MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Coordinate-wise Transformation of Probability Distributions to Achieve a Stein-type Identity

Tomonari SEI

METR 2017–04 February 2017

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Coordinate-wise Transformation of Probability Distributions to Achieve a Stein-type Identity

Tomonari SEI*

February 2017

Abstract

It is shown that for any given multi-dimensional probability distribution, there exists a unique coordinate-wise transformation such that the transformed distribution satisfies a Stein-type identity. The proof is based on an energy minimization problem over a subset of the Wasserstein space. The result is interpreted as a generalization of the diagonal scaling theorem established by Marshall and Olkin (1968).

Keywords: Copositive distribution, Copula, Energy minimization, Optimal transportation, Stein-type distribution, Wasserstein space.

1 Introduction

In their seminal paper [20], Marshall and Olkin proved the following diagonal scaling theorem. Let S be a $d \times d$ positive semi-definite matrix and assume that S is strictly copositive in the sense that

$$\inf_{w_1,\ldots,w_d>0} \frac{\sum_i \sum_j w_i S_{ij} w_j}{\sum_i w_i^2} > 0. \tag{1}$$

Then, there exists a unique positive diagonal matrix D such that the sum of each row of DSD is unity. Note that (1) is satisfied if S is positive definite. The theorem is interpreted in a probabilistic framework. Let X be a random column vector with mean zero and covariance matrix S. Then, since $\sum_{j=1}^d (DSD)_{ij} = 1$ for each i, the distribution μ of the transformed random vector DX satisfies an identity

$$\sum_{j=1}^d \int x_i x_j \mathrm{d}\mu = 1, \quad i = 1, \ldots, d. \tag{2}$$

This property is applied to summarize multivariate data. Refer to [28] for details.

In the present paper, we provide a nonlinear analogue of the result. We admit a nonlinear coordinate-wise transformation of a random vector to achieve a stronger condition than (2).

*Graduate School of Information Science and Technology, The University of Tokyo.

This condition will be referred to as the Stein-type identity. Under some mild conditions on μ, it is shown that there exists such a unique transformation. The proof is based on a variational formulation. The Marshall-Olkin theorem is, in fact, derived in a similar manner [20, 15]. The space we use in the proof is the Wasserstein space, a distance space induced from optimal transportation. Refer to [25, 32] for comprehensive studies of optimal transportation and its applications. Another generalization of the Marshall-Olkin theorem is considered by [3], where the dimension d is infinity but the transformation is linear.

As is well known, Sklar's theorem (see, e.g., [22]) states that any multi-dimensional distribution is transformed by the probability integral transformation into a distribution with uniform marginals. The resultant distribution is called a copula. Our result is considered as an alternative to Sklar's theorem.

The remainder of the present paper is organized as follows. In Section 2, we describe the existence and uniqueness theorem as well as a variational characterization theorem. In Section 3, we clarify the regularity properties of Stein-type distributions. In Section 4, we prove the main results using the theory of optimal transportation. In Section 5, tractable conditions for existence are considered. In Section 6, a numerical method to find the transformation for piecewise uniform distributions is proposed. Finally, we discuss open problems in Section 7.

2 Main results

We first define a class of distributions that satisfy a stronger condition than (2). Let $\mathcal{P}^2 = \mathcal{P}^2(\mathbb{R}^d)$ be the set of probability distributions μ on $\mathbb{R}^d$ with mean zero and finite variance such that each marginal distribution μ_i of μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Note that μ itself is not assumed to be absolutely continuous. The mean-zero condition is imposed only for simplicity. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous if there exists a locally integrable function f' such that $f(x) = f(0) + \int_0^x f'(y)\mathrm{d}y$ in Lebesgue's sense.

Definition 1. We say that a distribution $\mu \in \mathcal{P}^2$ is *Stein-type* if it satisfies

$$\int f(x_i) \left(\sum_{j=1}^{d} x_j \right) \mathrm{d}\mu = \int f'(x_i)\mathrm{d}\mu, \quad i = 1, \ldots, d, \tag{3}$$

for any absolutely continuous function $f : \mathbb{R} \to \mathbb{R}$ with essentially bounded derivative f'.

Note that the equation (2) is a special case of (3), where $f(x_i) = x_i$.

We refer to the equation (3) as the *Stein-type identity.* Indeed, if $d = 1$, it reduces to the Stein identity $\int f(x_1)x_1 \mathrm{d}\mu = \int f'(x_1)\mathrm{d}\mu$, which implies that μ is the standard normal distribution (see [29] and [5]). Similarly, if μ is completely independent in the sense that μ is the direct product of its marginal μ_i, then only the d-dimensional standard normal distribution satisfies (3). We focus on dependent cases.

For Gaussian random variables, we obtain the following lemma, where the expectation is denoted by E.

Lemma 1 (Theorem 5 of [28])**.** Let μ denote the d-dimensional normal distribution with mean zero and covariance matrix S. Then, μ is Stein-type if and only if $\sum_j S_{ij} = 1$ for each i.

Proof. Let $(X_1, \ldots, X_d)$ be distributed according to μ. Then, $\mathrm{E}[X_j|X_i] = S_{ij}X_i/S_{ii}$ and

$$\mathrm{E}\left[f(X_i)\sum_j X_j\right] = \frac{\sum_j S_{ij}}{S_{ii}}\mathrm{E}[f(X_i)X_i] = \left(\sum_j S_{ij}\right)\mathrm{E}[f'(X_i)].$$

The last equality follows from the Stein identity for the univariate normal distributions. □

The following example gives a rich class of Stein-type distributions.

Example 1. Let W be a random variable with the standard normal distribution and let U be any random variable independent of W such that $\mathrm{E}[U] = 0$ and $\mathrm{E}[U^2] < \infty$. Then, the joint distribution of two variables

$$X_1 = \frac{W+U}{\sqrt{2}} \quad \text{and} \quad X_2 = \frac{W-U}{\sqrt{2}}$$

is Stein-type. Indeed, we obtain

$$\mathrm{E}\left[f\left(\frac{W \pm U}{\sqrt{2}}\right)\sqrt{2}W\right] = \mathrm{E}\left[f'\left(\frac{W \pm U}{\sqrt{2}}\right)\right]$$

for any f by the Stein identity with respect to W conditional on U. For $d \geq 3$, define a random vector $(X_1, \ldots, X_d)$ by $X_i = (W + U_i)/\sqrt{d}$, where W has the standard normal distribution independent of $U_1, \ldots, U_{d-1}$ and $U_d = -\sum_{j=1}^{d-1} U_j$. Then, the distribution of $(X_1, \ldots, X_d)$ is Stein-type as long as $\mathrm{E}[U_i] = 0$ and $\mathrm{E}[U_i^2] < \infty$. □

The example does not cover the entire class of Stein-type distributions. Other examples are given in Section 6 and Appendix B.

If a random vector $(X_1, \ldots, X_d)$ has a Stein-type distribution, then the sum $\sum_j X_j$ is positively correlated with any monotone transformation of X_i due to (3). Refer to Section 8 of [28] for an application of this property.

For each $\mu \in \mathcal{P}^2$, let $\mathcal{T}_{\mathrm{cw}}(\mu)$ be the set of coordinate-wise transformations

$$T(x) = (T_1(x_1), \ldots, T_d(x_d)), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

such that each $T_i : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ is non-decreasing and $T_\sharp\mu$ belongs to $\mathcal{P}^2$. Here, $T_\sharp\mu$ is the push-forward measure defined by $(T_\sharp\mu)(A) = \mu(T^{-1}(A))$ for any measurable set A. The set $\mathcal{T}_{\mathrm{cw}}(\mu)$ depends only on the marginal distributions of μ. Two maps T and U in $\mathcal{T}_{\mathrm{cw}}(\mu)$ are identified if $\mu(T = U) = 1$. Note that T_i has discontinuous points if the support of $(T_\sharp\mu)_i$ is not connected.

We consider a problem to find a map $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$ such that $T_\sharp\mu$ is Stein-type. Let us call such a map *a Stein-type transformation* of μ. For example, if μ is the direct measure of one-dimensional continuous distributions μ_i, then the map T defined by $T_i(x_i) = \Phi^{-1} \circ \mu_i((-\infty, x_i])$ is the Stein-type transformation of μ, where Φ is the cumulative distribution function of the standard normal distribution.

The following lemma is immediate.

Lemma 2. Let μ be the normal distribution with a covariance matrix S. Then, μ has a Stein-type transformation if S is strictly copositive in the sense of (1).

Proof. Let D be the diagonal matrix with entries $w_1, \ldots, w_d$ satisfying (2). Set $T(x) = Dx$. Then, $T_\sharp\mu$ is Stein-type due to Lemma 1. $\square$

Denote the set of coordinate-wise transformed distributions of μ by

$$\mathcal{F}_\mu = \{T_\sharp\mu \mid T \in \mathcal{T}_{\mathrm{cw}}(\mu)\} \subset \mathcal{P}^2.$$

We refer to $\mathcal{F}_\mu$ as a *fiber*. The following lemma is a direct consequence of the one-dimensional optimal transportation. See Appendix A.

Lemma 3. For given $\mu \in \mathcal{P}^2$ and $\nu \in \mathcal{F}_\mu$, the map $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$ satisfying $\nu = T_\sharp\mu$ is uniquely determined μ-almost everywhere. Furthermore, the relation $\nu \in \mathcal{F}_\mu$ between two measures μ and ν is an equivalence relation. In particular, $\mathcal{P}^2$ is partitioned into mutually disjoint fibers.

Now, we state our three main theorems. All proofs are presented in Section 4.

The first theorem characterizes Stein-type distributions in terms of the variational principle. Define an energy functional $\mathcal{E}(\mu)$ of μ by

$$\mathcal{E}(\mu) = \sum_{i=1}^{d} \int p_i(x_i) \log p_i(x_i) \mathrm{d}x_i + \int \left(\sum_{j=1}^{d} x_j \right)^2 \mathrm{d}\mu, \tag{4}$$

where $p_i = \mathrm{d}\mu_i/\mathrm{d}x_i$ is the marginal density function. The first term of $\mathcal{E}(\mu)$ represents the negative entropy of the marginal distributions, and the second term is the variance of the diagonal part $\sum_j x_j$.

Theorem 1. A measure $\mu \in \mathcal{P}^2$ is Stein-type if and only if $\mathcal{E}(\mu)$ is finite and μ minimizes $\mathcal{E}$ over the fiber $\mathcal{F}_\mu$.

The second theorem is on the uniqueness of Stein-type transformations. A distribution μ on $\mathbb{R}^d$ is said to have a *regular support* if the support of μ is equal to the direct product of the supports of the marginal distributions μ_i. This property is invariant under coordinate-wise transformations. Note that the regular support condition does not imply absolute continuity of μ with respect to $\prod_{i=1}^d \mu_i$.

