{i!}\psi_i(\Phi^{-1}(u))\psi_i(\Phi^{-1}(v)).
\end{aligned}$$

□

Using the well-known maximal correlation property of the Gaussian distribution [25, Section 6 and references therein], the representation in equation (27) is the one obtained by canonical correlation and thus a decomposition in the sense of Section 2.6 [28].

Figure 8(c) shows the first transformed probabilist's Hermite polynomials ψ_i and Figure 8(d) the geometric dimension of Gaussian checkerboard copulas for various grid sizes n. The geometric dimension is bounded by $n - 1$ and increases with n, whereas the continuous Gaussian copula has an infinite geometric dimension. Figure 9 shows the distance between the piecewise integrated transformed Hermite polynomials (equations (19) and (20)) and the singular vectors of the Gaussian checkerboard copula for polynomial degrees 1 to 7. The distance decreases with n for all degrees. The smaller the parameter ρ and the degree, the faster the distance decreases.

3.4 Copula similarities

The difference between copulas can be quantified using the calculated measures of Section 2.5. Figures 10 and 11 show the examples of the similarity of copulas using the (normalized) Frobenius distance of discretizations

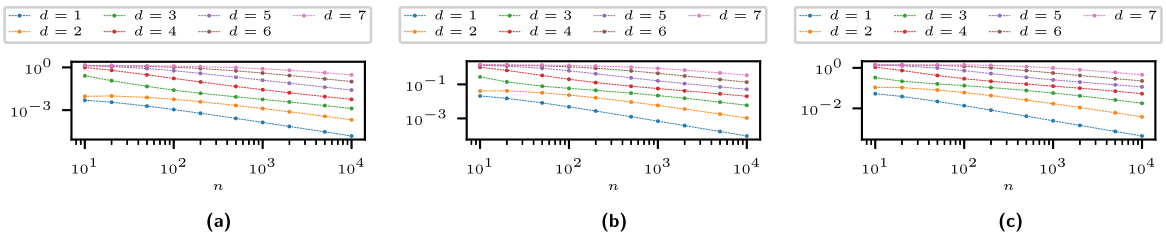


Figure 9: Distance between the d-th piecewise integrated continuous singular vectors and d-th singular vector of the discretized matrix for a Gaussian copula for increasing n and different values of ρ. Distance decreases for all degrees and parameters ρ considered with n: (a) $\rho = 0.25$, (b) $\rho = 0.50$, and (c) $\rho = 0.75$.

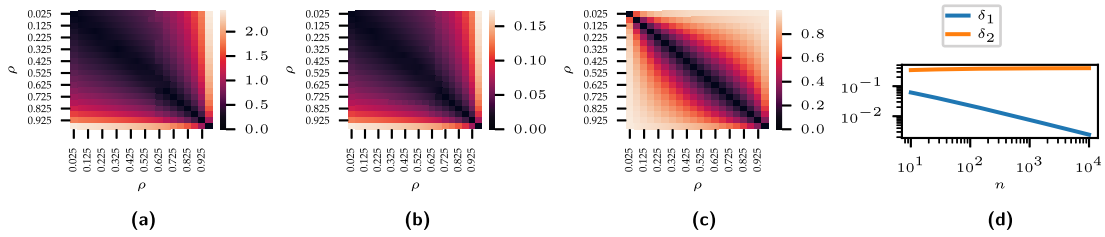


Figure 10: Comparison of the normalizations of Section 2.5 for a Gaussian copula with various copula correlations ρ and $n = 100$. Normalization δ_1 tends to shrink the distance with increasing n: (a) $\|\cdot\|_F$, (b) $\delta_1(\cdot)$. For small ρ, the copulas are assigned similar differences, (c) $\delta_2(\cdot)$, and (d) comparison the normalizations for two Gaussian copulas with $\rho = 0.4$ and $\rho = 0.6$, respectively, with increasing grid size n.

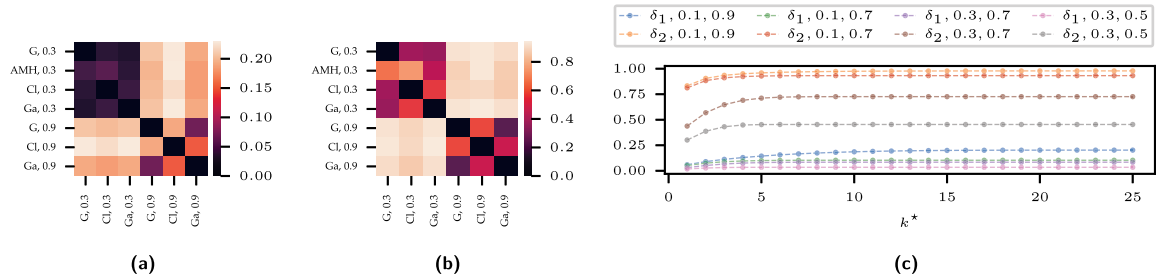
of Section 2.5. Figure 10 shows the distance between Gaussian copulas with different correlations. While Figure 10(a) shows the Frobenius distance, Figure 10(b) and (c) shows the results using the normalizations. In Figure 10(b), the distances are scaled by a common factor, which results in pairs of Gaussian copulas with large ρ being considered more dissimilar than pairs with small ρ with identical difference. The second normalization, δ_2, scales based on Pearson's ϕ^2 of the copulas and yields similar differences for pairs of the same ρ difference. With increasing discretization grid size, n, δ_1 tends to shrink the distance, whereas δ_2 maintains the value (Figure 10(d)). Figure 11 shows the two normalizations for various checkerboard approximations of parametric copulas for two values of τ. Using the Frobenius distance or δ_1 in Figure 11(a) generally produces higher distances if at least one copula has a high τ. Normalization δ_2 produces close distance values for similar values of τ as shown in Figure 11(b) and covers a wider range of possible distances between 0 and 1. Figure 11(c) shows the results of the truncated approximation of δ_1 and δ_2. Most of the pairs of Gaussian copulas considered already approximate the distance for small k^*.

4 Visual exploratory analysis of copulas with profile plots

A primary purpose of correspondence analysis is usually to generate visual representations of high-dimensional data by projecting row and column profiles into a low-dimensional space while maximizing the covered variation of the data (for an introduction, see, e.g., [18]). We start by describing the approach and identifying the characteristics of the copula visible in the graphs, and thus, the characteristics of the graphs to be analyzed in Section 4.1. In Section 4.2, we use empirical data plots from ranked pseudo-observations to analyze the dependence structure.

4.1 Understanding and interpreting profile plots

In profile plots, the similarity of the rows and columns of the checkerboard copula is shown. A row corresponds to the conditional distribution of u given the "row" value of v (where we use the standard notation of u being the horizontal coordinate and v being the vertical coordinate). The row profiles, $\mathbf{F}$, and the column profiles, $\mathbf{G}$, correspond to the singular value-weighted coordinates in the space spanned by the opposing singular vectors, i.e., in the notation of Section 2.2, $\mathbf{F} \coloneqq \mathbf{US}$ and $\mathbf{G} \coloneqq \mathbf{VS}$. All n row and column profiles are shown on the basis of their first two coordinates in the profile plot. Therefore, a row profile shows the two most significant coordinates of the rows with respect to the basis spanned by the columns and *vice versa*. The proximity of different row profiles reflects the similarity of the corresponding conditional distributions of u given the value of v. For example, for independent variables, the distribution of u given v does not change with the value of v, and all profile points in a profile plot would match. In a case with monotone dependence instead, the distribution of u given v changes with v, and the profile points referring to different values of v would not match, and their distance increases with the dissimilarity of the respective conditional distributions.



(a) (b) (c)

Figure 11: Comparison of the normalizations of Section 2.5 for a (G)umbel, (C)layton, and (Ga)ussian copula for two different values of τ and $n = 100$. Kendall's τ values refer to the copula to be discretized. 11(c) uses the truncated representation of δ_1 and δ_2 and shows the computed values for increasing truncation parameter: (a) $\delta_1(\cdot)$. For $\tau = 0.3$, the copulas are assigned similar differences, (b) $\delta_2(\cdot)$, and (c) the truncated, normalized distances δ_1 and δ_2 from Section 2.5 over truncation order $k^\star$ for pairs of Gaussian copulas with $n = 100$.

