

# Joint mixability and notions of negative dependence

Takaaki Koike\*      Liyuan Lin<sup>†</sup>      Ruodu Wang<sup>‡</sup>

January 1, 2024

## Abstract

A joint mix is a random vector with a constant component-wise sum. The dependence structure of a joint mix minimizes some common objectives such as the variance of the component-wise sum, and it is regarded as a concept of extremal negative dependence. In this paper, we explore the connection between the joint mix structure and popular notions of negative dependence in statistics, such as negative correlation dependence, negative orthant dependence and negative association. A joint mix is not always negatively dependent in any of the above senses, but some natural classes of joint mixes are. We derive various necessary and sufficient conditions for a joint mix to be negatively dependent, and study the compatibility of these notions. For identical marginal distributions, we show that a negatively dependent joint mix solves a multi-marginal optimal transport problem for quadratic cost under a novel setting of uncertainty. Analysis of this optimal transport problem with heterogeneous marginals reveals a trade-off between negative dependence and the joint mix structure.

*Keywords:* Joint mixability; negative dependence; optimal transport; extreme dependence; uncertainty

## 1 Introduction

Dependence among multiple sources of randomness has always been an active topic in operations research, statistics, transport theory, economics, and finance; see [Denuit et al. \(2005\)](#), [Joe \(2014\)](#), [Rüschendorf \(2013\)](#), [McNeil et al. \(2015\)](#) and [Galichon \(2016\)](#) for standard textbook treatment in different fields, and the recent work [Blanchet et al. \(2020\)](#) for relevant examples in operations research. In contrast to positive, which received much attention in the literature,

---

\*Graduate School of Economics, Hitotsubashi University, Japan. ✉ [takaaki.koike@r.hit-u.ac.jp](mailto:takaaki.koike@r.hit-u.ac.jp)

<sup>†</sup>Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ [1891lin@uwaterloo.ca](mailto:1891lin@uwaterloo.ca)

<sup>‡</sup>Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ [wang@uwaterloo.ca](mailto:wang@uwaterloo.ca)

considerably fewer studies are found on negative dependence, partially due to its more complicated mathematical nature. For a review and historical account on extremal positive and negative dependence concepts, we refer to [Puccetti and Wang \(2015\)](#).

In the past decade, the notion of joint mixability proposed by [Wang et al. \(2013\)](#), which generalizes complete mixability ([Wang and Wang, 2011](#)), has been shown useful for solving many optimization problems involving the dependence of multiple risks. In particular, joint mixability is essential to worst-case bounds on Value-at-Risk and other risk measures under dependence uncertainty ([Puccetti and Rüschendorf, 2013](#); [Embrechts et al., 2013](#); [Bernard et al., 2014](#)), as well as bottleneck assignment and scheduling problems ([Coffman and Yannakakis, 1984](#); [Hsu, 1984](#); [Haus, 2015](#); [Bernard et al., 2018](#)).

Joint mixability concerns, for given marginal distributions, the existence of a random vector which has a constant component-wise sum. Such a random vector is called a *joint mix* supported by the given marginal distributions, and it represents a very simple concept of dependence. A joint mix is commonly regarded as a notion of extremal negative dependence; see the review of [Puccetti and Wang \(2015\)](#). The reason why a joint mix represents negative dependence is that it minimizes many objectives which are maximized by comonotonicity. For instance, for fixed marginal distributions of the risks, comonotonicity maximizes the variance, the stop-loss premium, and the Expected Shortfall (ES) of the sum of the risks, whereas a joint mix, if it exists, minimizes these quantities; see e.g., [Rüschendorf \(2013\)](#). As such, joint mixability is seen as the safest dependence structure, as long as risk aggregation is concerned ([Embrechts et al., 2014](#)).

Although a joint mix has been treated as a concept of negative dependence, it remains unclear whether it is consistent with classic notions of negative dependence in statistics. Popular notions of negative dependence include negative correlation dependence (NCD), negative orthant dependence (NOD; [Block et al., 1982](#); [Lehmann, 1966](#)) and negative association (NA; [Alam and Saxena, 1981](#); [Joag-Dev and Proschan, 1983](#)). The connection between joint mixes and these negative dependence concepts is the main object that we address in this paper. We obtain some necessary and sufficient conditions for a joint mix to be NOD or NA in [Section 3](#). Some characterization results are obtained in [Section 5](#) for the class of elliptical distributions. In particular, among all elliptical classes, only the Gaussian family supports NOD joint mixes of any dimension.

Since a joint mix may be either negatively dependent or not, a natural question is whether there are special features of negatively dependent joint mixes which are useful in applications.

For this question, we consider a multi-marginal optimal transport problem under uncertainty on the set of components. A few optimality results on negatively dependent joint mixes are obtained, and they demonstrate an interesting interplay between joint mixes and negative dependence. In particular, for the special case of quadratic cost, we show that the optimizer has to be an NCD JM in some settings. This is the topic of Section 4.

The study of joint mixability was originally motivated by questions in risk management and operations research, and it has a strong connection to the theory of multi-marginal optimal transport (Santambrogio, 2015; Pass, 2015) and variance reduction in random sampling (Craiu and Meng, 2001, 2005); see also our Section 4. Recently, there is a growing spectrum of applications of joint mixability outside the above fields, including multiple hypothesis testing (Vovk et al., 2022), wireless communications (Besser and Jorswieck, 2020), labor market matching (Boerma et al., 2021), and resource allocation games (Perchet et al., 2022). Results in this paper connect the two topics of joint mixability and negative dependence, allowing us to bring tools from one area to the other.

This paper is organized as follows. Section 2 introduces the concepts of negative dependence and joint mixability, and summarizes their basic interrelationships. Section 3 explores conditions for joint mixes to be negatively dependent. Two results on decompositions of joint mixes into negatively dependent ones are also obtained. Section 4 studies a multi-marginal optimal transport problem as an application of negatively dependent joint mixes. Section 5 studies joint mixes within the elliptical family, and we obtain a new characterization of the Gaussian family as the only one supporting a negatively dependent elliptical distribution for every dimension. Section 6 concludes the paper with some open questions and potential directions for future research. All the proofs are deferred to Appendix A.

## 2 Notions of negative dependence

In this section we recall a few classic notions of negative dependence. Throughout, denote by  $[n] = \{1, \dots, n\}$  and  $\mathbf{1}_n$  the  $n$ -vector with all components being 1; the vector  $\mathbf{0}_n$  is defined analogously. All inequalities and equalities between (random) vectors are component-wise. For an  $n$ -dimensional random vector  $\mathbf{X} = (X_1, \dots, X_n)$ , denote by  $\mathbf{X}^\perp = (X_1^\perp, \dots, X_n^\perp)$  a random vector with independent components such that  $X_i \stackrel{d}{=} X_i^\perp$ ,  $i \in [n]$ , where  $\stackrel{d}{=}$  stands for equality in distribution. For a set  $A \subseteq [n]$ , we denote by  $\mathbf{X}_A = (X_k)_{k \in A}$ . A function  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  is called *supermodular* if  $\psi(\mathbf{x} \wedge \mathbf{y}) + \psi(\mathbf{x} \vee \mathbf{y}) \geq \psi(\mathbf{x}) + \psi(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , where  $\mathbf{x} \wedge \mathbf{y}$  and  $\mathbf{x} \vee \mathbf{y}$  are the component-wise minimum and maximum of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively.

**Definition 1.** Let  $\mathbf{X} = (X_1, \dots, X_n)$  be an  $n$ -dimensional random vector.

- (i)  $\mathbf{X}$  is *negative correlation dependent (NCD)* if  $\text{Cov}(X_i, X_j) \leq 0$  for all  $i, j \in [n]$  with  $i \neq j$ .
- (ii)  $\mathbf{X}$  is *negative upper orthant dependent (NUOD)* if  $\mathbb{P}(\mathbf{X} > \mathbf{t}) \leq \mathbb{P}(\mathbf{X}^\perp > \mathbf{t})$  for all  $\mathbf{t} \in \mathbb{R}^n$ ;  
 $\mathbf{X}$  is *negative lower orthant dependent (NLOD)* if  $\mathbb{P}(\mathbf{X} \leq \mathbf{t}) \leq \mathbb{P}(\mathbf{X}^\perp \leq \mathbf{t})$  for all  $\mathbf{t} \in \mathbb{R}^n$ .  
 If  $\mathbf{X}$  is both NLOD and NUOD, then it is *negative orthant dependent (NOD)*.
- (iii)  $\mathbf{X}$  is *negative supermodular dependent (NSD)* if  $\mathbb{E}[\psi(\mathbf{X})] \leq \mathbb{E}[\psi(\mathbf{X}^\perp)]$  for all supermodular functions  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  such that the expectations exist.
- (iv)  $\mathbf{X}$  is *negatively associated (NA)* if

$$\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leq 0, \quad (1)$$

for any disjoint subsets  $A, B \subseteq [n]$  and any real-valued, coordinate-wise increasing functions  $f$  and  $g$  such that  $f(\mathbf{X}_A)$  and  $g(\mathbf{X}_B)$  have finite second moments.

- (v)  $\mathbf{X}$  is *counter-monotonic (CT)* if each pair of its component  $(X_i, X_j)$  for  $i \neq j$  satisfies  $(X_i, X_j) = (f(Z), -g(Z))$  almost surely (a.s.) for some random variable  $Z$  and increasing functions  $f, g$ .
- (vi)  $\mathbf{X}$  is a *joint mix* (abbreviated as “ $\mathbf{X}$  is JM”) if  $\sum_{i=1}^n X_i = c$  a.s. for some constant  $c \in \mathbb{R}$ .

All abbreviations introduced in this section are also used as nouns to represent the corresponding dependence concept. The next definition concerns properties of the marginal distributions that allow for JM random vectors.

**Definition 2.** An  $n$ -tuple  $(F_1, \dots, F_n)$  of distributions on  $\mathbb{R}$  is called *jointly mixable* if there exists a joint mix  $\mathbf{X} = (X_1, \dots, X_n)$  such that  $X_i \sim F_i$ ,  $i \in [n]$ . The constant  $c = \sum_{i=1}^n X_i$  is called a *center* of  $\mathbf{X}$ . In this case, we also say that  $(F_1, \dots, F_n)$  *supports a joint mix  $\mathbf{X}$* . A distribution  $F$  is called  *$n$ -completely mixable* if the  $n$ -tuple  $(F, \dots, F)$  is jointly mixable.

The following implications hold between the above concepts of negative dependence. These implications is either checked directly by definition or shown in the literature, e.g., [Joag-Dev and Proschan \(1983\)](#), [Christofides and Vaggelatou \(2004\)](#) and [Lauzier et al. \(2023\)](#). The case of  $n \geq 3$  is different from the case  $n = 2$ .

$$n = 2 : \quad \text{JM} \implies \text{CT} \implies \text{NA} \iff \text{NSD} \iff \text{NOD} \iff \text{NUOD} \iff \text{NLOD} \implies \text{NCD}; \quad (2)$$

$$\text{general } n : \quad \text{CT} \implies \text{NA} \implies \text{NSD} \implies \text{NOD} \implies \text{NUOD or NLOD} \implies \text{NCD}. \quad (3)$$

All one-direction implications in (2) and (3) are strict for  $n \geq 3$ ; see [Amini et al. \(2013\)](#) for some examples. In contrast to the case  $n = 2$  in (2), JM no longer implies any of the properties in (3). This can be observed by the following properties of Gaussian random vectors.