Theorem 2 (Uniqueness)**.** Assume that $\mu \in \mathcal{P}^2$ has a regular support. Then, a Stein-type transformation of μ is unique if it exists.

We conjecture that the uniqueness follows without the regular support condition. See Section 7 for more details.

The third theorem is on existence. A measure $\mu \in \mathcal{P}^2$ is said to be *copositive* if

$$\beta(\mu) := \inf_{T \in \mathcal{T}_{\mathrm{cw}}(\mu)} \frac{\int (\sum_i T_i)^2 \mathrm{d}\mu}{\sum_i \int T_i^2 \mathrm{d}\mu} > 0. \tag{5}$$

For example, if μ is completely independent, then $\int (\sum_i T_i)^2 \mathrm{d}\mu = \sum_i \int T_i^2 \mathrm{d}\mu$ for any T, and therefore $\beta(\mu) = 1$. It is not difficult to see that $\beta(\mu) \leq 1$ for any μ. If μ is associated in the sense of [7], [8], and [16], then $\int T_i T_j \mathrm{d}\mu \geq 0$ for each pair of i and j, and therefore $\beta(\mu) \geq 1$. On the other hand, if $d = 2$ and $\mu(\{x \mid x_1 + x_2 = 0\}) = 1$, then $\beta(\mu) = 0$ because $\int (x_1 + x_2)^2 \mathrm{d}\mu = 0$. Sufficient conditions for copositivity are presented in Section 5.

Theorem 3 (Existence)**.** Let $\mu \in \mathcal{P}^2$ be copositive. Then, there exists a Stein-type transformation of μ.

We now present a few remarks before proceeding to the following section.

The uniqueness and existence results in Theorem 2 and Theorem 3 are consequences of the variational characterization in Theorem 1, as will be shown in Section 4. For

$d = 1$, the functional $\mathcal{E}(\mu)$ is the Kullback-Leibler divergence from μ to the standard normal density up to a constant term. For $d \geq 2$, however, $\mathcal{E}$ is not even bounded from below. Indeed, for each $t > 0$, let μ^t be the multivariate normal distribution with mean zero and covariance matrix $\Sigma_t = P + t(I - P)$, where I is the identity matrix, and P denotes the orthogonal projection to the direction $(1, \ldots, 1)^\top \in \mathbb{R}^d$. Then, each marginal distribution of μ^t is normal with variance $\sigma_t^2 = (1/d) + t(1 - 1/d)$. We can show that $\mathcal{E}(\mu^t) = -(d/2)\log(2\pi\sigma_t^2)$, which tends to $-\infty$ as $t \to \infty$. Therefore, it is not trivial if there is a minimizer of $\mathcal{E}$ over the fiber. Nevertheless, the existence and uniqueness theorems are obtained.

If μ has the joint density function $p(x)$, then the negative joint entropy is defined by

$$\mathcal{U}_d(\mu) = \int p(x) \log p(x) \mathrm{d}x.$$

In most cases, we can replace the marginal entropy term in $\mathcal{E}(\mu)$ with the joint entropy because the difference $\mathcal{U}_d(\mu) - \sum_{i=1}^d \mathcal{U}_1(\mu_i)$, which is referred to as the multi-information function or the measure of multivariate dependence, is invariant in each fiber (e.g., [11] and [30]). However, in some pathological cases, the difference diverges. Therefore, it is more appropriate to adopt the marginal entropy.

According to Sklar's theorem (e.g. [22]), any d-dimensional distribution μ is transformed by the probability integral transformation $T_i(x_i) = \int_{-\infty}^{x_i} \mathrm{d}\mu_i$ into the distribution $T_\sharp\mu$ with uniform marginals unless some μ_i has an atom. The resultant distribution $T_\sharp\mu$ is called a copula. The Stein-type distribution we defined is considered as an alternative representation of the copula. Copulas are also characterized by an energy minimization problem. Here, the potential term in (4) is replaced with $\int V(x)\mathrm{d}\mu$, where $V(x) = \infty$ if $x \notin [0,1]^d$ and 0 otherwise. In parallel, we have to remove the condition $\int x_i \mathrm{d}\mu_i = 0$ from the definition of $\mathcal{P}^2$. Maximum entropy copulas under a given diagonal section are discussed in [4], where, in contrast to the present paper, the marginals are fixed to be uniform.

3 Regularity of Stein-type distributions

The Stein-type identity forces regularity of marginal density functions. We first characterize this by an integral equation.

Theorem 4. Let $\mu \in \mathcal{P}^2$. Denote the marginal density functions of μ by $p_i(x_i)$. Then, μ

is a Stein-type distribution if and only if it satisfies a set of integral equations

$$p_i(a) = \int_a^\infty p_i(x_i) m_i(x_i) \mathrm{d}x_i, \quad a \in \mathbb{R}, \quad i = 1, \ldots, d, \tag{6}$$

where $m_i(x_i)$ denotes the conditional expectation of $\sum_{j=1}^d x_j$ given x_i with respect to μ.

Proof. First, note that $m_i(x_i)$ is finite μ_i-almost everywhere because μ belongs to $\mathcal{P}^2$.

Assume μ is Stein-type. For $-\infty < a < b < \infty$, let $h_{ab}(x) = (b-a)^{-1} \int_{-\infty}^x I_{(a,b)}(\xi) \mathrm{d}\xi$, where $I_{(a,b)}$ is the indicator function of (a, b). The Stein-type identity for h_{ab} implies

$$\frac{1}{b-a} \int_a^b \mathrm{d}\mu_i = \int h'_{ab}(x_i) \mathrm{d}\mu_i = \int h_{ab}(x_i) \sum_j x_j \mathrm{d}\mu = \int h_{ab}(x_i) m_i(x_i) \mathrm{d}\mu_i. \tag{7}$$

Letting $b \to a$ in (7), we obtain (6).

Conversely, assume (6). The right-hand side of (6) converges to zero as $a \to \pm\infty$ because $\int x_j \mathrm{d}\mu_j = 0$ for all j. Then, for any bounded and absolutely continuous function f with bounded derivative f', we obtain

$$\begin{aligned} \int f'(a) p_i(a) \mathrm{d}a &= \int_{-\infty}^\infty f'(a) \left(\int_a^\infty p_i(x_i) m_i(x_i) \mathrm{d}x_i \right) \mathrm{d}a \\ &= \int_{-\infty}^\infty f(x_i) p_i(x_i) m_i(x_i) \mathrm{d}x_i \\ &= \int f(x_i) \sum_j x_j \mathrm{d}\mu, \end{aligned}$$

where the second equality follows from the integral-by-parts formula. If f is not bounded, then let $f_M(x) = f(0) + \int_0^x f'(u) 1_{\{|u| \le M\}} \mathrm{d}u$ and take $M \to \infty$. □

As a corollary, the regularity of the marginal density functions is established.

Corollary 1. Let μ be Stein-type. Then, its marginal density functions $p_i(x_i)$ are bounded, absolutely continuous, and converge to zero as $x_i \to \pm\infty$.

Proof. From the formula (6), it is obvious that p_i is absolutely continuous and bounded by $\int \sum_i |x_i| \mathrm{d}\mu_i < \infty$. We also have $p_i(x_i) \to 0$ as $x_i \to \pm\infty$ because the right-hand side of (6) vanishes as $a \to \pm\infty$. □

Although the marginal density function of any Stein-type distribution is absolutely continuous, it can have non-differentiable points as shown in an example in Section 6. The continuous differentiability of $p_i(x_i)$ follows from the regularity of the pair-wise copula of μ along formula (6). We do not pursue this line of investigation here. On the other hand, we conjecture that the marginal density of any Stein-type distribution is positive everywhere. See Section 7 for more details.

The following corollary will be used in Section 4.

Corollary 2. Let μ be Stein-type. Then, its negative marginal entropy $\int p_i(x_i)\log p_i(x_i)\mathrm{d}x_i$ is finite.

Proof. Since the marginal density $p_i(x_i)$ is bounded, we have $\int p_i(x_i)\log p_i(x_i)\mathrm{d}x_i < \infty$. To prove $\int p_i(x_i)\log p_i(x_i)\mathrm{d}x_i > -\infty$, we use the non-negativity of the Kullback-Leibler information from p_i to the standard normal density $\phi(x_i) = e^{-x_i^2/2}/\sqrt{2\pi}$. Indeed,

$$\int p_i(x_i)\log p_i(x_i)\mathrm{d}x_i \geq \int p_i(x_i)\log \phi(x_i)\mathrm{d}x_i = \int \{-(1/2)\log(2\pi) - x_i^2/2\}\mathrm{d}\mu_i > -\infty$$

because $\int x_i^2 \mathrm{d}\mu_i < \infty$. □

As a remark, we also show that Stein-type distributions have finite Fisher information. The Fisher information of a density function q on $\mathbb{R}$ is defined by

$$I(q) = \int \left(\frac{q'(x)}{q(x)}\right)^2 q(x)\mathrm{d}x,$$

where q is assumed to be absolutely continuous, and $q'(x)/q(x)$ is set to 0 if q is not differentiable or not positive at x. See [13] for properties implied by finite Fisher information. Note that the Fisher information we defined is that of location family $\{q(x-\theta) \mid \theta \in \mathbb{R}\}$ in statistics (e.g., [18]).

Corollary 3. For any Stein-type distribution μ, the Fisher information $I(p_i)$ of each marginal density p_i is bounded by the dimension d. In particular, p_i has bounded variation.

Proof. From (6), the score function $p_i'(x_i)/p_i(x_i)$ is equal to $-m_i(x_i)$. Since $m_i(x_i)$ is the conditional expectation of $\sum_j x_j$ given x_i, we obtain

$$I(p_i) = \int m_i(x_i)^2 \mathrm{d}\mu \leq \int \left(\sum_j x_j\right)^2 \mathrm{d}\mu = d,$$

where the last equality follows from the Stein-type identity with $f(x_i) = x_i$. By the Cauchy-Schwarz inequality, we also have $\int |p_i'(x_i)|\mathrm{d}x_i \leq \sqrt{I(p_i)}$. Then, p_i has bounded variation. □

Other properties are given in Appendix C.

4 Proofs based on the theory of optimal transportation

In this section, we prove the three main theorems stated in Section 2. The proof is based on the theory of optimal transportation. Necessary facts about one-dimensional optimal transportation are summarized in Appendix A.

