A Wasserstein index of dependence
for random measures

Marta Catalano

Department of Economics and Finance, Luiss University

and

Hugo Lavenant, Antonio Lijoi, Igor Prünster

Department of Decision Sciences and BIDSA, Bocconi University

Abstract

Optimal transport and Wasserstein distances are flourishing in many scientific fields as a means for comparing and connecting random structures. Here we pioneer the use of an optimal transport distance between Lévy measures to solve a statistical problem. Dependent Bayesian nonparametric models provide flexible inference on distinct, yet related, groups of observations. Each component of a vector of random measures models a group of exchangeable observations, while their dependence regulates the borrowing of information across groups. We derive the first statistical index of dependence in $[0, 1]$ for (completely) random measures that accounts for their whole infinite-dimensional distribution, which is assumed to be equal across different groups. This is accomplished by using the geometric properties of the Wasserstein distance to solve a max-min problem at the level of the underlying Lévy measures. The Wasserstein index of dependence sheds light on the models' deep structure and has desirable properties: (i) it is 0 if and only if the random measures are independent; (ii) it is 1 if and only if the random measures are completely dependent; (iii) it simultaneously quantifies the dependence of $d \geq 2$ random measures, avoiding the need for pairwise comparisons; (iv) it can be evaluated numerically. Moreover, the index allows for informed prior specifications and fair model comparisons for Bayesian nonparametric models.

Keywords: Bayesian nonparametrics | Index of dependence | Lévy measure | Random measure | Wasserstein distance.

1   Introduction

Complex phenomena often yield data from different but related sources, which are ideally suited to Bayesian modeling because of its inherent borrowing of information. In a nonparametric setting this is regulated by the dependence between random measures, which provide the main building block for many dependent priors. This is witnessed by the multitude of contributions in the literature; see MacEachern (1999, 2000) for pioneering ideas and Quintana et al. (2022) for a recent review. The unknown distribution $\tilde{P}_i$ of each group of observations $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n_i})$

is flexibly modeled as $\tilde{P}_i = t(\tilde{\mu}_i)$, where $\tilde{\mu}_i$ is a random measure and t is a suitable transformation that typically maps it to a space of random *probability* measures. Notable examples, for which $\tilde{\mu}_i$ is a completely random measure, include normalization for random probability mass functions (Regazzini et al., 2003), kernel mixtures for densities (Lo, 1984) and for hazards (Dykstra and Laud, 1981; James, 2005), exponential transformations for survival functions (Doksum, 1974) and cumulative transformations for cumulative hazards (Hjort, 1990). When a priori the distribution for each group is assessed to be similar, it is often convenient to borrow information across different groups. One can allow for different levels of borrowing through the dependence structure of the random measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$, which regulates the interaction among $d \geq 2$ groups of observations; see, e.g., Nguyen (2016); Camerlenghi et al. (2019). This class of models is summarized as $\boldsymbol{X}_i | \tilde{\boldsymbol{\mu}} \sim t(\tilde{\mu}_i)$ independently for $i = 1, \ldots, d$, where $\tilde{\boldsymbol{\mu}}$ has all equal marginal distributions. One may distinguish two extreme situations: (a) when the random measures are completely dependent, that is, $\tilde{\mu}_1 = \cdots = \tilde{\mu}_d$ almost surely, there is no distinction between the different groups. In such case the observations are exchangeable, in the sense that their law is invariant with respect to permutations not only within the same group but also across different groups; (b) when the random measures are independent, the groups do not interact and consist of d independent groups of exchangeable observations. When performing Bayesian inference on this class of models, different levels of prior interaction between groups entail smaller or greater borrowing of information in the posterior update, with a crucial impact on the estimates for the distribution of each group. One should thus enable the practitioner to choose the hyperparameters of the prior for $\tilde{\boldsymbol{\mu}}$ so to include the desired level of interaction between groups. This leads to the need for a precise measure of dependence between d random measures that simultaneously quantifies the discrepancy from both extreme situations of exchangeability and independence.

Our proposal is rooted in the theory of optimal transport and Wasserstein distances, which are flourishing in many scientific fields as a means to compare and connect different random structures (Santambrogio, 2015; Villani, 2003; Panaretos and Zemel, 2019). We take three conceptual steps: 1) For most completely random measures the density of $\tilde{\mu}_i(A)$ is intractable and the distribution of $\tilde{\boldsymbol{\mu}}$ is specified in an indirect way in terms of a multivariate Lévy measure, which characterizes its distribution. In order to have closed form expressions for a measure of dependence the first key idea is thus to define it at the level of the Lévy measures. 2) We measure the dependence as distance from exchangeability, which corresponds to maximal dependence, by resorting to an extended Wasserstein distance between Lévy measures. This geometric distance was introduced by Figalli and Gigli (2010) in a different context and it remarkably allows for the comparison between measures with different and possibly infinite mass. We unravel key properties of this distance that allow one to find the optimal extended coupling for the distance from exchangeability, reducing its evaluation to a one-dimensional integral that can be computed numerically. 3) We use the distance to quantify the discrepancy with respect to the other extreme, independence, by proving that it achieves the maximum distance from exchangeability. This pivotal result requires to solve an intriguing max-min problem whose solution leverages the dual formulation of the extended Wasserstein distance. There are two crucial consequences: first, once we find the maximum of the distance from exchangeability, we are able to renormalize the distance and obtain an index between 0 and 1; second, since the index is equal to 0 if and only if the random measures are independent and equal to 1 if and only if they are completely dependent, the Wasserstein index of dependence provides an overall measure of discrepancy from both extremes.

The Wasserstein distance has been used to measure the dependence on Euclidean or Polish

spaces in several interesting settings; see Nies et al. (2021) for an up-to-date account. Specifically, both Nies et al. (2021) and Mordant and Segers (2022) propose to define an index of dependence by finding the supremum of the Wasserstein distance from an extremal dependence structure, which in their works (as in many others) is independence. Measuring the distance from independence allows to be more flexible in the definition of maximal dependence for a random vector (X, Y), going beyond almost sure equality ($Y = X$ a.s.). This is especially useful when the marginal distributions of X and Y differ: in such case the definition of maximal dependence usually boils down to $Y = f(X)$ a.s., for f in some class of functions. However, there is no universal consensus on which class of functions to use. As effectively underlined in Nies et al. (2021), such choice is context-specific and should rather be made on a case-by-case basis. Arguably, the most common classes found in the literature are: the whole set of measurable functions, the set of monotonic functions, and the set of linear functions (interestingly, Nies et al. (2021) define an index that is maximized on the set of α-Lipschitz functions). While this level of flexibility may be valuable in many settings, in our context not only it is not necessary, since we are considering equal marginal distributions, but it could also be harmful in future extensions to the case with unequal marginals. Indeed, our notion of complete dependence for random measures is rooted in the full homogeneity of the underlying groups of observations (exchangeability), which requires the random measures to be almost surely equal.

A similar idea to the one of the present contribution can be found in Catalano et al. (2021), where dependence is measured in terms of distance from exchangeability at the level of two random measures. This precludes exact calculations (only upper bounds are available) and ultimately does not provide any notion of discrepancy from independence. The distance from exchangeability alone can still be useful for relative comparisons between dependence structures ("$\tilde{\boldsymbol{\mu}}^1$ is more dependent than $\tilde{\boldsymbol{\mu}}^2$") but prevents absolute quantifications of dependence ("$\tilde{\boldsymbol{\mu}}$ has an *intermediate* dependence structure") and the assessment of closeness to independence. Summing up, our Wasserstein index of dependence $I_{\mathcal{W}}$ crucially overcomes this limitation and has the following properties: (i) it is equal to 0 if and only if the random measures are independent; (ii) it is equal to 1 if and only if the random measures are completely dependent; (iii) it simultaneously quantifies the dependence of $d \geq 2$ random measures, avoiding the need for pairwise comparisons; (iv) since it is defined at the level of the Lévy measures, it is possible to evaluate it numerically. An important additional merit of the proposed index is that it allows for a principled comparison of the inferential performance of different models: by tuning their prior parameters to achieve the same value of the index of dependence, one can make a fair assessment of their posterior performance under different scenarios.

The paper is structured as follows. In Section 2 we define the Wasserstein index of dependence and state our main result (Theorem 1), together with some intuition on both the statistical and the mathematical problems we address in this work. In Section 3 we define the extended Wasserstein distance between Lévy measures and highlight some important novel properties that provide further insights on this distance and that are needed to prove the results in Section 4. Here, we focus on the theoretical findings behind the evaluation of the index, which remarkably recover an explicit expression for the optimal transport coupling in this multivariate setting (Theorem 5). In Section 5 we evaluate the Wasserstein index of dependence in notable models in the literature, namely additive random measures (Müller et al., 2004; Lijoi et al., 2014; Lijoi and Nipoti, 2014) and compound random measures (Griffin and Leisen, 2017, 2018; Riva-Palacio and Leisen, 2021). We also investigate its behavior in the generic setup of multiple comonotone replicates of independent components. Finally in Section 6 we perform a simulation study to

showcase the relevance of the index to conduct principled and fair model comparisons. In the Supplementary Material we describe our proof techniques and the underlying optimal transport problem, which we believe are of interest beyond the present setup with natural applications to the theory of partial differential equations and of Lévy processes.

2   Main result

Most nonparametric models are built on random structures taking values on spaces of measures. Among this large class, completely random measures (Kingman, 1967) stand out for their ability of combining analytical tractability with a large support. We recall that a random measure $\tilde{\mu}$ is *completely random* whenever its evaluations $\{\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)\}$ on pairwise disjoint sets are mutually independent random variables on $[0, +\infty)$. Here and after, we use the term *set* to indicate a Borel set on a generic Polish space, and the symbol $\sim$ to underline the randomness of the measure. Following Catalano et al. (2021), we refer to a *completely random vector* as its multivariate extension.

Definition 1. *A vector of random measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ is a completely random vector (CRV) if for any $n \geq 2$, and for any family of pairwise disjoint sets $\{A_1, \ldots, A_n\}$, the set-wise evaluations $\{\tilde{\boldsymbol{\mu}}(A_1), \ldots, \tilde{\boldsymbol{\mu}}(A_n)\}$ are mutually independent random vectors on $[0, +\infty)^d$.*

The definition remarkably entails that $\tilde{\boldsymbol{\mu}} = \sum_{i=1}^{\infty} (J_i^1, \ldots, J_i^d)\delta_{Y_i}$ are almost surely discrete measures with jumps $\{(J_i^1, \ldots, J_i^d)\}$ and common atoms $\{Y_i\}$ (up to a potential deterministic drift). CRVs often arise in Bayesian nonparametrics to model the interaction across distinct groups of observations. Prior specifications are typically based on CRVs without fixed atoms and with equal marginals, where the jumps and the atoms are independent (*homogeneity*) and every random measure has an infinite number of jumps on bounded sets (*infinite activity*). We focus on this class and restrict our attention to CRVs with finite second moments, that is $\mathbb{E}(\|\tilde{\boldsymbol{\mu}}(A)\|^2) < +\infty$ for every set A, where we have used the compact notation $\tilde{\boldsymbol{\mu}}(A) = (\tilde{\mu}_1(A), \ldots, \tilde{\mu}_d(A))$.

The goal of our work is to provide a tractable index of dependence for CRVs. Since these multivariate random quantities live in a non-Euclidean space, a natural way to define the index is by introducing a distance $\mathcal{D}$ between the laws of CRVs. First, we highlight two extreme dependence structures: (a) complete dependence (or comonotonicity) $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$, where $\tilde{\mu}_1 = \cdots = \tilde{\mu}_d$ almost surely, (b) independence $\tilde{\boldsymbol{\mu}}^{\perp}$, where $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_d\}$ are independent random measures. Then, we define an index $I_{\mathcal{D}}$ between 0 and 1 in terms of distance from complete dependence:

$$I_{\mathcal{D}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{\mathcal{D}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^2}{\sup_{\tilde{\boldsymbol{\mu}}'} \mathcal{D}(\tilde{\boldsymbol{\mu}}', \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^2}, \tag{1}$$

where the supremum is taken over all CRVs $\tilde{\boldsymbol{\mu}}'$ with the same marginal distributions as $\tilde{\boldsymbol{\mu}}$, that is, $\tilde{\mu}_i' = \tilde{\mu}_i$ in distribution for $i = 1, \ldots, d$. Here, and in the sequel, we use the notation $\mathcal{D}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})$ to indicate the distance between the laws of $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$. We observe that the non-degeneracy of distances entails that $I_{\mathcal{D}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ in distribution. In order to evaluate this index in practice, we need to find a tractable distance $\mathcal{D}$ and to use its geometry to find the supremum of the distance from complete dependence. The underlying intuition is that the supremum should be achieved under independence and that the distance from any other dependence structure should be strictly smaller. We are able to make this intuition rigorous by building $\mathcal{D}$ on the Wasserstein distance, as defined below in (3). This leads to our main result which we now state.

4

Theorem 1. *Let $\tilde{\boldsymbol{\mu}}$ be a homogeneous infinitely active CRV without fixed atoms, with equal marginals and finite second moments. The Wasserstein index of dependence $I_{\mathcal{W}}$ defined in* (1) *with $\mathcal{D}$ equal to* (3) *satisfies the following properties:*

(i) $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \in [0, 1]$;

(ii) $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ in distribution;

(iii) $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$ in distribution.

Remark 1. By dropping the infinite activity assumption, one can prove that (i) and (ii) continue to hold, whereas (iii) is replaced by (iii') $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$ in distribution. Details are provided in the proof of Theorem 1.

Remark 2. When defining the index (1), $\mathcal{D}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^2$ can be replaced by $\mathcal{D}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^p$ for any $p > 0$, without compromising any of our main findings. We fix $p = 2$ because of an intuitive linearity property on the space of measures highlighted in Remark 4 and due to the use of the extended Wasserstein distance of order 2, as will be clear in Section 3. This allows one to draw a parallel with linear correlation in Section 5.1.

Before considering technical aspects, let us first provide some intuition on the definition of a tractable distance $\mathcal{D}$ on the laws of CRVs. We observe that thanks to the independence on disjoint sets, the distribution of a CRV $\tilde{\boldsymbol{\mu}}$ is characterized by the set-wise evaluations $\{\tilde{\boldsymbol{\mu}}(A)\}$, where A spans over all sets. The definition of $\mathcal{D}$ is then achieved through two conceptual steps. Since $\tilde{\boldsymbol{\mu}}(A)$ takes values in $\mathbb{R}^d$, the first step consists in reducing the dimensionality of the problem by expressing the distance as a supremum over distances between finite-dimensional random objects, $\mathcal{D}(\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2) = \sup_A \mathcal{D}_d(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A))$, where $\mathcal{D}_d$ indicates a distance between the laws of set-wise evaluations. The second step consists in choosing a distance $\mathcal{D}_d$ that allows for numerical evaluations. To this end, it is worth underlining that the density and the cumulative distribution function of $\tilde{\boldsymbol{\mu}}(A)$ are usually intractable and its (multivariate infinitely divisible) distribution is specified through a Lévy intensity $\nu_A(\cdot) = \alpha(A)\nu(\cdot)$ for some base measure α with finite mass and some Lévy measure ν, which characterizes the distribution of $\tilde{\boldsymbol{\mu}}(A)$ through its Laplace transform. More specifically, let $\Omega_d = [0, +\infty)^d \setminus \{\mathbf{0}\}$. Then ν_A is the only measure on Ω_d that satisfies

$$- \log\big(\mathbb{E}\big(e^{-\boldsymbol{\lambda}\tilde{\boldsymbol{\mu}}(A)}\big)\big) = \int_{\Omega_d} (1 - e^{-\boldsymbol{\lambda}\boldsymbol{s}})\nu_A(\mathrm{d}\boldsymbol{s}), \tag{2}$$

for every $\boldsymbol{\lambda} \in [0, +\infty)^d$ and for every set A. For this reason, the most natural choice is to define the distance directly on the Lévy intensities.

When restricting to Lévy intensities, the distance should allow for informative comparisons between measures with (i) unbounded mass, which is always the case under infinite activity, (ii) different support, which is crucial in our context since the Lévy intensity under complete dependence has a degenerate support on the bisecting line, see Figure 1. We show that the extended Wassertstein distance $\mathcal{W}_*$ (Definition 2 below), introduced by Figalli and Gigli (2010) and specialized to Lévy measures in Guillen et al. (2019), remarkably satisfies both these properties. This leads to the study of the following distance between the laws of CRVs:

$$\mathcal{D}(\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2) = \sup_A \mathcal{W}_*(\nu_A^1, \nu_A^2), \tag{3}$$

where ν_A^1 and ν_A^2 are the Lévy intensities of the corresponding CRVs, uniquely defined by (2), and A spans over all sets.