**Proposition 1.** *Let  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$  be a Gaussian random vector with mean vector  $\boldsymbol{\mu} \in \mathbb{R}^n$  and covariance matrix  $\Sigma = (\sigma_{ij})_{n \times n}$ .*

(a) *The followings are equivalent: (i)  $\mathbf{X}$  is NA; (ii)  $\mathbf{X}$  is NSD; (iii)  $\mathbf{X}$  is NUOD; (iv)  $\mathbf{X}$  is NLOD; (v)  $\mathbf{X}$  is NCD.*

(b) *The followings are equivalent: (i)  $\mathbf{X}$  is JM; (ii)  $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$ .*

(c) *For  $n = 2$ , the followings are equivalent: (i)  $\mathbf{X}$  is JM; (ii)  $\mathbf{X}$  is CT and  $\sigma_{11} = -\sigma_{12}$ .*

(d) *For  $n \geq 3$ ,  $\mathbf{X}$  is never CT unless at least  $n - 2$  components of  $\mathbf{X}$  are degenerate.*

Proposition 1 shows the convenient property of the Gaussian distribution that the concepts of NA, NSD, NOD, NLOD, NUOD and NCD are all equivalent for this class. Parts (a) and (b) immediately tell that, for  $n \geq 3$ , JM does not imply any of these concepts, and none of these concepts implies JM. We will focus mostly on NA, NOD and NCD given their popularity and relative strength in the chains (2) and (3). For some other notions of negative dependence, see [Joe \(2014\)](#).

### 3 JM and negative dependence

In this section, we explore the relation between JM and negative dependence concepts introduced in Section 2 by means of several theoretical results. We first show that a joint mix is NA under some properties of conditional independence and monotonicity.

**Theorem 1.** *Let  $\mathbf{X}$  be a joint mix and write  $S_A = \sum_{i \in A} X_i$  for  $A \subseteq [n]$ . Suppose that*

(a)  *$\mathbf{X}_A$  and  $\mathbf{X}_{[n] \setminus A}$  are independent conditionally on  $S_A$  for every  $A \subseteq [n]$ ;*

(b)  *$\mathbb{E}[f(\mathbf{X}_A) | S_A]$  is increasing in  $S_A$  for every increasing function  $f$  and  $A \subseteq [n]$ .*

*Then  $\mathbf{X}$  is NA.*

Theorem 1 can be compared with Theorem 2.6 of [Joag-Dev and Proschan \(1983\)](#), which says that if  $\mathbf{X}$  is independent and satisfies (b), then the conditional distribution of  $\mathbf{X}$  given  $S_{[n]}$  is NA. Since  $S_{[n]}$  is a constant for JM and (a) is implied by independence, Theorem 1 means that the NA condition in Theorem 2.6 of [Joag-Dev and Proschan \(1983\)](#) holds if the independence

assumption is weakened to conditional independence (a), and in addition we assume JM. Note, however, that JM and independence of  $\mathbf{X}$  conflict each other unless  $\mathbf{X}$  is degenerate.

Most existing examples of NA random vectors are presented by [Joag-Dev and Proschan \(1983\)](#). Although [Theorem 1](#) does not directly give new examples of NA random vectors, it can be used to check NA in popular examples.

**Example 1.** We use [Theorem 1](#) to check that the uniform distribution on the standard simplex  $\Delta_n = \{(x_1, \dots, x_n) \in [0, 1]^n : \sum_{i=1}^n x_i = 1\}$  is NA. Let  $\mathbf{X}$  follow the uniform distribution over  $\Delta_n$  which is JM. For every  $A \subseteq [n]$ , we can check that  $(\mathbf{X}_A, \mathbf{X}_{[n] \setminus A}) | \{S_A = s\}$  for  $s \in (0, 1)$  follows a uniform distribution on  $(s\Delta_n) \times ((1-s)\Delta_n)$  and condition (a) holds. Condition (b) follows by noting that  $\mathbf{X}_A | \{S_A = s\} \stackrel{d}{=} s\mathbf{X}_A | \{S_A = 1\}$  for  $s \in (0, 1)$  and thus  $\mathbf{X}_A | \{S_A = s\}$  is stochastically increasing in  $s$ .

**Example 2.** The multinomial distribution is known to be NA ([Joag-Dev and Proschan, 1983](#)). We show this by virtue of [Theorem 1](#). Let  $\mathbf{X} \sim \text{MN}_n(k, \mathbf{p})$  follow a multinomial distribution with  $k$  trials,  $n$  mutually exclusive events and event probabilities  $\mathbf{p} = (p_1, \dots, p_n)$ . For  $s \in \{0, \dots, k\}$  and every  $A \subseteq [n]$  with  $B = [n] \setminus A$ , it holds that  $\mathbf{X}_A | \{S_A = s\} \sim \text{MN}_{|A|}(s, \mathbf{p}_A / \sum_{i \in A} p_i)$  and  $\mathbf{X}_B | \{S_A = s\} \sim \text{MN}_{|B|}(k-s, \mathbf{p}_B / \sum_{i \in B} p_i)$ , where  $p_A = (p_i)_{i \in A}$  and  $p_B = (p_i)_{i \in B}$ . Then conditions (a) and (b) can be checked directly by calculation.

We next focus on exchangeable joint mixes, which exhibit some specific forms of negative dependence. A random vector  $\mathbf{X} = (X_1, \dots, X_n)$  (or its distribution) is called *exchangeable* if  $\mathbf{X} \stackrel{d}{=} \mathbf{X}^\pi$  for all  $\pi \in \mathfrak{S}_n$ , where  $\mathfrak{S}_n$  is the set of all permutations on  $[n]$  and  $\mathbf{X}^\pi = (X_{\pi(1)}, \dots, X_{\pi(n)})$ . First, we note that if  $\mathbf{X}$  is CT with identical marginal distributions equal to  $F$ , then the distribution of  $\mathbf{X}$  is explicitly given by  $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) = (F(x_1) + \dots + F(x_n) - n + 1)_+$  for  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ ; see, for example, [Theorem 3.3 of Puccetti and Wang \(2015\)](#). Clearly, this distribution is exchangeable. Moreover, for any given marginal distribution  $F$  which is  $n$ -completely mixable, there exists an exchangeable joint mix with marginals  $F$ ; see [Proposition 2.1 of Puccetti et al. \(2019\)](#). Note that an exchangeable joint mix is NCD because each bivariate correlation coefficient is equal to  $-1/(n-1)$ . The next proposition states that such an exchangeable joint mix is also negatively dependent in the sense of NSD, NUOD and NULD if so is  $\mathbf{X}$ .

**Proposition 2.** *If a univariate distribution function  $F$  supports an NSD  $n$ -joint mix, then  $F$  supports an exchangeable NSD  $n$ -joint mix. The statement holds true if NSD is replaced by NOD, NUOD or NLOD.*

One may wonder whether Proposition 2 holds with NSD replaced by NA. Unfortunately, this question remains open, as our proof for Proposition 2 does not extend to NA.

Next, we present a necessary condition for a tuple of distributions to support any negatively dependent joint mixes. This condition is also sufficient when the marginal distributions are Gaussian.

**Proposition 3.** *If the tuple of distributions  $(F_1, \dots, F_n)$  with finite variance vector  $(\sigma_1^2, \dots, \sigma_n^2)$  supports an NCD joint mix, then*

$$2 \max_{i \in [n]} \sigma_i^2 \leq \sum_{i \in [n]} \sigma_i^2. \quad (4)$$

Since NCD is weaker than NOD and NA, the necessary condition (4) is also necessary for NOD and NA joint mixes.

For a given tuple of distributions  $(F_1, \dots, F_n)$  with finite variance vector  $(\sigma_1^2, \dots, \sigma_n^2)$ , the condition (4) is not necessary for the existence of an NCD random vector, since an independent random vector supported by  $(F_1, \dots, F_n)$  always exists and it is NCD. More interestingly, the condition (4) is not necessary for the existence of a joint mix either. Indeed, as shown by Wang and Wang (2016, Corollary 2.2), a useful necessary condition for a joint mix supported by  $(F_1, \dots, F_n)$  to exist is

$$2 \max_{i \in [n]} \sigma_i \leq \sum_{i \in [n]} \sigma_i. \quad (5)$$

We note that (4) is strictly stronger than (5), because for any  $j \in [n]$ , (4) gives

$$\sigma_j^2 \leq \sum_{i \in [n] \setminus \{j\}} \sigma_i^2 \implies \sigma_j \leq \left( \sum_{i \in [n] \setminus \{j\}} \sigma_i^2 \right)^{1/2} \leq \sum_{i \in [n] \setminus \{j\}} \sigma_i,$$

which implies (5). It is clear that (4) and (5) are not equivalent; for example,  $(\sigma_1, \sigma_2, \sigma_3) = (2, 2, 3)$  satisfies (5) but not (4).

By Proposition 2.4 of Wang et al. (2013), if the marginal distributions  $F_1, \dots, F_n$  are Gaussian, then the condition (5) is necessary and sufficient for a joint mix supported by  $(F_1, \dots, F_n)$  to exist. Hence, the condition (4), which is strictly stronger than (5), is not necessary for a joint mix to exist. On the other hand, (4) is generally not sufficient for an NCD joint mix to exist either, since it is well known that a joint mix may not exist even if the marginal distributions are identical. Nevertheless, it turns out that (4) is necessary and sufficient for an NCD or NA joint mix to exist for Gaussian marginals.

**Theorem 2.** *A tuple of univariate Gaussian distributions with variance vector  $(\sigma_1^2, \dots, \sigma_n^2)$*

supports an NCD or NA joint mix if and only if (4) holds, that is,  $2 \max_{i \in [n]} \sigma_i^2 \leq \sum_{i \in [n]} \sigma_i^2$ . Moreover, such an NCD or NA joint mix can be chosen as a Gaussian random vector.

The negative dependence concepts of NA, NSD, NOD, NLOD, NUOD and NCD are all equivalent for multivariate Gaussian distributions, as we see in Proposition 1. Hence, (4) is also necessary and sufficient for an NSD, NOD, NLOD, or NUOD joint mix to exist for Gaussian marginals. For elliptical distributions (details in Section 5), the corresponding statement to Theorem 2 holds for NCD but not the other forms of negative dependence; see Proposition 7.

**Example 3.** In case  $n = 3$ , for any marginals with variance vector  $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ , the covariance of a joint mix  $\mathbf{X}$  is uniquely given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) & \sigma_3^2 \end{pmatrix};$$

see Xiao and Yao (2020, Corollary 6) for this statement. If  $\mathbf{X}$  is Gaussian, it is clear that  $\mathbf{X}$  is NA if and only if (4) holds. In case  $n \geq 4$ , for Gaussian marginals we can obtain an explicit covariance matrix of an NA joint mix from the proof of Theorem 2.

We end this section with two decomposition results of a joint mix into NA joint mixes, one through a random vector decomposition, and one through a mixture decomposition. We first establish a new result showing that any finitely supported discrete joint mix can be decomposed into a linear combination of binary multinomial random vectors. A binary multinomial random vector is a random vector  $(X_1, \dots, X_n)$  taking values in  $\{0, 1\}^n$  such that  $\sum_{i=1}^n X_i = 1$ ; that is, exactly one of  $X_1, \dots, X_n$  takes the value 1. By definition, binary multinomial random vectors are CT (hence NA) and JM.