4.1 Variational problem over a fiber of Wasserstein space

Let $\mathcal{F}$ be a fiber of $\mathcal{P}^2$ (see Section 2 for the definition) and choose two measures μ and $\nu = T_\sharp\mu$ in $\mathcal{F}$, where $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$. Define the geodesic, which is also referred to as the displacement interpolation [21], from μ to ν by

$$[\mu, \nu]_t = [(1-t)\mathrm{Id} + tT]_\sharp\mu, \quad t \in [0, 1],$$

where Id denotes the identity map. Based on the one-dimensional optimal transportation, it follows that $[\mu, \nu]_t \in \mathcal{F}$ and $[\mu, \nu]_t = [\nu, \mu]_{1-t}$ for each t.

Although a geodesic between any pair of distributions in $\mathcal{P}^2$ is similarly defined, we need only geodesics in a common fiber. It is known that a geodesic actually attains the minimum length of a path between two measures with respect to the L^2-Wasserstein distance (see e.g. [2] and [32]). Here the L^2-Wasserstein distance is the infimum of $\left(\int \|x-y\|^2 \mathrm{d}\gamma(x,y)\right)^{1/2}$ over the joint distribution γ on $\mathbb{R}^{2d}$ with the marginal distributions μ and ν. Note that each fiber $\mathcal{F}$ is totally geodesic in the sense of [31].

From a different perspective from ours, optimal transportation between two distributions sharing the same copula is considered in [1], where the various cost functions are the center of discussion.

Recall that μ is said to have a regular support if its support is the direct product of the supports of marginal distributions.

Lemma 4. Let $\mathcal{F}$ be a fiber and choose any two distributions μ and ν in $\mathcal{F}$, where $\mu \neq \nu$. Then, $\mathcal{E}([\mu, \nu]_t)$ is convex in t. Furthermore, $\mathcal{E}([\mu, \nu]_t)$ is strictly convex if one of the following conditions is satisfied:

(i) μ (and therefore ν) has a regular support, or

(ii) the supports of μ_i and ν_i are connected, respectively, for each i.

Proof. Let $\nu = T_\sharp\mu$, with $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$. Let p_i be the marginal density of μ. By the change-of-variable formula (Lemma 11 in Appendix A), we obtain

$$\mathcal{E}([\mu,\nu]_t) = \sum_i \int p_i(x_i) \log \frac{p_i(x_i)}{(1-t)+tT_i'(x_i)} \mathrm{d}x_i + \frac{1}{2} \int \left(\sum_i ((1-t)x_i + tT_i(x_i)) \right)^2 \mathrm{d}\mu, \tag{8}$$

where $T_i'(x_i)$ is the derivative of T_i if it exists, and $T_i'(x_i) = 0$ otherwise. Both terms in (8) are convex in t.

Assume (i) and that $\mathcal{E}([\mu,\nu]_t)$ is not strictly convex. Then, there is an interval over which $\mathcal{E}([\mu,\nu]_t)$ is linear. It is deduced from (8) that $\sum_{i=1}^d (T_i(x_i) - x_i) = 0$, μ-almost everywhere. Let I be the set of indices i such that $\mu_i(T_i(x_i) \neq x_i) > 0$. Then, I is not empty because $T \neq \mathrm{Id}$. For each $i \in I$, the probability $\mu_i(T_i(x_i) - x_i > 0)$ is positive because $\int (T_i(x_i) - x_i)\mathrm{d}\mu_i = 0$. Then, by the regular support condition, we have $\mu(\sum_{i \in I}(T_i(x_i) - x_i) > 0)$ is positive. However, this contradicts $\sum_{i=1}^d (T_i(x_i) - x_i) = 0$. Thus, $\mathcal{E}([\mu,\nu]_t)$ should be strictly convex under (i).

Next, assume (ii). Then, T_i has no discontinuous points. Assume $\mathcal{E}([\mu,\nu]_t)$ is not strictly convex. Then, if follows from (8) that $T_i'(x_i) = 1$ and, therefore, $T_i(x_i) = x_i$ by the connectedness of the support together with the condition $\int T_i \mathrm{d}\mu_i = 0$. However, this contradicts $\mu \neq \nu$. Thus, $\mathcal{E}([\mu,\nu]_t)$ is strictly convex. □

Example 2. The strict convexity of $\mathcal{E}([\mu,\nu]_t)$ can fail if neither condition (i) nor condition (ii) in Lemma 4 is satisfied. For example, let $d = 2$ and assume that μ is uniformly distributed over the region $([-1,0] \times [0,1]) \cup ([0,1] \times [-1,0])$. Define the map T by $T_i(x_i) = x_i + 1$ if $x_i > 0$, and $T_i(x_i) = x_i - 1$ otherwise for each i. Let $\nu = T_\sharp \mu$. Then, $\mathcal{E}([\mu,\nu]_t)$ is constant along $t \in [0,1]$ because $T_i'(x_i) = 1$ and $\sum_i T_i(x_i) = \sum_i x_i$, μ-almost everywhere. In this case, μ_i is supported on $[-1,1]$, whereas ν_i is supported on $[-2,-1] \cup [1,2]$. □

Convexity along a geodesic is referred to as displacement convexity [21]. Lemma 4 shows that $\mathcal{E}$ is displacement convex over each fiber. Refer to [2] for further details on displacement convexity.

4.2 Proof of Theorem 1

Let μ be a Stein-type distribution. Corollary 2 implies that μ belongs to $\mathrm{dom}\,\mathcal{E}$. From the convexity (Lemma 4), it is sufficient to show that

$$\left. \frac{\mathrm{d}}{\mathrm{d}t_+} \mathcal{E}([\mu,\nu]_t) \right|_{t=0} \geq 0$$

for any $\nu = T_\sharp\mu \in \mathcal{F}$, where $\mathrm{d}/\mathrm{d}t_+$ denotes the right derivative. It follows from formula (8) that

$$\frac{\mathrm{d}}{\mathrm{d}t_+}\mathcal{E}([\mu,\nu]_t)\Big|_{t=0} = -\sum_i \int p_i(x_i)(T_i'(x_i)-1)\mathrm{d}x_i + \sum_i\sum_j \int (T_i(x_i)-x_i)x_j\mathrm{d}\mu. \quad (9)$$

If T_i is absolutely continuous, the right-hand side vanishes by the Stein-type identity, where the boundedness of the derivatives T_i' can be assumed by a standard approximation argument, as in the proof of Theorem 4. If T_i is not absolutely continuous, T_i can be decomposed into an absolutely continuous part and a discontinuous part as $T_i = T_i^{\mathrm{ac}} + T_i^{\mathrm{d}}$. See Appendix A. The contribution of T_i^{ac} in (9) vanishes due to the Stein-type identity. It is sufficient to prove that $\sum_j \int T_i^{\mathrm{d}}(x_i)x_j\mathrm{d}\mu \geq 0$ for each i because $(T_i^{\mathrm{d}})' = 0$ by definition. We can take a sequence $\{f_{i,n}\}_{n=1}^\infty$ of non-decreasing differentiable functions with a bounded derivative such that $f_{i,n}(x_i)$ converges to $T_i^{\mathrm{d}}(x_i)$ μ-almost everywhere. More specifically, a step function $I_{[\xi,\infty)}(x_i)$ at each $\xi \in \mathbb{R}$ is approximated by a logistic function $1/(1+\exp(-n(x_i-\xi)))$. Then, by Lebesgue's dominated convergence theorem and the Stein-type identity, we obtain

$$\begin{aligned}\sum_j \int T_i^{\mathrm{d}}(x_i)x_j\mathrm{d}\mu &= \lim_{n\to\infty}\sum_j \int f_{i,n}(x_i)x_j\mathrm{d}\mu \\ &= \lim_{n\to\infty}\int f_{i,n}'(x_i)\mathrm{d}\mu \\ &\geq 0.\end{aligned}$$

Conversely, assume that $\mathcal{E}(T_\sharp\mu)$ is minimized at $T = \mathrm{Id}$. Let f be an absolutely continuous function with bounded derivative f'. Then, for sufficiently small $\varepsilon > 0$, both of $T(x) = x \pm \varepsilon f(x)$ belong to $\mathcal{T}_{\mathrm{cw}}(\mu)$. Thus, (9) is zero, and μ satisfies the Stein-type identity.

4.3 Proof of Theorem 2

Assume that μ has a regular support and admits a Stein-type transformation T. Then, Theorem 1 implies that $T_\sharp\mu$ minimizes $\mathcal{E}$ over the fiber $\mathcal{F}_\mu$. However, it is deduced from Lemma 4 that $\mathcal{E}$ is strictly convex over $\mathcal{F}_\mu$. Thus, the minimizer is unique.

4.4 Proof of Theorem 3

Assume that μ is copositive. Denote the functional $\mathcal{E}$ restricted to the fiber $\mathcal{F}_\mu$ by $\mathcal{E}_\mu$. From Theorem 1, it is sufficient to show that $\mathcal{E}_\mu$ has a minimum point. We first show that

$\mathcal{E}_\mu$ is bounded from below and that the level set $\{\nu \mid \mathcal{E}_\mu \leq c\}$ for each $c \in \mathbb{R}$ is tight. For any $\nu \in \mathcal{F}_\mu$, the copositivity condition implies

$$\mathcal{E}_\mu(\nu) \geq \sum_{i=1}^{d} \left\{ \int q_i(x_i) \log q_i(x_i) \mathrm{d}x_i + \frac{\beta}{2} \int x_i^2 \mathrm{d}\nu_i \right\},$$

where $q_i = \mathrm{d}\nu_i/\mathrm{d}x_i$ and $\beta = \beta(\nu) = \beta(\mu) > 0$. We obtain

$$\begin{aligned}\int q_i(x_i) \log q_i(x_i) \mathrm{d}x_i &= \int q_i(x_i) \log \frac{q_i(x_i)}{\sqrt{\beta/(4\pi)} e^{-\beta x_i^2/4}} \mathrm{d}x_i - \frac{\beta}{4} \int x_i^2 \mathrm{d}\nu_i + \frac{1}{2} \log \frac{\beta}{4\pi} \\ &\geq -\frac{\beta}{4} \int x_i^2 \mathrm{d}\nu_i + \frac{1}{2} \log \frac{\beta}{4\pi},\end{aligned}$$

where the last inequality follows from the nonnegativity of the Kullback-Leibler divergence. Then, $\mathcal{E}_\mu$ is bounded from below as

$$\mathcal{E}_\mu(\nu) \geq C + \frac{\beta}{4} \sum_{i=1}^{d} \int x_i^2 \mathrm{d}\nu_i,$$

where C is a constant independent of ν. This inequality also implies that the level set $\{\nu \mid \mathcal{E}_\mu(\nu) \leq c\}$ is tight.