In a 2D plot of the first two basis vectors, the v value of the row profile is not visible in the coordinates of the points. Thus, we color the profiles to reflect the position of the profile: the lighter the color, the closer the v value is to zero. Thus, the row profile for the conditional distribution u given $v \approx 0$ is the point with the lightest color, and the profile given $v \approx 1$ is the point with the darkest color.

The profiles of rows and columns, $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$, using the MAR can be computed analogously. We compare $\mathbf{F}$ and $\tilde{\mathbf{F}}$ for the Clayton copula below. For symmetric copulas, $\mathbf{F}$ and $\tilde{\mathbf{F}}$ differ only in singular values and not in singular vectors that lead to the same shapes but different profile scalings (Lemma 1). In traditional correspondence analysis, normalizations of the row and column profiles account for the variation of the frequencies of the individual profiles. All profiles have the same frequencies in the copula domain. Therefore, normalizations are not necessary in this setting. Although sometimes plotted in one figure, the distances between a row and a column profile cannot be interpreted directly because the representation is based on a different basis. The profile plots of rows and columns are identical if the underlying copula is symmetric positive definite; i.e., $\mathbf{C}^n$ and $(\mathbf{C}^n)^\top$ are equal, and $\mathbf{C}^n$ is positive definite. The matrix $\mathbf{C}^n$ is, in particular, for exchangeable copulas, symmetric but not necessarily positive definite. A copula is exchangeable if $(F_X(X), F_Y(Y))$ has the same distribution as $(F_Y(Y), F_X(X))$. The plot of several copula profiles in one plot displays differences between copulas.

Figure 12 shows the graphs for some of the copulas of Sections 3.1 and 3.2, visualizing the general characteristics depicted in the profile plot. Profiles of the raw model lying close to the zero point indicate approximately conditionally independent distributions since the most significant coordinates are close to zero. For an independence copula, all profiles lie close to zero. Significant deviations between the components in the raw model graph and the MAR model graph refer to strong characteristics of the comonotonicity copula. Figure 12(a) shows the examples for an independence and in Figure 12(b) and (c) for a comonotonicity copula. Through the points' colors, the plots also display how the profiles evolve and how rapidly the profiles change. Points of similar colors lying close together exhibit a smooth evolution of the copula, whereas varying distances show more extensive changes of the copula in certain areas. Increasing changes are evident, for example, in the case of tail dependence, where the profiles change rapidly in the area of the tail. The plot of the comonotonicity copula in the raw model in Figure 12(b) shows unordered profiles. The comonotonicity copulas SVD is ambiguous since any orthonormal set of vectors forms singular vectors of the diagonal matrix. Thus, the calculated basis is merely random, and the profiles are scattered. For the Gumbel copula, the profiles in Figure 12(d) and (e) evolve smoothly. Still, the differences become larger for higher values of θ and the profiles closer to one since the copula has an upper tail dependence that increases with θ. Using the MAR affects the profiles only slightly in Figure 12(e) as MAR only changes singular values and not singular vectors for symmetric copulas. Figure 12(f) shows the profiles of a Gaussian copula for different values of ρ in one chart. The similarity of profiles changes most pronounced in the tails of the profiles, whereas the profile differences corresponding to middle columns and rows remain similar. Figure 12(g) depicts the row and column profiles of an asymmetric copula, where the profiles do not coincide but are mirrored with respect to the horizontal axis.

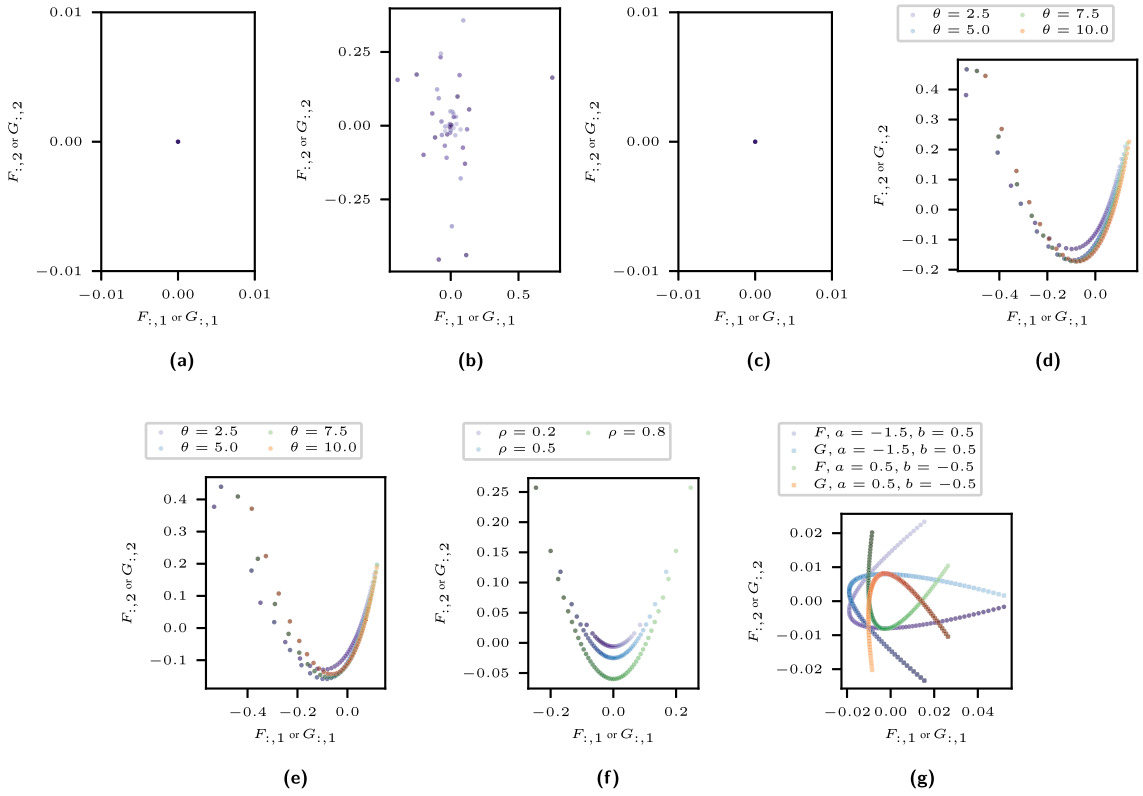


Figure 12: Row and column profiles for four copulas with various parameters, each with grid size $n = 50$. Except for the copula in (g), the displayed copulas are symmetric and have identical row and column profiles. The profiles reflect various copula characteristics, such as the strength of dependence, symmetry for different axes, and areas with high variation: (a) the independence copula in the raw model. All profiles lie close to zero, (b) the comonotonicity copula in the raw model. The profiles are scattered, (c) the comonotonicity copula in the MAR. All profiles lie close to zero, (d) the Gumbel copula in the raw model, (e) the Gumbel copula in the MAR. The axis limits vary slightly compared to (d) and (f) the Gaussian copula in the raw model, and (g) the asymmetric copula according to equation (26). Row and column profiles differ.

4.2 Profile plots illustrated with a data example

Using data from an engineering context, we apply the graphical dependence assessment to empirical data. Coblenz et al. [6] modeled the distribution of fuel drops that are generated by a fuel injector in a jet engine using vine copulas. The droplets are characterized by five variables $x_1, \ldots, x_5$, i.e., the size of the drop, the position in the x and y directions, and the velocity in the x and y directions. Data are generated using numerical simulations under different operating conditions of jet engines, specified by the air velocity, the air pressure, and the thickness of the atomizing edge. Coblenz et al. [6] published statistically simulated data for different operating conditions. We focus on one of the ten operating conditions modeled, i.e., an air velocity of 90 m s^{-1}, an air pressure of 5 bar, and a thickness of the atomizing edge of 230 µm since this is the largest of the provided datasets. It consists of 5,252 points in the five dimensions listed earlier.

The published data of Coblenz et al. [6] are available in the rank-transformed copula domain, which we denote by $u_1, \ldots, u_5$. Note that due to the rank transformation, all values of u_j are in the discrete set $\{1/5{,}252, 2/2{,}525, \ldots, 1\}$. The copula domain's relative frequency table, $\mathbf{C}^n$, is computed by counting the number of points per lattice box in $I^n \times I^n$. Observations lying precisely on a grid boundary are counted for the box below. We use $n = 26$ as a divisor of 5,252 for the analysis so that the resulting table has 202 observations in each row and column, and dividing each cell by 202 leads to a doubly stochastic matrix. For each distinct pair of dimensions, we plot the row profiles, the column profiles, and a checkerboard copula plot of the pseudo-

observations in Figure 13. We focus on five of the pairs here, and the graphs for the combinations (u_1, u_3), (u_1, u_5), (u_2, u_5), (u_3, u_5), and (u_4, u_5) are shown in Appendix D in Figure A1.