3 Wasserstein distance between Lévy measures

In this section we introduce the extended Wasserstein distance between Lévy measures, highlight its relation to the *classical* Wasserstein distance between probability measures and state some key properties underlying the Wasserstein index of dependence. We refer to the supplement for additional results on this optimal transport problem and in particular for the dual formulation, which is pivotal in the proof of Theorem 1. We first introduce the classical framework and refer to Santambrogio (2015); Villani (2003); Panaretos and Zemel (2019) for exhaustive accounts.

The definition of Wasserstein distance starts with the notion of coupling. To this end, for a point $(\boldsymbol{s}, \boldsymbol{s}') \in \mathbb{R}^{2d}$, we denote by $\pi_1(\boldsymbol{s}, \boldsymbol{s}') = \boldsymbol{s} \in \mathbb{R}^d$ and $\pi_2(\boldsymbol{s}, \boldsymbol{s}') = \boldsymbol{s}' \in \mathbb{R}^d$ its projections. Moreover, if μ is a measure on $\mathbb{X}$ and $f : \mathbb{X} \to \mathbb{Y}$, $f_{\#}\mu$ stands for the pushforward of μ by f, that is, the measure on $\mathbb{Y}$ defined by $(f_{\#}\mu)(A) = \mu(f^{-1}(A))$. If ν^1, ν^2 are two probability measures on $\mathbb{R}^d$, a coupling γ is a probability measure on $\mathbb{R}^{2d}$ such that $\pi_{i\#}\gamma = \nu^i$ for $i = 1, 2$. Equivalently, it can be seen as a law of a random vector (X, Y) such that $X \sim \nu^1$ and $Y \sim \nu^2$. Let $\Gamma(\nu^1, \nu^2)$ be the set of couplings. If ν^1, ν^2 are probability measures on $\mathbb{R}^d$ with finite second moments, the classical Wasserstein distance is defined as

$$\mathcal{W}(\nu^1, \nu^2)^2 = \inf_{\gamma \in \Gamma(\nu^1, \nu^2)} \iint_{\mathbb{R}^{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\boldsymbol{s}, \boldsymbol{s}') = \inf_{X \sim \nu^1, Y \sim \nu^2} \mathbb{E}\left[\|X - Y\|^2\right]. \qquad (4)$$

There always exists a coupling γ^* that realizes the infimum in (4) and it is termed an optimal transport coupling. If there exists $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\gamma^* = (\mathrm{id}, T)_{\#}\nu^1$, T is termed optimal transport map and the Wasserstein distance can be conveniently expressed in the form $\mathcal{W}(\nu^1, \nu^2)^2 = \int \|\boldsymbol{s} - T(\boldsymbol{s})\|^2 \mathrm{d}\nu^1(\boldsymbol{s})$.

A priori this definition requires ν^1 and ν^2 to be probability measures, or at least to have finite and equal mass. Following Figalli and Gigli (2010) and Guillen et al. (2019), we now extend it to measures with different or infinite mass. Let $\Omega_d = [0, +\infty)^d \setminus \{\mathbf{0}\}$ and let $\mathcal{M}_2(\Omega_d)$ denote the set of positive Borel measures ν on Ω_d with finite second moments, that is

$$\mathcal{M}_2(\Omega_d) = \left\{ \nu \text{ positive Borel measure on } \Omega_d \text{ s.t. } M_2(\nu) = \int_{\Omega_d} \|\boldsymbol{s}\|^2 \mathrm{d}\nu(\boldsymbol{s}) < +\infty \right\}.$$

For a measure $\gamma \in \mathcal{M}_2(\Omega_{2d})$ the projections $\pi_{i\#}\gamma$ are measures on $[0, +\infty)^d$. In the following definition we need to consider their restrictions to Ω_d, which we denote as $\pi_{i\#}\gamma\big|_{\Omega_d}$.

Definition 2 (Extended Wasserstein distance). *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$ and let $\overline{\Gamma}(\nu^1, \nu^2)$ be the set of $\gamma \in \mathcal{M}_2(\Omega_{2d})$ such that $\pi_{1\#}\gamma\big|_{\Omega_d} = \nu^1$ and $\pi_{2\#}\gamma\big|_{\Omega_d} = \nu^2$. We define*

$$\mathcal{W}_*(\nu^1, \nu^2)^2 = \inf_{\gamma \in \overline{\Gamma}(\nu^1, \nu^2)} \iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\boldsymbol{s}, \boldsymbol{s}'). \qquad (5)$$

"Extended" couplings $\overline{\Gamma}(\nu^1, \nu^2)$ are needed to prove the existence of an optimal coupling, that is, to prove that the infimum is attained in (5). To give an intuition to the reader, couplings $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ are defined on Ω_{2d}, which is strictly larger than $\Omega_d \times \Omega_d$ as it includes $\{\mathbf{0}\} \times \Omega_d$ and $\Omega_d \times \{\mathbf{0}\}$. Moreover, the mass that γ puts on $\{\mathbf{0}\} \times \Omega_d$ only contributes to $\pi_{2\#}\gamma\big|_{\Omega_d}$ and not to $\pi_{1\#}\gamma\big|_{\Omega_d}$ because we look at the marginal $\pi_{1\#}\gamma$, a priori defined on $[0, +\infty)^d$ but we restrict it to Ω_d. Intuitively, the point $\{\mathbf{0}\}$ behaves like an infinite reservoir and sink of mass: $\gamma(\Omega_d \times \{\mathbf{0}\})$ (resp. $\gamma(\{\mathbf{0}\} \times \Omega_d)$) correspond to the mass exchanged by ν^1 (resp. ν^2) with this reservoir. We

define optimal transport couplings in $\overline{\Gamma}(\nu^1, \nu^2)$ and optimal transport maps $T : \Omega_d \to \Omega_d$ as for the classical Wasserstein distance. In the Supplementary Material we provide a characterization of extended optimal couplings in terms of c-cyclically monotone support, which establishes an interesting link with the couplings studied in de Valk and Segers (2019) in the context of tail limits of regularly varying probability measures.

Let us further explain the link between this distance and the work of Catalano et al. (2021). Indeed, in the latter the authors introduce a distance based only on the Wasserstein distance between setwise evaluations of the CRV, but also provide an upper bound in terms of a quantity depending on Lévy measures (see Theorem 5 therein). We recover this result by a simpler proof in (7) below, which brings to a better understanding of our distance. We start from the following alternative expression of our extended Wasserstein distance.

Proposition 2. *Let $\tilde{\boldsymbol{\mu}}^i$ satisfy the assumptions of Theorem 1 and let ν_A^i indicate its Lévy intensities as in (2), for $i = 1, 2$. Then for any set A,*

$$\mathcal{W}_*(\nu_A^1, \nu_A^2)^2 = \min_{(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)} \mathbb{E}(\|\hat{\boldsymbol{\mu}}^1(A) - \hat{\boldsymbol{\mu}}^2(A)\|^2),$$

where the minimum is taken over all homogeneous CRV $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ such that $\hat{\boldsymbol{\mu}}^i = \tilde{\boldsymbol{\mu}}^i$ in distribution for $i = 1, 2$.

This proposition immediately yields an upper bound on the Wasserstein distance between the laws of the setwise evaluations $\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)$:

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A))^2 \leq \mathcal{W}_*(\nu_A^1, \nu_A^2)^2, \tag{6}$$

as the quantity in the left hand side corresponds to a minimum taken among all couplings between the random vectors $\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)$, while the right hand side restricts to couplings that derive from the law of a joint CRV.

Remark 3. The inequality in (6) is generically strict, clearly implying that the distance in Catalano et al. (2021) and our proposal are different. Indeed, for a homogeneous CRV $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ and a set A, the coupling $(\hat{\boldsymbol{\mu}}^1(A), \hat{\boldsymbol{\mu}}^2(A))$ is not deterministic (that is, $\hat{\boldsymbol{\mu}}^2(A)$ is not a deterministic function of $\hat{\boldsymbol{\mu}}^1(A)$). Thus by Brenier's theorem (Villani, 2003, Theorem 2.12) it cannot be the optimal coupling between the law of $\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)$.

The next step to recover the result of Catalano et al. (2021) is a rewriting of the extended Wasserstein distance as a limit of classical Wasserstein distances, at least under the assumption of infinite mass.

Proposition 3. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$ be Lévy measures with finite second moments such that $\nu^1(\Omega_d) = \nu^2(\Omega_d) = +\infty$. For each $r > 0$, assume that ν_r^1, ν_r^2 are two measures with finite mass r such that for each set B, $\nu_r^i(B) \to \nu^i(B)$ increasingly as $r \to +\infty$. Then*

$$\mathcal{W}_*(\nu^1, \nu^2) = \lim_{r \to +\infty} \sqrt{r}\, \mathcal{W}\left(\frac{\nu_r^1}{r}, \frac{\nu_r^2}{r}\right).$$

Remark 4. The factor $\sqrt{r}$ comes from the $1/2$-homogeneity of the Wasserstein distance with respect to the mass. Similarly, one can easily see that $\mathcal{W}_*(a\nu^1, a\nu^2)^2 = a\mathcal{W}_*(\nu^1, \nu^2)^2$ for any $a > 0$. This has an important consequence in our context because of the homogeneity of CRVs. Indeed, the Lévy intensities satisfy $\nu_A(\cdot) = \alpha(A)\nu(\cdot)$ for some base measure α with total mass

$\bar{\alpha}$ and some Lévy measure ν. Thus $\mathcal{W}_*(\nu_A^1, \nu_A^2)^2 = \alpha(A)\mathcal{W}_*(\nu^1, \nu^2)^2$ and as an immediate, yet remarkable consequence its supremum is achieved on the total space. This implies that

$$\mathcal{D}(\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2)^2 = \bar{\alpha}\mathcal{W}_*(\nu^1, \nu^2)^2$$

and the base measure α only acts as a scaling factor through its total mass $\bar{\alpha}$. In particular, when normalizing the distance to obtain the index of dependence, the scaling factor cancels out (see (8) below) and the base measure does not impact the value of the index. This corresponds to a desirable intuitive property, since all the dependence is introduced at the level of the jumps, whose joint distribution does not depend on α.

Putting together bound (6) and Proposition 3, as well as the scaling of Remark 4, we obtain that for every set A,

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)) \le \sqrt{\alpha(A)} \lim_{r \to +\infty} \frac{1}{\sqrt{r}} \mathcal{W}\left(\frac{\nu_r^1}{r}, \frac{\nu_r^2}{r}\right), \tag{7}$$

which is nothing else than Theorem 5 in Catalano et al. (2021), but now the right hand side has a much neater interpretation.

The last key result that we state is that, similarly to the classical case, the optimal transport coupling for measures lying on the one-dimensional axis is the unique non-decreasing one. Moreover for atomless measures with infinite mass, as in our context, we also find the expression of the optimal transport map. Recall that $\Omega_1 = (0, +\infty)$. For a measure $\nu \in \mathcal{M}_2(\Omega_1)$ with finite second moment we define its tail integral $U_\nu : x \in \Omega_1 \mapsto \nu((x, +\infty))$ and its generalized inverse $U_\nu^{-1} : t \mapsto \inf\{x \ge 0 \ : \ U_\nu(x) \le t\}$. If U_ν is injective, it coincides with the usual inverse. Moreover, we denote by $\text{Leb}(\Omega_1)$ the Lebesgue measure on Ω_1.

Proposition 4. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_1)$ be Lévy measures with finite second moment and let γ be the restriction of $(U_{\nu^1}^{-1}, U_{\nu^2}^{-1})_{\#}\text{Leb}(\Omega_1)$ to Ω_2. Then $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ is the unique optimal transport coupling and*

$$\mathcal{W}_*(\nu^1, \nu^2)^2 = \int_0^{+\infty} (U_{\nu^1}^{-1}(s) - U_{\nu^2}^{-1}(s))^2 \, \mathrm{d}s.$$

Moreover, if ν^1 is atomless and $\nu^1(\Omega_1) \ge \nu^2(\Omega_1)$, $T(x) = U_{\nu^2}^{-1}(U_{\nu^1}(x))$ is an optimal transport map.

4 Evaluation of the index

In this section we provide some guidance on how to use the properties of the extended Wasserstein distance in Section 3 to evaluate the Wasserstein index of dependence $I_{\mathcal{W}}$.

First of all it is worth underlying that both extreme dependence structures of complete dependence ($\tilde{\boldsymbol{\mu}}^{\text{co}}$) and independence ($\tilde{\boldsymbol{\mu}}^{\perp}$), as defined in Section 2, are CRVs and therefore their law is characterized by Lévy intensities $\nu_A^{\text{co}} = \alpha(A)\nu^{\text{co}}$ and $\nu_A^{\perp} = \alpha(A)\nu^{\perp}$, respectively, where α is a finite measure, while ν^{co} and $\nu^{\perp}$ are Lévy measures. We recall that under complete dependence the Lévy measure $\tilde{\boldsymbol{\mu}}^{\text{co}}$ is concentrated on the bisecting line, while under independence the Lévy measure $\nu^{\perp}$ is supported on the axes, that is

$$\mathrm{d}\nu^{\text{co}}(\mathbf{s}) = \left(\prod_{i=2}^d \mathrm{d}\delta_{s_1}(s_i)\right) \mathrm{d}\overline{\nu}(s_1), \qquad \mathrm{d}\nu^{\perp}(\mathbf{s}) = \sum_{j=1}^d \left(\prod_{i \ne j} \mathrm{d}\delta_0(s_i) \, \mathrm{d}\overline{\nu}(s_j)\right),$$

8
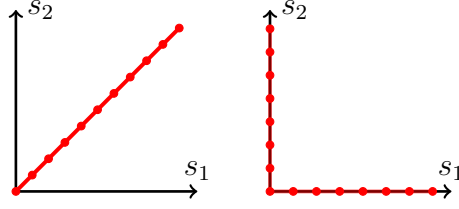
Figure 1: Support of ν^{co} and $\nu^\perp$ in Ω_2, respectively.

where δ is the Dirac measure and $\overline{\nu}$ is the Lévy measure of the marginal completely random measures. We refer to Figure 1 for intuition, and to Cont and Tankov (2004) and Catalano et al. (2021) for a proof in the context of multivariate Lévy processes and of completely random vectors, respectively.

Let now $\tilde{\boldsymbol{\mu}}$ be a homogeneous infinitely active CRV without fixed atoms, with equal marginals and finite second moments, whose dependence we wish to quantify. As recalled in (2), $\tilde{\boldsymbol{\mu}}$ is uniquely characterized by the Lévy intensities $\nu_A = \alpha(A)\nu$, where α is a finite measure and ν is a Lévy measure. Starting from (1), the homogeneity property highlighted in Remark 4 ensures that

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2}{\sup_{\nu'} \mathcal{W}_*(\nu', \nu^{\mathrm{co}})^2}, \tag{8}$$

where the supremum is taken over all Lévy measures on Ω_d with the same marginals as ν, that is, $\pi_{i\#}\nu = \pi_{i\#}\nu' = \overline{\nu}$ for $i = 1, \ldots, d$. The evaluation of $I_{\mathcal{W}}$ requires two steps: (i) to compute the numerator, we need to find an optimal extended coupling between ν and ν^{co}, so to have an integral expression for $\mathcal{W}_*(\nu, \nu^{\mathrm{co}})$; (ii) to compute the denominator, we have to find the supremum of $\mathcal{W}_*(\nu, \nu^{\mathrm{co}})$ over all possible dependence structures of ν, which involves solving a highly nontrivial max-min problem. The solutions to these two points are strongly intertwined. The integral expression for $\mathcal{W}_*(\nu, \nu^{\mathrm{co}})$ is provided

in Theorem 5 below and has two benefits: on the one hand it allows one to compute the numerator in explicit examples in the literature, and on the other it also provides the starting point to solve (ii). In the proof of Theorem 1 we show that the supremum is a maximum and it is achieved under independence, that is, $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) \leq \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$ with equality if and only if $\nu = \nu^\perp$. To this end, we heavily rely on the dual formulation of $\mathcal{W}_*$, whose details are provided in the supplementary material. Finally, to compute the denominator, we have to evaluate the distance between complete dependence and independence, which is done by resorting again to Theorem 5.

Remark 5. In principle, the Lévy measure ν corresponding to a CRV $\tilde{\boldsymbol{\mu}}$ is a positive Borel measure on Ω_d with finite second moments near the origin. Because of infinite activity, ν has infinite mass. Since $\tilde{\boldsymbol{\mu}}$ has fixed and equal marginal distributions, ν has d equal marginals, that is, $\pi_{i\#}\nu = \overline{\nu}$ for $i = 1, \ldots, d$, where $\overline{\nu}$ is a 1-dimensional Lévy measure on Ω_1. Since we restrict to $\tilde{\boldsymbol{\mu}}$ with finite second moments, $M_2(\overline{\nu}) = \int_{\Omega_1} s^2 \mathrm{d}\overline{\nu}(s)$ is finite and thus $\overline{\nu}$ belongs to $\mathcal{M}_2(\Omega_1)$. In particular, the Lévy measure $\nu \in \mathcal{M}_2(\Omega_d)$ has finite second moments.