Now there exists a weakly converging sequence ν_k such that $\mathcal{E}_\mu(\nu_k)$ converges to $\inf \mathcal{E}_\mu(\nu)$. Let ν_* be the weak limit. Then, Corollary 3.5 of [21] shows that $\nu_* \in \mathcal{P}^2$ and $\mathcal{E}_\mu(\nu_*) \leq \lim_k \mathcal{E}_\mu(\nu_k)$. The distribution ν_* gives a minimum point of $\mathcal{E}_\mu$. This completes the proof.

5 Sufficient conditions for copositivity

We now present the sufficient conditions for copositivity of a given distribution μ. In Subsection 5.1, we first take into account the measures with a non-zero mean as well as coordinate-wise transformations that are constant over an interval. We then present a lower bound of the quantity $\beta(\mu)$ in (5). Subsequent subsections are devoted to finding sufficient conditions for copositivity.

5.1 Extension of the definition and a lower bound

Let $\mathcal{P}_*^2$ be the set of measures on $\mathbb{R}^d$ such that each marginal μ_i is absolutely continuous and $\int x_i^2 \mathrm{d}\mu_i < \infty$ without assuming $\int x_i \mathrm{d}\mu_i = 0$. The set $\mathcal{T}_{\mathrm{cw}*}(\mu)$ for $\mu \in \mathcal{P}_*^2$ is defined by the set of coordinate-wise non-decreasing map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\int T_i \mathrm{d}\mu_i = 0$ and $\int T_i^2 \mathrm{d}\mu_i < \infty$ for each i.

The following lemma is useful to consider copositivity. Denote the inner product and norm of $L^2(\mu)$ by $\langle f, g\rangle = \int f(x)g(x)\mathrm{d}\mu$ and $\|f\| = \langle f, f\rangle^{1/2}$, respectively.

Lemma 5. If $\mu \in \mathcal{P}$, then

$$\beta(\mu) = \inf_{0 \neq T \in \mathcal{T}_{\mathrm{cw}*}(\mu)} \frac{\|\sum_i T_i\|^2}{\sum_i \|T_i\|^2}. \tag{10}$$

Proof. Denote the right-hand side of (10) by $\delta(\mu)$. Then, it is obvious that $\beta(\mu) \geq \delta(\mu)$ since $\mathcal{T}_{\mathrm{cw}}(\mu) \subset \mathcal{T}_{\mathrm{cw}*}(\mu)$. In order to prove the converse inequality, choose $0 \neq T \in \mathcal{T}_{\mathrm{cw}*}(\mu)$ such that $\|\sum_i T_i\|^2/(\sum_i \|T_i\|^2) \leq \delta(\mu) + \varepsilon$ for given ε. It follows from Lemma 13 in Appendix A that a map T^η defined by $T^\eta(x) = T(x) + \eta x$ belongs to $\mathcal{T}_{\mathrm{cw}}(\mu)$ for each $\eta > 0$. Then, we have

$$\frac{\|\sum_i T_i\|^2}{\sum_i \|T_i\|^2} = \lim_{\eta \to 0} \frac{\|\sum_i T_i^\eta\|^2}{\sum_i \|T_i^\eta\|^2},$$

implying $\beta(\mu) \leq \delta(\mu) + \varepsilon$. □

We extend the definition of $\beta(\mu)$ for any $\mu \in \mathcal{P}_*^2$ by (10). In the following, μ is a measure in $\mathcal{P}_*^2$ unless otherwise stated.

Let $L_0^2(\mu_i)$ be the set of functions $T_i : \mathbb{R} \to \mathbb{R}$ such that $\int T_i \mathrm{d}\mu_i = 0$ and $\int T_i^2 \mathrm{d}\mu_i < \infty$. The set $\mathcal{T}_{\mathrm{cw}*}(\mu)$ is a subset of $\prod_{i=1}^d L_0^2(\mu_i)$. The space $L_0^2(\mu_i)$ is considered to be a subspace of $L^2(\mu)$. More precisely, $T_i \in L_0^2(\mu_i)$ is identified with the function $x \mapsto T_i(x_i)$ in $L^2(\mu)$.

By relaxing the set $\mathcal{T}_{\mathrm{cw}*}(\mu)$ in (10), we obtain a lower bound of $\beta(\mu)$ as

$$\beta_{\mathrm{L}}(\mu) := \inf_{0 \neq T \in \prod_{i=1}^d L_0^2(\mu_i)} \frac{\|\sum_i T_i\|^2}{\sum_i \|T_i\|^2} \leq \beta(\mu).$$

Therefore, μ is copositive if $\beta_{\mathrm{L}}(\mu) > 0$.

It is shown that $\beta(\mu)$ and $\beta_{\mathrm{L}}(\mu)$ are invariant under coordinate-wise transformations. Thus, $\beta(\mu)$ and $\beta_{\mathrm{L}}(\mu)$ depend only on the copula of μ. Furthermore, they depend only on the set of two-dimensional marginal copulas of μ.

If $d = 2$, then the quantity $\beta_{\mathrm{L}}(\mu)$ is related to the Hirschfeld-Gebelein-Rényi maximal correlation coefficient (refer to [9], [26] and [19])

$$\gamma(\mu) = \sup_{0 \neq T_1 \in L_0^2(\mu_1), 0 \neq T_2 \in L_0^2(\mu_2)} \frac{\langle T_1, T_2 \rangle}{\|T_1\| \|T_2\|}.$$

Lemma 6. Let $d = 2$. Then, $\beta_{\mathrm{L}}(\mu) = 1 - \gamma(\mu)$. In particular, μ is copositive if $\gamma(\mu) < 1$.

Proof. Let $\gamma = \gamma(\mu)$. For any $T_1 \in L_0^2(\mu_1)$ and $T_2 \in L_0^2(\mu_2)$, we have

$$\begin{aligned}
\|T_1 + T_2\|^2 &= \|T_1\|^2 + \|T_2\|^2 + 2\langle T_1, T_2 \rangle \\
&\geq \|T_1\|^2 + \|T_2\|^2 - 2\gamma \|T_1\| \|T_2\| \\
&= \gamma(\|T_1\| - \|T_2\|)^2 + (1 - \gamma)(\|T_1\|^2 + \|T_2\|^2) \\
&\geq (1 - \gamma)(\|T_1\|^2 + \|T_2\|^2).
\end{aligned}$$

Thus, we have $\beta_{\mathrm{L}}(\mu) \geq 1-\gamma$. In order to prove the converse inequality, take sequences T_{1n} and T_{2n} satisfying $\|T_{1n}\| = \|T_{2n}\| = 1$ and $\lim_{n\to\infty}\langle T_{1n}, T_{2n}\rangle = \gamma$. Then,

$$\beta_{\mathrm{L}}(\mu) \leq \lim_{n\to\infty} \frac{\|T_{1n} - T_{2n}\|^2}{\|T_{1n}\|^2 + \|T_{2n}\|^2} = 1-\gamma.$$

□

In the literature, two subspaces H_1 and H_2 of a Hilbert space with the property

$$\sup_{h_1\in H_1, h_2\in H_2} \frac{\langle h_1, h_2\rangle}{\|h_1\|\|h_2\|} < 1$$

are said to satisfy the strengthened Cauchy-Schwarz inequality [6]. In our setting, μ is copositive if $L_0^2(\mu_1)$ and $L_0^2(\mu_2)$ satisfy the strengthened Cauchy-Schwarz inequality.

5.2 Gaussian case

We obtain an explicit expression of $\beta_{\mathrm{L}}(\mu)$ if μ is a multivariate normal distribution.

Lemma 7. Let μ be the multivariate normal distribution with mean vector 0 and covariance matrix S. Then, $\beta_{\mathrm{L}}(\mu)$ is the minimum eigenvalue of the correlation matrix of S. In particular, μ is copositive if S is non-singular.

Proof. The case of $d = 2$ has been proven by [17].

Assume that the marginal density of μ is the standard normal $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ without loss of generality. Then, the covariance matrix coincides with the correlation matrix $R = (\rho_{ij})$. We prove that $\beta_{\mathrm{L}}(\mu) = \lambda_{\min}(R)$, where the minimum eigenvalue of a positive definite matrix A is denoted by $\lambda_{\min}(A)$. Note that $\lambda_{\min}(R) \leq 1$ because $\mathrm{tr}(R) = d$.

Denote the Hermite polynomial of order k by $\eta_k(x) = (-1)^k\phi(x)^{-1}(\mathrm{d}^k/\mathrm{d}x^k)\phi(x)$ for $x \in \mathbb{R}$. Any function $T_i \in L_0^2(\mu_i)$ is expanded as

$$T_i(x_i) = \sum_{k\geq 1} c_{ik}\eta_k(x_i), \quad \sum_{k\geq 1} k! c_{ik}^2 < \infty.$$

Since $\int \eta_k(x_i)\eta_l(x_j)\mathrm{d}\mu = \delta_{kl}(k!\rho_{ij}^k)$, we obtain

$$\int\left(\sum_i T_i\right)^2 \mathrm{d}\mu = \sum_{k\geq 1} k! \sum_i \sum_j c_{ik}c_{jk}\rho_{ij}^k$$

and

$$\sum_i \int T_i^2 \mathrm{d}\mu_i = \sum_{k\geq 1} k! \sum_i c_{ik}^2.$$

For any $k \geq 1$, we can show that

$$\sum_i \sum_j c_{ik} c_{jk} \rho_{ij}^k \geq \lambda_{\min}(R) \sum_i c_{ik}^2.$$

Indeed, set $A_{ij} = \rho_{ij}$ and $B_{ij} = c_{ik} c_{jk} \rho_{ij}^{k-1}$ in an inequality $\mathrm{tr}(AB) \geq \lambda_{\min}(A)\mathrm{tr}(B)$ for any positive definite matrices A and B. Thus, we have

$$\int (\sum_i T_i)^2 \mathrm{d}\mu \geq \lambda_{\min}(R) \sum_i \int T_i^2 \mathrm{d}\mu_i,$$

which implies $\beta_{\mathrm{L}}(\mu) \geq \lambda_{\min}(R)$.

Conversely, let $(v_1, \ldots, v_d)$ be the eigenvector corresponding to $\lambda_{\min}(R)$ and $T_i(x_i) = v_i x_i$. Then, we have $\int (\sum_i T_i)^2 \mathrm{d}\mu = \lambda_{\min}(R) \sum_i \int T_i^2 \mathrm{d}\mu_i$. Thus, $\beta_{\mathrm{L}}(\mu) \leq \lambda_{\min}(R)$. □

We conjecture that $\beta(\mu)$ coincides with (1) if μ is Gaussian and S is its covariance matrix. See Section 7.

5.3 Rényi's condition of positive copula densities

The following theorem, which has been proven by [26] for $d = 2$, provides a checkable condition for copositivity.