As profile points are obtained from empirical data, they deviate from their theoretical counterparts. To visualize statistical noise in the plots, we show the typical minimal and maximal values of profiles for an independence copula by a gray rectangle in the plots. The gray rectangles are obtained by sampling 5,252 realizations from an independence copula and computing their row and column profiles. The procedure is

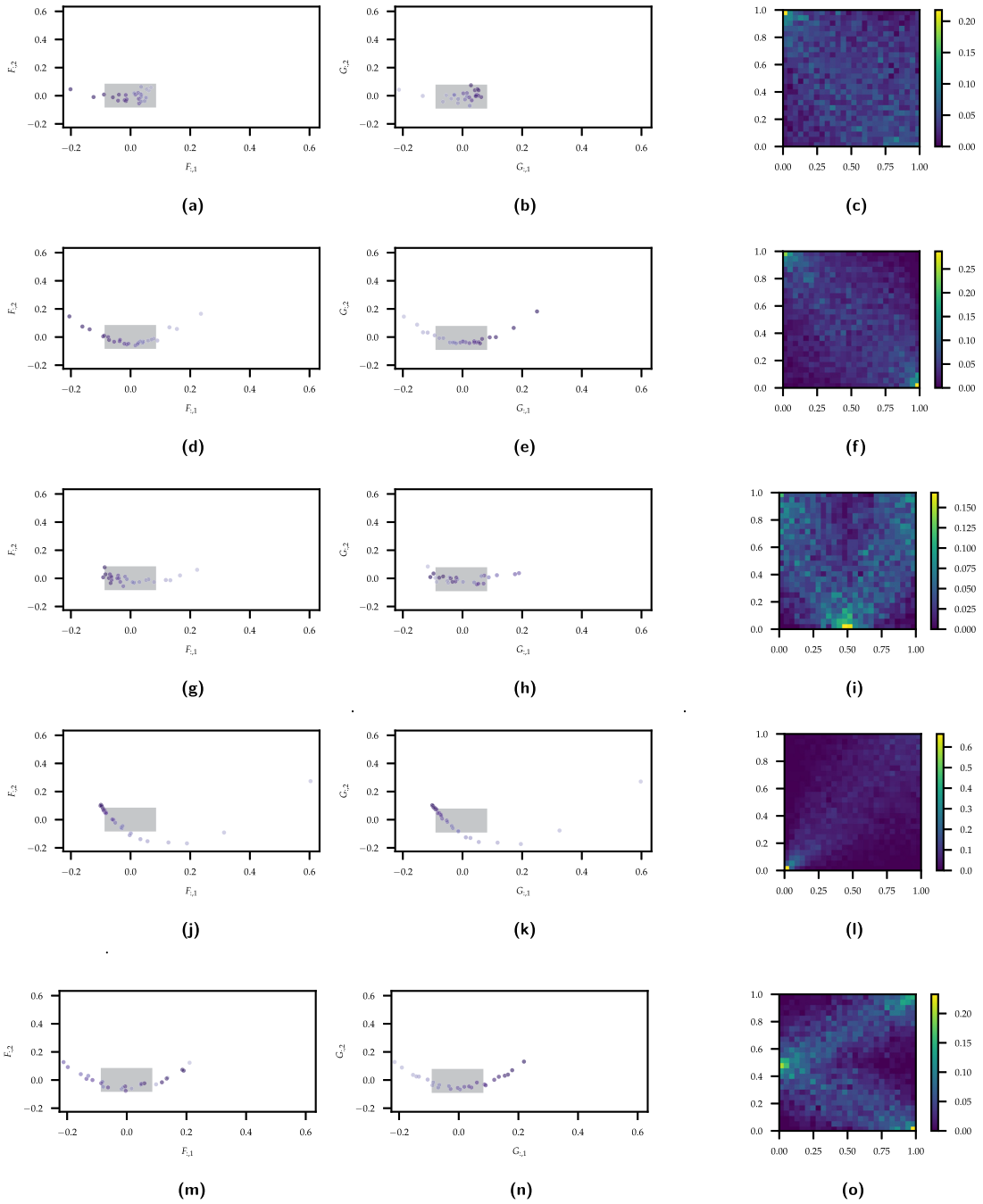


Figure 13: Profile and checkerboard plots of the fuel injector spray characteristics in jet engines from Coblenz et al. [6]. The physical interpretations of the variables are drop size (u_1), x-position (u_2), y-position (u_3), x-velocity (u_4), and y-velocity (u_5): (a) row profiles for variables u_2 and u_2, (b) column profiles for variables u_1 and u_2, (c) checkerboard plot for variables u_1 and u_2, (d) row profiles for variables u_1 and u_4, (e) column profiles for variables u_1 and u_4, (f) checkerboards plot for variables u_1 and u_4, (g) row profiles for variables u_2 and u_3, (h) column profiles for variables u_2 and u_3, (i) checkerboard plot for variables u_2 and u_3, (j) row profiles for variables u_2 and u_4. A profile at (0.60, 0.27) is out of scope, (k) column profiles for variables u_2 and u_4. A profile at (0.60, 0.27) is out of scope, (l) checkerboard copula plot for variables u_2 and u_4, (m) row profiles for variables u_3 and u_4, (n) column profiles for variables u_3 and u_4, and (o) checkerboard plot for variables u_3 and u_4.

repeated 100 times. The rectangles cover 95% of the minimal and maximal point coordinates of the 100 samples in every dimension. Thus, if the profiles are outside the gray box, the underlying copula is unlikely to be the independence copula. This approach aligns with Greenacre [18], who advocates resampling methods, for example, bootstrapping, over using asymptotic results for profile values. Again, the darker the point's color, the closer the conditional distribution's conditioning variable is to one.

The profile plots for variables u_1 and u_2 in Figure 13(a) and (b) show that some profiles deviate from others and that there is a continuous development with the conditioning variable. For row profiles, i.e., conditional distributions u_2 given u_1, with u_1 close to one (dark points in the row profile plot) and column profiles u_2 close to zero (light points in the column profile plot), the profiles change and indicate that the variables are not independent. The checkerboard plot in Figure 13(c) shows a peak at $(0, 1)$. Further information is covered by noise. The profiles in Figure 13(d) and (e) exhibit a U-shaped pattern and are mirrored with respect to the vertical axis. Thus, the dependence has a countermonotonic characteristic that is reflected loosely in the checkerboard copula plot in Figure 13(f). While high values are apparent near $(0, 1)$ and $(1, 0)$, the pattern in between is hard to distinguish. The row and column profiles of Figure 13(g) and (h) differ clearly. Whereas the row profiles evolve in a similar direction with stronger changes near 0, the column profiles undergo a cyclical transformation. The profiles corresponding to small and large values of u_3 are similar, and the profiles for u_3 near 0.5 are different. This pattern is a sign of U or hump-shaped dependence, which is also reflected in the checkerboard plot. For variables u_2 and u_4, Figure 13(j) to (l) show a typical tail-dependence behavior. The profiles change rapidly for small values of u_2 and u_4, whereas they evolve relatively smoothly for larger values. The behavior of the profile plots in Figure 13(m) and (n) is similar to variables u_2 and u_3, but is exchanged. The row profiles undergo a cyclical transformation, while the column profiles evolve smoothly. As the U-shaped form is more apparent than for variables u_1 and u_4, the profiles show a stronger pattern for variables u_3 and u_4. The u-shape is distinguishable in Figure 13(o).

Overall, the row and profile plots provide at least the same amount of information as the checkerboard plots, but they are more transparent and less cluttered than the checkerboard plots.