**Theorem 3.** *Suppose that the random vector  $\mathbf{X}$  takes values in a finite set. Then,  $\mathbf{X}$  is JM if and only if it can be represented as a finite linear combination of binary multinomial random vectors.*

Theorem 3 generalizes Theorem 2 of Wang (2015) which has a decomposition of a joint mix taking nonnegative integer values as the sum of binary multinomial random vectors. The assumption of finite support in Theorem 3 does not seem to be dispensable with the current proof techniques.

Next, using the fact that the distribution of a joint mix can be written as a mixture of discrete uniform (DU) distributions on  $n$  points in  $\mathbb{R}^n$ , we obtain the following decomposition.



**Proposition 4.** *The distribution of any exchangeable joint mix with center  $\mu$  can be written as a mixture of distributions of exchangeable NA joint mixes with center  $\mu$ .*

## 4 A multi-marginal optimal transport problem

To connect JM and negative dependence, a natural question is whether in some applications negatively dependent joint mixes have additional attractive properties that are not shared by other joint mixes. We present an optimal transport problem with uncertainty in this section where a combination of negative dependence and JM naturally appears.

In the multi-marginal optimal transport theory ([Santambrogio, 2015](#); [Pass, 2015](#)), a general objective is

$$\text{to minimize } \mathbb{E}[c(X_1, \dots, X_n)] \quad \text{subject to } X_i \sim F_i, i \in [n],$$

where  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  is a cost function and  $F_1, \dots, F_n$  are specified marginal distributions. This problem is referred to as the *Monge-Kantorovich problem*. In the context of this paper, the distributions  $F_1, \dots, F_n$  are on  $\mathbb{R}$ . In all optimization problems we discussed in this section, the constraint is always  $X_i \sim F_i$  for each  $i \in [n]$  with  $F_1, \dots, F_n$  given, and we assume that  $F_1, \dots, F_n$  have finite second moments throughout this section.

### 4.1 Optimal transport under uncertainty on the set of components

We will consider a special class of cost functions, leading to the Monge-Kantorovich problem

$$\text{to minimize } \mathbb{E} \left[ f \left( \sum_{i=1}^n X_i \right) \right] \quad \text{subject to } X_i \sim F_i, i \in [n], \quad (6)$$

where  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function. This special setting is important to JM because, assuming that a joint mix with marginal distributions  $F_1, \dots, F_n$  exists, then any joint mix is an optimizer of (6) due to Jensen's inequality, and conversely, any optimizer of (6) has to be a joint mix if  $f$  is strictly convex. As discussed by [Puccetti and Wang \(2015\)](#) and [Wang and Wang \(2016\)](#), one of the main motivations of JM is to solve optimization problems similar to (6).

Since joint mixes with given marginal distributions are not unique, we wonder whether a negatively dependent joint mix plays a special role among optimizers to (6). This is our main question to address.

Although each joint mix minimizes (6), their distributions can be quite different. For a concrete example, suppose that the marginal distributions are standard Gaussian, and let  $n$  be even. With these marginals,  $\mathbf{X}^E \sim N_n(\mathbf{0}_n, P_n^*)$  is an NA joint mix, where  $P_n^*$  is a matrix with diagonal entries being 1 and off-diagonal entries being  $-1/(n-1)$ , and  $\mathbf{X}^A = ((-1)^i Z)_{i \in [n]}$ ,

$Z \sim N_1(0, 1)$ , is another joint mix which is not NA. Here, “E” stands for “exchangeable” and “A” stands for “alternating”. These two joint mixes have the same value  $f(0)$  for (6). Nevertheless,  $\mathbf{X}^A$  may be seen as undesirable in some situations, because some subgroups of its components are comonotonic. Inspired by this, we consider the cost of a subset  $K \subseteq [n]$  of risks  $f\left(\sum_{i \in K} X_i\right)$ . If  $K$  is known to the decision maker, then we are back to (6) with  $(X_i)_{i \in [n]}$  replaced by  $(X_i)_{i \in K}$ .

In different applications, allowing a flexible choice of  $K$  may represent the absence of some risks in a risk aggregation pool, missing particles in a quantum system, an unspecified number of simulation sizes in a sampling program, or uncertainty on the participation of some agents in a risk-sharing game. In each context above, a decision maker may not know  $K$  a priori, and hence she may be interested in minimizing a weighted average of the cost, that is

$$C_\mu^f(X_1, \dots, X_n) := \sum_{K \subseteq [n]} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i \right) \right] \mu(K), \quad (7)$$

where  $\mu$  is a probability on the sample space  $2^{[n]}$ , the power set of  $[n]$ , and  $\sum_{i \in K} X_i$  is set to 0 if  $K$  is empty; here we use the notation  $\mu(K) = \mu(\{K\})$ , which should not lead to any confusion.

We consider the formulation of uncertainty as in the framework of [Gilboa and Schmeidler \(1989\)](#). With a probability on  $2^{[n]}$  uncertain, we consider a set  $\mathcal{M}$  of probabilities on  $2^{[n]}$ , called an uncertainty set. The formulation of (7) with uncertainty set  $\mathcal{M}$  is

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} C_\mu^f(X_1, \dots, X_n) \quad \text{subject to } X_i \sim F_i, i \in [n]. \quad (8)$$

The supremum represents a worst-case attitude towards uncertainty, which is axiomatized by [Gilboa and Schmeidler \(1989\)](#) in decision theory. We explain two simple special cases of (8).

First, by taking  $\mathcal{M}$  as the set of all probabilities on  $2^{[n]}$ , the objective in (8) becomes

$$\sup_{\mu \in \mathcal{M}} C_\mu^f(X_1, \dots, X_n) = \max_{K \subseteq [n]} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i \right) \right], \quad (9)$$

which represents the situation of having no information on  $K$ . Second, by taking  $\mathcal{M}$  as the set of all probabilities on  $2^{[n]}$  supported by sets  $K$  of cardinality  $|K| = k \in [n]$ , the objective in (8) becomes

$$\sup_{\mu \in \mathcal{M}} C_\mu^f(X_1, \dots, X_n) = \max_{K \subseteq [n], |K|=k} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i \right) \right], \quad (10)$$

which represents the situation where one knows how large the subset  $K$  is, but not precisely

how it is composed.

The problem (8) is generally difficult to solve. We will first focus on the homogeneous case where  $F = F_1 = \dots = F_n$ , and this will be relaxed in Section 4.3. With this interpretation, it is natural to consider uncertainty sets  $\mathcal{M}$  that are symmetric. We say that  $\mathcal{M}$  is *symmetric* if  $\mu \in \mathcal{M}$  implies  $\mu_\pi \in \mathcal{M}$  for  $\pi \in \mathfrak{S}_n$ , where  $\mu_\pi$  is a permutation of  $\mu$ , defined by  $\mu_\pi(K) = \mu(\{\pi(i) : i \in K\})$  for  $K \subseteq [n]$ .

Recall that we are interested in whether a negatively dependent joint mix plays a special role among other joint mixes. The next proposition provides a step in this direction. Also recall that  $P_n^*$  is the correlation matrix with off-diagonal entries equal to  $-1/(n-1)$ .

**Proposition 5.** *Suppose that  $\mathcal{M}$  is symmetric, and  $\mathbf{X}$  is a joint mix with identical marginals  $F$ . Then there exists an exchangeable NCD joint mix  $\mathbf{X}^E$  with marginals  $F$  and correlation matrix  $P_n^*$  such that*

$$\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^E) \leq \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X})$$

for all measurable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

Proposition 5 illustrates the intuition that among all joint mixes, the NCD ones with correlation matrix  $P_n^*$  are better choices under uncertainty. However, this does not answer whether such NCD joint mixes are optimizers to our main optimal transport problem (8). In the next section, we consider the quadratic cost, and show that indeed those NCD joint mixes are solutions to (8) for the quadratic cost.

## 4.2 Quadratic cost

We consider the quadratic cost given by  $f(x) = x^2$ . In this case, we denote by

$$C_\mu^2(X_1, \dots, X_n) := \sum_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \mu(K),$$

and (8) becomes, assuming homogeneous marginals,

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \mu(K), \quad \text{subject to } X_i \sim F, i \in [n]. \quad (11)$$

Two other formulations related to the quadratic cost, the repulsive harmonic cost problem and the variance minimization problem, are discussed below in Examples 4 and 5.

It is clear that, for  $\mu \in \mathcal{M}$ , its permutation  $\mu_\pi$  satisfies  $C_\mu^2(\mathbf{X}) = C_{\mu_\pi}^2(\mathbf{X})$ . We first show that the exchangeable NCD joint mix is a minimizer to (11) if  $\mathcal{M}$  is symmetric.

**Theorem 4.** *Suppose that  $F$  is  $n$ -completely mixable with finite variance and  $\mathcal{M}$  is symmetric. Then, each NCD joint mix with marginals  $F$  and correlation matrix  $P_n^*$  minimizes (11).*

If the marginal distribution  $F$  is Gaussian, then we can replace NCD by NA, NSD or NOD in Theorem 4, since for the Gaussian class NCD is equivalent to these notions (Proposition 1).

Theorem 4 does not state that all the minimizers to (11) are NCD joint mixes. For instance, if  $\mathcal{M}$  contains only measures concentrated on  $K$  with  $|K| = n$ , then any joint mix minimizes (11); see also Remark 1 below for other similar cases. Next, we study the uniqueness of the optimizers for two special choices of  $\mathcal{M}$  in (9) and (10), namely,

$$\text{to minimize } \max_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \quad \text{subject to } X_i \sim F, i \in [n], \quad (12)$$

and for a fixed  $k \in [n]$ ,

$$\text{to minimize } \max_{K \subseteq [n], |K|=k} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \quad \text{subject to } X_i \sim F, i \in [n]. \quad (13)$$

In Theorem 5 below, we will see that, assuming  $F$  has mean zero, negative dependence yields more stable costs in the presence of uncertainty. The exchangeable joint mix with correlation matrix  $P_n^*$  minimizes (13) for each  $k \in [n]$ , and this correlation matrix is unique for all minimizers for each  $k \in [n] \setminus \{1, n-1, n\}$ . As a consequence, all minimizers to (12) have the same correlation matrix  $P_n^*$  (this holds for  $n \geq 3$ ).

**Theorem 5.** *Suppose that  $n \geq 3$  and the distribution  $F$  is  $n$ -completely mixable with mean 0 and finite positive variance. A random vector is a minimizer to (12) if and only if it is an NCD joint mix with correlation matrix  $P_n^*$ . The same conclusion holds true if (12) is replaced by (13) with any  $k \in [n] \setminus \{1, n-1, n\}$ .*

**Remark 1.** We briefly comment on the three cases of  $k$  excluded from the statement regarding the unique minimizer of (13), and it will be clear that uniqueness cannot be expected in these cases. Recall that the marginal distributions of  $\mathbf{X}$  are assumed identical.

1. If  $k = 1$ , then  $\mathbb{E}[(\sum_{i \in K} X_i)^2] = \mathbb{E}[X_1^2]$  which does not depend on the dependence structure of  $\mathbf{X}$ , and hence any coupling minimizes (13).
2. If  $k = n$ , then  $K = [n]$  and thus any joint mix minimizes (13).

3. If  $k = n - 1$ , then  $\mathbb{E}[(\sum_{i \in K} X_i)^2] = \mathbb{E}[(c - X_1)^2]$  for any joint mix  $\mathbf{X}$  with center  $c$ . Hence, any joint mix has the same value for (13).