Theorem 5 ([26] for $d = 2$)**.** Assume that μ has a regular support (see Section 2 for the definition) and for each pair $i \neq j$, the two-dimensional marginal copula density function c_{ij} of μ is square integrable. Then, $\beta_{\mathrm{L}}(\mu) > 0$. In particular, μ is copositive.

Proof. We first prove that if $T \in \prod_{i=1}^d L_0^2(\mu_i)$ satisfies an equation $\sum_i T_i = 0$, then $T = 0$. Assume $\sum_i T_i = 0$. Let $I \subset \{1, \ldots, d\}$ be the set of indices i such that $\mu(T_i \neq 0) > 0$. Next, by contradiction, assume I is not empty. Let $A_i = \{x_i \mid T_i > 0\}$ for $i \in I$. Since $\int T_i \mathrm{d}\mu_i = 0$, we have $\mu_i(A_i) > 0$. However, based on the assumption about the support, we obtain $\mu(\cap_{i \in I} A_i) > 0$, which implies that $\mu(\sum_i T_i > 0) > 0$ and contradiction. Thus, I is empty, and $T = 0$.

Now, we prove that $\beta_{\mathrm{L}}(\mu) > 0$ using elementary cocepts of functional analysis (refer to [33]). Assume that μ_i is uniform over $[0, 1]$, i.e., μ is a copula distribution. Let $H = \prod_{i=1}^d L_0^2(\mu_i)$ be a Hilbert space of $\mathbb{R}^d$-valued functions and define the inner product of H as $\langle T, U \rangle_H = \sum_i \int T_i U_i \mathrm{d}x_i$. Let c_{ij} be the pairwise copula density and define an operator $C : H \to H$ by

$$(CT)_i = \sum_{j \neq i} \int_0^1 c_{ij} T_j \mathrm{d}x_j, \quad i = 1, \ldots, d.$$

Based on the assumption that $\iint c_{ij}^2 \mathrm{d}x_i \mathrm{d}x_j < \infty$, we deduce that C is a Hilbert-Schmidt operator. It is easy to see that C is self-adjoint. Now, we can write

$$\int (\sum_i T_i)^2 \mathrm{d}\mu = \langle T, (I + C)T \rangle_L \tag{11}$$

If $(I + C)T = 0$, then (11) implies $\sum_i T_i = 0$ and, therefore, $T = 0$. Thus, $I + C$ is injective. Since the operator $I + C$ is an injective Fredholm operator, it is surjective. By the continuous inverse theorem, we deduce that the inverse operator $(I+C)^{-1}$ is bounded. Therefore, we have

$$\langle T, (I + C)T \rangle_L \geq \frac{1}{\|(I + C)^{-1}\|} \langle T, T \rangle_L,$$

which means $\beta_{\mathrm{L}}(\mu) \geq \|(I + C)^{-1}\|^{-1} > 0$. □

Corollary 4. If μ has a positive and bounded copula density function, then μ is copositive.

By Theorem 5, we obtain an alternative proof of Lemma 7 without evaluating $\beta_{\mathrm{L}}(\mu)$ (details omitted). In Section 6, we deal with positive and piecewise uniform copula density functions.

Note that the support of μ is not determined from the support of two-dimensional marginal distributions. See the following example. Refer to [27] for related topics.

Example 3. Let $\mu \in \mathcal{P}^2(\mathbb{R}^4)$ be the uniform measure supported on the region

$$(+, +, -, -) \cup (+, -, +, -) \cup (+, -, -, +) \cup (-, +, +, -) \cup (-, +, -, +) \cup (-, -, +, +),$$

where $(+, +, -, -)$ denotes the set $[0, 1] \times [0, 1] \times [-1, 0] \times [-1, 0]$, and so on. Then μ is not copositive although each two-dimensional marginal distribution is supported on $[-1, 1]^2$. In order to demonstrate this point, let $T_i(x_i) = \mathrm{sign}(x_i)$ for each i. Then $\int T_i \mathrm{d}\mu_i = 0$ and $\int T_i^2 \mathrm{d}\mu_i > 0$ but $\int (\sum_i T_i)^2 \mathrm{d}\mu = 0$. Hence, $\beta(\mu) = 0$. □

5.4 A condition without regular supports

Theorem 5 assumes regularity of the support. Here, we present a result without assuming the regular support condition.

Theorem 6. Let μ be a d-dimensional copula with density $c(x)$. Assume that there exists a constant $0 < \delta \leq 1$ such that an inequality

$$c(x_1, \ldots, \min(x_i, y_i), \ldots, x_d) + c(y_1, \ldots, \max(x_i, y_i), \ldots, y_d) \geq \delta\{c(x) + c(y)\} \tag{12}$$

holds for any i and $x, y \in [0, 1]^d$. Then, μ is copositive.

Proof. We first prove the case $d = 2$. For each $i \in \{1, 2\}$, let $0 \neq T_i \in L_0^2(\mu_i)$, and let $T_i^{\pm}$ be the positive and negative parts of T_i such that $T_i = T_i^+ - T_i^-$. Let $I_i^{\pm}$ be the support of $T_i^{\pm}$. Assume that T_1 is non-decreasing. Then, we have $a_1 < b_1$ for any $a_1 \in I_1^-$ and $b_1 \in I_1^+$. Setting $Z_i = \int T_i^+(x_i)\mathrm{d}x_i = \int T_i^-(x_i)\mathrm{d}x_i > 0$, we obtain

$$
\begin{aligned}
&\langle T_1^-, T_2^- \rangle + \langle T_1^+, T_2^+ \rangle \\
&= \int_{I_1^- \times I_2^-} c(a_1, a_2) T_1^-(a_1) T_2^-(a_2) \mathrm{d}a_1 \mathrm{d}a_2 + \int_{I_1^+ \times I_2^+} c(b_1, b_2) T_1^+(b_1) T_2^+(b_2) \mathrm{d}b_1 \mathrm{d}b_2 \\
&= Z_1 Z_2 \int_{I_1^- \times I_2^- \times I_1^+ \times I_2^+} \{c(a_1, a_2) + c(b_1, b_2)\} \frac{T_1^-(a_1)}{Z_1} \frac{T_2^-(a_2)}{Z_2} \frac{T_1^+(b_1)}{Z_1} \frac{T_2^+(b_2)}{Z_2} \mathrm{d}a_1 \mathrm{d}a_2 \mathrm{d}b_1 \mathrm{d}b_2 \\
&\geq \delta Z_1 Z_2 \int_{I_1^- \times I_2^- \times I_1^+ \times I_2^+} \{c(b_1, a_2) + c(a_1, b_2)\} \frac{T_1^-(a_1)}{Z_1} \frac{T_2^-(a_2)}{Z_2} \frac{T_1^+(b_1)}{Z_1} \frac{T_2^+(b_2)}{Z_2} \mathrm{d}a_1 \mathrm{d}a_2 \mathrm{d}b_1 \mathrm{d}b_2 \\
&= \delta(\langle T_1^+, T_2^- \rangle + \langle T_1^-, T_2^+ \rangle).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|T_1 + T_2\|^2 &= \|T_1\|^2 + \|T_2\|^2 + 2(\langle T_1^+, T_2^+ \rangle - \langle T_1^+, T_2^- \rangle - \langle T_1^-, T_2^+ \rangle + \langle T_1^-, T_2^- \rangle) \\
&\geq \|T_1\|^2 + \|T_2\|^2 - 2(1 - \delta)(\langle T_1^+, T_2^- \rangle + \langle T_1^-, T_2^+ \rangle) \\
&= \delta(\|T_1\|^2 + \|T_2\|^2) + (1 - \delta)(\|T_1^+ - T_2^-\|^2 + \|T_1^- - T_2^+\|^2) \\
&\geq \delta(\|T_1\|^2 + \|T_2\|^2)
\end{aligned}
$$

and the result follows. Note that we did not use the monotonicity of T_2.

Now, we prove the case $d \geq 3$. In the same manner as above, we have

$$
\|T_1 + \cdots + T_d\|^2 \geq \delta\{\|T_1\|^2 + \|T_2 + \cdots + T_d\|^2\}.
$$

Since the condition (12) is invariant under marginalization, it is inductively shown that

$$
\|T_1 + \cdots + T_d\|^2 \geq \delta\|T_1\|^2 + \cdots + \delta^{d-1}\|T_{d-1}\|^2 + \delta^{d-1}\|T_d\|^2.
$$

Thus, μ is copositive, where $\beta(\mu) \geq \delta^{d-1}$. □

For example, if μ is the uniform distribution over the region $[-1, 1]^2 \setminus [-1, 0]^2$, then μ does not have a regular support but is copositive, where the constant δ in (12) is $1/2$.

5.5 Tail dependence

Many useful copulas in application exhibit tail dependence (e.g. [22], [12], [10]). The following lemma shows that, unfortunately, Theorem 5 is not helpful for this class of copulas.

Lemma 8. Let $d = 2$ and assume that the copula density $c(u_1, u_2)$ has lower-tail dependency

$$\lambda = \lim_{\delta \to 0} \frac{\int_0^\delta \int_0^\delta c(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2}{\delta} > 0.$$

Then, $\iint c(u_1, u_2)^2 \mathrm{d}u_1 \mathrm{d}u_2 = \infty$. Similar results hold for other types of tail dependency.

Proof. The Cauchy-Schwarz inequality implies that

$$\int_0^\delta \int_0^\delta c(u_1, u_2)^2 \mathrm{d}u_1 \mathrm{d}u_2 \geq \frac{1}{\delta^2} \left(\int_0^\delta \int_0^\delta c(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 \right)^2.$$

If c is square-integrable, then the left-hand side should converge to 0 as $\delta \to 0$, which is impossible. Thus, c is not square-integrable. □

We conjecture that many copulas with tail dependence are copositive. On the other hand, there is a non-copositive measure with positive copula density, as follows.