5 Conclusion

This article analyzes truncations of SVD and correspondence analysis of checkerboard copulas. Checkerboard copulas can be mapped to doubly stochastic matrices, making it straightforward to ensure copula properties for the approximations. We find that some common copulas, for example, comonotonicity-like, have high ranks and thus are poorly represented in the straightforward SVD and that truncations can have negative elements. To account for comonotonicity-like copulas with high ranks, we adapt a representation anchored with the comonotonicity copula and show its performance in examples. We compute the nearest valid doubly stochastic matrix to correct the truncations with negative entries. We analyze the representations of statistical characteristics of copulas, such as Kendall's τ, Spearman's ρ, or differences between copulas through the decomposition. The truncations can be used to compute discretized versions of continuous decompositions, linking our analysis to continuous decompositions. We derive a decomposition of the Gaussian copula into transformed Hermite polynomials and show that the discretized singular vectors draw closer to the transformed Hermite polynomials with increasing grid size. We analyze correspondence analysis profile plots for copulas and show that they reveal asymmetries and non-monotonic dependence. Profile plots for various copulas are shown, and the graphical analysis is illustrated on a dataset on fuel injector spray characteristics in jet engines.

Other approaches for reducing the comonoticity-like characteristics of the copula are possible, such as using rook copulas [7] and, for empirical data, sample-dependent grid sizes [22] or anchoring with respect to other copulas while varying the sample size [13]. They need more complex fitting of the parameters and components and might use different grid sizes. Thus, we leave the comparison of these methods for further research.

In this article, we do not expand on the empirical estimation of the model. It is well known that the empirical checkerboard copula converges to the theoretical checkerboard copula. Perturbation theory

analyzes the effect of noise on the results of the SVD (for a concise overview, see, e.g., [46]). The singular vectors can suffer from large fluctuations for small noise; the singular values, however, are estimated more robustly. Thus, the visual analysis in Section 4 is less prone to noise than plotting the singular vectors directly.

Although the approach can be extended to larger dimensions, it is not straightforward. The concept of the checkerboard copula is viewed in a higher dimension, for example, in Carley and Taylor [5]. There is no direct analog of SVD in three and higher dimensions, but various approaches exist (see, for example, Kolda and Bader [26] for an introduction). Copula-specific methods for modeling high-dimensional data include vine copulas [3,14,23,36] and nested Archimedean copulas [20,42], where the copulas involved could be analyzed using the methods presented here.

Acknowledgements: We thank Johan Segers for his thoughtful and constructive discussions, as well as his insightful feedback, which have greatly enhanced the quality of this article. We extend our gratitude to the two anonymous reviewers whose valuable insights significantly improved this manuscript.

Funding information: We gratefully acknowledge the financial support provided by the Bischöfliche Studienförderung Cusanuswerk to JR and by the KIT publication fund for open access publishing.

Author contributions: Both authors have accepted responsibility for the entire content of this manuscript and consented to its submission to the journal and have reviewed all the results and approved the final version of the manuscript. OG: conceptualization, methodology, writing, supervision. JR: conceptualization, methodology, writing, software, simulation.

Conflict of interest: The authors have no conflicts of interest related to this publication.

Appendix

A Calculations for the algorithms of Section 2.3

We consider the problem

$$P_1(\mathbf{A}) \coloneqq \operatorname*{arg\,min}_{\mathbf{B}\in\mathbb{R}^{n\times n}} \|\mathbf{A}-\mathbf{B}\|_F^2 \quad \text{s.t. } \mathbf{B}\mathbf{1}=\mathbf{1}, \mathbf{B}^\top\mathbf{1}=\mathbf{1},$$

with a symmetric matrix $\mathbf{A}$ in Appendix A.1 and an asymmetric $\mathbf{A}$ in Appendix A.2.

The solution of $P_1(\mathbf{A}_1)$ has a closed form, if the matrix $\mathbf{A}_1$ is symmetric, i.e.,

$$P_1^{\mathrm{sym}}(\mathbf{A}_1) = \mathbf{A}_1 + \left(\frac{1}{n}I_n - \frac{1}{n}\mathbf{A}_1 + \frac{1}{n^2}\mathbf{1}^\top\mathbf{A}_1\mathbf{1}I_n\right)\mathbf{1}\mathbf{1}^\top - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{A}_1.$$

In the case of an asymmetric matrix $\mathbf{A}_1$, the problem $P_1(\mathbf{A}_1)$ boils down to a linear system (Appendix A.2).

For $P_2(\mathbf{A}_2)$,

$$\begin{aligned} P_2(\mathbf{A}_2) = \operatorname*{arg\,min}_{\mathbf{B}\in\mathbb{R}^{n\times n}} \ & \|\mathbf{A}_2-\mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}_{i,j} \geq 0, \quad \forall i,j\in[n], \end{aligned}$$

there exists a closed-form solution independent of the symmetry of $\mathbf{A}_2$. As the Frobenius norm can be minimized elementwise,

$$\|\mathbf{B}-\mathbf{A}_2\|_F^2 = \sum_{i,j:(\mathbf{A}_2)_{ij}<0} (\mathbf{B}-\mathbf{A}_2)_{ij}^2 + \sum_{i,j:(\mathbf{A}_2)_{ij}\geq 0} (\mathbf{B}-\mathbf{A}_2)_{ij}^2 \quad B\in\mathbb{R}^{n\times n},$$

the solution of $P_2(\mathbf{A}_2)$ is the elementwise positive part of A_2. Algorithm 1 combines P_1 and P_2.

A.1 Symmetric copula

The proof is analogous to Zass and Shashua [47] for symmetric $\mathbf{A}$. We provide it here for completeness and to emphasize its inapplicability to asymmetric matrices. Let $\mathbf{A} = \mathbf{A}^\top$ and $P_1^*(\mathbf{A})$ be the relaxation

$$P_1^*(\mathbf{A}) \coloneqq \underset{\mathbf{B}\in\mathbb{R}^{n\times n}}{\arg\min} \|\mathbf{A}-\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \mathbf{B1}-\mathbf{1}+\mathbf{B}^\top\mathbf{1}-\mathbf{1}=0.$$

If $\mathbf{B}=\mathbf{B}^\top, \mathbf{B1}-\mathbf{1}+\mathbf{B}^\top\mathbf{1}-\mathbf{1}=0 \Rightarrow \mathbf{B1}=\mathbf{1}=\mathbf{B}^\top\mathbf{1}$. Thus, if $\mathbf{B}\coloneqq P_1^*(X)=\mathbf{B}^\top$, $\mathbf{B}$ is also the solution of $P_1(X)$. Let $L(\mathbf{A},\mu)$ be the Lagrangian of $P_1^*(\mathbf{A})$ with

$$L(\mathbf{B},\mu)=\operatorname{trace}(\mathbf{B}^\top\mathbf{B})-\operatorname{trace}(2\mathbf{A}^\top\mathbf{B})-2\mu^\top(\mathbf{B1}+\mathbf{B}^\top\mathbf{1}-2\mathbf{1}).$$

Then,

$$\begin{aligned}\frac{\partial L(\mathbf{B},\mu)}{\partial \mathbf{B}} &= 2\mathbf{B}-2\mathbf{A}-2\mu\mathbf{1}^\top-2\mathbf{1}\mu^\top \overset{!}{=} 0 \Leftrightarrow \\ \mathbf{B} &= \mathbf{A}+\mu\mathbf{1}^\top+\mathbf{1}\mu^\top \qquad \Leftrightarrow |(\cdot)^\top \\ \mathbf{B}^\top &= \mathbf{A}^\top+\mathbf{1}\mu^\top+\mu\mathbf{1}^\top\end{aligned} \tag{A1}$$

and

$$\begin{aligned}\mathbf{B}+\mathbf{B}^\top &= (\mathbf{A}+\mathbf{A}^\top)+2\cdot\mu\mathbf{1}^\top+2\cdot\mathbf{1}\mu^\top \quad \Leftrightarrow \\ \underbrace{\mathbf{B1}+\mathbf{B}^\top\mathbf{1}=2\cdot\mathbf{1}}_{\text{follows from } \frac{\partial L(\mathbf{B},\mu)}{\partial\mu}=0} &= (\mathbf{A}+\mathbf{A}^\top)\mathbf{1}+2\cdot\mu\underbrace{\mathbf{1}^\top\mathbf{1}}_{=n}+2\cdot\mathbf{1}\mu^\top\mathbf{1} \Leftrightarrow \ldots\end{aligned} \tag{A2}$$

$$\mu=\frac{1}{n}\left(I_n-\frac{1}{2n}\mathbf{1}\mathbf{1}^\top\right)\left(I_n-\frac{1}{2}(\mathbf{A}+\mathbf{A}^\top)\right)\mathbf{1}, \tag{A3}$$