Theorem 5 implies that for a standard Gaussian  $F$ , the exchangeable joint mix  $\mathbf{X}^E \sim N_n(\mathbf{0}_n, P_n^*)$  is a minimizer to both (12) and (13) for each  $k \in [n]$ . If  $n \geq 3$ , this minimizer is unique among Gaussian vectors in both cases of (12) and (13) with  $k \in [n] \setminus \{1, n - 1, n\}$ .

**Remark 2.** As we have seen in Remark 1, if  $n = 3$ , then any joint mix minimizes (13) for each  $k \in [n]$ . The uniqueness statement in Theorem 5 implies that the covariance structure of a joint mix is unique for  $n = 3$ , as we see in Example 3.

Below we discuss two specific optimal transport problems related to the quadratic cost.

**Example 4** (Variance minimization). The quadratic cost minimization problem is equivalent to variance minimization with given marginals. It is clear that

$$C_\mu^2(X_1, \dots, X_n) = \sum_{K \subseteq [n]} \text{Var} \left( \sum_{i \in K} X_i \right) \mu(K) + \sum_{K \subseteq [n]} \left( \sum_{i \in K} \mathbb{E}[X_i] \right)^2 \mu(K),$$

and the second term does not depend on the dependence structure of  $(X_1, \dots, X_n)$ . If  $F_1, \dots, F_n$  have zero mean, then the problem (11) can be written as

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \text{Var} \left( \sum_{i \in K} X_i \right) \mu(K) \quad \text{subject to } X_i \sim F, i \in [n].$$

Variance minimization is a classic problem in Monte Carlo simulation (Craiu and Meng, 2001, 2005) and risk management (Rüschendorf, 2013). The above arguments show that the statements in Theorems 4 and 5 hold true if the objective of quadratic cost  $\mathbb{E}[(\sum_{i \in K} X_i)^2]$  is replaced by the variance  $\text{Var}(\sum_{i \in K} X_i)$ .

**Example 5** (Repulsive harmonic cost). The *repulsive harmonic cost* function is defined by

$$c(x_1, \dots, x_n) = - \sum_{i,j=1}^n (x_i - x_j)^2, \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

This cost function originates from the so-called weak interaction regime in Quantum Mechanics; see e.g., Di Marino et al. (2017). Any joint mix minimizes the expected repulsive harmonic cost.

To see this, we can rewrite

$$\mathbb{E}[c(X_1, \dots, X_n)] = -2n \sum_{i=1}^n \mathbb{E}[X_i^2] + 2\mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^2 \right]. \quad (14)$$

Since the first terms on the right-hand side of (14) do not depend on the dependence structure of  $(X_1, \dots, X_n)$ , minimizing  $\mathbb{E}[c(X_1, \dots, X_n)]$  is equivalent to minimizing  $\mathbb{E}[(\sum_{i=1}^n X_i)^2]$ , which is clearly minimized if  $(X_1, \dots, X_n)$  is a joint mix. Let  $c_K(x_1, \dots, x_n) = -\sum_{i,j \in K} (x_i - x_j)^2$ ,  $(x_1, \dots, x_n) \in \mathbb{R}^n$ , for  $K \subseteq [n]$ . The problem (11) can be written as

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \left( \frac{1}{2} \mathbb{E}[c_K(\mathbf{X})] + |K| \sum_{i \in K} \mathbb{E}[X_i^2] \right) \mu(K) \quad \text{subject to } X_i \sim F, i \in [n].$$

The statement in Theorem 4 remains true if the objective of quadratic cost  $\mathbb{E}[(\sum_{i \in K} X_i)^2]$  is replaced by the cost  $\mathbb{E}[c_K(\mathbf{X})]$ .

### 4.3 Discussions on heterogeneous marginals

In Theorem 5, we assumed that the marginal distributions are identical. This assumption is not dispensable, as the situation for heterogeneous marginals is drastically different and we do not have general results. In this section, we present a result in the simple case  $n = 3$  and provide several examples to discuss some subtle issues and open questions. To illustrate these issues, we focus on the problems (12) and (13) for  $n = 3$ . In all examples, we explain with Gaussian marginal distributions, but this assumption can be replaced as long as the covariance matrices in the examples are compatible with the marginals.

**Proposition 6.** *Let  $n = 3$ . For any tuple of marginal distributions with finite variance vector and zero means, any joint mix, if it exists, minimizes (12). If an NCD joint mix exists, then no random vector with any positive bivariate covariance can minimize (12).*

**Remark 3.** Proposition 6 states that, if a Gaussian triplet supports an NCD joint mix, then it minimizes (12), and all Gaussian minimizers must be NCD. It is not clear whether this observation can be extended to  $n \geq 4$ .

Unlike the situation in Theorem 5, uniqueness of the covariance matrix does not hold in the setting of Proposition 6, as illustrated in the following example.

**Example 6.** Consider Gaussian marginal distributions with variance vector  $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (2, 1, 1)$ . In this case, (4) holds, and an NCD joint mix exists by Theorem 2. Both the co-

variance matrices  $\Sigma$  and  $\Sigma'$  defined by

$$\Sigma = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma' = \begin{pmatrix} 2 & -1/2 & -1 \\ -1/2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

minimize (12) subject to the marginal distributions. We can see that  $\Sigma$  corresponds to an NCD joint mix, whereas  $\Sigma'$  corresponds to an NCD random vector, but not a joint mix.

The next example illustrates that, although a joint mix generally minimizes (12) in case  $n = 3$ , NCD may be more relevant than joint mixes for minimizing (13) with some  $k \neq n$  when the two dependence requirements cannot be simultaneously achieved.

**Example 7.** Consider Gaussian marginal distributions with variance vector  $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (4, 1, 1)$ . In this case, (4) does not hold, and no NCD joint mix exists. Both the covariance matrices  $\Sigma$  and  $\Sigma'$  defined by

$$\Sigma = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 1 & 1 \\ -2 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma' = \begin{pmatrix} 4 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

minimize (12) subject to the marginal distributions. The covariance matrix  $\Sigma$  corresponds to a joint mix, but not NCD. The covariance matrix  $\Sigma'$  corresponds to an NCD random vector, but not a joint mix. Thus, the problem (12) admits an NCD minimizing distribution  $N_3(\mathbf{0}_3, \Sigma')$ . Moreover, for (13) with  $k = 2$ , the NCD distribution  $N_3(\mathbf{0}_3, \Sigma')$  has a maximum of 3 which is strictly better than the joint mix distribution  $N_3(\mathbf{0}_3, \Sigma)$  with a maximum of 4.

Example 7 suggests, informally, that there is a trade-off between a joint mix and NCD when both cannot be attained simultaneously, with a joint mix minimizing (13) for  $k = n$ , and an NCD random vector improving (13) from the case of a joint mix for some  $1 < k < n$ . In fact, (13) is not always minimized by NCD random vectors as seen in the following example.

**Example 8.** Consider marginal distributions with variance vector  $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (\sigma^2, 1, 1)$  and zero means, where  $\sigma > 3$ . For any  $(X_1, X_2, X_3)$  with the given marginals, we have

$$\mathbb{E} [(X_2 + X_3)^2] \leq 4 < (\sigma - 1)^2 \leq \min(\mathbb{E} [(X_1 + X_2)^2], \mathbb{E} [(X_1 + X_3)^2]),$$

and hence

$$\max_{K \subseteq [3], |K|=2} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] = \sigma^2 + 1 + 2\sigma \max(\rho_{12}, \rho_{13}), \quad (15)$$

where  $\rho_{ij}$ ,  $i, j \in [3]$  is the correlation coefficient of  $(X_i, X_j)$ . For Gaussian marginals, the minimum of (15) is attained if and only if  $\rho_{12} = \rho_{13} = -1$ . In this case,  $\rho_{23} = 1$  is the only possible correlation, and thus the minimizer to (15) cannot be NCD.

On the other hand, the next example shows that, if  $n = 3$ , there always exists an NCD minimizer to (12) for Gaussian marginals.

**Example 9.** Let  $(X_1, X_2, X_3)$  follow a multivariate Gaussian distribution with equicorrelation matrix  $P_3^*$ ; i.e., all pairwise correlation coefficients are  $-1/2$ . The variances  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_3^2$  are assumed to satisfy  $\sigma_1 \leq \sigma_2 \leq \sigma_3$  without the loss of generality and the means for prescribed marginal distributions are assumed to be zero. We can easily verify that each of  $\mathbb{E}[(X_1 + X_2 + X_3)^2]$  and  $\mathbb{E}[(X_i + X_j)^2]$ ,  $i, j \in [3]$ , is smaller than or equal to  $\sigma_3^2$ . Hence,  $(X_1, X_2, X_3)$  attains the lower bound

$$\max_{K \subseteq [3]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] = \sigma_3^2 = \max_{i \in [3]} \sigma_i^2,$$

and thus it minimizes (12).

**Remark 4.** For  $n \geq 4$ , it is not clear whether there always exists an NCD minimizer to (12) under a general heterogeneous marginal constraint.

## 5 Elliptical distributions

Elliptical distributions form a tractable class of joint mixes for arbitrary dimensions. In this section, we investigate negative dependence properties of such elliptical joint mixes.

An  $n$ -dimensional *elliptical distribution* is a family of multivariate distributions defined through the characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E} \left[ \exp \left( \mathbf{it}^\top \mathbf{X} \right) \right] = \exp \left( \mathbf{it}^\top \boldsymbol{\mu} \right) \psi(\mathbf{t}^\top \Sigma \mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^n, \quad (16)$$

for some location parameter  $\boldsymbol{\mu} \in \mathbb{R}^n$ ,  $n \times n$  positive semi-definite symmetric matrix  $\Sigma \in \mathbb{R}^{n \times n}$  and the so-called *characteristic generator*  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ , where  $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ . See Section 6 of McNeil et al. (2015) for more properties. We denote an elliptical distribution by  $E_n(\boldsymbol{\mu}, \Sigma, \psi)$  and refer to  $\boldsymbol{\mu}$  as the *location vector* and  $\Sigma$  the *dispersion matrix*. We say that



an elliptical distribution  $E_n(\boldsymbol{\mu}, \Sigma, \psi)$  is non-degenerate if all its marginals are non-degenerate (i.e., not a point-mass). Equivalently, the diagonal entries of  $\Sigma$  are positive. As presented in Proposition 6.27 of McNeil et al. (2015), a random vector  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$  with  $\text{rank}(\Sigma) = k$  admits the stochastic representation  $\mathbf{X} = \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$ , where  $\mathbf{S}$  is the uniform distribution on the unit sphere on  $\mathbb{R}^k$ , the radial random variable  $R \geq 0$  is independent of  $\mathbf{S}$ , and  $A \in \mathbb{R}^{n \times k}$  is such that  $AA^\top = \Sigma$ . With this representation, we have that  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  and  $\text{Cov}(\mathbf{X}) = \mathbb{E}[R^2] \Sigma/k$  provided  $\mathbb{E}[R^2] < \infty$ .

We first present a simple lemma on elliptical joint mixes which will be useful for later discussions.

**Lemma 1.** *An  $n$ -dimensional elliptically distributed random vector  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$  is a joint mix if and only if  $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$  or  $\psi = 1$  on  $\mathbb{R}_+$ .*

Negative dependence of such an elliptical joint mix is the main topic of this section. We next provide a characterization for NCD-JM as an extension to Theorem 2.

