ISSN 2279-9362

Research
Education
Outreach

CCA

A Wasserstein index of dependence for random measures

Marta Catalano, Hugo Lavenant, Antonio Lijoi, Igor Prünster

No. 659
October 2021

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

A Wasserstein index of dependence for random measures

Marta Catalano
ESOMAS Department and Collegio Carlo Alberto, University of Torino
E-mail: marta.catalano@unito.it

Hugo Lavenant
Department of Decision Sciences and BIDSA, Bocconi University
E-mail: hugo.lavenant@unibocconi.it

Antonio Lijoi
Department of Decision Sciences and BIDSA, Bocconi University & Collegio Carlo Alberto
E-mail: antonio.lijoi@unibocconi.it

Igor Prünster
Department of Decision Sciences and BIDSA, Bocconi University & Collegio Carlo Alberto
E-mail: igor@unibocconi.it

Abstract

Nonparametric latent structure models provide flexible inference on distinct, yet related, groups of observations. Each component of a vector of $d \geq 2$ random measures models the distribution of a group of exchangeable observations, while their dependence structure regulates the borrowing of information across different groups. Recent work has quantified the dependence between random measures in terms of Wasserstein distance from the maximally dependent scenario when $d = 2$. By solving an intriguing max-min problem we are now able to define a Wasserstein index of dependence $I_{\mathcal{W}}$ with the following properties: (i) it simultaneously quantifies the dependence of $d \geq 2$ random measures; (ii) it takes values in [0,1]; (iii) it attains the extreme values $\{0, 1\}$ under independence and complete dependence, respectively; (iv) since it is defined in terms of the underlying Lévy measures, it is possible to evaluate it numerically in many Bayesian nonparametric models for partially exchangeable data.

Keywords: Completely random vector; index of dependence; Lévy measure; min-max; optimal transport; random measure; Wasserstein distance.

1 Introduction

Dependent random measures provide the main building block for many nonparametric latent structure models for distinct, yet related, groups of observations. The distribution P_i of each group $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n_i})$ is flexibly modeled as $P_i = t(\tilde{\mu}_i)$, where $\tilde{\mu}_i$ is a random measure and t is a suitable transformation. Notable examples include normalization for random probabilities (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). Additionally, the dependence among the random measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ regulates the interaction among $d \geq 2$ different groups of observations, allowing for different levels of borrowing of information; 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 $i = 1, \ldots, d$. Their conditional formulation

makes them ideal for performing Bayesian nonparametric inference on the distribution of the distinct groups, as witnessed by the multitude of contributions in the literature; see Quintana et al. (2021) for a recent review.

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, a greater prior interaction between groups entails a 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 translates in the need for a precise measure of dependence between d random measures. In Catalano et al. (2020) the authors focused on $d = 2$ and proposed to measure the dependence in terms of the Wasserstein distance of the random measures from complete dependence $\tilde{\boldsymbol{\mu}}^{\text{co}}$, namely $\sup_A \mathcal{W}(\tilde{\boldsymbol{\mu}}(A), \tilde{\boldsymbol{\mu}}^{\text{co}}(A))$. The main limitations of this measure are the following. First, since the law of $\tilde{\boldsymbol{\mu}}$ is specified through a Lévy measure, one needs to express the distance in terms of the Lévy measure in order to compute it numerically. For this reason, Catalano et al. (2020) derive tight bounds of the distance in terms of the Lévy measure and evaluate them in terms of the hyperparameters of popular models. Second, measuring dependence through a distance allows to perform relative comparisons between dependence structures ("$\tilde{\boldsymbol{\mu}}^1$ is more dependent than $\tilde{\boldsymbol{\mu}}^2$") but does not allow for absolute quantifications ("$\tilde{\boldsymbol{\mu}}$ has an *intermediate* dependence structure"), unless we find the maximum possible value for the distance. In this work we overcome these two limitations in the following way. First, we work directly at the level of the Lévy measures with the extended Wasserstein distance of (Figalli and Gigli, 2010), which allows for the comparison between measures with different and possibly infinite mass. As an additional result, we prove that the upper bounds in Catalano et al. (2020) can be expressed in terms of this distance. Second, we leverage the dual formulation of the extended Wasserstein distance to find its maximum and create a Wasserstein index of dependence $\mathcal{I}_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ with the following properties: (i) it simultaneously quantifies the dependence of $d \geq 2$ random measures; (ii) it takes values in [0,1]; (iii) it attains the extreme values $\{0, 1\}$ *only* under independence and complete dependence, respectively; (iv) since it is defined in terms of the underlying Lévy measures, it is possible to evaluate it numerically. We compute the index in notable models in the Bayesian nonparametric literature, namely additive random measures (Lijoi et al., 2014; Lijoi and Nipoti, 2014) and compound random measures (Griffin and Leisen, 2017). 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 so to achieve the same value of the index of dependence, one can make a fair assessment of their posterior performance under different scenarios.