using $(nI_n+\mathbf{1}\mathbf{1}^\top)^{-1}=\frac{1}{n}(I_n+\frac{1}{2n}\mathbf{1}\mathbf{1}^\top)$. Plugging μ from equation (A3) into (A1) yields

$$\begin{aligned}\mathbf{B} &= \mathbf{A}+\left(\frac{1}{n}\left(I_n-\frac{1}{2n}\mathbf{1}\mathbf{1}^\top\right)\left(I_n-\frac{1}{2}(\mathbf{A}+\mathbf{A}^\top)\right)\mathbf{1}\right)\mathbf{1}^\top+\mathbf{1}\left(\frac{1}{n}\left(I_n-\frac{1}{2n}\mathbf{1}\mathbf{1}^\top\right)(I_n-\frac{1}{2}(\mathbf{A}+\mathbf{A}^\top))\mathbf{1}\right)^\top \\ &= \mathbf{A}+\left(\frac{1}{n}I_n-\frac{1}{2n}(\mathbf{A}+\mathbf{A}^\top)+\frac{1}{2n^2}\mathbf{1}^\top(\mathbf{A}+\mathbf{A}^\top)\mathbf{1}I_n\right)\mathbf{1}\mathbf{1}^\top-\frac{1}{2n}\mathbf{1}\mathbf{1}^\top(\mathbf{A}+\mathbf{A}^\top) \\ &\overset{\mathbf{A}=\mathbf{A}^\top}{=} \mathbf{A}+\left(\frac{1}{n}I_n-\frac{1}{n}\mathbf{A}+\frac{1}{n^2}\mathbf{1}^\top\mathbf{A}\mathbf{1}I_n\right)\mathbf{1}\mathbf{1}^\top-\frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{A}.\end{aligned} \tag{A4}$$

The result for $\mathbf{B}$ in equation (A4) is symmetric, and thus, also solution for $P_1(\mathbf{A})$:

$$\begin{aligned}\mathbf{B}^\top &= \left(\mathbf{A}+\left(\frac{1}{n}I_n-\frac{1}{n}\mathbf{A}+\frac{1}{n^2}\mathbf{1}^\top\mathbf{A}\mathbf{1}I_n\right)\mathbf{1}\mathbf{1}^\top-\frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{A}\right)^\top \\ &= \mathbf{A}+\left(\frac{1}{n}I_n-\frac{1}{n}\mathbf{A}+\frac{1}{n^2}\mathbf{1}^\top\mathbf{A}\mathbf{1}I_n\right)\mathbf{1}\mathbf{1}^\top-\frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{A} \\ &= \mathbf{B},\end{aligned}$$

using

$$(\mathbf{A}\mathbf{1}\mathbf{1}^\top)_{ij}=\left(\sum_{k=1}^n \mathbf{A}_{ik}\right)=\left(\sum_{k=1}^n \mathbf{A}_{kj}\right)=(\mathbf{1}\mathbf{1}^\top\mathbf{A})_{ij}, \quad i,j\in[n].$$

Algorithm 1: Algorithm to compute the nearest doubly stochastic matrix in terms of the Frobenius error following Zass and Shashua [47] for symmetric matrices $\mathbf{A}$ and $\varepsilon > 0$. The stopping criterion $c \le c_{\max}$ ensures the termination of the algorithm. The solutions of P_1 and P_2 can be found in Appendix A.

input Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\varepsilon > 0$, number of maximal iterations $c_{\max}$
output: nearest doubly stochastic matrix $\mathbf{B}$
Set $\mathbf{B} = \mathbf{A}$ and $c = 1$;
Update $\mathbf{B} = P_1(\mathbf{B})$;
while $\exists\, i, j : \mathbf{B}_{i,j} < -\varepsilon \wedge c \le c_{\max}$ **do**
  Update $\mathbf{B} = P_2(\mathbf{B})$;
  Update $\mathbf{B} = P_1(\mathbf{B})$;
  Update $c = c + 1$;
end

A.2 Asymmetric copula

For asymmetric $\mathbf{A}$, the result of $P_1^*(\mathbf{A})$ is not symmetric and thus is not a solution to the original problem $P_1(\mathbf{A})$.

Instead, the solution of a Karush-Kuhn-Tucker equation system yields the solution for P_1. The problem

$$P_1(\mathbf{A}) \coloneqq \underset{\mathbf{B}\in\mathbb{R}^{n\times n}}{\arg\min} \|\mathbf{A} - \mathbf{B}\|_F^2 \quad \text{s.t. } \mathbf{B1} = \mathbf{1}, \mathbf{B}^\top\mathbf{1} = \mathbf{1}$$

with the Lagrange function and its derivative

$$L(\mathbf{B}, \lambda, \mu) = \operatorname{trace}(\mathbf{B}^\top\mathbf{B}) - \operatorname{trace}(2\mathbf{A}^\top\mathbf{B}) - \lambda^\top(\mathbf{B1} - \mathbf{1}) - \mu^\top(\mathbf{B}^\top\mathbf{1} - \mathbf{1})$$
$$\frac{\partial L}{\partial \mathbf{B}_{ij}}(\mathbf{B}, \lambda, \mu) = 2\mathbf{B}_{ij} - 2\mathbf{A}_{ij} - \lambda_j - \mu_i, \quad \text{for } i, j \in [n]$$

yields the system

$$2\mathbf{B}_{ij} - \lambda_j - \mu_i = 2\mathbf{A}_{ij}, \quad \forall (i,j) \in [n]\times[n], \tag{A5}$$

$$\sum_{i=1}^{n} \mathbf{B}_{ij} = 1, \quad \forall j \in [n], \tag{A6}$$

$$\sum_{j=1}^{n} \mathbf{B}_{ij} = 1, \quad \forall i \in [n]. \tag{A7}$$

The solution of the Karush-Kuhn-Tucker equation system is the solution of the linear equation system $Kb = a$ with

$$\tilde{K} = \begin{pmatrix} 2I_n[n^2 \times n^2] & [\mathbf{1} \otimes I_n[n\times n]] & [I_n[n\times n] \otimes \mathbf{1}] \\ [\mathbf{1}^\top \otimes I_n[n\times n]] & 0 & 0 \\ [I_n[n\times n] \otimes \mathbf{1}^\top] & 0 & 0 \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} \vec{\mathbf{B}} \\ \lambda \\ \mu \end{pmatrix}, \quad \tilde{a} = \begin{pmatrix} 2\vec{A} \\ 0^{n\times 1} \\ 0^{n\times 1} \end{pmatrix}$$

and $K = \tilde{K}_{1:n^2+2n-1,1:n^2+2n-1}$, $b = \tilde{b}_{1:n^2+2n-1}$, and $a = \tilde{a}_{1:n^2+2n-1}$. Thereby, $\otimes$ denotes the Kronecker product of the matrices, and $\vec{\cdot}$ denotes the column-wise stacking of a matrix into a vector. The last row and column are excluded, as the matrix K is singular and the constraint $\sum_{j=1}^{n}\mathbf{B}_{nj} = 1$ is guaranteed by the remaining constraints (A6) and (A7). Then, the first n^2 elements of the solution b rearranged as matrix $\mathbf{B}$ are the solution of P_1.

Additionally, Algorithm 1 includes a deflection component to account for the more general setting [17,47], as summarized in Algorithm 2.