**Proposition 7.** *Suppose that  $\psi$  is the characteristic generator of an  $n$ -dimensional elliptical distribution. A tuple of univariate distributions  $(E_1(\mu_i, \sigma_i^2, \psi), i \in [n])$  supports an NCD joint mix if and only if (4) holds, that is,  $2 \max_{i \in [n]} \sigma_i^2 \leq \sum_{i \in [n]} \sigma_i^2$ . Moreover, such an NCD joint mix can be chosen to follow an elliptical distribution.*

As we showed in Proposition 1, for Gaussian random vectors, NA, NSD and NOD are all equivalent to NCD; that is, the bivariate correlations are non-positive. Recall that  $P_n^*$  is an  $n \times n$  matrix whose diagonal entries are 1 and off-diagonal entries are  $-1/(n-1)$ . Together with Lemma 1, the matrix  $P_n^*$  is the only choice of  $\Sigma$  with diagonal entries being 1 such that  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$  is an exchangeable NA (and thus NSD and NOD) joint mix.

One may hope that non-Gaussian elliptical distributions can represent NOD, NSD and NA joint mixes for  $n \geq 3$ . The following result states that Gaussian family is characterized as the only elliptical family which admits such a negatively dependent  $n$ -joint mix for all  $n$ . For a characteristic generator  $\psi$ , denote by  $\mathcal{E}(\psi)$  the class of all non-degenerate random vectors following an elliptical distribution with characteristic generator  $\psi$ . In what follows, a class  $\mathcal{E}(\psi)$  is a *Gaussian variance mixture family* if there exists a nonnegative random variable  $W$  such that each member  $\mathbf{X}$  admits the stochastic representation  $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$ , where  $\boldsymbol{\mu} \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{n \times k}$ , and  $\mathbf{Z}$  is a  $k$ -dimensional standard Gaussian independent of  $W$ .

**Theorem 6.** *Let  $\psi$  be a characteristic generator.*

- (i) *The class  $\mathcal{E}(\psi)$  contains an NCD  $n$ -joint mix for all  $n \geq 2$  if and only if the class  $\mathcal{E}(\psi)$  is a Gaussian variance mixture family.*
- (ii) *The class  $\mathcal{E}(\psi)$  contains an NOD, NSD, or NA  $n$ -joint mix for all  $n \geq 2$  if and only if the class  $\mathcal{E}(\psi)$  is Gaussian.*

Theorem 6 shows a clear contrast between NCD and other concepts of negative dependence. As seen in the proof of Theorem 6, NCD does not restrict the class of elliptical distributions since  $\psi$  generates an  $n$ -dimensional elliptical distribution for every  $n \in \mathbb{N}$  if and only if the corresponding elliptical class is a Gaussian variance mixture family (Fang et al., 1990, Section 2.6). Note that a class of multivariate  $t$  distributions (with a common degree of freedom) is an example of a Gaussian variance mixture family. On the other hand, NOD, NSD and NA characterize Gaussian. This result stems from the fact that multivariate Gaussian distribution is the only one among elliptical distributions such that independence is equivalent to uncorrelatedness.

## 6 Conclusion

The paper has focused on the relationship between JM and classic notions of negative dependence such as NOD and NA. Various connections between these concepts are obtained, and some conditions for a joint mix to be negatively dependent are derived. In particular, an exchangeable negatively dependent joint mix solves a multi-marginal optimal transport problem for quadratic cost under uncertainty on the participation of agents.

Negative dependence is always studied with many technical challenges. Although our main questions are addressed or partially addressed in this paper, they give rise to many questions that remain open. We list a few of them that we find particularly interesting.

1. Under what conditions, possibly stronger than exchangeability and NOD, is a joint mix NA? An example of NOD joint mix that is not NA can be found in Section 3.2 of Malinovsky and Rinott (2023) in the context of knockout tournaments with a nonrandom draw.
2. Under what general conditions, other than Gaussian, do we know a tuple of distributions supports an NA joint mix? For a fixed  $n \geq 3$ , this question is not clear even within the elliptical class.
3. It is unclear whether the decomposition result in Theorem 3 can be generalized to joint mixes that take infinitely many values or that are continuously distributed.

4. Assuming homogeneous marginals, does an exchangeable joint mix solve problem (8) for a general convex cost function  $f$ ? In Theorem 4, we showed that this holds true for quadratic cost. We also know that a joint mix is an optimizer for the general convex cost problem without uncertainty. These observations seem to hint at the possible optimality of some exchangeable joint mix for general convex cost under uncertainty, but we do not have a proof.
5. Do negatively dependent joint mixes play an important role in optimization problems other than the ones considered in Section 4? It is also unclear how results in Section 4 can be extended to heterogeneous marginal distributions with dimension higher than 3. Two unsolved questions have already been mentioned in Remarks 3 and 4.

These questions yield new challenges to dependence theory and require future research.

## Acknowledgements

Takaaki Koike was supported by JSPS KAKENHI Grant Number JP21K13275. Ruodu Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

## References

- Alam, K. and Saxena, K. M. L. (1981). Positive dependence in multivariate distributions. *Communications in Statistics-Theory and Methods*, **10**(12), 1183–1196.
- Amini, M., Nili Sani, H. R. and Bozorgnia, A. (2013). Aspects of negative dependence structures. *Communications in Statistics-Theory and Methods*, **42**(5), 907–917.
- Bernard, C., Bondarenko, O. and Vanduffel, S. (2018). Rearrangement algorithm and maximum entropy. *Annals of Operations Research*, **261**(1), 107–134.
- Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.
- Besser, K. and Jorswieck, E. A. (2020). Reliability bounds for dependent fading wireless channels. *IEEE Transactions on Wireless Communications*, **59**(9), 5833–5845.
- Blanchet, J., Lam, H., Liu, Y. and Wang, R. (2020). Convolution bounds on quantile aggregation. *arXiv: 2007.09320*
- Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *Annals of Probability*, **10**(3), 765–772.

- Boerma, J., Tsyvinski, A. and Zimin, A. P. (2021). Sorting with team formation. *arXiv: 2109.02730*.
- Christofides, T. C. and Vaggelatou, E. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, **88**(1), 138–151.
- Coffman, E. G. and Yannakakis, M. (1984). Permuting elements within columns of a matrix in order to minimize maximum row sum. *Mathematics of Operations Research*, **9**(3), 384–390.
- Craiu, R. V. and Meng, X. L. (2001). Chance and fractals. *Chance*, **14**(2), 47–52.
- Craiu, R. V. and Meng, X. L. (2005). Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Annals of Statistics*, **33**(2), 661–697.
- Dall’Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, **9**, 131–150.
- Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley: Chichester, UK.
- Di Marino, S., Gerolin, A. and Nenna, L. (2017). Optimal transportation theory with repulsive costs. In: *Topological Optimization and Optimal Transport* (Eds: Bergounioux et al.), 204–256.
- Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraaj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.
- Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. London: Chapman & Hall.
- Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press.
- Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141–153.
- Haus, U. (2015). Bounding stochastic dependence, complete mixability of matrices, and multi-dimensional bottleneck assignment problems. *Operations Research Letters*, **43**(1), 74–79.
- Hsu, W. L. (1984). Approximation algorithms for the assembly line crew scheduling problem. *Mathematics of Operations Research*, **9**(3), 376–383.
- Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, **11**, 286–295.
- Joe, H. (2014). *Dependence Modeling with Copulas*. London: Chapman & Hall.
- Lauzier, J. G., Lin, L. and Wang, R. (2023). Pairwise counter-monotonicity. *Insurance: Mathe-*

- matics and Economics*, **111**, 279–287.
- Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5), 1137–1153.
- Malinovsky, Y. and Rinott, Y. (2023). On tournaments and negative dependence. *Journal of Applied Probability*, **60**(3), 945–954.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley.
- Pass, B. (2015). Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, **49**(6), 1771–1790.
- Perchet, V., Rigollet, P. and Gouic, T. L. (2022). An algorithmic solution to the Blotto game using multi-marginal couplings. *arXiv*: 2202.07318.
- Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, **32**(3), 1482–1501.
- Puccetti, G. and Rüschendorf, L. (2013). Sharp bounds for sums of dependent risks. *Journal of Applied Probability*, **50**(1), 42–53.
- Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.
- Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.
- Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*. Springer, New York.
- Vovk, V., Wang, B. and Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *Annals of Statistics*, **50**(1), 351–375.
- Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, **102**(10), 1344–1360.
- Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.
- Wang, R. (2015). Current open questions in complete mixability. *Probability Surveys*, **12**, 13–32.
- Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.
- Xiao, Y. and Yao, J. (2020). A note on joint mix random vectors. *Communications in Statistics-Theory and Methods*, **49**(12), 3063–3072.
- Yin, C. (2021). Stochastic orderings of multivariate elliptical distributions. *Journal of Applied*

## A Proofs of all results

Some notions of negative dependence introduced in Section 2 are related to stochastic orders. We first introduce some concepts of stochastic order.

For two  $n$ -dimensional random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ ,  $\mathbf{X}$  is said to be less than  $\mathbf{Y}$  in *lower concordance order* (denoted by  $\mathbf{X} \leq_{cL} \mathbf{Y}$ ) if  $\mathbb{P}(\mathbf{X} \leq \mathbf{t}) \leq \mathbb{P}(\mathbf{Y} \leq \mathbf{t})$  for all  $\mathbf{t} \in \mathbb{R}^n$ , *upper concordance order* (denoted by  $\mathbf{X} \leq_{cU} \mathbf{Y}$ ) if  $\mathbb{P}(\mathbf{X} > \mathbf{t}) \leq \mathbb{P}(\mathbf{Y} > \mathbf{t})$  for all  $\mathbf{t} \in \mathbb{R}^n$ , *concordance order* (denoted by  $\mathbf{X} \leq_c \mathbf{Y}$ ) if  $\mathbf{X} \leq_{cL} \mathbf{Y}$  and  $\mathbf{X} \leq_{cU} \mathbf{Y}$ , and in *supermodular order* (denoted by  $\mathbf{X} \leq_{sm} \mathbf{Y}$ ) if  $\mathbb{E}[\psi(\mathbf{X})] \leq \mathbb{E}[\psi(\mathbf{Y})]$  for all supermodular functions  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  such that the expectations exist. Using these notations of stochastic order, the notions of negative dependence NLOD, NUOD, NOD and NSD for an  $n$ -dimensional random vector  $\mathbf{X} = (X_1, \dots, X_n)$  are denoted by  $\mathbf{X} \leq_{cL} \mathbf{X}^\perp$ ,  $\mathbf{X} \leq_{cU} \mathbf{X}^\perp$ ,  $\mathbf{X} \leq_c \mathbf{X}^\perp$  and  $\mathbf{X} \leq_{sm} \mathbf{X}^\perp$ , respectively, where we recall that  $\mathbf{X}^\perp = (X_1^\perp, \dots, X_n^\perp)$  is a random vector with independent components such that  $X_i \stackrel{d}{=} X_i^\perp$ ,  $i \in [n]$ .