Example 4 (Tail counter-comonotonic copula)**.** It is known that there is a positive copula density function with the property

$$\lim_{\delta \to 0} P(X_2 < \delta | X_1 < \delta) = 1,$$

which is equivalent to $\lambda = 1$ in Lemma 8. Such a copula is referred to as a lower tail comonotonic copula (see Section 2.21 of [12]). Let μ be the induced measure of $Y_1 = X_1$ and $Y_2 = 1 - X_2$. Then, μ is not copositive. Indeed, define a map $T \in \mathcal{T}_{\mathrm{cw}*}(\mu)$ by

$$T_1(y_1) = \delta I_{(\delta,1)}(y_1) - (1-\delta) I_{(0,\delta)}(y_1), \quad T_2(y_2) = (1-\delta) I_{(1-\delta,1)}(y_2) - \delta I_{(0,1-\delta)}(y_2),$$

where I_A denotes the indicator function of a set A. Then, $\|T_1\| = \|T_2\| = \sqrt{\delta(1-\delta)}$ and

$$\langle T_1, T_2 \rangle = -\delta(1-\delta)^2 P(Y_2 > 1 - \delta | Y_1 < \delta) + O(\delta^2), \quad \delta \to 0.$$

Therefore,

$$\lim_{\delta \to 0} \frac{\langle T_1, T_2 \rangle}{\|T_1\| \|T_2\|} = -1.$$

In a similar manner to Lemma 6, we deduce that $\beta(\mu) = 0$. □

6 Piecewise uniform densities

In this section, it is shown that if μ has piecewise uniform density function, then the Stein-type transformation of μ is obtained by finite-dimensional optimization. Here, we do not impose the zero mean condition on measures μ as the preceding section.

We say that a probability density function $c(u)$ on $[0,1]^d$ is piecewise uniform if its two-dimensional marginal densities are written as

$$c_{ij}(u_i,u_j) = n^2\pi_{ab}^{ij} \quad \text{if} \quad (u_i,u_j) \in (\tfrac{a-1}{n},\tfrac{a}{n}] \times (\tfrac{b-1}{n},\tfrac{b}{n}], \quad a,b \in \{1,\dots,n\}, \tag{13}$$

for some n, where π_{ab}^{ij} is a positive number such that

$$\sum_{a=1}^{n}\sum_{b=1}^{n}\pi_{ab}^{ij} = 1.$$

Let $\pi_a^i = \sum_{b=1}^{n}\pi_{ab}^{ij}$. Note that c is not necessarily a copula density. However, it is transformed by a piecewise linear transform into a copula density. Then, Corollary 4 guarantees the existence of a Stein-type transformation.

By solving Equation (6), we obtain an expression of the Stein-type transformation of c as follows. Denote the cumulative distribution function and density function of the standard normal distribution by Φ and ϕ, respectively.

Lemma 9. Let c satisfy (13), and let p be the Stein-type density corresponding to c. Then, there exist real constants $\alpha_{1i},\dots,\alpha_{ni}$ and $\xi_{1i} < \cdots < \xi_{n-1,i}$ such that

$$p_i(x_i) = \pi_a^i \frac{\phi(x_i-\alpha_{ai})}{Z_{ai}} \quad \text{for} \quad \xi_{a-1,i} < x_i \le \xi_{ai} \tag{14}$$

where $\xi_{0i} = -\infty$, $\xi_{ni} = \infty$, and $Z_{ai} = \Phi(\xi_{ai}-\alpha_{ai}) - \Phi(\xi_{a-1,i}-\alpha_{ai})$. The Stein-type transformation is

$$x_i = T_i(u_i) = \alpha_{ai} + \Phi^{-1}\left(\Phi(\xi_{a-1,i}-\alpha_{ai}) + n(u_i - \tfrac{a-1}{n})Z_{ai}\right), \quad u_i \in (\tfrac{a-1}{n},\tfrac{a}{n}], \tag{15}$$

and the two-dimensional marginal density is

$$p_{ij}(x_i,x_j) = \pi_{ab}^{ij}\frac{\phi(x_i-\alpha_{ai})}{Z_{ai}}\frac{\phi(x_j-\alpha_{bj})}{Z_{bj}}, \quad (x_i,x_j) \in (\xi_{a-1,i},\xi_{ai}] \times (\xi_{b-1,j},\xi_{bj}]. \tag{16}$$

Furthermore, the following identity is satisfied:

$$\alpha_{ai} = -\sum_{j\neq i}\sum_{b}\frac{\pi_{ab}^{ij}}{\pi_a^i}\int_{\xi_{b-1,j}}^{\xi_{bj}}\frac{x_j\phi(x_j-\alpha_{bj})}{Z_{bj}}\mathrm{d}x_j. \tag{17}$$

Proof. Equation (6) implies that $\partial_i p_i(x_i) = -(x_i + \sum_{j\neq i} E[X_j|x_i])p_i(x_i)$, where $\partial_i = \partial/\partial x_i$. Since the conditional expectation $E[X_j|x_i]$ has to be piecewise constant, $p_i(x_i)$ is piecewise Gaussian up to a normalizing constant. Since the mass of each piece is preserved under a coordinate-wise transformation, we obtain the form (14). Then, the unique monotone transformation (15) is derived from $c_i(u_i)\mathrm{d}u_i = p_i(x_i)\mathrm{d}x_i$. Equation (16) results from the transformation of $c_{ij}(u_i,u_j)$. Finally, Equation (17) is obtained from $\partial_i \log p_i(x_i) = -(x_i + \sum_{j\neq i} E[X_j|x_i])$. □

The parameters α_{ai} and ξ_{ai} are determined by the continuity of (14) at $x_i = \xi_{ai}$ and the identity (17). However, instead of solving the simultaneous equations directly, we adopt an optimization approach.

Assume the density of a distribution μ obeys the parametric form given by Equation (14). Then, the energy function $\mathcal{E}(\mu)$ defined in Section 4 is a function of α and ξ, which is denoted by $F(\alpha, \xi)$ and is obtained as follows:

$$\begin{aligned}
F(\alpha,\xi) &= \sum_i \int p_i(x_i)\log p_i(x_i) + \frac{1}{2}\sum_i \int x_i^2 p_i(x_i)\mathrm{d}x_i + \sum_{i<j}\int x_i x_j p_{ij}(x_i,x_j)\mathrm{d}x_i\mathrm{d}x_j \\
&= \sum_i\sum_a \pi_a^i \int_{\xi_{a-1,i}}^{\xi_{ai}} \frac{\phi(x_i-\alpha_{ai})}{Z_{ai}}\left(\log\frac{\pi_a^i}{\sqrt{2\pi}} - \frac{(x_i-\alpha_{ai})^2}{2} - \log Z_{ai} + \frac{x_i^2}{2}\right) \\
&\quad + \sum_{i<j}\sum_a\sum_b \frac{\pi_{ab}^{ij}}{Z_{ai}Z_{bj}}\int_{\xi_{a-1,i}}^{\xi_{ai}}\int_{\xi_{b-1,j}}^{\xi_{bj}} x_i x_j \phi(x_i-\alpha_{ai})\phi(x_j-\alpha_{bj})\mathrm{d}x_i\mathrm{d}x_j \\
&= \sum_i\sum_a \pi_a^i \log\frac{\pi_a^i}{\sqrt{2\pi}} + \sum_i\sum_a \pi_a^i\left(-\frac{\alpha_{ai}^2}{2} + \alpha_{ai}M_{ai} - \log Z_{ai}\right) \\
&\quad + \sum_{i<j}\sum_a\sum_b \pi_{ab}^{ij} M_{ai}M_{bj},
\end{aligned}$$

where

$$\begin{aligned}
M_{ai} &= \frac{1}{Z_{ai}}\int_{\xi_{a-1,i}}^{\xi_{ai}} x_i\phi(x_i-\alpha_{ai})\mathrm{d}x_i \\
&= \alpha_{ai} + \frac{1}{Z_{ai}}\left(-\phi(\xi_{ai}-\alpha_{ai}) + \phi(\xi_{a-1,i}-\alpha_{ai})\right).
\end{aligned}$$

Since Z_{ai} and M_{ai} are functions of three parameters α_{ai}, ξ_{ai}, and $\xi_{a-1,i}$, we denote the corresponding partial derivative by D_1, D_2, and D_3. The derivatives of F are

$$\begin{aligned}
\frac{\partial F}{\partial \alpha_{ai}} &= \pi_a^i\left(-\alpha_{ai} + M_{ai} + \alpha_{ai}D_1M_{ai} - \frac{D_1Z_{ai}}{Z_{ai}}\right) + \sum_{j\neq i}\sum_b \pi_{ab}^{ij}(D_1M_{ai})M_{bj} \\
&= \pi_a^i\left(\alpha_{ai} + \sum_{j\neq i}\sum_b \frac{\pi_{ab}^{ij}}{\pi_a^i}M_{bj}\right)(D_1M_{ai}), \qquad (18) \\
\frac{\partial F}{\partial \xi_{ai}} &= \pi_a^i\left(\alpha_{ai}D_2M_{ai} - \frac{D_2Z_{ai}}{Z_{ai}}\right) + \pi_{a+1}^i\left(\alpha_{a+1,i}D_3M_{a+1,i} - \frac{D_3Z_{a+1,i}}{Z_{a+1,i}}\right) \\
&\quad + \sum_{j\neq i}\sum_b\left\{\pi_{ab}^{ij}(D_2M_{ai}) + \pi_{a+1,b}^{ij}(D_3M_{a+1,i})\right\}M_{bj} \\
&= \pi_a^i\left(\alpha_{ai} + \sum_{j\neq i}\sum_b \frac{\pi_{ab}^{ij}}{\pi_a^i}M_{bj}\right)D_2M_{ai} + \pi_{a+1}^i\left(\alpha_{a+1,i} + \sum_{j\neq i}\sum_b\frac{\pi_{a+1,b}^{ij}}{\pi_{a+1}^i}M_{bj}\right)D_3M_{a+1,i} \\
&\quad - \pi_a^i\frac{D_2Z_{ai}}{Z_{ai}} - \pi_{a+1}^i\frac{D_3Z_{a+1,i}}{Z_{a+1,i}}. \qquad (19)
\end{aligned}$$

By using these formulas, we obtain the following theorem.

Theorem 7. A stationary point of F together with formula (14) provides the global minimum point of the energy functional $\mathcal{E}(\mu)$ over the fiber. In other words, F has a unique stationary point that corresponds to the Stein-type density.

Proof. Since $M_{ai} = \int x_i \phi(x_i - \alpha_{ai}) \mathrm{d}x_i / Z_{ai}$ is the expectation parameter of an exponential family $\phi(x_i - \alpha_{ai})/Z_{ai}$, it is an increasing function of α_{ai} (e.g., [18]). Therefore, $D_1 M_{ai} > 0$. Thus, the stationary condition $\partial F / \partial \alpha_{ai} = 0$ is equivalent to

$$\alpha_{ai} + \sum_{j \neq i} \sum_{b} \frac{\pi_{ab}^{ij}}{\pi_a^i} M_{bj} = 0,$$

which is equivalent to (17) and solves the integral equation (6) except at boundary points ξ_{ai}. Furthermore, substituting this relation into (19), we obtain

$$\begin{aligned} \frac{\partial F}{\partial \xi_{ai}} &= -\pi_a^i \frac{D_2 Z_{ai}}{Z_{ai}} - \pi_{a+1}^i \frac{D_3 Z_{a+1,i}}{Z_{a+1,i}} \\ &= -p_i(\xi_{ai}-) + p_i(\xi_{ai}+). \end{aligned}$$

Therefore, $\partial F / \partial \xi_{ai} = 0$ is equivalent to the continuity of p_i at ξ_{ai}. Then, the density p is the Stein-type density, which is unique due to Theorem 2. □

The minimization problem of $F(\alpha, \xi)$ over $\alpha_{ai} \in \mathbb{R}$ and $\xi_{1i} < \cdots < \xi_{n-1,i}$ is performed using a standard optimization package (e.g., the function `optim` in R [24]) when the coordinate $\tau_{ai} = \xi_{ai} - \xi_{a-1,i}$, rather than ξ_{ai}, is used for $2 \leq a \leq n-1$.