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.

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 evaluation on disjoint sets $\{\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)\}$ are independent random variables on $\mathbb{R}$. Following Catalano et al. (2020), we refer to a *completely random vector* as its multivariate extension, that is a random vector of measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ such that its evaluation on disjoint sets $\{\tilde{\boldsymbol{\mu}}(A_1), \ldots, \tilde{\boldsymbol{\mu}}(A_n)\}$ are independent random vectors on $\mathbb{R}^d$. 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 atoms $\{Y_i\}$ (up to a potential deterministic drift). Completely random vectors often arise in

Bayesian nonparametrics to model the interaction across distinct groups of observations. Prior specifications are typically based on completely random vectors 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 small jumps (*infinite activity*). We focus on this class and restrict our attention to completely random vectors with finite second moments $\mathbb{E}(\tilde{\boldsymbol{\mu}}(A)^2) < +\infty$.

The goal of our work is to provide a tractable index of dependence for completely random vectors. Since these multivariate random quantities live a non-Euclidean space, a natural way to define the index is by introducing a distance $\mathcal{D}$ between the laws of completely random vectors. First, we highlight two extreme dependence structures: (a) complete dependence (or comonotonicity) $\tilde{\boldsymbol{\mu}}^{\text{co}}$, where $\mu_1 = \cdots = \mu_d$ almost surely, (b) independence $\tilde{\boldsymbol{\mu}}^{\perp}$, where $\{\mu_1, \ldots, \mu_d\}$ are independent random measures. Then, we define an index $I_{\mathcal{D}}$ between 0 and 1 in terms of distance from complete dependence:

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

Here and after, we use the notation $\mathcal{D}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\text{co}})$ to indicate the distance between the laws of $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^{\text{co}}$. We observe that the non-degeneracy of distances entails that $I_{\mathcal{D}}(\tilde{\boldsymbol{\mu}}) = 1$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\text{co}}$. 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 bulding $\mathcal{D}$ on the Wasserstein distance, as defined in (2). This leads to our main result that we now state.

Theorem 1. *Let $\tilde{\boldsymbol{\mu}}$ be a homogeneous infinitely active completely random vector 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* (2) *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 $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\text{co}}$; (iii) $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 0$ if and only if $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$.

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 $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\perp}$. Details are provided in the proof of Theorem 1.

Remark 2. When defining the index (1), $\mathcal{D}(\tilde{\boldsymbol{\mu}}', \tilde{\boldsymbol{\mu}}^{\text{co}})^2$ can be replaced by $\mathcal{D}(\tilde{\boldsymbol{\mu}}', \tilde{\boldsymbol{\mu}}^{\text{co}})^p$ 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 3 and due to the use of the extended Wasserstein distance of order 2, as will be clear in Section 3. This allows to draw a parallel with linear correlation in Section 5.1.

We conclude this section with some intuition on the definition of a tractable distance $\mathcal{D}$ on the laws of completely random vectors. We observe that thanks to the independence on disjoint sets, the distribution of a completely random vector $\tilde{\boldsymbol{\mu}}$ is characterized by the set-wise evaluations $\{\tilde{\boldsymbol{\mu}}(A)\}$, where A spans over all Borel 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 (infinitely divisible) distribution is specified through a Lévy intensity $\nu_A(\cdot) = \alpha(A)\nu(\cdot)$ for some base measure α and some Lévy measure ν. 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 1 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 completely random vectors:

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

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 to support the intuition 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 references.

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 $f : \mathbb{X} \to \mathbb{Y}$, we denote by $f_{\#}\mu$ 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

$$\mathcal{W}(\nu^1, \nu^2)^2 = \inf_{\gamma \in \Gamma(\nu^1, \nu^2)} \iint_{\mathbb{R}^{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\boldsymbol{s}, \boldsymbol{s}') = \inf_{X \sim \nu^1, Y \sim \nu^2} \mathbb{E}\left[\|X - Y\|^2\right]. \tag{3}$$

There always exists a coupling γ^* that realizes the infimum in (3) 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); 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 moment, that is $M_2(\nu) = \int_{\Omega_d} \|\boldsymbol{s}\|^2 \mathrm{d}\nu(\boldsymbol{s}) < +\infty$.