Theorem 5. *Let* $\nu \in \mathcal{M}_2(\Omega_d)$ *be a Lévy measure with finite second moments and equal marginals* $\pi_{i\#}\nu = \overline{\nu}$ *for* $i = 1, \ldots, d$. *Denote by* $\nu^+ = \Sigma_{\#}\nu \in \mathcal{M}_2(\Omega_1)$, *where* $\Sigma(\mathbf{s}) = \sum_{i=1}^d s_i$. *Then*

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = 2d M_2(\overline{\nu}) - 2 \int_0^{+\infty} U_{\nu^+}^{-1}(s)\, U_{\overline{\nu}}^{-1}(s)\, \mathrm{d}s,$$

9

where $M_2(\bar{\nu}) = \int_{\Omega_1} s^2 \mathrm{d}\bar{\nu}(s)$. Moreover, when $\nu = \nu^\perp$,

$$\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2 = 2d\big(M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(d\, U_{\bar{\nu}}(s))\, \mathrm{d}\bar{\nu}(s)\big).$$

When ν^+ is atomless, Theorem 5 amounts to showing that $\mathbf{s} \mapsto (T(\sum_{i=1}^d s_i), \ldots, T(\sum_{i=1}^d s_i))$ is an optimal transport map between ν and ν^{co}, with $T(s) = U_{\bar{\nu}}^{-1}(U_{\nu^+}(s))$. Thanks to Proposition 4, T is the optimal transport map from ν^+ to $\bar{\nu}$. Thus, the optimal way to transport the mass of ν to ν^{co} first sends each point $\mathbf{s} \mapsto (\sum_{i=1}^d s_i, \ldots, \sum_{i=1}^d s_i)$, so to concentrate the mass onto the bisecting line. This is then optimally transported to ν^{co}, reducing to an optimal transport problem on a one-dimensional subspace of Ω_d and thus crucially ending up with a tractable computation. In short, we use the geometry of the support of the measure ν^{co} to find the explicit expression of the optimal transport map, whereas in the general case one has to solve a nonlinear partial differential equation for which there is no explicit solution, see Villani (2003, Chapter 4).

5 Examples

5.1 Additive models

Additive models first appeared in the Bayesian nonparametric literature to borrow information across distinct groups of observations (Müller et al., 2004; Lijoi et al., 2014; Lijoi and Nipoti, 2014). The dependence between random measures is introduced in a natural way through a superposition of independent components. In this section we find the corresponding Lévy measure and use it to evaluate the Wasserstein index of dependence in terms of the hyperparameter of the model. When restricting to 2-dimensional vectors, this brings to interesting links with linear correlation.

Let $\tilde{\xi}_0, \tilde{\xi}_1, \ldots, \tilde{\xi}_d$ be independent completely random measures whose Lévy measures satisfy $\nu_0 = z\bar{\nu}$ and $\nu_i = (1 - z)\bar{\nu}$, where $z \in [0, 1]$ and $\bar{\nu} \in \mathcal{M}_2(\Omega_1)$ is a fixed Lévy measure, for $i = 1, \ldots, d$. A CRV $\tilde{\boldsymbol{\mu}}$ is said to be *additive* or GM-dependent if its marginals satisfy $\tilde{\mu}_i = \tilde{\xi}_i + \tilde{\xi}_0$ in distribution, for $i = 1, \ldots, d$. The parameter z adjusts for dependence linearly with respect to the Lévy measures, reaching complete dependence $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ as $z \to 1$ and independence $\tilde{\boldsymbol{\mu}}^\perp$ as $z \to 0$.

Lemma 6. *The Lévy measure of an additive CRV is $\nu = z\nu^{\mathrm{co}} + (1 - z)\nu^\perp$.*

It follows that the Lévy measure of an additive CRV has mass both on the bisecting line and on the axes, differently weighted according to the parameter z. In particular, the marginals are not affected by z since $\pi_{i\#}\nu = z\pi_{i\#}\nu^{\mathrm{co}} + (1 - z)\pi_{i\#}\nu^\perp = \bar{\nu}$, for $i = 1, \ldots, d$. Let

$$U_{\nu_z^+}(s) = d(1 - z)U_{\bar{\nu}}(s) + zU_{\bar{\nu}}(sd^{-1}), \qquad \nu_z^+(s) = d(1 - z)\bar{\nu}(s) + zd^{-1}\bar{\nu}(sd^{-1}).$$

Proposition 7. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional additive CRV of parameter z such that $M_2(\bar{\nu}) = \int_{\Omega_1} s^2 \mathrm{d}\bar{\nu}(s) < +\infty$. Then $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq z$ and*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{dM_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(U_{\nu_z^+}(s))\nu_z^+(s)\, \mathrm{d}s}{dM_2(\bar{\nu}) - d\int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(dU_{\bar{\nu}}(s))\bar{\nu}(s)\, \mathrm{d}s}.$$
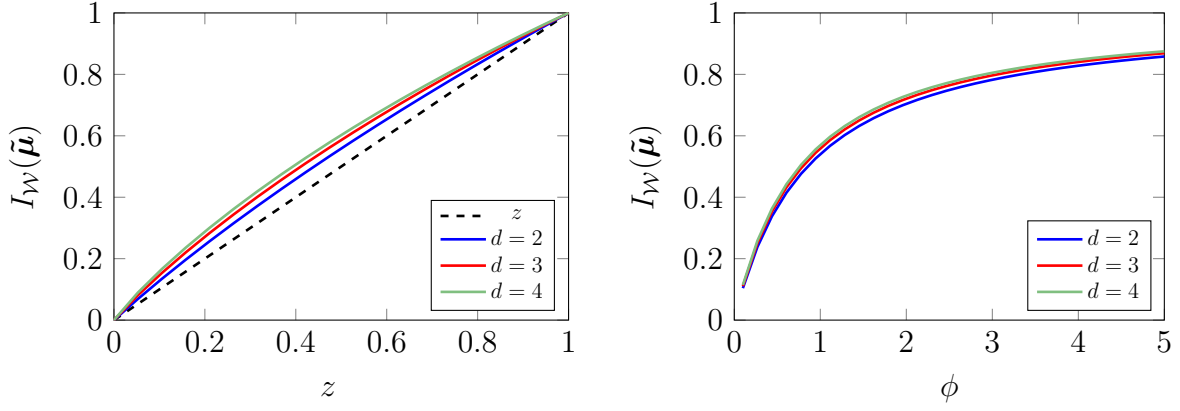
Figure 2: Left: $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ for $\tilde{\boldsymbol{\mu}}$ additive gamma CRV of parameter z and dimension d. Right: $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ for $\tilde{\boldsymbol{\mu}}$ gamma compound random vector of parameter ϕ and dimension d.

Proposition 7 provides the exact expression of the index of dependence and shows that it is always larger than the parameter z, for any choice of Lévy measure $\bar{\nu}$ and dimension d. When restricting to two groups of observations, $z = \mathrm{cor}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$, which does not depend on $A \in \mathcal{X}$. One proves this by using Proposition 6 and Campbell's theorem (see, e.g., (9.5.2) in Daley and Vere-Jones (2007)), so that

$$\mathrm{cov}(\tilde{\mu}_1(A), \tilde{\mu}_1(A)) = \alpha(A) \int_0^{+\infty} \int_0^{+\infty} s_1 s_2 \, \mathrm{d}\nu(s_1, s_2) = z\alpha(A)M_2(\bar{\nu})$$

$$= z \, \mathrm{var}(\mu_1(A))^{\frac{1}{2}} \mathrm{var}(\mu_1(A))^{\frac{1}{2}}.$$

Thus when $d = 2$, Proposition 7 guarantees that $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq \mathrm{cor}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$. In Figure 2 (left) we plot the value of $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ when the marginal is a gamma random measure and we see that the lower bound appears to be tight. This is a very desirable property of the index, since correlation is the most well-established measure of linear dependence between two random variables and additive CRVs introduce dependence linearly at the level of the Lévy measures.

5.2 Compound random measures

Compound random measures (Griffin and Leisen, 2017, 2018; Riva-Palacio and Leisen, 2021) provide a flexible way to model dependence between different families of completely random measures. A CRV $\tilde{\boldsymbol{\mu}}$ is a *compound random vector* if its Lévy density takes the form

$$\nu(\boldsymbol{s}) = \int_{(0,+\infty)} \frac{1}{u^d} h\left(\frac{s_1}{u}, \dots, \frac{s_d}{u}\right) \mathrm{d}\nu^*(u),$$

where h is a density function on $(0, +\infty)^d$ and ν^* is a Lévy measure. A widely used specification takes h the density of d independent gamma$(\phi, 1)$ random variables and $\nu^*(u) = (1 - u)^{\phi-1}u^{-1}\mathbb{1}_{(0,1)}(u)$, for $\phi > 0$. The marginals $\tilde{\mu}_i$ are then gamma completely random measures and that ϕ only accounts for dependence. Under these specifications $\tilde{\boldsymbol{\mu}}$ is a gamma compound random vector of parameter ϕ. Let

$$U_{\nu_\phi^+}(s) = \frac{1}{\Gamma(d\phi)} \int_0^1 \Gamma\left(d\phi, \frac{s}{u}\right) \frac{(1-u)^{\phi-1}}{u} \, \mathrm{d}u, \qquad \nu_\phi^+(s) = \frac{s^{d\phi-1}}{\Gamma(d\phi)} \int_0^1 e^{-\frac{s}{u}} \frac{(1-u)^{\phi-1}}{u^{d\phi+1}} \mathrm{d}u,$$

11

where $\Gamma(a, s) = \int_s^{+\infty} e^{-t} t^{a-1} dt$ is the upper incomplete gamma function. Moreover, we indicate by $E_1(s) = \Gamma(0, s)$ the exponential integral and by E_1^{-1} its inverse function.

Proposition 8. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional gamma compound random vector of parameter ϕ. Then,*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{d - \int_0^{+\infty} s\, E_1^{-1}(U_{\nu_\phi^+}(s))\, \nu_\phi^+(s)\, \mathrm{d}s}{d - d \int_0^{+\infty} E_1^{-1}(d E_1(s)) e^{-s}\, \mathrm{d}s}.$$

We use Proposition 8 to analyze the dependence structure induced by gamma compound random measures, as in Figure 2 (right). In particular, we observe that large values of ϕ favour highly dependent completely random measures and already with $\phi = 1$, $I_{\mathcal{W}}$ is slightly larger than 0.5. Finally, we observe that for both classes of models the dependence increases with the dimension d. Hence, one should take the dimension into account when fixing a value or an hyperprior for z and ϕ. This is an example of the use of the index for an informed prior specification of the dependence structure in presence of an arbitrary number of groups of observations.

5.3 Comonotone replicates of independent components

Additive and compound random measures are symmetric laws for a CRV $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}^1, \ldots, \tilde{\mu}^d)$, in the sense that $\tilde{\mu}^1, \ldots, \tilde{\mu}^d$ are exchangeable. This implies, for example, that all pairs of random measures have the same distribution and thus dependence structure. When this condition is not met, our index provides a valuable quantification of the overall dependence, which can not be grasped with pairwise comparisons. A prototype situation is the one of a d-dimensional CRV with m independent components and n comonotone replicates each, that is,

$$\tilde{\boldsymbol{\mu}} = (\overbrace{\tilde{\mu}^1, \ldots, \tilde{\mu}^1}^{n}, \overbrace{\tilde{\mu}^2, \ldots, \tilde{\mu}^2}^{n}, \ldots, \overbrace{\tilde{\mu}^m, \ldots, \tilde{\mu}^m}^{n}), \tag{9}$$

where $\tilde{\boldsymbol{\mu}}^\perp = (\tilde{\mu}^1, \tilde{\mu}^2, \ldots, \tilde{\mu}^m) \in \mathcal{M}_2(\Omega_m)$ is an independent CRV with equal marginals $\bar{\nu}$.

Proposition 9. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional CRV as in (9), with m independent components and n comonotone replicates such that $M_2(\bar{\nu}) = \int_{\Omega_1} s^2 \mathrm{d}\bar{\nu}(s) < +\infty$. Then,*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(m U_{\bar{\nu}}(s)) \bar{\nu}(s)\, \mathrm{d}s}{M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(d U_{\bar{\nu}}(s)) \bar{\nu}(s)\, \mathrm{d}s}.$$

The previous result has the merit of reducing the evaluation of the Wasserstein distance to a 1-dimensional integral, and this quantity may be easily evaluated numerically also in presence of a large number of groups. Moreover, its analytical expression can be studied as the number of groups $d = nm$ diverges, which can happen if the number either of independent components or of comonotone replicates increases.

Proposition 10. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional completely random vector as in (9) with m independent components and n comonotone replicates such that $M_2(\bar{\nu}) = \int_{\Omega_1} s^2 \mathrm{d}\bar{\nu}(s) < +\infty$. If n is fixed and $m \to +\infty$,*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \to 0$$

monotonically from above. If m is fixed and $n \to +\infty$,

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \to \frac{\int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(m U_{\bar{\nu}}(s)) \bar{\nu}(s)\, \mathrm{d}s}{M_2(\bar{\nu})}$$
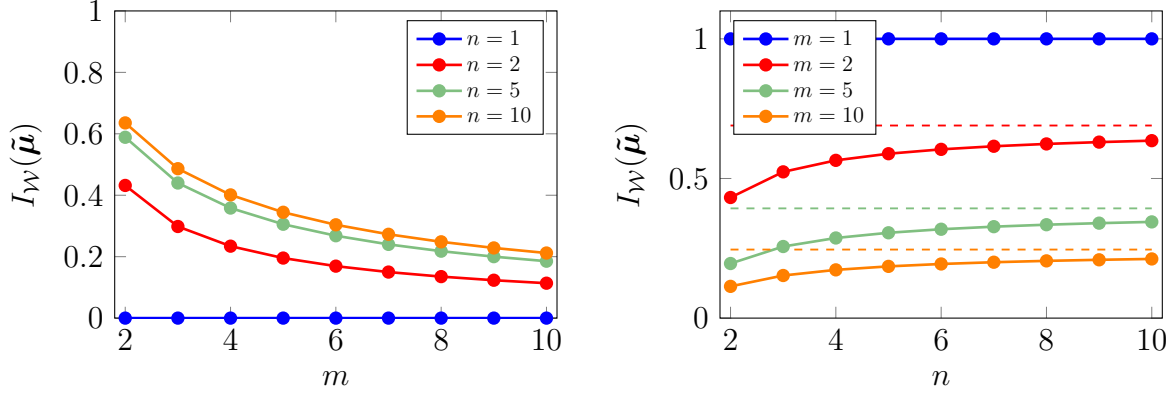
monotonically from below.

Figure 3: Evaluation of the index for a CRV with m independent components and n comonotone replicates, in the case of gamma marginals. The dashed lines are the limits in Proposition 10.

Proposition 10 shows that if the number of independent component diverges, the index goes to zero, whereas if the number of comonotone replicates diverges, the index increases and converges to a quantity that depends on the number m of independent components. In particular, if $m = 1$ the index is equal to 1 since in such case,

$$\int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(m U_{\bar{\nu}}(s))\bar{\nu}(s)\,\mathrm{d}s = \int_0^{+\infty} s^2 \bar{\nu}(s)\,\mathrm{d}s = M_2(\bar{\nu}).$$

These limiting behaviors confirm what one would intuitively expect and actually provide further evidence of the principled nature of the proposed index of dependence. Figure 3 further illustrates our findings by specializing them to the case of gamma marginals.