Example 5. We numerically obtain the Stein-type densities of discretized copulas. The result is shown in Figure 1. The copula used here is the Clayton copula

$$C_\theta(x_1, x_2) = \left[\max(x_1^{-\theta} + x_2^{-\theta} - 1, 0)\right]^{-1/\theta}.$$

The discretized copula density of $n \times n$ cells is given by (13) with

$$\pi_{ab}^{12} = C_\theta(\tfrac{a}{n}, \tfrac{b}{n}) - C_\theta(\tfrac{a-1}{n}, \tfrac{b}{n}) - C_\theta(\tfrac{a}{n}, \tfrac{b-1}{n}) + C_\theta(\tfrac{a-1}{n}, \tfrac{b-1}{n}).$$

7 Discussion

In the present paper, we showed that a class of multi-dimensional distributions has a unique representation via the Stein-type identity. Now, we describe areas for future study and some open problems.

In Section 3, we derived some properties of Stein-type distributions. The author could not find any counter-example against the following conjecture.

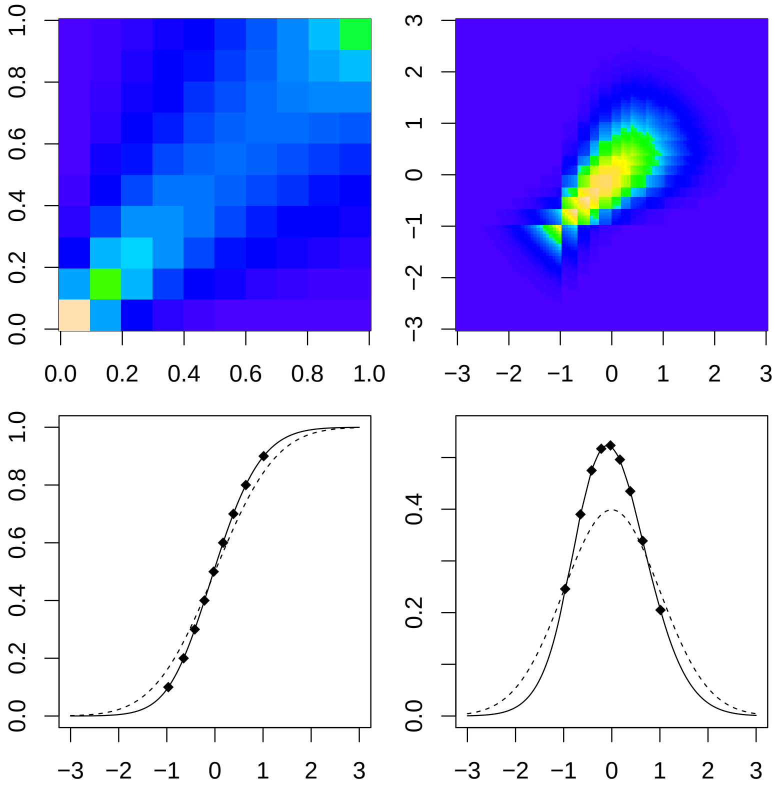

Figure 1: Coordinate-wise transformation for the two-dimensional discretized Clayton copula with $\theta = 2$ and $n = 10$ is shown. The joint density function (top-left) is transformed into a Stein-type density function (top-right) by the inverse function of a cumulative distribution (bottom-left), the density of which is piece-wise Gaussian (bottom-right). The dashed curve in the bottom figures represents the standard normal distribution.

Conjecture 1. The marginal density function of any Stein-type distribution is positive everywhere.

A partial answer to Conjecture 1 is given in the following lemma.

Lemma 10. Let μ be a Stein-type distribution. If the copula of μ has pair-wise marginal densities c_{ij} such that

$$D = \sup_{i,j:i\neq j} \sup_{u_i\in[0,1]} \int_0^1 c_{ij}(u_i,u_j)^2 \mathrm{d}u_j < \infty,$$

then each marginal density p_i of μ is positive everywhere. In particular, if the copula density of μ is bounded, then the same consequence follows.

Proof. The density $p_i(x_i)$ satisfies $\partial_i p_i(x_i) + p_i(x_i)m_i(x_i) = 0$ with $m_i(x_i) = E[\sum_j X_j|x_i]$ by Theorem 4. The conditional expectation satisfies

$$|\mathrm{E}[X_j|x_i]| = \left|\int x_j c_{ij}(F_i(x_i), F_j(x_j))p_j(x_j)\mathrm{d}x_j\right| \leq (D\mathrm{E}[X_j^2])^{1/2},$$

where $F(x_i) = \int_{-\infty}^{x_i} p_i(\xi)\mathrm{d}\xi$. Let $D_* = \sum_{j\neq i}(D\mathrm{E}[X_j^2])^{1/2}$. Then, we obtain an inequality

$$-(x_i + D_*)p_i(x_i) \leq \partial_i p_i(x_i) \leq -(x_i - D_*)p_i(x_i)$$

Let $a \in \mathbb{R}$ be a point at which $p_i(a) > 0$. Then, Gronwall's lemma shows that $p_i(x_i) \geq p_i(a)e^{-(x_i+D_*)^2/2+(a+D_*)^2/2} > 0$ for $x_i > a$, and similarly $p_i(x_i) > 0$ for $x_i < a$. □

If Conjecture 1 is positively solved, then the following conjecture, which is based on Theorem 2, is also positive according to Lemma 4 (ii).

Conjecture 2. A Stein-type transformation is unique if it exists.

We state a relevant conjecture that is the converse of Theorem 3.

Conjecture 3. A distribution is copositive if it has a Stein-type transformation.

In Section 4, we showed that a Stein-type distribution is characterized by the stationary point of an energy functional $\mathcal{E}$ over a fiber $\mathcal{F}$. From the perspective of optimal transportation, we can construct the gradient flow of the energy functional with respect to the L^2-Wasserstein space ([14], [23] and [32]). The formal equation is as follows

$$\partial_t p_i = \partial_i(\partial_i p_i + m_i p_i), \quad t \geq 0, \quad i = 1, \ldots, d, \tag{20}$$

where $m_i(x_i) = \mathrm{E}[\sum_j X_j|x_i]$. Although this appears to be an independent system of one-dimensional Fokker-Planck equations, the equations interact with each other via $m_i(x_i)$.

Moreover, the physical meaning of the equation is not clear. From Theorem 4, it follows that each Stein-type density is a stationary point of (20). The time evolution will be theoretically of interest.

In Section 5, we presented sufficient conditions for copositivity of distributions. In particular, a Gaussian distribution is copositive if its covariance matrix is not degenerated. Conversely, if a Gaussian distribution is copositive, then the covariance matrix must, by definition, be strictly copositive (see Equation (1)). The following conjecture naturally arises but is not proven. This is positively solved if Conjecture 3 is correct, due to Lemma 2.

Conjecture 4. A Gaussian distribution is copositive if the covariance matrix is strictly copositive.

As stated in Subsection 5.5, tail-dependent copulas do not satisfy the sufficient condition in Theorem 5. The copositivity of tail-dependent copulas remains unclear.

In the present paper, we did not consider statistical models that explain a given data set. A statistical model involving a Stein-type distribution is essentially equivalent to a copula model because such models correspond to each other through coordinate-wise transformations, whereas the marginal distributions are not of much interest in copula modelling. The class given in Example 1 provides a flexible model because the distribution of U_i's in the construction can be selected arbitrarily.

Finally, it is expected that there is a coordinate-wise transformation to satisfy

$$\mathrm{E}[f(X_i)g(X_1 + \cdots + X_d)] > 0, \quad i = 1, \ldots, d, \tag{21}$$

for any monotone increasing functions f and g. Although the condition (21) appears to be too strong, how to deal with this problem remains unclear.

Acknowledgments

The present study was supported by JSPS KAKENHI Grant Numbers JP26108003 and JP26540013.

Appendix

A One-dimensional optimal transportation

Necessary information about one-dimensional optimal transportation is summarized. Refer to [25] and [32] for further details.

Let $\mathcal{P}^2(\mathbb{R})$ be the set of absolutely continuous probability distributions μ on $\mathbb{R}$ such that $\int x\mathrm{d}\mu = 0$ and $\int x^2\mathrm{d}\mu < \infty$. For given $\mu \in \mathcal{P}^2(\mathbb{R})$, let $\mathcal{T}(\mu)$ be the set of non-decreasing functions $T : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ such that $T_\sharp\mu \in \mathcal{P}^2(\mathbb{R})$.

For given μ and ν in $\mathcal{P}^2(\mathbb{R})$, there exists $T \in \mathcal{T}(\mu)$ such that $\nu = T_\sharp\mu$. The map is uniquely determined μ-almost everywhere. More explicitly, T is given by $T = G^- \circ F$, where $F(x) = \int_{-\infty}^x \mathrm{d}\mu$, $G(x) = \int_{-\infty}^x \mathrm{d}\nu$, and $G^-(u) = \inf\{x \in \mathbb{R} \mid G(x) > u\}$. The map T is called the optimal transportation from μ to ν because this map minimizes the functional $\int (T(x) - x)^2\mathrm{d}\mu$ over $\{T \mid T_\sharp\mu = \nu\}$. Since μ and ν are absolutely continuous, T is decomposed into an absolutely continuous part, T^{ac}, and a discontinuous part, T^{d}, without a singular continuous part. This is because G^- constructed above has the same property. The decomposition is unique up to a μ-negligible set.

The following lemmas are used in Section 4 and Section 5. These lemmas were originally proven for multi-dimensional measures but here we simplify them for the one-dimensional case.

Lemma 11 (Theorem 4.4 of [21])**.** For given μ and ν in $\mathcal{P}^2(\mathbb{R})$, let T be a unique monotone map such that $\nu = T_\sharp\mu$. Let p and q be density functions of μ and ν, respectively. Let $X \subset \mathbb{R}$ denote the set of points where the derivative T' is defined and positive. Then, $\mu(X) = 1$. Furthermore,

$$\int A(q(y))dy = \int_X A\left(\frac{p(x)}{T'(x)}\right) T'(x)\mathrm{d}x$$

for any measurable function A on $[0, \infty)$ with $A(0) = 0$.