*Proof of Proposition 1.* In part (a), the implication from (v) to (i) is shown by Joag-Dev and Proschan (1983). The other implications follow from (3). Parts (b) and (c) can be easily checked by definition. Finally, part (d) follows from the fact that a CT random vector for  $n \geq 3$  cannot have continuous marginal distributions (Dall'Aglio, 1972; Puccetti and Wang, 2015).  $\square$

*Proof of Theorem 1.* Note that to show NA, it suffices to show (1) for  $A, B$  that form a partition of  $[n]$ , as we can choose increasing functions in (1) that only depend on a subset of  $A$  and  $B$ . Let  $f$  and  $g$  be two increasing functions on  $\mathbb{R}^d$  and  $\mathbb{R}^{n-d}$ , respectively, where  $d$  is the cardinality of  $A$ . Note that

$$\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) = \mathbb{E}[\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)|S_A)] + \text{Cov}(\mathbb{E}[f(\mathbf{X}_A)|S_A], \mathbb{E}[g(\mathbf{X}_B)|S_A]);$$

see (1.1) of Joag-Dev and Proschan (1983). Using conditional independence (a), we get

$$\text{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) = \text{Cov}(\mathbb{E}[f(\mathbf{X}_A)|S_A], \mathbb{E}[g(\mathbf{X}_B)|S_A]).$$

Since  $S_A + S_B$  is a constant, condition (b) implies that  $\mathbb{E}[f(\mathbf{X}_A)|S_A]$  is an increasing function of  $S_A$ , and  $\mathbb{E}[g(\mathbf{X}_B)|S_A] = \mathbb{E}[g(\mathbf{X}_B)|S_B]$  is a decreasing function of  $S_A$ . This shows that their covariance is non-positive. Therefore,  $\mathbf{X}$  is NA.  $\square$

*Proof of Proposition 2.* Denote by  $\mathbf{X} = (X_1, \dots, X_n)$  the NSD  $n$ -joint mix with joint distribution  $F_{\mathbf{X}}$ . Let  $\mathbf{X}^{\Pi} = (X_{\Pi(1)}, \dots, X_{\Pi(n)})$  be an exchangeable joint mix, where  $\Pi$  follows a uniform distribution on  $\mathfrak{S}_n$  and is independent of  $\mathbf{X}$ . Obviously  $\mathbf{X}^{\Pi}$  is a joint mix, and has the same marginal distributions as  $\mathbf{X}$ . Let  $\bar{F} = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} F_{\mathbf{X}^{\pi}}$  be the distribution function of  $\mathbf{X}^{\Pi}$ . Then  $\bar{F}$  is exchangeable. Moreover,  $\bar{F}$  is NSD since

$$\mathbb{E}[\psi(\mathbf{X}^{\Pi})] = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathbb{E}[\psi(\mathbf{X}^{\pi})] \leq \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathbb{E}[\psi(\mathbf{X}^{\perp})] = \mathbb{E}[\psi(\mathbf{X}^{\perp})]$$

for every supermodular function  $\psi$  such that the expectations above exist. Other cases of NOD, NUOD and NLOD are shown analogously.  $\square$

*Proof of Proposition 3.* Without loss of generality, assume  $\sigma_n^2$  is the maximum of  $\{\sigma_1^2, \dots, \sigma_n^2\}$ . Note that NCD implies that the bivariate correlations are non-positive. If  $(X_1, \dots, X_n)$  is an NCD joint mix where  $X_i \sim F_i$ ,  $i \in [n]$ , then

$$\sigma_n^2 = \text{Var}(X_n) = \text{Var}(X_1 + \dots + X_{n-1}) \leq \sum_{i=1}^{n-1} \text{Var}(X_i) = \sum_{i=1}^{n-1} \sigma_i^2,$$

which yields (4) by adding  $\sigma_n^2$  to both sides.  $\square$

*Proof of Theorem 2.* The necessity follows from Proposition 3, and below we show sufficiency. Suppose that (4) holds. Without loss of generality, we can assume  $\sigma_n \geq \sigma_{n-1} \geq \dots \geq \sigma_1$ . It suffices to consider  $n \geq 3$  and  $\sigma_{n-1} > 0$ , and otherwise the problem is trivial. Moreover, the location parameters of the Gaussian distributions are not relevant, and they are assumed to be 0.

Let  $\lambda$  be a constant such that

$$\lambda^2 \sum_{i=1}^{n-1} \sigma_i^2 + (1 - \lambda^2) \sigma_{n-1}^2 = \sigma_n^2. \quad (17)$$

By (4), we have  $\sum_{i=1}^{n-1} \sigma_i^2 \geq \sigma_n^2 \geq \sigma_{n-1}^2$ , and this ensures that we can take  $\lambda \in [0, 1]$ .

Let  $P_n^*$  be a matrix with diagonal entries being 1 and off-diagonal entries being  $-1/(n-1)$ , and let  $P_n^{\perp}$  be a matrix with diagonal entries being 1 and off-diagonal entries being 0. Take  $\mathbf{Y} = (Y_1, \dots, Y_{n-1}) \sim N_{n-1}(\mathbf{0}_{n-1}, P_{n-1}^{\perp})$  and

$$\mathbf{Z}^{(m)} = (Z_n^{(m)}, \dots, Z_1^{(m)}) \sim N_{n-m+1}(\mathbf{0}_{n-m+1}, P_{n-m+1}^*), \quad m = 1, \dots, n-1,$$

such that  $\mathbf{Y}, \mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(n-1)}$  are independent. Note that  $\mathbf{Z}^{(n-1)} = (Z_{n-1}^{(n-1)}, Z_n^{(n-1)}) \sim N_2(\mathbf{0}_2, P_2^*)$  is 2-dimensional, and each  $\mathbf{Z}^{(m)}$  is a joint mix.

For notational simplicity, let the function  $d$  be given by  $d(a, b) = (a^2 - b^2)^{1/2}$  for  $a \geq b \geq 0$ . Note that  $a^2 = d(a, b)^2 + b^2$ . Moreover, for  $k = 1, \dots, n-1$ , let

$$\alpha_k = d(\sigma_k, \sigma_{k-1}) = (\sigma_k^2 - \sigma_{k-1}^2)^{1/2},$$

with  $\sigma_0 = 0$ , and thus  $\alpha_1 = \sigma_1$ . For  $k = 1, \dots, n-1$ , let

$$X_k = \lambda \sigma_k Y_k + d(1, \lambda) \sum_{j=1}^k \alpha_j Z_k^{(j)}.$$

Moreover, let

$$X_n = -\lambda Y^* + d(1, \lambda) \sum_{j=1}^{n-1} \alpha_j Z_n^{(j)}, \quad \text{where } Y^* = \sum_{k=1}^{n-1} \sigma_k Y_k.$$

For  $k = 1, \dots, n-1$ , using independence among  $Z_k^{(1)}, \dots, Z_k^{(k)}$ , we get

$$\text{Var} \left( \sum_{j=1}^k \alpha_j Z_k^{(j)} \right) = \sum_{i=1}^k \alpha_i^2 = \sigma_1^2 + d(\sigma_2, \sigma_1)^2 + \dots + d(\sigma_k, \sigma_{k-1})^2 = \sigma_k^2.$$

Hence,  $\sum_{j=1}^k \alpha_j Z_k^{(j)} \sim N_1(0, \sigma_k^2)$ , and again using independence of  $Y_k$  and  $\sum_{j=1}^k \alpha_j Z_k^{(j)}$ , we get  $X_k \sim N_1(0, \sigma_k^2)$ . By (17), we have

$$\text{Var}(X_n) = \text{Var} \left( \lambda \sum_{k=1}^{n-1} \sigma_k Y_k \right) + \text{Var} \left( d(1, \lambda) \sum_{j=1}^{n-1} \alpha_j Z_n^{(j)} \right) = \lambda^2 \sum_{i=1}^{n-1} \sigma_i^2 + (1 - \lambda^2) \sigma_{n-1}^2 = \sigma_n^2.$$

Hence,  $X_n \sim N_1(0, \sigma_n^2)$ .

Next, we show that  $(X_1, \dots, X_n)$  is a joint mix. We can directly compute

$$\begin{aligned} \sum_{k=1}^n X_k &= \sum_{i=k}^{n-1} \lambda \sigma_k Y_k + d(1, \lambda) \sum_{k=1}^{n-1} \sum_{j=1}^k \alpha_j Z_k^{(j)} - \lambda \sum_{k=1}^{n-1} \sigma_k Y_k + d(1, \lambda) \sum_{j=1}^{n-1} \alpha_j Z_n^{(j)} \\ &= d(1, \lambda) \sum_{j=1}^{n-1} \sum_{k=j}^n \alpha_j Z_k^{(j)} = 0, \end{aligned}$$

where the last equality follows from the fact that  $\mathbf{Z}^{(j)}$  is a joint mix for each  $j = 1, \dots, n-1$ .

We check that  $(X_1, \dots, X_n)$  is NA. This follows from the fact that  $(X_1, \dots, X_n)$  is the weighted sum of several independent NA random vectors  $(\sigma_1 Y_1, \dots, \sigma_{n-1} Y_{n-1}, -Y^*)$  and  $(\mathbf{0}_{n-1}, \mathbf{Z}^{(m)})$



for  $m = 1, \dots, n-1$ . Alternatively, one can check that all non-zero terms in  $\text{Cov}(X_i, X_j)$  are negative for  $i \neq j$  as follows. For  $X_k, X_l$  with  $k, l \leq n-1$  and  $k < l$ , we have

$$\begin{aligned}
\text{Cov}(X_k, X_l) &= \text{Cov} \left( \lambda \sigma_k Y_k + d(1, \lambda) \sum_{j=1}^k \alpha_j Z_k^{(j)}, \lambda \sigma_l Y_l + d(1, \lambda) \sum_{i=1}^l \alpha_i Z_l^{(i)} \right) \\
&= \lambda^2 \sigma_k \sigma_l \text{Cov}(Y_k, Y_l) + \lambda \sigma_k d(1, \lambda) \sum_{i=1}^l \alpha_i \text{Cov}(Z_l^{(i)}, Y_k) \\
&\quad + \lambda \sigma_l d(1, \lambda) \sum_{j=1}^k \alpha_j \text{Cov}(Z_k^{(j)}, Y_l) + d^2(1, \lambda) \sum_{j=1}^k \sum_{i=1}^l \alpha_j \alpha_i \text{Cov}(Z_k^{(j)}, Z_l^{(i)}) \\
&= -d^2(1, \lambda) \sum_{j=1}^k \frac{\alpha_j^2}{n-j} \leq 0.
\end{aligned}$$

For  $X_k, X_n$  for all  $k \leq n-1$ , we have

$$\begin{aligned}
\text{Cov}(X_k, X_n) &= \text{Cov} \left( \lambda \sigma_k Y_k + d(1, \lambda) \sum_{j=1}^k \alpha_j Z_k^{(j)}, -\lambda \sum_{i=1}^{n-1} \sigma_i Y_i + d(1, \lambda) \sum_{i=1}^{n-1} \alpha_i Z_n^{(i)} \right) \\
&= -\lambda^2 \sigma_k \sum_{i=1}^{n-1} \sigma_i \text{Cov}(Y_k, Y_i) + \lambda \sigma_k d(1, \lambda) \sum_{i=1}^{n-1} \alpha_i \text{Cov}(Z_n^{(i)}, Y_k) \\
&\quad - \lambda d(1, \lambda) \sum_{j=1}^k \sum_{i=1}^{n-1} \sigma_i \alpha_j \text{Cov}(Z_k^{(j)}, Y_i) + d^2(1, \lambda) \sum_{j=1}^k \sum_{i=1}^{n-1} \alpha_j \alpha_i \text{Cov}(Z_k^{(j)}, Z_n^{(i)}) \\
&= -\lambda^2 \sigma_k^2 - d^2(1, \lambda) \sum_{j=1}^k \frac{\alpha_j^2}{n-j} \leq 0.
\end{aligned}$$