6 Model comparison

The proposed index has another important merit: by tuning the prior parameters of different models to achieve the same value of the dependence index, one can design a principled comparison of their inferential properties. While there is a multitude of dependent priors in the literature, a tool for matching their level of dependence a priori was still missing, preventing a fair comparison of their posterior performance under different scenarios. To illustrate this point, in Figure 4, we compare the dependence structure of an additive gamma completely random vector of parameter z with a gamma compound random vector of parameter ϕ, when $d = 3$. We associate the values of $I_{\mathcal{W}}$ to the corresponding parameter, z and ϕ respectively. For instance, the same level of high dependence, say 0.8, corresponds to setting $z = 0.75$ for the additive model and $\phi = 3$ for the compound random vector. It should be clear that a fair posterior comparison requires to match their a priori strength of dependence. Moreover, note that $I_{\mathcal{W}}$ grows almost linearly with z and non-linearly with ϕ. This may provide valuable guidance on the choice of hyperpriors for the parameters in the model: in the first case a uniform prior on z will imply a roughly uniform prior on the dependence structure, whereas with compound random measures standard priors on $[0, +\infty)$, unless carefully parametrized, would implicitly favor highly correlated marginals.
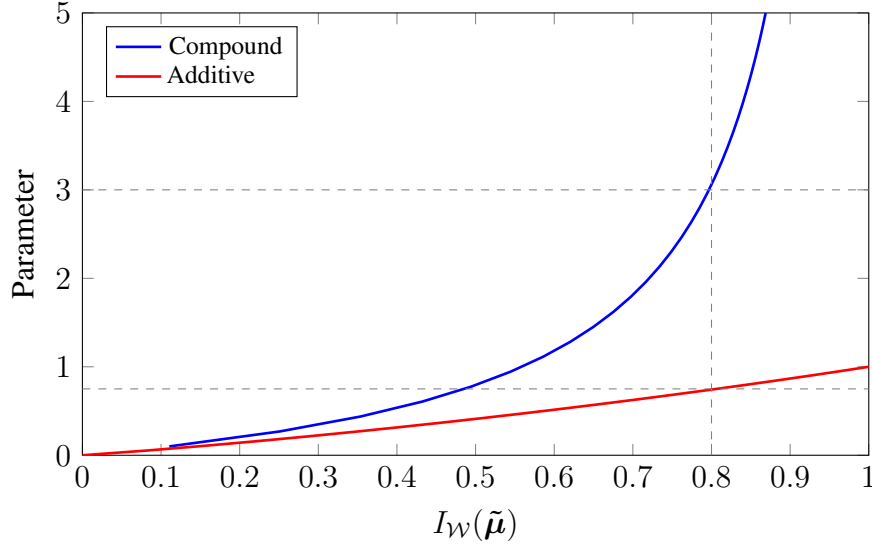
Figure 4: Parameter $\phi(z)$ of a compound (additive) completely random vector when $d = 3$, as $I_{\mathcal{W}}$ varies.

We now provide an illustration of how the choice of dependence structure impacts posterior inference, supporting our understanding of these models and highlighting the relevance of dependence matching when performing model comparisons. Simulations are performed using the BNPmix package (Corradin et al., 2021) in R (R Core Team, 2022), the POT package (Flamary et al., 2021) in Python (Van Rossum and Drake, 2009) and code kindly made available by Riccardo Corradin.

A highly popular use of dependent random measures is to model dependent random densities as location-scale Gaussian mixtures over the normalized random measures. These models allow to simultaneously perform joint density estimation and cluster analysis within and across populations; see Müller et al. (2015); Foti and Williamson (2015); Quintana et al. (2022) and references therein. For simplicity we focus on the case of two groups of observations, though the analysis can be readily extended to an arbitrary number of groups. Let $(X_{1,j_1})_{j_1 \geq 1}$ and $(X_{2,j_2})_{j_2 \geq 1}$ be two sequences of random variables such that

$$(X_{1,j_1}, X_{2,j_2})|(\tilde{\mu}_1, \tilde{\mu}_2) \overset{\text{iid}}{\sim} \tilde{f}_1 \times \tilde{f}_2$$

where $\tilde{f}_1$ and $\tilde{f}_2$ are location-scale Gaussian mixtures, that is,

$$\tilde{f}_i(y) = \int_{\mathbb{R} \times (0,+\infty)} \mathcal{N}(y; m, \sigma^2) \, \mathrm{d}\tilde{p}_i(m, \sigma^2)$$

with $\mathbb{X} = \mathbb{R} \times (0, +\infty)$, $\tilde{p}_i = \tilde{\mu}_i / \tilde{\mu}_i(\mathbb{X})$ are random probability measures, and $\mathcal{N}(\cdot; m, \sigma^2)$ indicates the density of a Gaussian distribution $\mathcal{N}(m, \sigma^2)$ with mean m and variance σ^2.

We consider two different priors on the vector of dependent random measures $(\tilde{\mu}_1, \tilde{\mu}_2)$, namely an additive and a compound CRV. In both cases we keep the same specification of the marginals random measures, which are equal to gamma completely random measures with a normal-inverse gamma base probability measure, i.e., α has density

$$\alpha(m, \sigma^2) = \mathcal{N}\left(m; m_0, \frac{\sigma^2}{k_0}\right) \times \text{InvGamma}(\sigma^2; a_0, b_0),$$

14

with hyperparameters $m_0 \in \mathbb{R}$ and $k_0, a_0, b_0 > 0$. Given the data $\boldsymbol{X}_1 = (X_{1,1}, \ldots, X_{1,n_1})$ from the first group and $\boldsymbol{X}_2 = (X_{2,1}, \ldots, X_{2,n_2})$ from the second, each specification yields an estimate for the random densities a posteriori, which we informally indicate as $(\tilde{f}_1^{\mathrm{a}}, \tilde{f}_2^{\mathrm{a}})|$data when we use an additive prior and $(\tilde{f}_1^{\mathrm{c}}, \tilde{f}_2^{\mathrm{c}})|$data when we use a compound prior. We consider two different scenarios for the data generating process:

a. The observations in each group are i.i.d. draws from a mixture of two Gaussians and one mixture component is shared between groups, namely, $\boldsymbol{X}_1 \stackrel{\mathrm{iid}}{\sim} 0.5\,\mathcal{N}(-5, 1) + 0.5\,\mathcal{N}(0, 1)$ and $\boldsymbol{X}_2 \stackrel{\mathrm{iid}}{\sim} 0.5\,\mathcal{N}(0, 1) + 0.5\,\mathcal{N}(5, 1)$.

b. The observations in each group are i.i.d. draws from a Gaussian distribution, but the two Gaussian distributions have different means, namely, $\boldsymbol{X}_1 \stackrel{\mathrm{iid}}{\sim} \mathcal{N}(-1, 2)$ and $\boldsymbol{X}_2 \stackrel{\mathrm{iid}}{\sim} \mathcal{N}(1, 2)$.

Since the borrowing of information is particularly useful in presence of unbalanced groups of observations, for both scenarios the first group has many more observations ($n_1 \in \{100, 200\}$) than the second ($n_2 = 10$). Our analysis proceeds as follows: i) We consider three different values of the index, namely $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \in \{0.1, 0.5, 0.9\}$, which correspond to the situation of almost independence, intermediate dependence and almost exchangeability, respectively. ii) We find the hyperparameters for the additive CRV and the compound random measures matching these values of the index, namely $z \in \{0.08, 0.44, 0.88\}$ for the former and $\phi \in \{0.1, 0.85, 8\}$ for the latter. iii) For both scenarios a. and b., for each value of the index, for both additive and compound random measures, we estimate the densities of the two groups a posteriori. Figure 5 displays the graphical output for scenario a. while Figure 6 for scenario b. We observe that in both scenarios the densities of the two groups differ the most when close to independence ($I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0.1$) and are more similar when close to exchangeability ($I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0.9$). There clearly are differences in the estimates between the two nonparametric models that are due to the specific amount of dependence rather than to the chosen prior.

This qualitative intuition can also be confirmed quantitively in the following way. Since for both scenarios a. and b. the first group has many more observations than the second, as the index varies we correctly observe a greater impact on the estimation of the second group. For the sake of compactness we thus focus on the estimation of the density of the second group, which is more interesting. We estimate the Wasserstein distance between the two mean posterior densities $\mathcal{W}(\mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\mathrm{data}), \mathbb{E}(\tilde{f}_2^{\mathrm{c}}|\mathrm{data}))$ as the index of dependence varies in both models. The heatmap in Figure 7 shows that in both scenarios a. and b. the estimates tend to be closer for the same value of the index than for different ones, i.e., the values near the diagonal tend to be smaller than the values far from the diagonal.

Finally we push the comparison even further and show that the distance between two additive models with different value for the index of dependence is bigger than the distance between an additive and a compound with the same index of dependence. This is pictured in the heatmap in Figure 8, where on the antidiagonal one finds $\mathcal{W}(\mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\mathrm{data}), \mathbb{E}(\tilde{f}_2^{\mathrm{c}}|\mathrm{data}))$ with the same index of dependence, while on the off-antidiagonal one finds $\mathcal{W}(\mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\mathrm{data}), \mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\mathrm{data}))$ with different index of dependence. This is repeated for both scenarios a. and b. Similar results can be obtained for compound random measures.

Figure 5: Mean posterior densities with additive random measures (left) and compound random measures (right) for three different values of the index (0.1, 0.5, 0.9, from top to bottom). Both models have gamma marginals with normal-inverse gamma base measure of parameters $m_0 = 0, k_0 = 0.1, a_0 = 2, b_0 = 1$. Group 1 has $n_1 = 100$ observations and Group 2 has $n_2 = 10$ observations, according to scenario a.
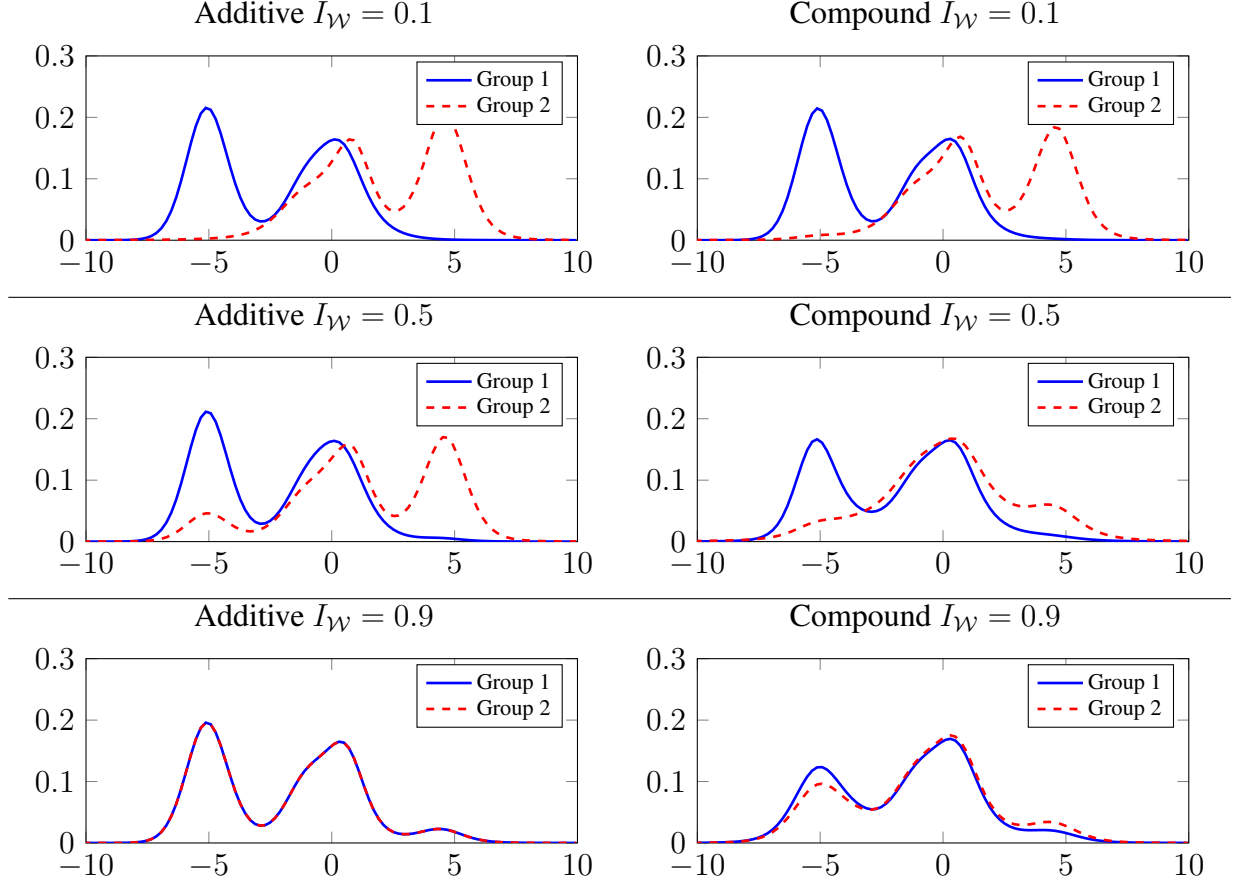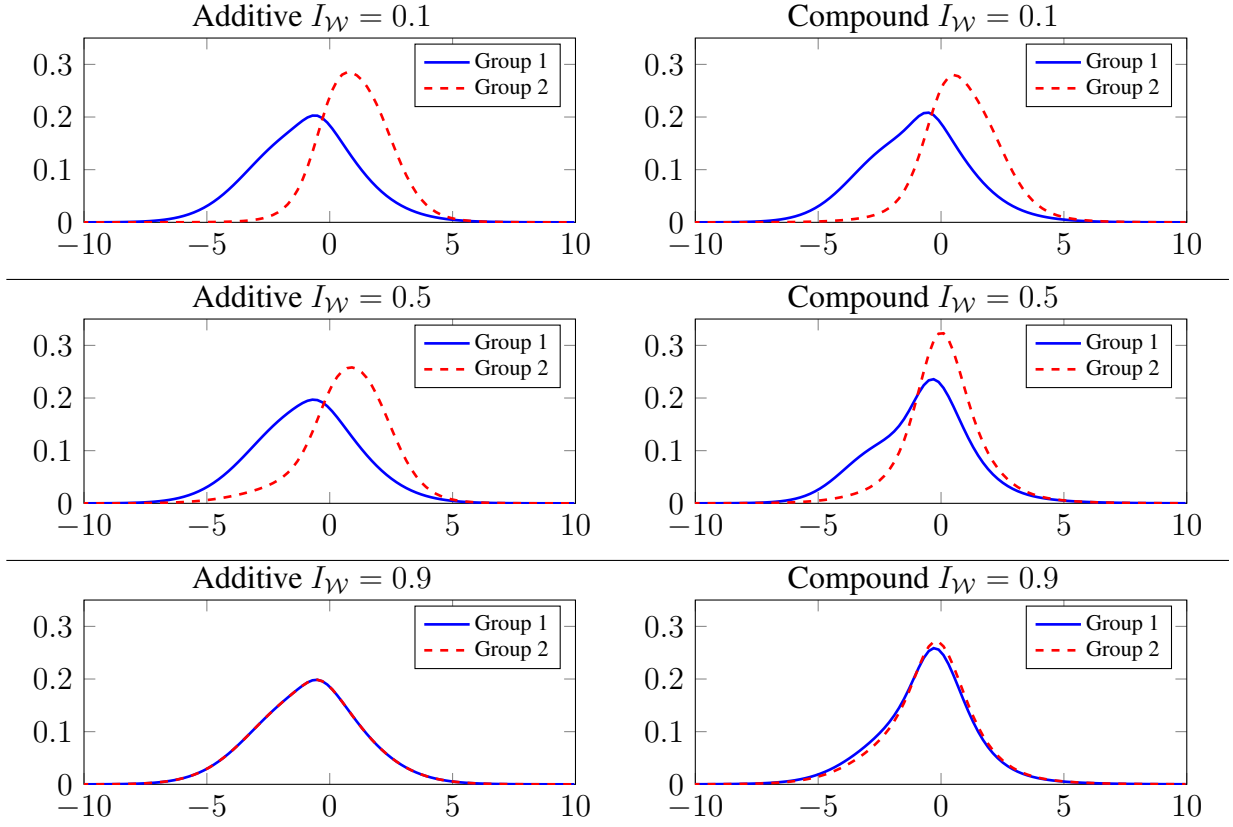
Figure 6: Mean posterior densities with additive random measures (left) and compound random measures (right) for three different values of the index (0.1, 0.5, 0.9, from top to bottom). Both models have gamma marginals with normal-inverse gamma base measure of parameters $m_0 = 0, k_0 = 1, a_0 = 2, b_0 = 1$. Group 1 has $n_1 = 200$ observations and Group 2 has $n_2 = 10$ observations, according to scenario b.