Algorithm 2: Algorithm to compute the nearest doubly stochastic matrix in terms of the Frobenius error following Zass and Shashua [47] for asymmetric matrices $\mathbf{A}$ and $\varepsilon > 0$. The stopping criterion $c \leq c_{\max}$ ensures the termination of the algorithm. The solutions of P_1 and P_2 can be found in A.2. $\mathbf{B}_{c,d}$ denotes the solution in iteration c in the subproblem d and $I_{c,d}$ the corresponding incremental change

input: Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\varepsilon > 0$, number of maximal iterations $c_{\max}$
output: Nearest doubly stochastic matrix $\mathbf{B}$

Set $\mathbf{B} = \mathbf{A}$ and $c = 1$;
Set $\mathbf{B}_{1,1} = P_1(\mathbf{B})$ and $I_{1,1} = \mathbf{B}_{1,1} - \mathbf{B}$;
Set $\mathbf{B}_{1,2} = P_2(\mathbf{B} + I_{1,1})$ and $I_{1,2} = \mathbf{B}_{c,2} - (\mathbf{B} + I_{c,1})$;
repeat
  Update $c = c + 1$;
  Set $\mathbf{B}_{c,1} = P_1(\mathbf{B} + I_{c-1,2})$ and $I_{c,1} = \mathbf{B}_{c,1} - (\mathbf{B} + I_{c-1,2})$;
  Set $\mathbf{B}_{c,2} = P_2(\mathbf{B} + I_{c,1})$ and $I_{c,2} = \mathbf{B}_{c,2} - (\mathbf{B} + I_{c,1})$;
$(\mathbf{B} + I_{c,1} + I_{c,2}) \geq -\varepsilon \wedge |(\mathbf{B} + I_{c,1} + I_{c,2})\mathbf{1} - \mathbf{1}| < \varepsilon \wedge |(\mathbf{B} + I_{c,1} + I_{c,2})^\top \mathbf{1} - \mathbf{1}| < \varepsilon \wedge c \leq c_{\max}$

B Decomposition in terms of the Hellinger distance

The SVD and the algorithms of Section 2.3 yield minimal errors in terms of the Frobenius norm. The SVD is also the best low-rank approximation considering the spectral norm [34]. In statistics, the Hellinger distance is often used to assess the proximity of densities (see [1,32], for two recent contributions). In this section, we analyze Hellinger distance-based decompositions for two different versions of the Hellinger distance for matrices, as, to our knowledge, there is no agreed definition in the matrix case yet. For a matrix square root-based Hellinger distance, the decomposition generalizes from the Frobenius case, while it is of a different and more complicated structure for an elementwise square root Hellinger distance.

For discrete probability distributions $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$, the Hellinger norm d_H is computed by

$$d_H(p, q) = \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2 \right]^{1/2} = \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} (p_i + q_i) - 2 \sum_{i=1} \sqrt{p_i q_i} \right]^{1/2}.$$

For matrices, there are different notions of the Hellinger distance in the literature. We consider a formulation based on the matrix square root first [4] and then turn to an elementwise square root method [39].

Bhatia et al. [4] started from the decomposition of the Hellinger distance for densities into an arithmetic and geometric mean. As the geometric mean for matrices can be interpreted in various ways, different notions of the distance can be obtained. We use their most straightforward generalization yielding the Hellinger distance for positive semidefinite, and thus, symmetric, matrices A and B

$$d_H(A, B) = ||A^{1/2} - B^{1/2}||_F = [\text{trace}(A + B) - 2\text{trace}(A^{1/2}B^{1/2})]^{1/2}.$$

Thereby, $A^{1/2}$ denotes the matrix square root with $(A^{1/2})^\top A^{1/2} = A^{1/2}A^{1/2} = A$.

Lemma 3. *The low-rank approximation problem of a positive definite matrix A yields the same eigenvectors and eigenvalues for the Frobenius and the Hellinger distance.*

Proof. Let $P_F(A, k)$ the low-rank approximation problem in terms of the Frobenius distance

$$P_F(A, k) \coloneqq \min_{B \in \mathbb{R}^{n \times n}} ||A - B||_F \quad \text{s.t. } \operatorname{rank}(B) \leq k$$

and $P_H(A, k)$ in terms of the Hellinger distance

$$P_H(A, k) = \min_{B \in \mathbb{R}^{n \times n}} ||A^{1/2} - B^{1/2}||_F \quad \text{s.t. } \operatorname{rank}(B) \leq k, B \text{ positive semidefinite}, \tag{A8}$$

as the matrix square root is unique and $\operatorname{rank}(B) = \operatorname{rank}(B^{1/2})$ [21, Theorem 7.2.6].

Thus, the minimizing argument, $B^{1/2}$, of Problem $P_H(A, k)$ is the k-truncated SVD of $A^{1/2}$. Due to the positive definiteness of A, left and right singular values are identical, and the eigenvalue decomposition

$$A^{1/2} \coloneqq U\Lambda(U)^\top \tag{A9}$$

exists with eigenvector matrix U and diagonal eigenvalue matrix Λ. This yields for the matrix square root of the minimizing argument, $B^{1/2}$, of $P_H(A, k)$ and the minimizing argument, B,

$$B^{1/2} = U_{[k]}\Lambda_{[k]}(U_{[k]})^\top \quad \text{and} \quad B = B^{1/2}B^{1/2} = U_{[k]}\Lambda^2_{[k]}(U_{[k]})^\top.$$

The eigenvectors of $A^{1/2}$ and A are identical and the singular vectors are squared, as

$$A = A^{1/2}A^{1/2} = (U\Lambda U^\top)(U\Lambda U^\top)^\top = U\Lambda^2 U^\top.$$

Thus, the minimizing argument of $P_F(A, k)$ is

$$U_{[k]}\Lambda^2_{[k]}(U_{[k]})^\top$$

and equal to the minimizing argument of $P_H(A, k)$. □

This definition of the Hellinger distance obtains the same decomposition as with the Frobenius distance. The coefficient η in the MAR can also be computed with the Hellinger distance instead of the Frobenius distance in equation (8). However, the definition of the Hellinger distance only for positive semidefinite copulas is restrictive, as valid bistochastic matrices do not need to be positive definite, for example, unsymmetric.

Rao [39] and Cuadras and Cuadras [11] defined the Hellinger distance in terms of an elementwise square root, thus considering only matrices with non-negative elements. Let $\sqrt{A}$ denote the elementwise square root of a matrix A. Then, the decomposition based on the elementwise Hellinger decomposition is for a symmetric checkerboard copula $\mathbf{C}^n$

$$\sqrt{\mathbf{C}^n} = U^H\Lambda^H(U^H)^\top. \tag{A10}$$

Truncations $T_{n^*}(\sqrt{\mathbf{C}^n})$ have to be squared elementwise to obtain a low-rank approximation of the Hellinger decomposition. Note that the squared decomposition does not keep the rank of $T_{n^*}(\sqrt{\mathbf{A}^n})$. The MAR could be used in the elementwise Hellinger decomposition, and the optimization in equation (8) could be adapted and solved by a general optimization problem solver. To our knowledge, no optimizations similar to those in the Frobenius case are available for the elementwise Hellinger scenario, as either the objective function contains square roots or the constraints are non-linear. While the elements of the squared decomposition are non-negative, the row and column sums are not one, in general, and thus, the squared decomposition is not doubly stochastic.

All in all, the elementwise Hellinger decomposition is not as straightforward as the Frobenius decomposition, as the squared decomposition does not keep the rank of the truncation, and the attached optimization problems obtain more complex. Through the elementwise square root, the influence of peaks in the checkerboard copula on the objective function is reduced compared to the Frobenius case. It is a modeling choice, whether this is desired or not. Rao [39] and Cuadras and Cuadras [11] pointed out that elementwise Hellinger-based decomposition's main advantage is the independence from the row and column marginals. However, the marginals are constant in the checkerboard copula setting; thus, the correspondence analysis does not depend on them. Thus, we do not expand on the Hellinger decompositions in the main part of this article.

C Computations for Spearman's ρ and Kendall's τ in Section 2.4

As in Section 2.4, $\mathbf{A}^n = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ denotes the centered copulas SVD. Let additionally be $\mathbf{u}_0 = \mathbf{v}_0 = 1/\sqrt{n} \cdot \mathbf{1}$ and $s_0 = 1$ to ease the notation. The equations for Spearman's ρ in Section 2.4 follow from

$$\begin{aligned}
\rho(\mathbf{C}^n) &\coloneqq \frac{3}{n}\operatorname{trace}(\mathbf{\Omega}\mathbf{C}^n) - 3 \\
&= \frac{3}{n}\operatorname{trace}\left(\Omega \sum_{k=0}^{n-1} \mathbf{u}_k s_k \mathbf{v}_k^\top\right) - 3 \\
&= \frac{3}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}(\Omega \mathbf{u}_k \mathbf{v}_k^\top) - 3 \\
&= \frac{3\|\check{\boldsymbol{\omega}}\|^2}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}(\boldsymbol{\omega}\boldsymbol{\omega}^\top \mathbf{u}_k \mathbf{v}_k^\top) - 3 \\
&= \frac{3 \cdot (\frac{4}{3}n - \frac{1}{3n})}{n}\sum_{k=0}^{n-1} s_k \operatorname{trace}(\boldsymbol{\omega}^\top \mathbf{u}_k \mathbf{v}_k^\top \boldsymbol{\omega}) - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\sum_{k=0}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\left(\langle * \rangle \boldsymbol{\omega}, \frac{1}{\sqrt{n}} \cdot \mathbf{1}^{\,2} + \sum_{k=1}^{n^*} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle\right) - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle + 3 - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\sum_{k=1}^{n-1} s_k \langle \boldsymbol{\omega}, \mathbf{u}_k\rangle\langle \mathbf{v}_k, \boldsymbol{\omega}\rangle.
\end{aligned}$$