Finally, the joint mix can be chosen as multivariate Gaussian by the construction of  $(X_1, \dots, X_n)$  as the sum of Gaussian vectors.  $\square$

*Proof of Theorem 3.* The “if” statement is straightforward, and we will check the “only if” statement. Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a joint mix and denote by  $c = \sum_{i=1}^n X_i \in \mathbb{R}$ . First, suppose that each component of  $\mathbf{X}$  is positive. Denote by  $V \subset \mathbb{R}$  the set of all possible values taken by random variables of the form  $\sum_{i=1}^j X_i$  for  $j = 0, \dots, n$ , with the convention that  $\sum_{j=1}^0 X_j = 0$ . Clearly,  $V$  is finite. The elements of  $V$  are denoted by  $v_0, v_1, \dots, v_K$  such that  $v_0 < v_1 < \dots < v_K$ . Our assumptions imply that  $v_0 = 0$ ,  $v_1 > 0$  and  $v_K = c$  because  $\sum_{i=1}^n X_i = c$ . For  $k \in [K]$  and  $i \in [n]$ , let

$$Y_{k,i} = \mathbb{1}_{\{\sum_{j=1}^i X_j \geq v_k\}} - \mathbb{1}_{\{\sum_{j=1}^{i-1} X_j \geq v_k\}}$$

and let  $\mathbf{Y}_k = (Y_{k,1}, \dots, Y_{k,n})$ . Since each  $X_j$  is positive, the value of  $Y_{k,i}$  is either 0 or 1, and

$$\sum_{i=1}^n Y_{k,i} = \mathbb{1}_{\{\sum_{j=1}^n X_j \geq v_k\}} - \mathbb{1}_{\{0 \geq v_k\}} = \mathbb{1}_{\{c \geq v_k\}} = 1.$$

Therefore,  $\mathbf{Y}_k$  follows a binary multinomial distribution for each  $k \in [K]$ . Let  $\mathbf{X}_k = (v_k - v_{k-1})\mathbf{Y}_k$  for  $k \in [K]$  with  $v_0 = 0$ . Note that for  $i \in [n]$ ,

$$\sum_{k=1}^K X_{k,i} = \sum_{k=1}^K (v_k - v_{k-1}) \left( \mathbb{1}_{\{\sum_{j=1}^i X_j \geq v_k\}} - \mathbb{1}_{\{\sum_{j=1}^{i-1} X_j \geq v_k\}} \right) = \sum_{j=1}^i X_j - \sum_{j=1}^{i-1} X_j = X_i,$$

where we used the identity  $\sum_{k=1}^K (v_k - v_{k-1}) \mathbb{1}_{\{x \geq v_k\}} = x$  for  $x \in V$ . Therefore,  $\sum_{k=1}^K \mathbf{X}_k = \mathbf{X}$ , showing that  $\mathbf{X}$  can be represented as a finite linear combination of binary multinomial random vectors.

If some components of  $\mathbf{X}$  are not positive, we can take  $m \in \mathbb{R}$  such that  $X_i > m$  for each  $i \in [n]$ . Applying the above result, we know that  $(X_1 - m, \dots, X_n - m)$  can be decomposed as the sum of JM joint mixes. Note that  $\mathbf{X} = (X_1 - m, \dots, X_n - m) + (m, \dots, m)$ , and  $(m, \dots, m)$  is  $m$  times the sum of  $n$  binary multinomial random vectors  $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ . Hence,  $\mathbf{X}$  admits a finite linear combination of binary multinomial random vectors.  $\square$

*Proof of Proposition 4.* Let  $G$  be the joint distribution of an exchangeable joint mix. Let us write

$$\tilde{G}(A) = \int_{\mathbb{R}^n} \delta_{\mathbf{a}}(A) dG(\mathbf{a}), \quad A \in \mathcal{B}(\mathbb{R}^n),$$

where  $\delta_{\mathbf{a}}$  is the point-mass at  $\mathbf{a}$ . By exchangeability, we have  $G(A^\pi) = G(A)$  for  $\pi \in \mathfrak{S}_n$  and  $A \in \mathcal{B}(\mathbb{R}^n)$ , where  $A^\pi$  is  $\pi$  applied to elements of  $A$ . Therefore,

$$G(A) = \int_{\mathbb{R}^n} \delta_{\mathbf{a}^\pi}(A) dG(\mathbf{a}).$$

Taking an average of the above formula over  $\mathfrak{S}_n$ , we have

$$G(A) = \int_{\mathbb{R}^n} U_{\mathbf{a}}(A) dG(\mathbf{a}).$$

It is known that each  $U_{\mathbf{a}}$  is NA (Joag-Dev and Proschan, 1983, Theorem 2.11). Moreover, the center of the joint mix distributed as  $U_{\mathbf{a}}$  is  $\mu$  since  $G$  is supported on  $\{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 + \dots + x_n = \mu\}$ .  $\square$

*Proof of Proposition 5.* As  $\mathcal{M}$  is symmetric, we have  $\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}) = \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\pi)$  for all

$\pi \in \mathfrak{S}_n$ . Let  $\Pi$  be uniformly distributed on  $\mathfrak{S}_n$  and independent of  $\mathbf{X}$ . Plugging  $\mathbf{X}^\Pi$  in the objective (7), we have

$$\begin{aligned}
\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\Pi) &= \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i^\Pi \right) \right] \mu(K) \\
&= \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i^\pi \right) \right] \mu(K) \\
&\leq \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E} \left[ f \left( \sum_{i \in K} X_i^\pi \right) \right] \mu(K) \\
&= \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\pi) = \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}).
\end{aligned}$$

Hence,  $\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\Pi) \leq \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X})$ . Furthermore, as  $\mathbf{X}$  is a joint mix, we have that  $\mathbf{X}^\Pi$  is an exchangeable NCD joint mix with marginals  $F$  and correlation matrix  $P_n^*$ .  $\square$

*Proof of Theorem 4.* Without loss of generality, we assume that the variance of  $F$  is 1. Using the same argument in the proof of Proposition 5, for any  $\mathbf{X}$  with identical marginals  $F$ , we have  $\sup_{\mu \in \mathcal{M}} C_\mu^2(\mathbf{X}^\Pi) \leq \sup_{\mu \in \mathcal{M}} C_\mu^2(\mathbf{X})$ , where  $\Pi$  is uniformly distributed on  $\mathfrak{S}_n$ . Let  $\mathbf{X}_\rho$  be a random vector with identical marginals  $F$  and a correlation matrix whose off-diagonal entries are all  $\rho$ . Since correlation matrices are positive semi-definite, we have  $\rho \in [-1/(n-1), 1]$ , with  $\rho = -1/(n-1)$  attainable since  $F$  is  $n$ -completely mixable. The value of  $\sup_{\mu \in \mathcal{M}} C_\mu^2(\mathbf{X})$  only depends on the correlation matrix. Therefore, it suffices to find an optimizer of the form  $\mathbf{X}_\rho$  for some  $\rho \in [-1/(n-1), 1]$ . Note that

$$\begin{aligned}
\sup_{\mu \in \mathcal{M}} C_\mu^2(\mathbf{X}_\rho) &= \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \left( \text{Var} \left( \sum_{i \in K} X_i \right) + \left( \mathbb{E} \left[ \sum_{i \in K} X_i \right] \right)^2 \right) \mu(K) \\
&= \sup_{\mu \in \mathcal{M}} \sum_{k=1}^n \sum_{K \subseteq [n], |K|=k} (k + (k^2 - k)\rho + k\mathbb{E}[X_1]) \mu(K).
\end{aligned}$$

It is clear that  $\sup_{\mu \in \mathcal{M}} C_\mu^2(\mathbf{X}_\rho)$  increases in  $\rho$ . Therefore, the minimum is achieved at  $\rho^* = -1/(n-1)$ , which implies that  $\mathbf{X}_{\rho^*}$  is an NCD joint mix with correlation matrix  $P_n^*$ . As the value of (11) only depends on the correlation matrix, we have the desired result.  $\square$

*Proof of Theorem 5.* The “if” part is shown by Theorem 4 by choosing  $\mathcal{M}$  as both (12) and (13) are special cases of (11). Next, we show the “only if” part by showing that the correlation matrix of the minimizer to (12) or (13) with any  $k \in [n] \setminus \{1, n-1, n\}$  is  $P_n^*$ .

Without loss of generality, we assume that the variance of  $F$  is 1. As the mean of  $F$  is zero, we have  $\mathbb{E}[(\sum_{i \in K} X_i)^2] = \text{Var}(\sum_{i \in K} X_i)$ . By plugging an NCD joint mix  $\mathbf{X}^E$  with correlation matrix  $P_n^*$  into (12), the optimal value for (12) is

$$\max_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] = \max_{K \subseteq [n]} \text{Var} \left( \sum_{i \in K} X_i^E \right) = \max_{k \in [n]} \frac{k(n-k)}{n-1} = \frac{k^*(n-k^*)}{n-1}, \quad (18)$$

where  $k^* = \lfloor n/2 \rfloor$ .

First, we consider the case  $n = 3$ . In this case,  $[n] \setminus \{1, n-1, n\}$  is empty, and we only need to show that  $P_n^*$  is the unique correlation matrix of the minimizer to (12). Suppose that  $\mathbf{X}$  with covariance matrix  $\Sigma$  is a minimizer to (12). By (18), optimal value for (12) is 1. Hence,

$$\text{Var} \left( \sum_{i \in K} X_i \right) \leq 1 \quad \text{for each } K \text{ with } |K| = 2,$$

and this implies

$$\sigma_{ij} \leq -1/2, \quad \text{for } i \neq j. \quad (19)$$

Since  $\Sigma$  is positive semi-definite, we have  $\sum_{i,j \in [3]} \sigma_{ij} \geq 0$ , which leads to  $3 + 2\sigma_{12} + 2\sigma_{13} + 2\sigma_{23} \geq 0$ , implying  $\sigma_{12} + \sigma_{13} + \sigma_{23} \geq -3/2$ . Together with (19), we get  $\sigma_{12} = \sigma_{13} = \sigma_{23} = -1/2$ , and hence  $\Sigma = P_n^*$ .

Next, we consider the case  $n \geq 4$ . We first show that the correlation matrix of the minimizer to (13) is unique for  $1 < k < n-1$ . Fix  $k \in [n] \setminus \{1, n-1, n\}$ . Suppose that  $\mathbf{X}$  with covariance matrix  $\Sigma$  is a minimizer to (13). Our goal is to show  $\Sigma = P_n^*$ .