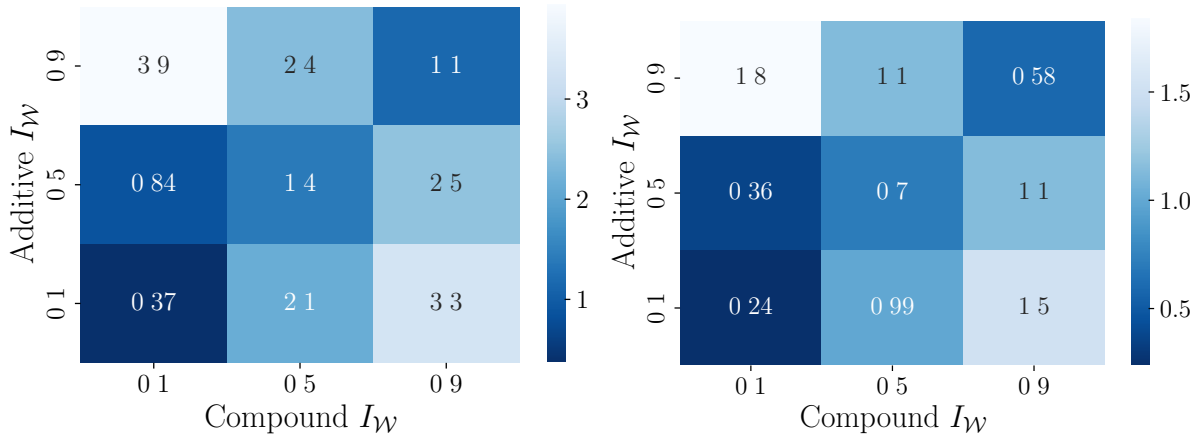Figure 7: Estimated Wasserstein distance $\mathcal{W}(\mathbb{E}(\tilde{f}_2^a|\text{data}), \mathbb{E}(\tilde{f}_2^c|\text{data}))$ between the mean posterior densities of additive and compound random measures for three different values of the index (0.1, 0.5, 0.9). The setups and data are the ones of Figure 5 (left) and of Figure 6 (right).
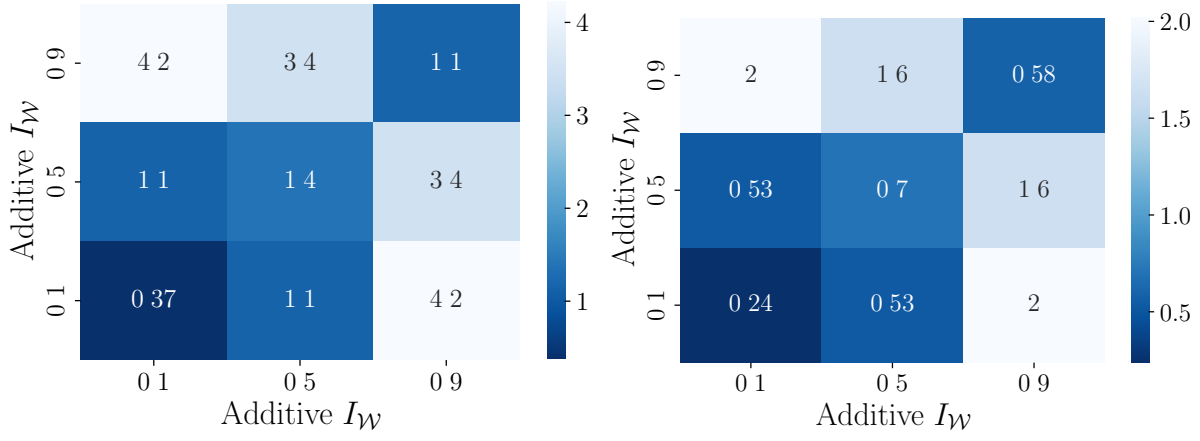
Figure 8: Antidiagonal: Estimated Wasserstein distance $\mathcal{W}(\mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\text{data}), \mathbb{E}(\tilde{f}_2^{\mathrm{c}}|\text{data}))$ between the mean posterior densities of the additive and compound models with the same value of the index, taking values in $(0.1, 0.5, 0.9)$. Off-antidiagonal: Estimated Wasserstein distance $\mathcal{W}(\mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\text{data}), \mathbb{E}(\tilde{f}_2^{\mathrm{a}}|\text{data}))$ between the mean posterior densities of the additive model for different values of the index. The setups are the same as in Figure 5 (left) and in Figure 6 (right).

SUPPLEMENTARY MATERIAL

The Supplementary Material contains our proof techniques and the underlying optimal transport problem, which we believe are of interest beyond the present setup with natural applications to the theory of partial differential equations and of Lévy processes.

ACKNOWLEDGEMENTS

References

Camerlenghi, F., A. Lijoi, P. Orbanz, and I. Prünster (2019). Distribution theory for hierarchical processes. *The Annals of Statistics 47*, 67–92.

Catalano, M., A. Lijoi, and I. Prünster (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *The Annals of Statistics 49*, 2916–2947.

Cont, R. and P. Tankov (2004). *Financial Modeling with Jump Processes*. Chapman & Hall/CRC.

Corradin, R., A. Canale, and B. Nipoti (2021). BNPmix: An R package for Bayesian nonparametric modeling via Pitman-Yor mixtures. *Journal of Statistical Software 100*, 1–33.

Daley, D. and D. Vere-Jones (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer New York.

de Valk, C. and J. Segers (2019). Tails of optimal transport plans for regularly varying probability measures. *arXiv: https://arxiv.org/abs/1811.12061*.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability 2*, 183–201.

Dykstra, R. L. and P. Laud (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics 9*, 356–367.

Figalli, A. and N. Gigli (2010). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées 94*, 107–130.

Flamary, R., N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer (2021). POT: Python Optimal Transport. *Journal of Machine Learning Research 22*(78), 1–8.

Foti, N. J. and S. A. Williamson (2015). A survey of non-exchangeable priors for bayesian nonparametric models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*(2), 359–371.

Griffin, J. E. and F. Leisen (2017). Compound random measures and their use in Bayesian nonparametrics. *Journal of the Royal Statistical Society Series B 79*, 525–545.

Griffin, J. E. and F. Leisen (2018). Modelling and Computation Using NCoRM Mixtures for Density Regression. *Bayesian Analysis 13*, 897 – 916.

Guillen, N., C. Mou, and A. Święch (2019). Coupling Lévy measures and comparison principles for viscosity solutions. *Transactions of the American Mathematical Society 372*, 7327–70.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics 18*, 1259–1294.

James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics 33*, 1771–1799.

Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics 21*, 59–78.

Lijoi, A. and B. Nipoti (2014). A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association 109*, 802–814.

Lijoi, A., B. Nipoti, and I. Prünster (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli 20*, 1260–1291.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics 12*, 351–357.

MacEachern, S. N. (1999). Dependent nonparametric processes. *In ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

MacEachern, S. N. (2000). Dependent Dirichlet processes. *Tech. Report*, Ohio State University.

Mordant, G. and J. Segers (2022). Measuring dependence between random vectors via optimal transport. *Journal of Multivariate Analysis 189*, 104912.

Müller, P., F. Quintana, A. Jara, and T. Hanson (2015). *Bayesian Nonparametric Data Analysis*. Springer Series in Statistics. Springer International Publishing.

Müller, P., F. Quintana, and G. Rosner (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society Series B 66*, 735–749.

Nguyen, X. (2016). Borrowing strengh in hierarchical bayes: Posterior concentration of the dirichlet base measure. *Bernoulli 22*, 1535–1571.

Nies, T. G., T. Staudt, and A. Munk (2021). Transport dependency: Optimal transport based dependency measures. *arXiv: https://doi.org/10.48550/arXiv.1811.12061*.

Panaretos, V. M. and Y. Zemel (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Applications. 6*, 405–431.

Quintana, F. A., P. Müller, A. Jara, and S. N. MacEachern (2022). The dependent Dirichlet process and related models. *Statistical Science 37*, 24–41.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Regazzini, E., A. Lijoi, and I. Prünster (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics 31*, 560–585.

Riva-Palacio, A. and F. Leisen (2021). Compound vectors of subordinators and their associated positive Lévy copulas. *Journal of Multivariate Analysis 183*, 104728.

Santambrogio, F. (2015). *Optimal transport for applied mathematicians*. Birkhäuser, NY.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Society.

Supplementary Material for "A Wasserstein index of dependence for random measures"

Marta Catalano

Department of Economics and Finance, Luiss University

and

Hugo Lavenant, Antonio Lijoi, Igor Prünster

Department of Decision Sciences and BIDSA, Bocconi University

Abstract

The Supplementary Material is organized as follows. In Section 1 we state some preliminary notions about the extended Wasserstein distance. We prove the results of the main manuscript in a slightly rearranged order to ease and streamline the exposition of the proofs. In particular, Sections 2, 3, 4, 5 contain the proofs of the results of, respectively, Sections 3, 4, 1, 5 of the main manuscript. To ease cross-reading between the main manuscript and the supplement, here we use a prefix SM for the numbering of results and definitions (e.g., Proposition SM1, Equation (SM1)).

1 Preliminaries

In this section we state some preliminary results on the extended Wasserstein distance that are needed for the proofs in the next sections. The main focus is on the notion of c-cyclically monotone set and on the dual formulation of the extended Wasserstein distance.

We recall that $\Omega_d = [0, +\infty)^d \setminus \{\mathbf{0}\}$ and $\mathcal{M}_2(\Omega_d)$ denotes the set of positive Borel measures with finite second moment. Given $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$, we define the set $\overline{\Gamma}(\nu^1, \nu^2)$ of (extended) couplings $\gamma \in \mathcal{M}_2(\Omega_{2d})$ such that $\pi_{1\#}\gamma\big|_{\Omega_d} = \nu^1$ and $\pi_{2\#}\gamma\big|_{\Omega_d} = \nu^2$. Following Figalli and Gigli (2010), we define the extended Wasserstein distance as follows.

Definition SM1. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$. We define the extended Wasserstein distance as*

$$\mathcal{W}_*(\nu^1, \nu^2)^2 = \inf_{\gamma \in \overline{\Gamma}(\nu^1, \nu^2)} \iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}'). \tag{SM1}$$

A coupling γ^* that realizes the infimum in (SM1) *is termed an optimal transport coupling.*

Proposition SM1 (Theorem A.5 in Guillen et al. (2019))**.** *If $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$, there exists at least one optimal transport coupling.*

Note in particular that we do not need the total mass of ν^1 to match the one of ν^2 for this result to hold. If there exists $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\gamma^* = (\mathrm{id}, T)_{\#}\nu^1$, T is termed optimal transport map between ν^1 and ν^2. Similarly to the classical optimal transport theory, c-cyclically monotone sets may be used to characterize optimal couplings.

1

Definition SM2. *A set $\mathcal{A} \subset [0, +\infty)^d \times [0, +\infty)^d$ is c-cyclically monotone if for every $N \in \mathbb{N}$, for every finite family of points $(\mathbf{s}_1, \mathbf{s}_1'), \ldots, (\mathbf{s}_N, \mathbf{s}_N') \in \mathcal{A}$ and for every permutation σ of $\{1, 2, \ldots, N\}$ there holds*

$$\sum_{i=1}^{N} \|\mathbf{s}_i - \mathbf{s}_i'\|^2 \leq \sum_{i=1}^{N} \|\mathbf{s}_i - \mathbf{s}_{\sigma(i)}'\|^2.$$

A proof of the following result is provided in Figalli and Gigli (2010, Proposition 2.3) and in Guillen et al. (2019, Theorem A.13). We recall that the support of a measure ν on Ω_d is the smallest closed set of $\overline{\Omega_d} = [0, +\infty)^d$ on which it is concentrated, that is, $\text{supp}(\nu) = \cap\{C \subset [0, +\infty)^d$ closed set such that $\nu(C^c \cap \Omega_d) = 0\}$.

Proposition SM2. *A coupling $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ is optimal if and only if $\text{supp}(\gamma) \cup \{\mathbf{0}\}$ is c-cyclically monotone.*

Proof. We apply Guillen et al. (2019, Theorem A.13), noticing that the set $\mathcal{K}$ coincides with $\mathbb{R}^{2d}$ in this case. $\square$

The next result shows that c-cyclical monotonicity is preserved under weak* convergence. We recall that $\mu_n \to \mu$ weakly* if and only if $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ for every $f \in C_c(\Omega_d)$ continuous function with compact support. This will be used in the proof of Proposition 3 in the next section.

Lemma SM3. *Let $(\gamma_n)_{n \in \mathbb{N}}$ a sequence of non-negative measures on Ω_d and assume that it weakly* converges to γ. If $\text{supp}(\gamma_n)$ is c-cyclically monotone, then so is $\text{supp}(\gamma)$.*

Proof. First, we argue that we can restrict to compact sets. As it is clear from the definition, a set $A \subset \Omega_d$ is c-cyclically monotone if and only if for any compact set $K \subset \Omega_d$, $K \cap A$ is c-cyclically monotone. Let us fix $K \subset \Omega_d$. As argued as in the proof of Santambrogio (2015, Theorem 1.50), since $\gamma_n|_K$ converges weakly* to $\gamma|_K$, $\text{supp}(\gamma_n|_K)$ converges in the Haussdorff topology to $\text{supp}(\gamma|_K)$. Since c-cyclically monotonicity is preserved under Hausdorff limit, we deduce that $\text{supp}(\gamma|_K)$ is c-cyclically monotone. As K is arbitrary, this concludes the proof. $\square$

We now introduce the dual formulation of $\mathcal{W}_*$. We first define a pair of *Kantorovich potentials* as any (φ, ψ) upper semi-continuous (hence measurable) functions $[0, +\infty)^d \to \mathbb{R} \cup \{-\infty\}$ such that $\varphi(\mathbf{0}) = \psi(\mathbf{0}) = 0$ and

$$\varphi(\mathbf{s}) + \psi(\mathbf{s}') \leq \frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2}. \tag{SM2}$$

We denote with $\mathcal{K}_D$ the set of pairs of Kantorovich potentials.

Theorem SM4 (Lemma 3.6 in Guillen et al. (2019)). *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$. Then*

$$\frac{1}{2} \mathcal{W}_*(\nu^1, \nu^2)^2 = \max_{(\varphi, \psi) \in \mathcal{K}_D} \left\{ \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s}) + \int_{\Omega_d} \psi(\mathbf{s}') \, \mathrm{d}\nu^2(\mathbf{s}') \right\}.$$

A pair $(\varphi, \psi) \in \mathcal{K}_D$ that realizes the supremum in the right hand side is termed an *optimal* pair of Kantorovich potentials. Note that Theorem SM4 entails that an optimal pair of Kantorovich potentials always exists. The pair of primal and dual problem gives an effective criterion to check the optimality of a solution, as summarized by the following result.

Proposition SM5. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$. Let $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ and (φ, ψ) a pair of Kantorovich potentials. Then the followings are equivalent:*

(i) The coupling γ is optimal and the pair (φ, ψ) is an optimal pair of Kantorovich potentials.

(ii) There holds

$$\iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s}) + \int_{\Omega_d} \psi(\mathbf{s}') \, \mathrm{d}\nu^2(\mathbf{s}'). \tag{SM3}$$

(iii) For every $(\mathbf{s}, \mathbf{s}') \in \mathrm{supp}(\gamma)$ there is equality in (SM2), that is,

$$\varphi(\mathbf{s}) + \psi(\mathbf{s}') = \frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2}. \tag{SM4}$$

Proof. That (i) and (ii) are equivalent is a straightforward consequence of Theorem SM4: indeed the left hand side of (SM3) is always larger than $\mathcal{W}_*(\nu^1, \nu^2)/2$, and the right hand side is smaller than this quantity by the aforementioned theorem. If there is equality if and only if everything coincides with $\mathcal{W}_*(\nu^1, \nu^2)^2/2$, which means that both γ as well as (φ, ψ) are optimal.

On the other hand, to show that (ii) and (iii) are equivalent, we write

$$\frac{1}{2} \iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') - \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s}) - \int_{\Omega_d} \psi(\mathbf{s}') \, \mathrm{d}\nu^2(\mathbf{s}')$$
$$= \iint_{\Omega_{2d}} \left(\frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2} - \varphi(\mathbf{s}) - \psi(\mathbf{s}') \right) \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}'),$$

thanks to the marginal property of γ and $\varphi(\mathbf{0}) = \psi(\mathbf{0}) = 0$. On the right hand side, we see that the integrand $\frac{\|\mathbf{s}-\mathbf{s}'\|^2}{2} - \varphi(\mathbf{s}) - \psi(\mathbf{s}')$ is non-negative and lower semi-continuous as $(\varphi, \psi) \in \mathcal{K}_D$, thus its integral with respect to γ vanishes if and only if it is identically 0 on the support of γ. $\square$

As we are in the case of quadratic cost, by the usual double convexification trick Villani (2003, Lemma 2.10), we can always assume the following form for the optimal Kantorovich potentials.

Lemma SM6. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$. There exists an optimal pair of Kantorovich potentials and a convex lower semi-continuous function u on Ω_d such that $u(\mathbf{0}) = u^*(\mathbf{0}) = 0$ and*

$$\varphi(\mathbf{s}) = \frac{1}{2} \sum_{i=1}^{d} s_i^2 - u(\mathbf{s}), \qquad \psi(\mathbf{s}) = \frac{1}{2} \sum_{i=1}^{d} s_i^2 - u^*(\mathbf{s}), \tag{SM5}$$

where u^* is the Legendre transform of u.