Similarly, follows for the MAR decomposition, $\tilde{\mathbf{A}}^n = \tilde{\mathbf{U}}\tilde{\mathbf{S}}(\tilde{\mathbf{V}})^\top$,

$$\begin{aligned}
\rho(\mathbf{C}^n) &= \frac{3}{n}\operatorname{trace}\left(\mathbf{\Omega}\left(\sum_{k=1}^{n-1} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\top - \eta I_n + (1+\eta)\mathbf{\Pi}^n\right)\right) - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1} \tilde{s}_k \operatorname{trace}(\boldsymbol{\omega}\boldsymbol{\omega}^\top \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^\top) - \eta \operatorname{trace}(\boldsymbol{\omega}\boldsymbol{\omega}^\top I_n) + (1+\eta)\operatorname{trace}(\boldsymbol{\omega}\boldsymbol{\omega}^\top \mathbf{\Pi}^n)\right] - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1} \tilde{s}_k \langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle \tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - \eta\|\boldsymbol{\omega}\|^2 + (1+\eta)\langle * \rangle \boldsymbol{\omega}, \frac{1}{\sqrt{n}} \cdot \mathbf{1}^{\,2}\right] - 3 \\
&= \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1} \tilde{s}_k \langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle\langle \tilde{\mathbf{v}}_k, \boldsymbol{\omega}\rangle - \left(4 - \frac{1}{n^2}\right)\eta + \frac{3}{n}(1+\eta)\right] - 3 \\
&\overset{(*)}{=} \left(4 - \frac{1}{n^2}\right)\left[\sum_{k=1}^{n-1} (s_k + \eta)\langle \tilde{\mathbf{u}}_k, \boldsymbol{\omega}\rangle^2 - \left(4 - \frac{1}{n^2}\right)\eta + \frac{3}{n}(1+\eta)\right] - 3,
\end{aligned}$$

where (*) is only valid for symmetric $\tilde{\mathbf{A}}^n$.

The respective computations for Kendall's τ are

$$\begin{aligned}
\tau(\mathbf{C}^n) &\coloneqq 1 - \frac{1}{n^2}\operatorname{trace}(\mathbf{E}\mathbf{C}^n\mathbf{E}(\mathbf{C}^n)^\top) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1} s_{k_2} \operatorname{trace}(\mathbf{E}\mathbf{u}_{k_1}\mathbf{v}_{k_1}^\top \mathbf{E}\mathbf{v}_{k_2}\mathbf{u}_{k_2}^\top) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1} s_{k_2} \operatorname{trace}(\mathbf{u}_{k_2}^\top \mathbf{E}\mathbf{u}_{k_1}\mathbf{v}_{k_1}^\top \mathbf{E}\mathbf{v}_{k_2}) \\
&= 1 - \frac{1}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} s_{k_1} s_{k_2} \langle \mathbf{u}_{k_2}, \mathbf{E}\mathbf{u}_{k_1}\rangle\langle \mathbf{v}_{k_1}, \mathbf{E}\mathbf{v}_{k_2}\rangle
\end{aligned}$$

and for the MAR analogously.

D Further figures for Section 4.2

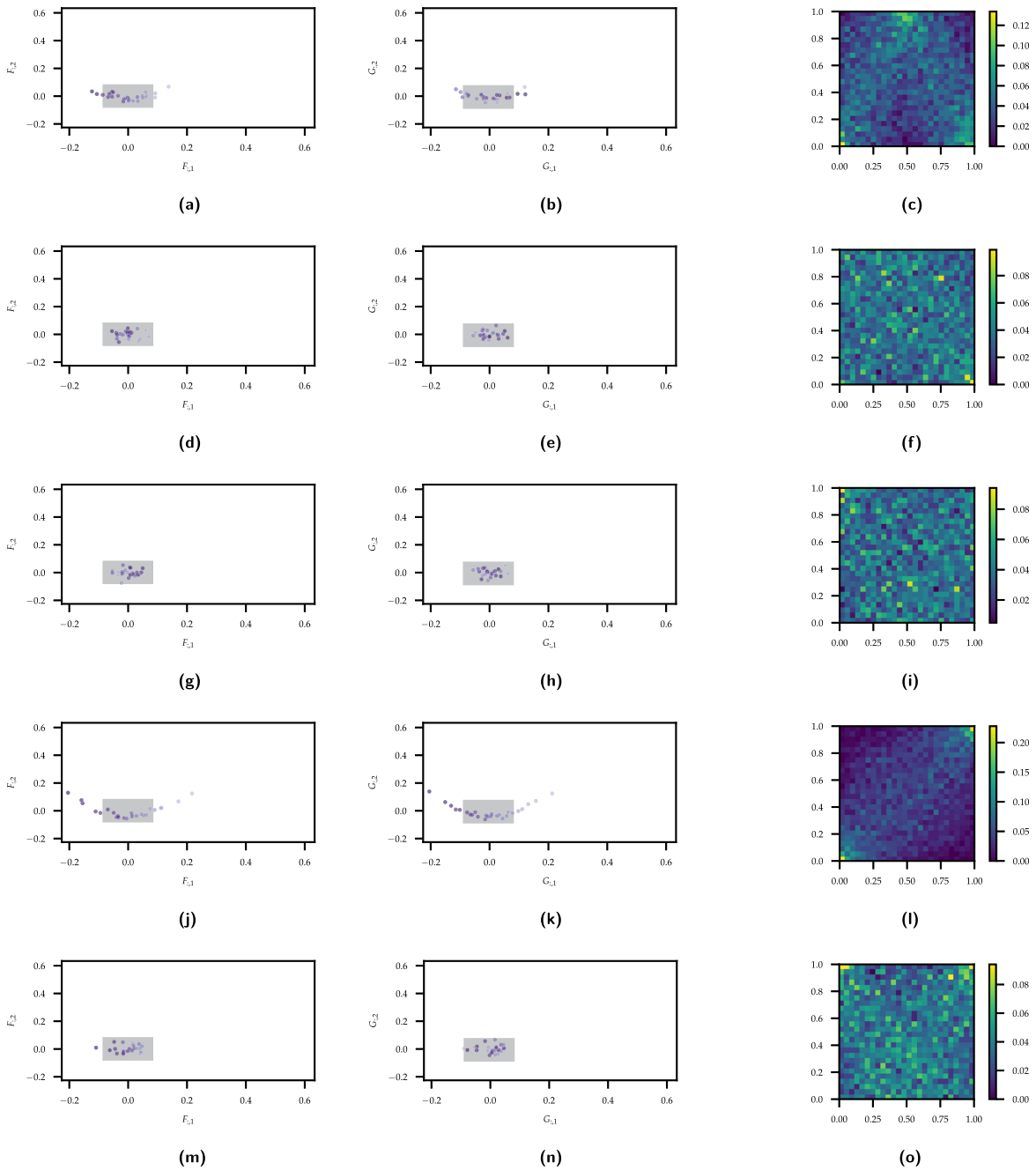


Figure A1: Remaining profile and checkerboard plots of the fuel injector spray characteristics in jet engines from Coblenz et al. [6] from Section 4.2. The other dimension combinations are shown in Figure 13. The physical interpretations of the variables are drop size (u_1), x-position (u_2), y-position (u_3), x-velocity (u_4), and y-velocity (u_5). For the variable pairs (u_1, u_5) and (u_2, u_5), no deviation from independence is discernible. A weak hump-shape can be observed for variables u_1 and u_3. Again, the course of column profiles is reversed in the middle of the profiles. The plots show a Gaussian-like behavior for variables u_3 and u_5. The profile plots for variables u_4 and u_5 show a weak deviation from independence for the profiles near $u_4 = 1$ and extreme values of u_5: (a) row profiles for variables u_1 and u_3, (b) column profiles for variables u_1 and u_3, (c) checkerboard plot for variables u_1 and u_3, (d) row profiles for variables u_1 and u_5, (e) column profiles for variables u_1 and u_5, (f) checkerboard plot for variables u_1 and u_5, (g) row profiles for variables u_2 and u_5, (h) column profiles for variables u_2 and u_5, (i) checkerboard plot for variables u_2 and u_5, (j) row profiles for variables u_3 and u_5, (k) column profiles for variables u_3 and u_5, (l) checkerboard plot for variables u_3 and u_5, (m) Row profiles for variables u_4 and u_5, (n) column profiles for variables u_4 and u_5, and (o) checkerboard plot for variables u_4 and u_5.