Lemma 12 (Proposition 1.3 of [21])**.** Let $\mu \in \mathcal{P}^2$ and $T \in \mathcal{T}(\mu)$. Then, $(1-t)\mathrm{Id} + tT \in \mathcal{T}(\mu)$ for each $t \in [0, 1]$.

Lemma 13 (Proposition 4.2 of [21])**.** Let $\mu \in \mathcal{P}^2$. If $T : \mathbb{R} \to \mathbb{R}$ is a non-decreasing function written as $T = T^{\mathrm{ac}} + T^{\mathrm{d}}$ and the derivative $(T^{\mathrm{ac}})'$ of the absolutely continuous part is strictly positive μ-almost everywhere, then $T_\sharp\mu$ is absolutely continuous.

B Explicit expression of Stein-type distributions

We formally derive an explicit expression of the Stein-type distributions.

Assume that $\mu \in \mathcal{P}^2$ has a smooth density function p with decay at infinity. Then, μ is Stein-type if and only if there exists a function $r(x)$ such that

$$\sum_{j=1}^{p} \left(\frac{\partial p(x)}{\partial x_j} + x_j p(x)\right) = r(x), \quad \int_{\mathbb{R}^{d-1}} r(x)\mathrm{d}x_{-i} = 0 \quad \text{for all } i, \tag{22}$$

where $\mathrm{d}x_{-i}$ means $\mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\mathrm{d}x_{i+1} \cdots \mathrm{d}x_d$. In fact, formula (22) is rewritten as $\partial p_i/\partial x_i + p_i(x_i)m_i(x_i) = 0$, where $m_i(x_i)$ is the conditional expectation of $\sum_j x_j$ given x_i, and this equation is equivalent to (6).

Equation (22) is explicitly solved if $r(x)$ is given. Let Q be a fixed orthogonal matrix such that $(Q^\top x)_1 = \sum_j x_j/\sqrt{d}$, where $Q^\top$ denotes the matrix transpose of Q. Then, (22) is written as

$$\frac{\partial p(Qw)}{\partial w_1} + w_1 p(Qw) = \frac{r(Qw)}{\sqrt{d}}, \quad w = Q^\top x.$$

The general solution is

$$p(Qw) = \frac{1}{\sqrt{2\pi}} e^{-w_1^2/2} \left(q(w_2, \ldots, w_d) + \int_0^{w_1} \frac{\sqrt{2\pi}}{\sqrt{d}} e^{v_1^2/2} r(Q(v_1, w_2, \ldots, w_d)) dv_1 \right),$$

where q is any probability density function on $\mathbb{R}^{d-1}$.

In particular, if $r(x) = 0$, we obtain a simple formula

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-w_1^2/2} q(w_2, \ldots, w_d), \quad w = Q^\top x. \tag{23}$$

Example 1 in Section 2 is this solution. The class of densities (23) is characterized by a stronger condition than the Stein-type identity, i.e.,

$$\int f(x) \sum_j x_j \mathrm{d}\mu = \int \sum_i \frac{\partial f(x)}{\partial x_i} \mathrm{d}\mu$$

for any function $f(x) = f(x_1, \ldots, x_d)$.

C Closedness properties of Stein-type distributions

Let $\mathcal{S}$ be the set of Stein-type distributions on $\mathbb{R}^d$. We prove that $\mathcal{S}$ is closed under mixture, normalized convolution, and weak limit.

Lemma 14 (Mixture)**.** If μ and ν are two distributions in $\mathcal{S}$, then $(1-t)\mu + t\nu$ belongs to $\mathcal{S}$ for any $t \in [0, 1]$.

Proof. This follows from the linearity of the Stein-type identity (3) with respect to μ. □

Lemma 15 (Normalized convolution)**.** Let $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_d)$ be independent random vectors with Stein-type distributions. Let a and b be real numbers with $a^2 + b^2 = 1$. Then, $aX + bY$ has a Stein-type distribution.

Proof. The Stein-type identity with respect to X implies that

$$\mathrm{E}\left[f(aX_i+bY_i)\left(\sum_j X_j\right)\right]=a\mathrm{E}\left[f'(aX_i+bY_i)\right]$$

for each i, because X and Y are independent. By changing the roles of X and Y, we have

$$\mathrm{E}\left[f(aX_i+bY_i)\left(\sum_j Y_j\right)\right]=b\mathrm{E}\left[f'(aX_i+bY_i)\right]$$

Their average is

$$\mathrm{E}\left[f(aX_i+bY_i)\left(\sum_j (aX_j+bY_j)\right)\right]=(a^2+b^2)\mathrm{E}\left[f'(aX_i+bY_i)\right].$$

Thus, the Stein-type identity for $aX+bY$ holds if and only if $a^2+b^2=1$. □

The set $\mathcal{S}$ is also closed under weak limit in the following sense. Denote the Euclidean norm on $\mathbb{R}^d$ by $\|x\|$ for $x\in\mathbb{R}^d$.

Lemma 16 (Weak convergence). Let $\mu^{(n)}$ be a sequence in $\mathcal{S}$. If $\mu^{(n)}$ converges to μ in law and $\int\|x\|^2\mathrm{d}\mu^{(n)}$ converges to $\int\|x\|^2\mathrm{d}\mu<\infty$, then μ belongs to $\mathcal{S}$.

Proof. These conditions imply that $\int\varphi\mathrm{d}\mu^{(n)}\to\int\varphi\mathrm{d}\mu$ for any continuous function φ such that $|\varphi(x)|\le C(1+\|x\|^2)$ for some $C>0$. (Refer to Theorem 7.12 of [32].) Letting $\varphi(x)$ be $f(x_i)\sum_j x_j$ and $f'(x_i)$, respectively, we obtain the Stein-type identity for μ. Absolute continuity of μ_i is shown in the same manner as in the proof of Theorem 4. □

The condition regarding moment convergence in Lemma 16 is necessary. Indeed, we can construct a sequence $(W,U^{(n)})$ of Stein-type random variables in the same manner as in Example 1 of Section 2 such that $U^{(n)}$ converges in law to a random variable U with $\mathrm{E}[U^2]=\infty$.

By Lemma 15 and Lemma 16 together with the central limit theorem, if we have independent and identically distributed samples $X^1,\ldots,X^n$ according to a Stein-type distribution μ, then the limit distribution of $(X^1+\cdots+X^n)/\sqrt{n}$ is a Stein-type normal distribution that is characterized by Lemma 1.

Note that the set of copulas satisfies the same consequence as Lemma 14 and Lemma 16. If we modify the definition of the copulas in such a way that the marginal distribution is standard normal, then the same consequence as Lemma 15 also follows.

References

[1] Alfonsi, A., Jourdain, B. (2014). A remark on the optimal transport between two probability measures sharing the same copula, *Statist. Probab. Let.*, **84**, 131–134.

[2] Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient Flows – in Metric Spaces and in the Space of Probability Measures*, Birkhäuser.

[3] Borwein, J. M., Lewis, A. S., and Nussbaum, R. D. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels, *J. Funct. Anal.*, **123**, 264–307.

[4] Butucea, C., Delmas, J., Dutfoy, A., and Fischer, R. (2015). Maximum entropy copula with given diagonal section, *J. Multivariate Anal.*, **137**, 61–81.

[5] Chen, L. H. Y., Goldstein, L., and Shao, Q. (2011). *Normal Approximation by Stein's Method*, Springer.

[6] De Rossi, A. and Rodino, L. (2004). Strengthened Cauchy-Schwarz inequality for biorthogonal wavelets in Sobolev spaces, *J. Math. Anal. Appl.*, **299**, 49–60.

[7] Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N., and Zwiernik, P. (2016). Total positivity in Markov structures, *Ann. Statist.*, to appear.

[8] Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets, *Comm. Math. Phys.*, **22**, 89–103.

[9] Gebelein, H. (1941). Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung, *Z. angew. Math. Mech.*, **21** (6), 364–379.

[10] Hua, L. (2012). Multivariate Extremal Dependence and Risk Measures, Ph. D. Thesis in the University of British Columbia.

[11] Joe, H. (1989). Relative entropy measures of multivariate dependence, *J. Amer. Statist. Assoc.*, **84**, 157–164.

[12] Joe, H. (2014). *Dependence Modeling with Copulas*, CRC Press.

[13] Johnson, O. and Barron, A. (2004). Fisher information inequalities and the central limit theorem, *Probab. Theory Relat. Fields*, **129**, 391–409.

[14] Jordan, R., Kinderlehrer, D., and Otto, F. (1998). The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.*, **29** (1), 1–17.

[15] Kalantari, B. (1996). A theorem of the alternative for multihomogeneous functions and its relationship to diagonal scaling of matrices. Linear Algebra Appl. 236, 1–24.

[16] Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. multivariate totally positive distributions, *J. Multivariate Anal.*, **10**, 467–498.

[17] Lancaster, H. O. (1957). Properties of the bivariate normal distribution considered in the form of a contingency table, *Biometrika*, **44**, 289–292.

[18] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, Springer.

[19] Lopez-Paz, D., Hennig, P., and Schölkopf, B. (2013). The randomized dependence coefficient, In *Advances in Neural Information Processing Systems*, **16**, 1–9.

[20] Marshall, A. W., Olkin, I., (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12, 83–90.

[21] McCann, R. J. (1997). A convexity principle for interacting gases, *Adv. Math.*, **128**, 153–179.

[22] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer.

[23] Otto, F. (2001). The geometry of dissipative evolution equations: the porus medium equation, *Comm. Partial Diff. Eq.*, **26**, 101–174.

[24] R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundations for Statistical Computing, Vienna, Austria. http://www.R-project.org/

[25] Rachev, S. T. and Rüschendorf, L. (1998). *Mass Transportation Problems I: Theory*, Springer-Verlag.

[26] Rényi (1959), On measures of dependence, *Acta Math. Acad. Sci. Hungar.*, **10**, 441–451.

[27] Rüschendorf, L. (2013). *Mathematical Risk Analysis*, Springer.

[28] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, 147, 247–264.

[29] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, Vol. 2, 583–602.

[30] Studený, M. (2005). *Probabilistic Conditional Independence Structures*, Springer.

[31] Takatsu, A. (2011). Wasserstein geometry of Gaussian measures, *Osaka J. Math.*, **48** (4), 1005–1026.

[32] Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society.

[33] Zeidler, E. (1995). *Applied Functional Analysis – Main Principles and their Applications*, Applied Mathematical Sciencs, 109, Springer.