Proof. We start with $(\tilde{\varphi}, \tilde{\psi})$ an optimal pair of Kantorovich potentials. We then define

$$\varphi(\mathbf{s}) = \inf_{\mathbf{s}' \in \Omega_d} \frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2} - \tilde{\psi}(\mathbf{s}'), \qquad \psi(\mathbf{s}') = \inf_{\mathbf{s} \in \Omega_d} \frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2} - \varphi(\mathbf{s}). \tag{SM6}$$

As we started from $(\tilde{\varphi}, \tilde{\psi})$ admissible, there holds $(\varphi, \tilde{\psi}), (\varphi, \psi) \in \mathcal{K}_D$. Moreover, still because $(\tilde{\varphi}, \tilde{\psi})$ admissible there holds $\varphi \geq \tilde{\varphi}$ and $\psi \geq \tilde{\psi}$, thus (φ, ψ) is an optimal pair of Kantorovich potentials.

Then we define u and v (which will coincide with u^*) via (SM5), that is, $u(\mathbf{s}) = 1/2 \sum_i s_i^2 - \varphi(\mathbf{s})$ and $v(\mathbf{s}) = 1/2 \sum_i s_i^2 - \psi(\mathbf{s})$. Thus (SM6) reads

$$u(\mathbf{s}) = \sup_{\mathbf{s}' \in \Omega_d} \mathbf{s} \cdot \mathbf{s}' - \left(\frac{\|\mathbf{s}'\|^2}{2} - \tilde{\psi}(\mathbf{s}') \right), \qquad v(\mathbf{s}') = \sup_{\mathbf{s} \in \Omega_d} \mathbf{s} \cdot \mathbf{s}' - u(\mathbf{s}).$$

The second equation yields $v = u^*$ by definition; while the first one shows that u is a supremum of linear function, thus is convex. $\square$

*Remark SM*1. In particular, assuming that (φ, ψ) is given as above, the constraint (SM2) simply amounts to $u(\mathbf{s}) + u^*(\mathbf{s}') \geq \mathbf{s} \cdot \mathbf{s}'$, which holds by Young's inequality. Thus (SM4) is equivalent to $\mathbf{s}'$ belonging to the subdifferential of u evaluated in $\mathbf{s}$, see Villani (2003, Proposition 2.4). If u is differentiable at $\mathbf{s}$ with gradient $\nabla u(\mathbf{s})$, then $(\mathbf{s}, \mathbf{s}')$ satisfies (SM4) if and only if $\mathbf{s}' = \nabla u(\mathbf{s})$.

Contrary to the optimal transport problem in terms of couplings, the existence of a transport map is difficult, even under the assumption that ν^1 is absolutely continuous with respect to the Lebesgue measure on Ω_d. This is because one has to exclude the possibility of mass being sent from $\{\mathbf{0}\}$ to more than one point of the support of ν^2, which presents some difficulties. As we do not rely on the existence of transport maps in the sequel, we will not investigate this question. However we do prove the converse result, that is, that any transport map that is the gradient of a convex function is optimal.

Lemma SM7. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$ and let $T : \Omega_d \to \Omega_d$ be a measurable map such that $T_{\#}\nu_1 = \nu_2$ and $T = \nabla u$ for some lower semi-continuous convex function u. If $u(\mathbf{0}) = u^*(\mathbf{0}) = 0$ then T is an optimal transport map between ν_1 and ν_2.*

Proof. This follows the same proof as for classical optimal transport. Let $\gamma_T = (\mathrm{id}, T)_{\#}\nu^1$ be the coupling generated by T that belongs to $\overline{\Gamma}(\nu^1, \nu^2)$. We define (φ, ψ) according to (SM5). Since $u(\mathbf{0}) = u^*(\mathbf{0}) = 0$, $(\varphi, \psi) \in \mathcal{K}_D$. In addition, thanks to Remark SM1 we know that the triple $(\gamma_T, \varphi, \psi)$ satisfies (iii) of Proposition SM5. This is enough to ensure optimality of γ_T. $\qquad\square$

2 Proofs of Section 3: Wasserstein distance between Lévy measures

Proposition 2. *Let $\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2$ satisfy the assumptions of Theorem 1. Then for any set A*

$$\mathcal{W}_*(\nu_A^1, \nu_A^2)^2 = \min_{(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)} \mathbb{E}(\|\hat{\boldsymbol{\mu}}^1(A) - \hat{\boldsymbol{\mu}}^2(A)\|^2),$$

where the minimum is taken over all homogeneous CRV $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ such that $\hat{\boldsymbol{\mu}}^i = \tilde{\boldsymbol{\mu}}^i$ in distribution for $i = 1, 2$.

Proof. We will re-express the right hand side to show it coincides with the left hand side.
Take an homogeneous CRV $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ such that $\hat{\boldsymbol{\mu}}^i = \tilde{\boldsymbol{\mu}}^i$ in distribution for $i = 1, 2$. As $\tilde{\boldsymbol{\mu}}^1$ and $\tilde{\boldsymbol{\mu}}^2$ do not have atoms, the same holds for $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$. Thus it is characterized by a Lévy measure γ over Ω_{2d}, and its base measure must be α by the marginal conditions. The condition that $\hat{\boldsymbol{\mu}}^i = \tilde{\boldsymbol{\mu}}^i$ in distribution for $i = 1, 2$ directly translates in $\pi_{1\#}\gamma\big|_{\Omega_d} = \nu^1$ and $\pi_{2\#}\gamma\big|_{\Omega_d} = \nu^2$, that is, $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$. Here the fact that we restrict $\pi_{1\#}\gamma$ and $\pi_{2\#}\gamma$ on Ω_d has a very simple interpretation: the mass of $\pi_{1\#}\gamma$ and $\pi_{2\#}\gamma$ given to $\{\mathbf{0}\}$ corresponds to jumps of size 0, thus they do not modify the law of the corresponding CRV. Thus the infimum over $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ can be parametrized as an infimum over $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$.
Eventually, we can use Campbell's formula: for any $\gamma \in \mathcal{M}_2(\Omega_{2d})$, if $(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)$ is a CRV with base measure α and Lévy measure γ, then

$$\|\hat{\boldsymbol{\mu}}^1(A) - \hat{\boldsymbol{\mu}}^2(A)\|^2 = \iiint_{A \times \Omega_{2d}} \|\mathbf{s} - \mathbf{s}'\|^2 \, \mathrm{d}\alpha(x)\mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \alpha(A) \iint_{\Omega_{2d}} \|\mathbf{s} - \mathbf{s}'\|^2 \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}').$$

Thus we can rewrite our right hand side as

$$\min_{(\hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\mu}}^2)} \|\hat{\boldsymbol{\mu}}^1(A) - \hat{\boldsymbol{\mu}}^2(A)\|^2 = \alpha(A) \min_{\gamma \in \overline{\Gamma}(\nu^1, \nu^2)} \iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\boldsymbol{s}, \boldsymbol{s}'),$$

and the right hand side coincides with $\mathcal{W}_*(\nu_A^1, \nu_A^2)^2$, see Definition 1 and Remark 4. □

Proposition 3. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$ such that $\nu^1(\Omega_d) = \nu^2(\Omega_d) = +\infty$. For each $r > 0$, assume that ν_r^1, ν_r^2 are two measures with finite mass r such that for each set B, $\nu_r^i(B) \to \nu^i(B)$ increasingly as $r \to +\infty$. Then*

$$\mathcal{W}_*(\nu^1, \nu^2) = \lim_{r \to +\infty} \sqrt{r} \, \mathcal{W}\left(\frac{\nu_r^1}{r}, \frac{\nu_r^2}{r}\right).$$

Proof. The idea of the proof is to extract a converging subsequence from the classical optimal couplings $\gamma_n^* \in \overline{\Gamma}(\nu_n^1, \nu_n^2)$ and prove that its limit is an extended optimal coupling. To this end, we heavily rely on the characterization of optimality with c-cyclically monotone sets in Proposition SM2 and on its preservation under weak* limits in Lemma SM3. We then express the Wasserstein distance in terms of the optimal coupling and conclude by the monotone convergence theorem thanks to the increasing convergence.

First, we prove that there exists a subsequence $\{\gamma_{n_k}\}_k$ and $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ such that $\gamma_{n_k} \to \gamma$ according to the weak* convergence. For every $f \in C_c(\Omega_{2d})$,

$$\left| \int_{\Omega_{2d}} f(\mathbf{s}, \mathbf{s}') \, \mathrm{d}\gamma_n(\mathbf{s}, \mathbf{s}') \right| \leq |f|_\infty \frac{1}{\inf_{\mathrm{supp}(f)} \|\mathbf{s}\|^2 + \|\mathbf{s}'\|^2} \int_{\mathrm{supp}(f)} \|\mathbf{s}\|^2 + \|\mathbf{s}'\|^2 \, \mathrm{d}\gamma_n(\mathbf{s}, \mathbf{s}').$$

The integral on the right hand side is bounded from above by $M_2(\nu_n^1) + M_2(\nu_n^2) \leq M_2(\nu^1) + M_2(\nu^2)$ by the increasing convergence hypothesis. Since $\mathrm{supp}(f)$ does not contain the origin, $\int_{\Omega_{2d}} f(\mathbf{s}, \mathbf{s}') \, \mathrm{d}\gamma_n(\mathbf{s}, \mathbf{s}')$ is uniformly bounded in n and thus there exists a subsequence $\{\gamma_{n_k}\}_k$ such that $\lim_{k \to +\infty} \int f \, \mathrm{d}\gamma_{n_k}$ exists and it is finite. By diagonal extraction and separability of $C_c(\Omega)$, this holds for every $f \in C_c(\Omega)$. The functional $f \mapsto \lim_{k \to +\infty} \int f \, \mathrm{d}\gamma_{n_k}$ is positive and linear, so that by Riesz Representation Theorem there exists a regular Borel measure γ on Ω_{2d} such that, for every $f \in C_c(\Omega_{2d})$, $\lim_{k \to +\infty} \int_{\Omega_{2d}} f \, \mathrm{d}\gamma_{n_k} = \int_{\Omega_{2d}} f \, \mathrm{d}\gamma$. We need to prove that $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$, that is that for every $\phi \in C_c(\Omega_d)$, $\int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \phi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s})$ and $\int_{\Omega_{2d}} \phi(\mathbf{s}') \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \phi(\mathbf{s}') \, \mathrm{d}\nu^2(\mathbf{s}')$. The main obstacle in the proof is that we may not directly take the limit of the integral with respect to $(\mathbf{s}, \mathbf{s}') \mapsto \phi(\mathbf{s})$ because it is not continuous and compactly supported on Ω_{2d}. We therefore multiply $\phi(\mathbf{s})$ by a suitable function $g_m(\mathbf{s}')$ as follows. Let $A_m = [-m, m]^d$. By Urysohn's Lemma, for every m and every open set $V_m \supset A_m$, there exists $g_m \in C_c(\Omega_d)$ such that $0 \leq g_m \leq 1$, $g_m = 1$ on A_m and $g_m = 0$ on V_m^c. One can assume that $\{g_m\}_m$ are non-decreasing in the sense that for every $\mathbf{s}' \in \Omega_d$ and every $m \in \mathbb{N}$, $g_m(\mathbf{s}') \leq g_{m+1}(\mathbf{s}')$. First we observe that

$$\sup_k \left| \int_{\Omega_{2d}} \phi(\mathbf{s}) g_m(\mathbf{s}') \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}') - \int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}') \right|$$

$$\leq \sup_k \int_{\Omega_{2d}} |\phi(\mathbf{s})| \mathbb{1}_{A_m^c}(\mathbf{s}') \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}'),$$

5

which is bounded from above by $|\phi|_\infty m^{-2} \int_{\Omega_d} \|\mathbf{s}'\|^2 \, \mathrm{d}\nu^2(\mathbf{s}')$ and thus goes to zero as $m \to +\infty$. Moreover, by the monotone convergence theorem,

$$\int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \lim_{m \to +\infty} \int_{\Omega_{2d}} \phi(\mathbf{s}) g_m(\mathbf{s}') \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}').$$

Since $\mathrm{supp}(\phi) \times \mathrm{supp}(g_m)$ is a compact subset of Ω_{2d},

$$\lim_{k \to +\infty} \int_{\Omega_{2d}} \phi(\mathbf{s}) g_m(\mathbf{s}') \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}') = \int_{\Omega_{2d}} \phi(\mathbf{s}) g_m(\mathbf{s}') \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}').$$

By taking the limit as $m \to +\infty$ and switching the order of the limits thanks to the uniform convergence of the first integral, we find that $\int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \phi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s})$. We reason in the same way to prove the second marginal condition.

We now prove that γ^* is an optimal coupling. Since γ_n are optimal couplings for the standard Wasserstein distance, their support is c-cyclically monotone. Thus by Proposition SM3 the measure γ^* has a c-cyclically monotone support. Next we claim that $\{\mathbf{0}\} \in \mathrm{supp}(\gamma^*)$, so that by Proposition SM2 γ^* is an optimal coupling. Indeed, if $\{\mathbf{0}\} \notin \mathrm{supp}(\gamma^*)$, there exists $\epsilon > 0$ such that $\gamma^*([0, \epsilon)^{2d} \setminus \{\mathbf{0}\}) = 0$. Then, $+\infty = \gamma^*(\Omega_{2d}) \le \gamma^*([\epsilon, +\infty)^d \times [0, +\infty)^d) + \gamma^*([0, +\infty)^d \times [\epsilon, +\infty)^d) = \nu^1([\epsilon, +\infty)^d) + \nu^2([\epsilon, +\infty)^d) < +\infty$. Thus, there is a contradiction.

To conclude, it suffices to show that

$$\int_{\Omega_{2d}} \|\mathbf{s} - \mathbf{s}'\|^2 \, \mathrm{d}\gamma^*(\mathbf{s}, \mathbf{s}') = \liminf_{k \to +\infty} \int_{\Omega_{2d}} \|\mathbf{s} - \mathbf{s}'\|^2 \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}').$$

Once again, we may not directly pass to the limit because the support of $\|\mathbf{s} - \mathbf{s}'\|^2$ is not bounded. We reason similarly to the proof of the marginal constraints for γ^*, by introducing $A_m = [-m, m]^{2d} \setminus (-1/m, 1/m)^{2d}$ compact set of Ω_{2d} and interchanging the limits thanks to the bounded second moments of ν^1, ν^2. $\qquad\square$

We now move to the one-dimensional case, for which we closely follow the techniques of Santambrogio (2015, Chapter 2). We recall that the tail integral of a measure is defined as $U_\nu(x) = \nu((x, +\infty))$. Moreover, we define its generalized inverse as $U_\nu^{-1}(t) = \inf\{x \ge 0 : U_\nu(x) \le t\}$. The function U_ν is non-increasing and right-continuous. Moreover, one can easily check that for any $a \ge 0$,

$$U_\nu^{-1}(t) > a \iff t < U_\nu(a). \tag{SM7}$$

We start with the following Lemma, whose statement and proof are adapted from Santambrogio (2015, Lemma 2.4).

Lemma SM8. *Let $\nu \in \mathcal{M}_2(\Omega_1)$ atomless. Then $U_{\nu\#}\nu = \mathrm{Leb}((0, U_\nu(0)))$. Moreover, for every $r > 0$, the set $\{x > 0 : U_\nu(x) = r\}$ is ν-negligible.*

Proof. Note that $U_\nu(0)$ is nothing else than the total mass of ν. Also, note that ν atomless translates in U_ν continuous. Let us take $a \in (0, U_\nu(0))$. Then the set $\{x > 0 : U_\nu(x) < a\}$ is a interval of the form $(x_a, +\infty)$ with $U_\nu(x_a) = a$. Thus $\nu(\{x > 0 : U_\nu(x) < a\}) = U_\nu(x_a) = a$. The conclusion that $\nu(\{x > 0 : U_\nu(x) = r\}) = 0$ follows exactly the same line as Santambrogio (2015, Lemma 2.4): if it were not the case, then $U_{\nu\#}\nu$ would have an atom. $\qquad\square$

We can now prove the Proposition 4 of the main manuscript.

Proposition 4. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_1)$ and let γ be the restriction of $(U_{\nu^1}^{-1}, U_{\nu^2}^{-1})_\# \mathrm{Leb}(\Omega_1)$ to Ω_2. Then $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ is the unique optimal transport coupling and*

$$\mathcal{W}_*(\nu^1, \nu^2)^2 = \int_0^{+\infty} (U_{\nu^1}^{-1}(s) - U_{\nu^2}^{-1}(s))^2 \, \mathrm{d}s.$$

Moreover, if ν^1 is atomless and $\nu^1(\Omega_1) \geq \nu^2(\Omega_1)$, $T(x) = U_{\nu^2}^{-1}(U_{\nu^1}(x))$ is an optimal transport map.