Let  $K_\ell$ ,  $\ell = 1, \dots, n_k$ , be all subsets of  $[n]$  with cardinality  $k$ , where  $n_k = \binom{n}{k}$ . Then we have

$$\frac{1}{n_k} \sum_{\ell=1}^{n_k} \text{Var} \left( \sum_{i \in K_\ell} X_i \right) \geq k - \frac{k(k-1)}{n-1} = \frac{k(n-k)}{n-1}. \quad (20)$$

As  $\mathbf{X}$  is a minimizer, for each  $K$  with  $|K| = k$ , we have

$$\text{Var} \left( \sum_{i \in K} X_i \right) = \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \leq \max_{K \subseteq [n], |K|=k} \mathbb{E} \left[ \left( \sum_{i \in K} X_i^E \right)^2 \right] = \frac{k(n-k)}{n-1}. \quad (21)$$

Combining (20) and (21), we have

$$\text{Var} \left( \sum_{i \in K} X_i \right) = \frac{k(n-k)}{n-1} \quad \text{for each } K \text{ with } |K| = k.$$

Take  $k = 2$ . For any  $i, j \in [n]$  such that  $i \neq j$ , the above equation implies

$$\text{Var}(X_i + X_j) = \text{Var}(X_i) + \text{Var}(X_j) + 2\text{Cov}(X_i, X_j) = \frac{2(n-2)}{n-1}.$$

As a result, we have  $\text{Cov}(X_i, X_j) = -1/(n-1)$  for all  $i, j \in [n]$  such that  $i \neq j$ . Hence, we conclude that  $\Sigma = P_n^*$ .

Finally, note that  $k^*$  in (18) satisfies  $1 < k^* < n-1$  for  $n \geq 4$ . We have justified that the correlation matrix for optimizers to (13) with  $k = k^*$  is unique. Therefore, by using (18), the correlation matrix for optimizers to (12) is also unique.

The above arguments show that, for  $n \geq 3$ , if  $\mathbf{X}$  is a minimizer to (12), then the correlation matrix of  $\mathbf{X}$  is  $P_n^*$ , which implies that  $\mathbf{X}$  is an NCD joint mix with correlation matrix  $P_n^*$ . The same conclusion holds true if (12) is replaced by (13) with any  $k \in [n] \setminus \{1, n-1, n\}$ .  $\square$

*Proof of Proposition 6.* As the given marginal distributions have zero means, for any  $(Y_1, \dots, Y_n)$  with variance vector  $(\sigma_1^2, \dots, \sigma_n^2)$ ,

$$\max_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} Y_i \right)^2 \right] = \max_{K \subseteq [n]} \text{Var} \left( \sum_{i \in K} Y_i \right) \geq \max_{i \in [n]} \text{Var}(Y_i) = \max_{i \in [n]} \sigma_i^2.$$

In case  $n = 3$ , a joint mix  $\mathbf{X}$  with variance vector  $(\sigma_1^2, \dots, \sigma_n^2)$  satisfies

$$\max_{K \subseteq [3], |K|=1} \text{Var} \left( \sum_{i \in K} X_i \right) = \max_{K \subseteq [3], |K|=2} \text{Var} \left( \sum_{i \in K} X_i \right) = \max_{i \in [3]} \text{Var}(X_i) = \max_{i \in [3]} \sigma_i^2,$$

and  $\text{Var}(X_1 + X_2 + X_3) = 0$ . Hence, the joint mix minimizes (12).

To show that no positive covariance is allowed, suppose that  $(X_1, X_2, X_3)$  is a minimizer to (12) and  $\text{Cov}(X_i, X_j) > 0$  for some  $i \neq j$ . We have

$$\max_{K \subseteq [n]} \mathbb{E} \left[ \left( \sum_{i \in K} X_i \right)^2 \right] \geq \mathbb{E} \left[ (X_i + X_j)^2 \right] = \text{Var}(X_i + X_j) > \sigma_i^2 + \sigma_j^2 \geq \max(\sigma_1^2, \sigma_2^2, \sigma_3^2), \quad (22)$$

where the last inequality follows from the necessary condition (4) of the existence of an NCD joint mix. Since we have seen that the optimal value of (12) is  $\max_{i \in [3]} \sigma_i^2$ , (22) implies that  $(X_1, X_2, X_3)$  does not minimize (12).  $\square$

*Proof of Lemma 1.* One of the key properties of elliptical distributions is that they are closed under linear transformations, which is clear from (16). Hence, for  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$ , the random

variable  $\sum_{i=1}^n X_i$  follows  $E_n(\mathbf{1}_n^\top \boldsymbol{\mu}, \mathbf{1}_n^\top \Sigma \mathbf{1}_n, \psi)$ , which is degenerate if and only if  $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$ . Hence,  $\mathbf{X}$  is a joint mix if and only if  $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$  or  $\psi = 1$  on  $\mathbb{R}_+$ .  $\square$

*Proof of Proposition 7.* Necessity follows from Proposition 3. To show sufficiency, let  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$ , where  $\Sigma$  is the dispersion matrix of the multivariate Gaussian distribution constructed in the proof of Theorem 2. Then  $X_i \sim E_1(\mu_i, \sigma_i, \psi)$ ,  $i \in [n]$ . Moreover, it is checked in the proof of Theorem 2 that  $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$  and  $\sigma_{ij} \leq 0$  for  $i, j \in [n]$  such that  $i \neq j$ . Therefore,  $\mathbf{X}$  is the desired NCD joint mix.  $\square$

*Proof of Theorem 6.* The statement (i) immediately follows from the facts that  $\psi$  generates an  $n$ -dimensional elliptical distribution for every  $n \in \mathbb{N}$  if and only if the corresponding elliptical class is a Gaussian variance mixture family (Fang et al., 1990, Section 2.6), and that  $E_n(\mathbf{0}_n, P_n^*, \psi)$  is an NCD joint mix, where  $P_n^*$  is an  $n \times n$  matrix whose diagonal entries are 1 and off-diagonal entries are  $-1/(n-1)$ .

To show (ii), we need two lemmas.

**Lemma 2** (Corollary 4 of Yin (2021)). *Let  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$  and  $\mathbf{Y} \sim E_n(\boldsymbol{\mu}, \Sigma', \psi)$  be two elliptical distributions such that  $\Sigma = (\sigma_{ij})$  and  $\Sigma' = (\sigma'_{ij})$  satisfy  $\sigma_{ii} = \sigma'_{ii}$  for all  $i \in [n]$ . Then  $\mathbf{X} \leq_{cU} \mathbf{Y}$  if and only if  $\sigma_{ij} \leq \sigma'_{ij}$  for all  $i \neq j$ .*

Let  $P_n^\perp$  be the identity matrix, which is the correlation matrix of an independent random vector. Although Lemma 2 implies that  $E_n(\boldsymbol{\mu}, P_n^*, \psi) \leq_c E_n(\boldsymbol{\mu}, P_n^\perp, \psi)$  for general elliptical distributions,  $E_n(\boldsymbol{\mu}, P_n^*, \psi)$  is not necessarily NOD in general since  $E_n(\mathbf{0}_n, P_n^\perp, \psi)$  does not have independent components. In fact, an elliptical distribution  $E_n(\mathbf{0}_n, P_n^\perp, \psi)$  is not NOD unless it is Gaussian.

**Lemma 3.** *The elliptical distribution  $E_n(\boldsymbol{\mu}, \Sigma, \psi)$  where  $\Sigma$  is diagonal is not NOD unless it is Gaussian.*

*Proof.* Assume that  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$  is NOD. Since NOD is location invariant, it suffices to show the case when  $\boldsymbol{\mu} = \mathbf{0}_n$ . When  $\mathbf{X}$  is NOD, then so is  $(X_1, X_2)$ , that is,

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \leq \mathbb{P}(X_1 \leq x_1)\mathbb{P}(X_2 \leq x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

Since  $(X_1, X_2)$  and  $(-X_1, X_2)$  are identically distributed, we have

$$\mathbb{P}(X_1 \geq x_1, X_2 \leq x_2) \leq \mathbb{P}(X_1 \geq x_1)\mathbb{P}(X_2 \leq x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2,$$

and similarly, by symmetry,

$$\mathbb{P}(X_1 \geq x_1, X_2 \geq x_2) \leq \mathbb{P}(X_1 \geq x_1)\mathbb{P}(X_2 \geq x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2,$$

$$\mathbb{P}(X_1 \leq x_1, X_2 \geq x_2) \leq \mathbb{P}(X_1 \leq x_1)\mathbb{P}(X_2 \geq x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

Adding the above four inequalities together, we get  $1 \leq 1$ . Hence, each of them is an equality. However,  $(X_1, X_2)$  follows a bivariate elliptical distribution with generator  $\psi$ , and thus  $X_1$  and  $X_2$  are not independent unless it is Gaussian; see Theorem 4.11 of [Fang et al. \(1990\)](#). Therefore,  $\mathbf{X}$  cannot be NOD unless it is Gaussian.  $\square$

Now we are ready to prove Theorem 6. The “if” statement follows from Proposition 1. It remains to show the “only if” statement. Let  $\psi$  be a characteristic generator different from that of the Gaussian distribution. For  $n \geq 2$ , let  $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \psi)$  be an NOD joint mix where  $\Sigma$  has positive diagonal entries. We start by observing from Lemma 2 that if  $\sigma_{ij} > 0$  for  $i \neq j$ , then the bivariate projection  $(X_i, X_j)$  of  $\mathbf{X}$  satisfies  $(X_i, X_j) \geq_c (X'_i, X'_j)$  where  $(X'_i, X'_j) \sim E_n(\boldsymbol{\mu}, \Sigma'_{ij}, \psi)$  with

$$\Sigma'_{ij} = \begin{pmatrix} \sigma_{ii} & 0 \\ 0 & \sigma_{jj} \end{pmatrix}.$$

Using Lemma 3, we know that  $(X'_i, X'_j)$  is not NOD, that is, there exists  $(x_i, x_j) \in \mathbb{R}^2$  such that

$$\mathbb{P}(X'_i \leq x_i, X'_j \leq x_j) > \mathbb{P}(X'_i \leq x_i)\mathbb{P}(X'_j \leq x_j); \quad (23)$$

note that it suffices to consider the inequality needed for NLOD (not NUOD) by symmetry of the elliptical distribution and location invariance of NOD. Therefore, we have that

$$\mathbb{P}(X_i \leq x_i, X_j \leq x_j) > \mathbb{P}(X'_i \leq x_i, X'_j \leq x_j). \quad (24)$$

The two inequalities (23) and (24) imply that

$$\mathbb{P}(X_i \leq x_i, X_j \leq x_j) > \mathbb{P}(X_i \leq x_i)\mathbb{P}(X_j \leq x_j),$$

that is,  $(X_i, X_j)$  is not NOD. This leads to a contradiction.

Next, we assume  $\sigma_{ij} \leq 0$  for all  $i \neq j$ . Since  $\mathbf{a}^\top \Sigma \mathbf{a} \geq 0$  for all  $\mathbf{a} \in \mathbb{R}^n$  and  $\Sigma$  has positive

diagonal entries, we can take  $\mathbf{a} = (1/\sqrt{\sigma_{11}}, \dots, 1/\sqrt{\sigma_{nn}})$ , and this yields

$$\sum_{i,j=1}^n \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = n + \sum_{i \neq j} \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \geq 0.$$

Hence, there exist  $i, j$  with  $i \neq j$  such that

$$\rho_{ij} := \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \geq -\frac{1}{n-1}.$$

Since NOD is location-scale invariant, the NOD of  $(X_i, X_j)$  implies that  $E_2(\mathbf{0}_2, P_{ij}, \psi)$  is NOD, where

$$P_{ij} = \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix}.$$

Taking a limit as  $n \rightarrow \infty$ , and noting that NOD is closed under weak convergence (Muller and Stoyan, 2002), we conclude that  $E_2(\mathbf{0}_2, P_2^\perp, \psi)$  is also NOD, which contradicts Lemma 3 if  $\mathcal{E}(\psi)$  is not Gaussian.

Finally, by Proposition 1, this characterization result follows if NOD is replaced by NSD or NA. □