Definition 1 (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 $\pi_{1\#}\gamma|_{\Omega_d} = \nu^1$ and $\pi_{2\#}\gamma|_{\Omega_d} = \nu^2$. We define*

$$\mathcal{W}_*(\nu^1, \nu^2)^2 = \inf_{\gamma \in \overline{\Gamma}(\nu^1, \nu^2)} \iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \, \mathrm{d}\gamma(\boldsymbol{s}, \boldsymbol{s}'). \tag{4}$$

"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 (4). 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 $\pi_{2\#}\gamma|_{\Omega_d}$ and not to $\pi_{1\#}\gamma|_{\Omega_d}$ 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 case of measures with infinite mass, which is always the case for infinitely active completely random vectors, we provide a new expression of this distance that makes the link with the upper bound studied in Catalano et al. (2020), see in particular Theorem 5. Moreover, the next result highlights a novel connection with the classical Wasserstein distance, which is specific to the infinite mass assumption.

Proposition 1. *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^1_r, ν^2_r are two measures with finite mass r such that for each Borel set B, $\nu^i_r(B) \to \nu^i(B)$ increasingly as $r \to +\infty$. Then*

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

Remark 3. 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 completely random vectors. Indeed, the Lévy intensities satisfy $\nu_A(\cdot) = \alpha(A)\nu(\cdot)$ for some base measure α with mass $\bar{\alpha}$ and some Lévy measure ν. Thus, $\mathcal{D}(\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2)^2 = \bar{\alpha}\mathcal{W}_*(\nu^1, \nu^2)^2$ and the base measure α only acts as a scaling factor.

Remark 4. Let $\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\mu}}^2$ satisfy the assumptions of Theorem 1. In Catalano et al. (2020) tight upper bounds of the classical Wasserstein distance between $\tilde{\boldsymbol{\mu}}^1(A)$ and $\tilde{\boldsymbol{\mu}}^2(A)$ were found through suitable compound Poisson approximations. Thanks to Proposition 1 we are now able to reinterpret those bounds in terms of extended Wasserstein distance between Lévy measures, yielding

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)) \leq \mathcal{W}_*(\nu^1_A, \nu^2_A). \tag{5}$$

Since $\tilde{\boldsymbol{\mu}}^i(A)$ has a multivariate infinitely divisible distribution with Lévy measure ν^i_A, (5) provides an upper bound of the classical Wasserstein distance between infinitely divisible distributions in terms of the extended Wasserstein distance between their corresponding Lévy measures.

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 $\nu \in \mathcal{M}_2(\Omega_1)$ we define its tail integral $U_\nu : x \in \Omega_1 \mapsto \nu((x, +\infty))$ and its generalized inverse $U^{-1}_\nu : t \mapsto \inf_{x \geq 0}\{x \,:\, U_\nu(x) \geq t\}$. If U_ν is injective, it coincides with the usual inverse. Moreover, we denote by $\mathrm{Leb}(\Omega_1)$ the Lebesgue measure on Ω_1.

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

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

Moreover, if ν^1 is atomless and $\nu^1(\Omega_1) \geq \nu^2(\Omega_1)$, $T(x) = U^{-1}_{\nu^2}(U_{\nu^1}(x))$ 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 $I_{\mathcal{W}}$. Starting from (1), the homogeneity property highlighted in Remark 3 ensures that

$$I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) = 1 - \frac{\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2}{\sup_\nu \mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2}. \tag{6}$$

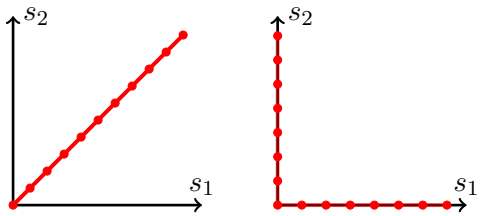

Figure 1: Support of ν^{co} and $\nu^{\perp}$ in Ω_2, respectively.