Proof. Thanks to Proposition SM1 there exists an optimal coupling γ^*. As proved in Santambrogio (2015, Theorem 2.9), since the support of γ^* is c-cyclically monotone we know that if (x, y) and $(x', y') \in \mathrm{supp}(\gamma^*)$ with $x \leq x'$, then $y \leq y'$. Then we reason as in Santambrogio (2015, Lemma 2.8). Let $a, b \geq 0$ and let us compute $\gamma^*((a, +\infty) \times (b, +\infty))$. By the property on the support we just mentioned, γ^* cannot give mass to both $A = (a, +\infty) \times [0, b]$ and $B = [0, a] \times (b, +\infty)$. Thus

$$
\begin{aligned}
\gamma^*((a, +\infty) \times (b, +\infty)) &= \min\left[\gamma^*((a, +\infty) \times (b, +\infty) \cup A), \gamma^*((a, +\infty) \times (b, +\infty) \cup B)\right] \\
&= \min\left[\gamma^*((a, +\infty) \times [0, +\infty)), \gamma^*([0, +\infty) \times (b, +\infty))\right] \\
&= \min[U_{\nu^1}(a), U_{\nu^2}(b)].
\end{aligned}
$$

On the other hand, by the marginal property if $a > 0$ there holds $\gamma^*((a, +\infty) \times [0, +\infty)) = \nu^1((a, +\infty))$ and similarly for the second marginal. In conclusion, the measure γ^* satisfies:

$$
\begin{cases}
\gamma^*((a, +\infty) \times (b, +\infty)) = \min[U_{\nu^1}(a), U_{\nu^2}(b)] & \forall a, b > 0, \\
\gamma^*((a, +\infty) \times [0, +\infty)) = U_{\nu^1}(a) & \forall a > 0, \\
\gamma^*([0, +\infty) \times (b, +\infty)) = U_{\nu^2}(b) & \forall b > 0,
\end{cases}
\tag{SM8}
$$

and these three set of equalities are enough to characterize a measure on Ω_2. Thus the optimal transport coupling is unique and any measure on Ω_2 which satisfies (SM8) is in fact the optimal coupling.

We now prove that γ in the statement satisfies these properties. Let $a, b \geq 0$. By definition,

$$\gamma((a, +\infty) \times (b, +\infty)) = \mathrm{Leb}(\Omega_1)\left\{t \geq 0 \; : \; U_{\nu^1}^{-1}(t) > a \text{ and } U_{\nu^2}^{-1}(t) > b\right\}.$$

From this we can use (SM7) to rewrite the quantity of interest as

$$
\begin{aligned}
\gamma((a, +\infty) \times (b, +\infty)) &= \mathrm{Leb}(\Omega_1)\left\{t \geq 0 \; : \; t < U_{\nu^1}(a) \text{ and } t < U_{\nu^2}(b)\right\} \\
&= \mathrm{Leb}(\Omega_1)\left\{t \geq 0 \; : \; t < \min[U_{\nu^1}(a), U_{\nu^2}(b)]\right\} = \min[U_{\nu^1}(a), U_{\nu^2}(b)].
\end{aligned}
$$

In addition, as

$$\gamma((a, +\infty) \times [0, +\infty)) = \mathrm{Leb}(\Omega_1)\left\{t \geq 0 \; : \; U_{\nu^1}^{-1}(t) > a \text{ and } U_{\nu^2}^{-1}(t) \geq 0\right\}$$

and the second inequality is always satisfied, we obtain thanks to (SM7) that $\gamma((a, +\infty) \times [0, +\infty)) = U_{\nu^1}(a)$. The case where we exchange the marginals is similar. We conclude that γ satisfies (SM8) and thus it is the optimal coupling.

Eventually, let's turn to the case of when ν^1 is atomless and $\nu^1(\Omega_1) \geq \nu^2(\Omega_1)$, and let us define $T(x) = U_{\nu^2}^{-1}(U_{\nu^1}(x))$. The map T is non-decreasing as the composition of two non-increasing functions. Let us define $\gamma_T = (\mathrm{id}, T)_\# \nu^1$, again we only have to show that γ_T satisfies (SM8). Let's take $a, b \geq 0$ and we write:

$$\begin{aligned}
\gamma_T((a, +\infty) \times (b, +\infty)) &= \nu^1(\{x \geq 0 \ : \ x > a \text{ and } T(x) > b\}) \\
&= \nu^1(\{x \geq 0 \ : \ x > a \text{ and } U_{\nu^2}^{-1}(U_{\nu^1}(x)) > b\}) \\
&= \nu^1(\{x \geq 0 \ : \ x > a \text{ and } U_{\nu^1}(x) < U_{\nu^2}(b)\}),
\end{aligned}$$

where again we have use (SM7) for the last equality. So if $U_{\nu^1}(a) < U_{\nu^2}(b)$ then indeed the second constraint is always satisfied so $\gamma_T((a, +\infty) \times (b, +\infty)) = U_{\nu^1}(a)$. On the other hand if $U_{\nu^1}(a) > U_{\nu^2}(b)$ then the first constraint is always satisfied and thus $\gamma_T((a, +\infty) \times (b, +\infty))$ coincides with $U_{\nu^1 \# \nu^1}((0, U_{\nu^2}(b))) = U_{\nu^2}(b)$: to show this we use Lemma SM8 and the assumption $U_{\nu^2}(b) \leq U_{\nu^2}(0) \leq U_{\nu^1}(0)$. Eventually, if $U_{\nu^1}(a) = U_{\nu^2}(b)$ then we can remove all the points x such that $U_{\nu^1}(x) = U_{\nu^1}(a)$, as it is a ν^1-negligible set by Lemma SM8. It shows that $\gamma_T((a, +\infty) \times (b, +\infty)) = \min[U_{\nu^1}(a), U_{\nu^2}(b)]$. Then let us take $a > 0$ and we look at $\gamma_T((a, +\infty) \times [0, +\infty))$. For this one it is clear that it coincides with $\nu^1((a, +\infty)) = U_{\nu^1}(a)$. Eventually, if $b > 0$ then

$$\begin{aligned}
\gamma_T([0, +\infty) \times (b, +\infty)) &= \nu^1(\{x \geq 0 \ : \ x \geq 0 \text{ and } T(x) > b\}) \\
&= \nu^1(\{x \geq 0 \ : U_{\nu^2}^{-1}(U_{\nu^1}(x)) > b\}) \\
&= \nu^1(\{x \geq 0 \ : U_{\nu^1}(x) < U_{\nu^2}(b)\}) \\
&= \mathrm{Leb}(\Omega_1)\{t \geq 0 \ : \ t \leq U_{\nu^2}(b)\} = U_{\nu^2}(b),
\end{aligned}$$

where the second to last equality follow from Lemma SM8 and the assumption $U_{\nu^2}(b) \leq U_{\nu^2}(0) \leq U_{\nu^1}(0)$. We conclude that γ_T satisfies the set of identities (SM8), and thus it is the unique optimal coupling. Lastly, the formula for the expression of the distance is clear from the definition of γ and T. $\qquad\square$

3   Proofs of Section 4: Evaluation of the index

Theorem 5 in the main manuscript is implied by the following Theorem SM9 and Corollary SM10.

Theorem SM9. *Let $\nu \in \mathcal{M}_2(\Omega_d)$ be a Lévy measure with equal marginals $\pi_{i\#}\nu = \overline{\nu}$ on Ω_1, for $i = 1, \ldots, d$. Denote by $\nu^+ = \Sigma_\# \nu \in \mathcal{M}_2(\Omega_1)$, where $\Sigma(\mathbf{s}) = \sum_{i=1}^d s_i$. Then*

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = 2d M_2(\overline{\nu}) - 2 \int_0^{+\infty} U_{\nu^+}^{-1}(s) \, U_{\overline{\nu}}^{-1}(s) \, \mathrm{d}s.$$

Moreover, there exist an optimal pair (φ, ψ) of Kantorovich potentials and a convex function u on Ω_1 such that

$$\varphi(\mathbf{s}) = \frac{1}{2} \sum_{i=1}^d s_i^2 - u(\Sigma(\mathbf{s})), \qquad \psi(\mathbf{s}) = \frac{1}{2} \sum_{i=1}^d s_i^2 - u^*(\max(\mathbf{s})),$$

where u^* is the Legendre transform of u and $\max(\mathbf{s}) = \max(s_1, \ldots, s_d)$.

8

Proof. Though not necessary, we first focus on the expression for the Wasserstein distance when ν_+ is atomless because the proof is more straightforward. In such case as clearly $\nu^+(\Omega_1) \geq \overline{\nu}(\Omega)$, by Proposition 2 there exists an optimal transport map $T^+ = \nabla u$ between ν^+ and $\overline{\nu}$. We prove that $T(\mathbf{s}) = (T^+(\Sigma(\mathbf{s})), \ldots, T^+(\Sigma(\mathbf{s})))$ is a transport map between ν and ν^{co}. We define $\Delta(\mathbf{s}) = (s, \ldots, s)$ and observe that $T = \Delta \circ T^+ \circ \Sigma$. By associativity of the pushforward operator $T_{\#}\nu = (\Delta \circ T^+)_{\#}\nu^+ = \Delta_{\#}(T_{\#}^+\nu^+)$. Since T^+ is the optimal transport map between ν^+ and $\overline{\nu}$, $T_{\#}\nu = \Delta_{\#}\overline{\nu} = \nu^{\mathrm{co}}$ by definition of comonotonic Lévy measure. Moreover, T is the gradient of $u'(\mathbf{s}) = u(\Sigma(\mathbf{s}))$, which is convex since u is convex. By Lemma SM7, T is the optimal transport map between ν and ν^{co}.

We now focus on the general case which follows the same idea but relies on disintegration to palliate the absence of transport map. By Proposition 4 there exists a unique optimal transport plan $\gamma^+ = (U_{\nu^+}^{-1}, U_{\overline{\nu}}^{-1})_{\#}\mathrm{Leb}(\Omega_1)$ between ν^+ and $\overline{\nu}$. We consider the disintegration of γ^+ with respect to ν^+, so that one can write $\gamma^+ = \int_{\Omega_1 \times \Omega_1} \gamma_{\mathbf{s}}^+ \mathrm{d}\nu^+(\mathbf{s})$, where $\gamma_{\mathbf{s}}^+$ is a probability measure on Ω_1 for every $\mathbf{s} \in \mathrm{supp}(\nu^+)$. We claim that an optimal transport map between ν and ν^{co} is

$$\gamma = \int_{\Omega_d \times \Omega_d} \Delta_{\#}\gamma_{\Sigma(\mathbf{s})}^+ \mathrm{d}\nu(\mathbf{s}).$$

First we prove that $\gamma \in \overline{\Gamma}(\nu, \nu^{\mathrm{co}})$, that is that γ has the right marginals. The first marginal ν derives from the definition of disintegration of measures. As for the second, we must show that for every $f \in C_c(\Omega_d)$,

$$\int_{\Omega_d \times \Omega_d} f(\mathbf{s}') \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} f(\mathbf{s}') \, \mathrm{d}\nu^{\mathrm{co}}(\mathbf{s}').$$

This follows by observing that the term on the left hand side is equal to

$$\int_{\Omega_1 \times \Omega_1} f(\Delta(s')) \, \mathrm{d}\gamma_s^+(s') \, \mathrm{d}\Sigma_{\#}\nu(s) = \int_{\Omega_1 \times \Omega_1} f(\Delta(s')) \, \mathrm{d}\gamma^+(s, s') = \int_{\Omega_1} f(\Delta(s')) \, \mathrm{d}\overline{\nu}(s'),$$

by definition of disintegration and because $\gamma^+ \in \overline{\Gamma}(\nu^+, \overline{\nu})$. We conclude by the definition of pushforward map since $\Delta_{\#}\overline{\nu} = \nu^{\mathrm{co}}$.

Consider the convex function u, u^* arising from the dual formulation of the transport problem between ν^+ and $\overline{\nu}$, as in Lemma SM6. We claim that (φ, ψ) in the statement is an optimal pair of Kantorovich potentials and γ is the optimal transport coupling. We will use Proposition SM5. We observe that $(\mathbf{s}, \mathbf{s}') \in \mathrm{supp}(\gamma)$ if and only if the following conditions hold: (i) $\mathbf{s} \in \mathrm{supp}(\nu)$; (ii) $\mathbf{s}' = (s', \cdots, s')$ for some $s' \in \Omega_1$; (iii) $(\Sigma(\mathbf{s}), s') \in \mathrm{supp}(\gamma^+)$. Thus in this case $\mathbf{s} \cdot \mathbf{s}' = \Sigma(\mathbf{s})s'$. Using the implication (i) $\Rightarrow$ (iii) from Proposition SM5 for the transport from ν^+ onto $\overline{\nu}$ (combined with Remark SM1), we see that $u(\Sigma(\mathbf{s})) + u^*(s') = \Sigma(\mathbf{s})s'$. But with the explicit expression that we have for (φ, ψ) it implies that (SM4) actually holds for the $(\mathbf{s}, \mathbf{s}')$ that we chose in $\mathrm{supp}(\gamma)$. Using this time the implication (iii) $\Rightarrow$ (i) from Proposition SM5 for the transport from ν onto ν^{co} yields optimality of γ for the primal problem and (φ, ψ) for the dual one.

We now show that γ induces the expression of the extended Wasserstein distance in the statement. Indeed, since $\mathbf{s} \cdot \mathbf{s}' = \Sigma(\mathbf{s})s'$ on the support of γ, $\int \|\mathbf{s} - \mathbf{s}'\|^2 \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}')$ is equal to

$$2dM_2(\overline{\nu}) - 2 \int_{\Omega_d \times \Omega_d} \Sigma(\mathbf{s})s' \, \mathrm{d}\Delta_{\#}\gamma_{\Sigma(\mathbf{s})}^+(\mathbf{s}') \, \mathrm{d}\nu(\mathbf{s}).$$

9

By definition of pushforward map, $\int_{\Omega_d \times \Omega_d} \Sigma(\mathbf{s}) s' \, \mathrm{d}\Delta_{\#}\gamma^+_{\Sigma(\mathbf{s})}(\mathbf{s}') \, \mathrm{d}\nu(\mathbf{s})$ is equal to

$$\int_{\Omega_1 \times \Omega_1} ss' \mathrm{d}\gamma_s^+(s') \, \mathrm{d}\nu^+(s) = \int_{\Omega_1 \times \Omega_1} ss' \, \mathrm{d}\gamma^+(s, s').$$

We conclude by substituting $\gamma^+ = (U_{\nu^+}^{-1}, U_{\bar{\nu}}^{-1})_{\#}\mathrm{Leb}(\Omega_1)$. $\qquad\square$

Corollary SM10. *Let $\nu^\perp \in \mathcal{M}_2(\Omega_d)$ be an independent Lévy measure with equal marginals $\pi_{i\#}\nu = \bar{\nu}$ on Ω_1, for $i = 1, \ldots, d$. Then,*

$$\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2 = 2d\left(M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(d\, U_{\bar{\nu}}(s)) \, \mathrm{d}\bar{\nu}(s)\right).$$

Proof. Let $\nu = \nu^\perp$ indicate the Lévy measure. Thanks to the support of ν and its marginal constraint, we observe that

$$\begin{aligned}
U_{\nu^+}(s) &= \nu(\{(s_1, \ldots, s_d) : s_1 + \cdots + s_d \leq s\}) \\
&= \sum_{i=1}^d \nu(\{(s_1, \ldots, s_d) : s_i \leq s\}) \\
&= \sum_{i=1}^d \bar{\nu}(\{s_i : s_i \leq s\}) \\
&= d\, U_{\bar{\nu}}(s)
\end{aligned}$$

By absolute continuity of $\bar{\nu}$, it follows that $\nu^+(s) = -\partial/\partial s\, U_{\nu^+}(s) = -d\, \partial/\partial s\, U_{\bar{\nu}}(s) = d\,\bar{\nu}(s)$ is atomless. Thus, with a change of variable $s = U_{\nu^+}^{-1}(t)$, the expression in Theorem SM9 becomes

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = 2dM_2(\bar{\nu}) - 2\int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(U_{\nu^+}(s))\nu^+(s) \, \mathrm{d}s. \qquad \text{(SM9)}$$

We conclude by substituting the expression of ν^+ and U_{ν^+} above. $\qquad\square$

4 Proofs of Section 2: Main result

Given the expression of the index in (7) of the main manuscript, Theorem 1 and Remark 1 are easily implied by the following result.