References

[1] Aya-Moreno, C., Geenens, G., Penev, S. (2018). Shape-preserving wavelet-based multivariate density estimation. *Journal of Multivariate Analysis*, *168*, 30–47, DOI: https://doi.org/10.1016/j.jmva.2018.07.002.

[2] Bakam, Y. I. N., Pommeret, D. (2022). *K-Sample test for equality of Copulas*.

[3] Bedford, T., Cooke, R. M. (2002). Vines-a new graphical model for dependent random variables. *The Annals of Statistics*, *30*(4), 1031–1068, DOI: https://doi.org/10.1214/aos/1031689016.

[4] Bhatia, R., Gaubert, S., Jain, T. (2019). Matrix versions of the Hellinger distance. *Letters in Mathematical Physics*, *109*(8), 1777–1804.

[5] Carley, H., Taylor, M. D. (2002). A new proof of Sklar's theorem. In: *Distributions with given marginals and statistical modelling* (pp. 29–34), Netherlands: Springer.

[6] Coblenz, M., Holz, S., Bauer, H.-J., Grothe, O., Koch, R. (2020). Modelling fuel injector spray characteristics in jet engines by using vine copulas. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *69*(4), 863–886.

[7] Cottin, C., Pfeifer, D. (2014). From Bernstein polynomials to Bernstein copulas. *Journal of Applied Functional Analysis*, *9*, 277–288.

[8] Cuadras, C. M. (2015). Contributions to the diagonal expansion of a bivariate copula with continuous extensions. *Journal of Multivariate Analysis*, *139*, 28–44.

[9] Cuadras, C. M. (2002). Correspondence analysis and diagonal expansions in terms of distribution functions. *Journal of Statistical Planning and Inference*, *103*(1–2), 137–150.

[10] Cuadras, C. M., Augé, J. (1981). A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, *10*(4), 339–353.

[11] Cuadras, C. M., Cuadras, D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, *417*(1), 64–74.

[12] Cuadras, C. M., Díaz, W. (2012). Another generalization of the bivariate FGM distribution with two-dimensional extensions. *Acta et Commentationes Universitatis Tartuensis de Mathematica 16*(1), 3–12.

[13] Cuberos, A., Masiello, E., Maume-Deschamps, V. (2020). Copulas checker-type approximations: Application to quantiles estimation of sums of dependent random variables. *Communications in Statistics - Theory and Methods*, *49*(12), 3044–3062.

[14] Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas*, vol. 222. Cham: Springer International Publishing.

[15] Durante, F., Sempi, C. (2015). *Principles of Copula Theory*. New York: Chapman and Hall/CRC.

[16] Durrleman, V., Nikeghbali, A., Roncalli, T. (2000). *Copulas Approximation and New Families*. DOI: https://doi.org/10.2139/ssrn.1032547.

[17] Dykstra, R. L. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, *78*(384), 837–842.

[18] Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

[19] Hill, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Applied Statistics*, *23*(3), 340–354.

[20] Hofert, M., Hofert, M., Mächler, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, *39*(9), 1–20.

[21] Horn, R. A., Johnson, C. R. (2012). *Matrix Analysis*. 2nd edition, Cambridge; New York: Cambridge University Press.

[22] Janssen, P., Swanepoel, J., Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, *142*(5), 1189–1197, DOI: https://doi.org/10.1016/j.jspi.2011.11.020.

[23] Joe, H. (1996). Families of m-variate distributions with given margins and $m(m–1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, *28*, 120–141.

[24] Kazmierczak, J.-B. (1978). Migrations interurbaines dans la banlieue sud de paris. *Cahiers de laanalyse des données*, *3*(2), 203–218.

[25] Klaassen, C. A. J., Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, *3*(1), 55, DOI: https://doi.org/10.2307/3318652.

[26] Kolda, T. G., Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3), 455–500.

[27] Kolesárová, A., Mesiar, R., Mordelová, J., Sempi, C. (2006). Discrete copulas. *IEEE Transactions on Fuzzy Systems*, *14*(5), 698–705.

[28] Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, *44*(1–2), 289–292, DOI: https://doi.org/10.1093/biomet/44.1-2.289.

[29] Li, X., Mikusiński, P., Sherwood, H., Taylor, M. D. (1997). On approximation of copulas. In Beneš, V., Štěpán, J., editors, *Distributions with given Marginals and Moment Problems* (pp. 107–116). Netherlands: Springer.

[30] Masuhr, A., Trede, M. (2020). Bayesian estimation of generalized partition of unity copulas. *Dependence Modeling*, *8*(1), 119–131, DOI: https://doi.org/10.1515/demo-2020-0007.

[31] Mayor, G., Suner, J., Torrens, J. (2005). Copula-like operations on finite settings. *IEEE Transactions on Fuzzy Systems*, *13*(4), 468–477, DOI: https://doi.org/10.1109/TFUZZ.2004.840129.

[32] Meier, C., Kirch, C., Meyer, R. (2018). Bayesian nonparametric analysis of multivariate time series: a matrix gamma process approach. *Journal of Multivariate Analysis*, *175*, 104560, DOI: https://doi.org/10.1016/j.jmva.2019.104560.

[33] Mesiar, R., Najjari, V. (2014). New families of symmetric/asymmetric copulas. *Fuzzy Sets and Systems*, *252*, 99–110.

[34] Mirsky, L. (1960). Symmetric Gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, *11*(1), 50–59.

[35] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. New York, NY: Springer New York.

[36] Panagiotelis, A., Czado, C., Joe, H., Stöber, J. (2017). Model selection for discrete regular vine copulas. *Computational Statistics & Data Analysis*, *106*, 138–152.
[37] Perfect, H., Mirsky, L. (1965). Spectral properties of doubly-stochastic matrices. *Monatshefte für Mathematik*, *69*(1), 35–57, DOI: https://doi.org/10.1007/BF01313442.
[38] Pfeifer, D., Tsatedem, H. A., Mändle, A., Girschig, C. (2016). New copulas based on general partitions-of-unity and their applications to risk management. *Dependence Modeling*, *4*(1), 000010151520160006, DOI: https://doi.org/10.1515/demo2016-0006.
[39] Rao, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestiió*, *19*(1-2-3), 23–63.
[40] Rodríguez-Lallena, J. (2004). A new class of bivariate copulas. *Statistics & Probability Letters*, *66*(3), 315–325.
[41] Rontsis, N., Goulart, P. (2020). Optimal approximation of doubly stochastic matrices. In: Chiappa, S., Calandra, R., editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, *volume 108 of Proceedings of Machine Learning Research* (pp. 3589–3598).
[42] Savu, C., Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance*, *10*(3), 295–304.
[43] Savu, C., Trede, M. (2008). Goodness-of-fit tests for parametric families of Archimedean copulas. *Quantitative Finance*, *8*(2), 109–116, DOI: https://doi.org/10.1080/14697680701207639.
[44] Schmid, F., Schmidt, R., Blumentritt, T., Gaißer, S., Ruppert, M. (2010). Copula-based measures of multivariate association. In Jaworski, P., Durante, F., Härdle, W. K., Rychlik, T., editors, *Copula Theory and Its Applications* (vol. 198, pp. 209–236). Springer Berlin Heidelberg.
[45] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de L'Institut de Statistique de L'Université de Paris*, *8*, 229–231.
[46] Stewart, G. W. (1991). Perturbation theory for the singular value decomposition. In: Vaccaro, R. J., editor, and University of Rhode Island, *SVD and Signal Processing, II: Algorithms, Analysis, and Applications*, Amsterdam; New York: New York, N.Y., U.S.A: Elsevier; Distributors for the U.S.A. and Canada, Elsevier Science Pub. Co., (pp. 99–109).
[47] Zass, R., Shashua, A. (2007). Doubly stochastic normalization for spectral clustering. In: Schölkopf, B., Platt, J., Hofmann, T., editors, *Advances in Neural Information Processing Systems* (vol. 19, pp. 1569–1576). The MIT Press.