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 $\mathcal{W}_*(\nu, \nu^{\text{co}})$; (ii) to compute the denominator, we have to find the supremum of $\mathcal{W}_*(\nu, \nu^{\text{co}})$ over all possible dependence structures of ν, which involves solving a highly non-trivial max-min problem. The solutions to these two points are strongly intertwined. The integral expression for $\mathcal{W}_*(\nu, \nu^{\text{co}})$ is provided in Proposition 3 below and has two benefits: on one hand it allows 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, $\mathcal{W}_*(\nu, \nu^{\text{co}}) \leq \mathcal{W}_*(\nu^{\perp}, \nu^{\text{co}})$ 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 Proposition 3.

Remark 5. In principle, the Lévy measure ν corresponding to a completely random vector $\tilde{\boldsymbol{\mu}}$ 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 $i = 1, \dots, d$, where $\overline{\nu}$ is a 1-dimensional Lévy measure on Ω_1. Since we restrict to $\tilde{\boldsymbol{\mu}}$ with finite second moments, the second moments $M_2(\overline{\nu})$ are finite and thus $\overline{\nu}$ belongs to $\mathcal{M}_2(\Omega_1)$. In particular, $\nu \in \mathcal{M}_2(\Omega_d)$. Moreover, under complete dependence the Lévy measure $\tilde{\boldsymbol{\mu}}^{\text{co}}$ is concentrated on the bisecting line, while under independence $\tilde{\boldsymbol{\mu}}^{\perp}$ is supported on the axes; see Figure 1. More specifically,

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

where δ is the Dirac measure. A proof may be found in Cont and Tankov (2004) and Catalano et al. (2020).

Proposition 3. *Let $\nu \in \mathcal{M}_2(\Omega_d)$ be a Lévy measure with equal marginals $\pi_{i\#}\nu = \overline{\nu}$ for $i = 1, \dots, d$. Denote by $\nu^+ = \Sigma_{\#}\nu \in \mathcal{M}_2(\Omega_1)$, where $\Sigma(\mathbf{s}) = \sum_{i=1}^{d} s_i$. Then*

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

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

When ν_+ is atomless, Proposition 3 amounts to showing that $\mathbf{s} \mapsto (T(\sum_{i=1}^{d} s_i), \dots, T(\sum_{i=1}^{d} s_i))$ is an optimal transport map between ν and ν^{co}, with $T(s) = U_{\overline{\nu}}^{-1}(U_{\nu^+}(s))$. Thanks to Proposition 2, 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 $\mathbf{s} \mapsto (\sum_{i=1}^{d} s_i, \dots, \sum_{i=1}^{d} s_i)$, 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 were introduced in the Bayesian nonparametric literature by Müller et al. (2004); Lijoi et al. (2014); Lijoi and Nipoti (2014) to borrow information across distinct groups of observations. 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 completely random vector $\tilde{\boldsymbol{\mu}}$ 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 $\tilde{\boldsymbol{\mu}}^{\text{co}}$ as $z \to 1$ and independence $\tilde{\boldsymbol{\mu}}^{\perp}$ as $z \to 0$.

Lemma 1. *The Lévy measure of an additive completely random vector is* $\nu = z\nu^{\text{co}} + (1-z)\nu^{\perp}$.

It follows that the Lévy measure of an additive completely random vector 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 $\pi_{i\#}\nu = z\pi_{i\#}\nu^{\text{co}} + (1-z)\pi_{i\#}\nu^{\perp} = \bar{\nu}$, for $i = 1, \ldots, d$. Let

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

Proposition 4. *Let* $\tilde{\boldsymbol{\mu}}$ *be a* d*-dimensional additive completely random vector of parameter* z*. Then* $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq z$ *and*

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

Proposition 4 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, $z = \text{cor}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$, which does not depend on $A \in \mathcal{X}$. One proves it by using Proposition 1 and Campbell's theorem, so that

$$\text{cov}(\tilde{\mu}_1(A), \tilde{\mu}_1(A)) = \alpha(A)\int_0^{+\infty}\int_0^{+\infty} s_1 s_2\,\mathrm{d}\nu(s_1, s_2) = z\alpha(A)M_2(\bar{\nu}) = z\,\text{var}(\mu_1(A))^{\frac{1}{2}}\text{var}(\mu_1(A))^{\frac{1}{2}}.$$

Thus when $d = 2$, Proposition 4 guarantees that $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq \text{cor}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$. 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 completely random vectors introduce dependence linearly at the level of the Lévy measures.