Theorem SM11. *Let $\nu \in \mathcal{M}_2(\Omega_d)$ with equal marginals $\pi_{i\#}\nu = \bar{\nu}$ on Ω_1, for $i = 1, \ldots, d$. Then $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) \leq \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$. If $\bar{\nu}(\Omega_1) = +\infty$, there is equality if and only if $\nu = \nu^\perp$.*

Proof. We first prove that $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) \leq \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$. Let (φ, ψ) be the optimal pair of Kantorovich potentials for the transport between ν and ν^{co} in Theorem SM9. Then,

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}\nu(\mathbf{s}) + \int_{\Omega_d} \psi(\mathbf{s}') \, \mathrm{d}\nu^{\mathrm{co}}(\mathbf{s}'),$$

$$\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2 \geq \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}\nu^\perp(\mathbf{s}) + \int_{\Omega_d} \psi(\mathbf{s}') \, \mathrm{d}\nu^{\mathrm{co}}(\mathbf{s}').$$

In particular this yields

$$\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2 - \mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 \geq \int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}(\nu^\perp - \nu)(\mathbf{s}).$$

By Theorem SM9, $\varphi(\mathbf{s}) = \frac{1}{2} \sum_{i=1}^d s_i^2 - u(\Sigma(\mathbf{s}))$, where u is convex. The second moments in the right hand side of the last inequality cancel out because ν and $\nu^\perp$ satisfy the same marginal constraints. Thus, we can rewrite the right hand side as $\int_{\Omega_d} u(\Sigma(\mathbf{s})) \, \mathrm{d}(\nu - \nu^\perp)(\mathbf{s})$. Since $\nu^\perp$ is supported on the axis and thanks to the marginal constraints of ν,

$$\int_{\Omega_d} u(\Sigma(\mathbf{s})) \, \mathrm{d}\nu^\perp(\mathbf{s}) = \sum_{i=1}^d \int_{\Omega_1} u(s_i) \, \mathrm{d}\overline{\nu}(s_i) = \int_{\Omega_d} \sum_{i=1}^d u(s_i) \, \mathrm{d}\nu(\mathbf{s}).$$

Thus,

$$\int_{\Omega_d} \varphi(\mathbf{s}) \, \mathrm{d}(\nu^\perp - \nu)(\mathbf{s}) = \int_{\Omega_d} \left(u(\Sigma(\mathbf{s})) - \sum_{i=1}^d u(s_i) \right) \mathrm{d}\nu(\mathbf{s}), \qquad \text{(SM10)}$$

which is non-negative thanks to superadditivity of convex functions, see Lemma SM12 below. This proves that $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) \leq \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$.

We now assume that $\overline{\nu}(\Omega_1) = +\infty$ and we show that equality holds only if $\nu = \nu^\perp$. Let ν satisfy $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) = \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$. Then by (SM10), the support of ν is contained in the set $D_u = \{\mathbf{s} \in \Omega_d : u(\Sigma(\mathbf{s})) = \sum_{i=1}^d u(s_i)\}$. We prove by contradiction that D_u only contains the axis. Without loss of generality let $\mathbf{s} \in D_u$ such that $s_1, s_2 > 0$. Then by Lemma SM12 below, u is linear on $[0, s_1 + s_2]$. Let $\lambda \geq 0$ be the slope of u on $[0, s_1 + s_2]$. On $(0, s_1 + s_2)$ the subdifferential of u coincides with the gradient of u, that is $\{\lambda\}$. Let $\gamma^+ = (U_{\nu^+}^{-1}, U_{\overline{\nu}}^{-1})_{\#}\mathrm{Leb}(\Omega_1)$ be the optimal transport coupling between ν^+ and $\overline{\nu}$. By Proposition SM5 and Remark SM1 we know the support of γ is concentrated on the graph of the subdifferential of u. Thus for $\varepsilon > 0$ there holds

$$\mathrm{supp}(\gamma^+) \cap ([\varepsilon, s_1 + s_2 - \varepsilon] \times [0, +\infty)) = [\varepsilon, s_1 + s_2 - \varepsilon] \times \{\lambda\}.$$

Note that the assumption of infinite mass also implies that $\nu^+(\Omega_1) = +\infty$ and in particular $U_{\overline{\nu}}^{-1}$ and $U_{\nu^+}^{-1}$ take strictly positive values. Thus λ cannot be equal to 0. On the other hand, if $\lambda > 0$ then by the marginal property of γ^+ we see that

$$\nu^+([\varepsilon, s_1 + s_2 - \varepsilon]) = \gamma^+([\varepsilon, s_1 + s_2 - \varepsilon] \times \{\lambda\}) \geq \overline{\nu}(\{\lambda\}).$$

Letting $\varepsilon \to 0$, as $\nu^+ \in \mathcal{M}_2(\Omega_1)$ is also a measure with infinite mass we see that $\overline{\nu}(\{\lambda\}) = +\infty$. This a contradiction since $\lambda > 0$ and the second moment of $\overline{\nu}$ is finite. $\qquad \square$

During the proof we have used the following elementary Lemma about convex functions.

Lemma SM12. *Let* $u : [0, +\infty) \to [0, +\infty)$ *a convex function such that* $u(0) = 0$. *Then if* $\mathbf{s} \in (0, +\infty)^d$, $u(\Sigma(\mathbf{s})) \geq \sum_i u(s_i)$ *and equality holds if and only if* u *is linear on* $[0, \Sigma(\mathbf{s})]$.

Proof. We reason by induction on d. For $d = 2$, as the slopes of u are non-decreasing,

$$\frac{u(s_1 + s_2) - u(s_2)}{(s_1 + s_2) - s_2} \geq \frac{u(s_1) - u(0)}{s_1},$$

11

which is equivalent to our claim. Moreover, it is clear that equality holds if and only if u is linear on $[0, s_1 + s_2]$. For $d \geq 3$ we rewrite

$$u(\Sigma(\mathbf{s})) - \sum_{i=1}^{d} u(s_i) = \left[u\left(\sum_{i=1}^{d-1} s_i + s_d \right) - u\left(\sum_{i=1}^{d-1} s_i \right) - u(s_d) \right] + \left[u\left(\sum_{i=1}^{d-1} s_i \right) - \sum_{i=1}^{d-1} u(s_i) \right],$$

and we use the case $d = 2$ for the first term in the sum and the case $d - 1$ for the second term. The equality case is implied by the equality case for the first term. $\qquad\square$

5 Proofs of Section 5: Examples

Lemma 6 easily follows from the following.

Lemma SM13. *Let $\tilde{\boldsymbol{\mu}}^1$ and $\tilde{\boldsymbol{\mu}}^2$ be independent homogeneous CRVs with same base measure α and Lévy measures ν^1, ν^2, respectively. Then $\tilde{\boldsymbol{\mu}}^1 + \tilde{\boldsymbol{\mu}}^2$ is a CRV with base measure α and Lévy measure $\nu^1 + \nu^2$.*

Proof. The law of a CRV with base measure α and Lévy measure ν is characterized by the joint Laplace exponent

$$\mathbb{E}(e^{-\boldsymbol{\lambda}\tilde{\boldsymbol{\mu}}(A)}) = e^{-\alpha(A)\int_{\Omega_d}(1 - e^{-\boldsymbol{\lambda}\boldsymbol{s}})\,\mathrm{d}\nu(\boldsymbol{s})},$$

for every set A and every $\boldsymbol{\lambda} \in \Omega_d$. By linearity of the integral,

$$e^{-\alpha(A)\int_{\Omega_d}(1 - e^{-\boldsymbol{\lambda}\boldsymbol{s}})\,\mathrm{d}(\nu^1 + \nu^2)(\boldsymbol{s})} = \mathbb{E}(e^{-\boldsymbol{\lambda}\tilde{\boldsymbol{\mu}}^1(A)})\,\mathbb{E}(e^{-\boldsymbol{\lambda}\tilde{\boldsymbol{\mu}}^2(A)}).$$

Since $\tilde{\boldsymbol{\mu}}^1$ and $\tilde{\boldsymbol{\mu}}^2$ are independent, this is equal to $\mathbb{E}(e^{-\boldsymbol{\lambda}(\tilde{\boldsymbol{\mu}}^1 + \tilde{\boldsymbol{\mu}}^2)(A)})$. $\qquad\square$

We define the following quantities.

$$U_{\nu_z^+}(s) = d(1 - z)U_{\overline{\nu}}(s) + zU_{\overline{\nu}}(sd^{-1}), \qquad \nu_z^+(s) = d(1 - z)\overline{\nu}(s) + zd^{-1}\overline{\nu}(sd^{-1}).$$

Proposition 7. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional additive CRV of parameter z. Then $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq z$ and*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{dM_2(\overline{\nu}) - \int_0^{+\infty} s\, U_{\overline{\nu}}^{-1}(U_{\nu_z^+}(s))\nu_z^+(s)\,\mathrm{d}s}{dM_2(\overline{\nu}) - d\int_0^{+\infty} s\, U_{\overline{\nu}}^{-1}(dU_{\overline{\nu}}(s))\overline{\nu}(s)\,\mathrm{d}s}.$$

Proof. Let $\nu = \nu_z$. The expression in the denominator follows by Corollary SM10. Since ν is supported on the bisecting line and on the axis, we observe that $U_{\nu^+}(s) = d(1 - z)U_{\overline{\nu}}(s) + zU_{\overline{\nu}}(sd^{-1})$. We derive the expression of ν^+ by differentiation. To prove the inequality $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq z$, we restrict to transport maps that acts as the identity on the mass on the bisecting line, so that

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = \mathcal{W}_*(z\nu^{\mathrm{co}} + (1 - z)\nu^\perp, \nu^{\mathrm{co}})^2 \leq \mathcal{W}_*((1 - z)\nu^\perp, (1 - z)\nu^{\mathrm{co}})^2,$$

which is equal to $(1 - z)\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2$ by homogeneity of the squared Wasserstein distance. Thus $\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2$ cancels out with the denominator and one gets $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq 1 - (1 - z) = z$. $\qquad\square$

We define the following quantities.

$$U_{\nu_\phi^+}(s) = \frac{1}{\Gamma(d\phi)}\int_0^1 \Gamma\left(d\phi, \frac{s}{u}\right)\frac{(1 - u)^{\phi-1}}{u}\,\mathrm{d}u, \qquad \nu_\phi^+(s) = \frac{s^{d\phi-1}}{\Gamma(d\phi)}\int_0^1 e^{-\frac{s}{u}}\frac{(1 - u)^{\phi-1}}{u^{d\phi+1}}\,\mathrm{d}u.$$

Proposition 8. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional gamma compound random vector of parameter ϕ. Then,*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{d - \int_0^{+\infty} a\, E_1^{-1}(U_{\nu_\phi^+}(s))\, \nu_\phi^+(s)\, \mathrm{d}s}{d - d\int_0^{+\infty} E_1^{-1}(dE_1(s))e^{-s}\, \mathrm{d}s}.$$

Proof. Since $M_2(\bar{\nu}) = 1$, by Corollary SM10,

$$\mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})^2 = 2d\left(1 - \int_0^{+\infty} E_1^{-1}(dE_1(s))e^{-s}\, \mathrm{d}s\right).$$

Let $\nu = \nu^\phi$. Since ν is diffuse, ν^+ is atomless and we may apply (SM9). We first find the expression of U_{ν^+}. Let $p_\phi(z) = \Gamma(\phi)^{-1}z^{\phi-1}e^{-z}\mathbb{1}_{(0,+\infty)}(z)$ indicate the density of a gamma$(\phi, 1)$, so that

$$\nu(\boldsymbol{s}) = \int_0^1 \frac{(1-u)^{\phi-1}}{u^{d+1}}\prod_{i=1}^d p_\phi\left(\frac{s_i}{u}\right)\mathrm{d}u.$$

By definition of pushforward measure,

$$U_{\nu^+}(t) = \int_{(0,+\infty)^d} \mathbb{1}_{(t,+\infty)}(s_1 + \cdots + s_d)\nu(\boldsymbol{s})\, \mathrm{d}s_1 \ldots \mathrm{d}s_d$$

$$= \int_0^1 \frac{(1-u)^{\phi-1}}{u}\left(\int_{(0,+\infty)^d} \mathbb{1}_{\left(\frac{t}{u},+\infty\right)}(v_1 + \cdots + v_d)\prod_{i=1}^d p_\phi(v_i)\, \mathrm{d}v_1 \ldots \mathrm{d}v_d\right)\mathrm{d}u,$$

with a change of variable $\boldsymbol{v} = \boldsymbol{s}/u$. The expression in the parenthesis coincides with the survival function of the sum of d independent gamma$(\phi, 1)$ random variables, evaluated in t/u. Since the sum of d independent gamma$(\phi, 1)$ random variables is a gamma$(d\phi, 1)$,

$$U_{\nu^+}(t) = \frac{1}{\Gamma(d\phi)}\int_0^1 \Gamma\left(d\phi, \frac{t}{u}\right)\frac{(1-u)^{\phi-1}}{u}\mathrm{d}u.$$

The expression of v^+ easily derives by differentiating U_{ν^+}. $\qquad\square$

We now consider a d-dimensional CRV with m independent components and n comonotone replicates each, that is,

$$\tilde{\boldsymbol{\mu}} = (\overbrace{\tilde{\mu}^1, \ldots, \tilde{\mu}^1}^{n}, \overbrace{\tilde{\mu}^2, \ldots, \tilde{\mu}^2}^{n}, \ldots, \overbrace{\tilde{\mu}^m, \ldots, \tilde{\mu}^m}^{n}), \tag{SM11}$$

where $\tilde{\boldsymbol{\mu}}^\perp = (\tilde{\mu}^1, \tilde{\mu}^2, \ldots, \tilde{\mu}^m) \in \mathcal{M}_2(\Omega_m)$ is an independent CRV with equal marginals $\bar{\nu}$.

Proposition 9. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional CRV as in* (SM11)*, with m independent components and n comonotone replicates. Then,*

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(mU_{\bar{\nu}}(s))\bar{\nu}(s)\, \mathrm{d}s}{M_2(\bar{\nu}) - \int_0^{+\infty} s\, U_{\bar{\nu}}^{-1}(dU_{\bar{\nu}}(s))\bar{\nu}(s)\, \mathrm{d}s}.$$

13

Proof. Let ν indicate the corresponding Lévy measure. Then,

$$
\begin{aligned}
U_{\nu^+}(s) &= \nu(\{(s_1, \ldots, s_d) : s_1 + \cdots + s_d \leq s\}) \\
&= \sum_{i=1}^{m} \nu(\{(s_1, \ldots, s_d) : s_{(i-1)n+1} + \cdots + s_{in} \leq s\}) \\
&= \sum_{i=1}^{m} \nu(\{(s_1, \ldots, s_d) : n s_{in} \leq s\}) \\
&= \sum_{i=1}^{m} \bar{\nu}\left(\left\{s_{in} : s_{in} \leq \frac{s}{n}\right\}\right) \\
&= m U_{\bar{\nu}}\left(\frac{s}{n}\right)
\end{aligned}
$$

By absolute continuity of $\bar{\nu}$, it follows that

$$
\nu^+(s) = -\frac{\partial}{\partial s} U_{\nu^+}(s) = \frac{m}{n} \bar{\nu}\left(\frac{s}{n}\right)
$$

is atomless. Thus, with a change of variable $s = U_{\nu^+}^{-1}(t)$,

$$
\mathcal{W}_*(\nu, \nu^{co})^2 = 2d M_2(\bar{\nu}) - 2 \int_0^{+\infty} s \, U_{\bar{\nu}}^{-1}(U_{\nu^+}(s)) \nu^+(s) \, \mathrm{d}s.
$$

The result follows by substituting the expression of ν^+ and U_{ν^+} above in the numerator and by using Corollary SM10 for the denominator. $\qquad\square$

Proposition 10. *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional CRV as in* (SM11) *with m independent components and n comonotone replicates. If n is fixed and $m \to +\infty$,*

$$
I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \to 0
$$

monotonically from above. If m is fixed and $n \to +\infty$,

$$
I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \to I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = \frac{\int_0^{+\infty} s \, U_{\bar{\nu}}^{-1}(m U_{\bar{\nu}}(s)) \bar{\nu}(s) \, \mathrm{d}s}{M_2(\bar{\nu})}
$$

monotonically from below.

Proof. Thanks to Proposition 9 we only need to show that

$$
\lim_{k \to +\infty} \int_0^{+\infty} s \, U_{\bar{\nu}}^{-1}(k U_{\bar{\nu}}(s)) \bar{\nu}(s) = 0
$$

monotonically in k. Since $U_{\bar{\nu}}$ is a decreasing function, so is $U_{\bar{\nu}}^{-1}$. Moreover $\lim_{s \to +\infty} U_{\bar{\nu}}^{-1}(s) = 0$ by the infinite activity assumption. Thus for every $s > 0$, $U_{\bar{\nu}}^{-1}(k U_{\bar{\nu}}(s)) \to 0$ monotonically. We then apply the monotone convergence theorem and conclude. $\qquad\square$

References

Figalli, A. and N. Gigli (2010). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées 94*, 107–130.

Guillen, N., C. Mou, and A. Święch (2019). Coupling Lévy measures and comparison principles for viscosity solutions. *Transactions of the American Mathematical Society 372*, 7327–7370.

Santambrogio, F. (2015). *Optimal transport for applied mathematicians.* Birkäuser, NY.

Villani, C. (2003). *Topics in optimal transportation.* American Mathematical Society.