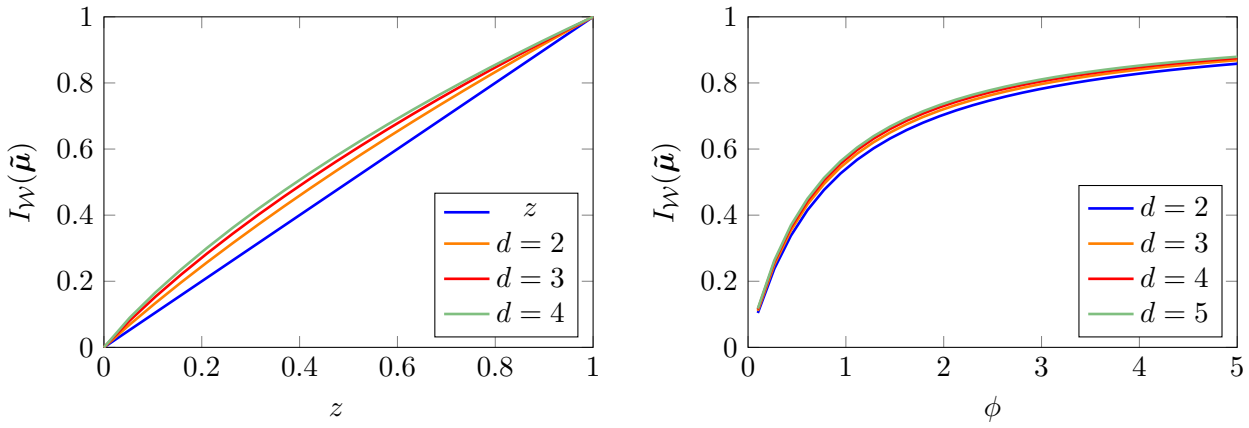

Figure 2: Left: $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ for $\tilde{\boldsymbol{\mu}}$ additive gamma completely random vector of parameter z and dimension d. Right: $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}})$ for $\tilde{\boldsymbol{\mu}}$ gamma compound random vector of parameter ϕ and dimension d.

5.2 Compound random measures

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

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

where h is a density function on $(0,+\infty)^d$ and ν^* is a Lévy measure. The predominant 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 $\phi > 0$. One proves that the marginals $\tilde{\mu}_i$ are gamma completely random measures and that ϕ only accounts for dependence. Under these specifications we say that $\tilde{\boldsymbol{\mu}}$ is a gamma compound random vector. Let

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

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

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

We use Proposition 5 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.

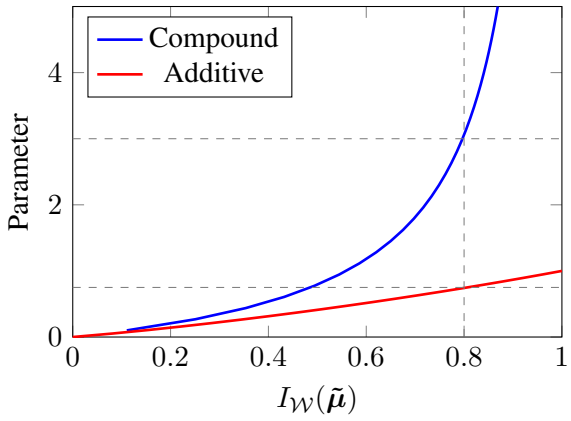

Figure 3: Parameter ϕ (z) of a compound (additive) completely random vector when $d = 3$, as $I_{\mathcal{W}}$ varies.

5.3 Model comparison

The proposed index has another important merit: by tuning the prior parameters of different models so 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 3, 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 it will implicitly favor highly correlated marginals.

Acknowledgements

The authors would like to thank Giuseppe Savaré for helpful discussions. A. Lijoi and I. Prünster are partially supported by MIUR, PRIN Project 2015SNS29B.

References

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019). Distribution theory for hierarchical processes. *The Annals of Statistics*, 47(1):67–92.

Catalano, M., Lijoi, A., and Prünster, I. (2020). Measuring the dependence in the Wasserstein distance for Bayesian nonparametric models. *Annals of Statistics, to appear*.

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

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

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics*, 9(2):356–367.

Figalli, A. and Gigli, N. (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(2):107–130.

Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in bayesian non-parametrics. *J. Royal Stat. Soc. Ser. B*, 79(2):525–545.

Guillen, N., Mou, C., and Święch, A. (2019). Coupling Lévy measures and comparison principles for viscosity solutions. *Transactions of the American Mathematical Society*, 372(10):7327–7370.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294.

James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics*, 33(4):1771–1799.

Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

Lijoi, A. and Nipoti, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc.*, 109(506):802–814.

Lijoi, A., Nipoti, B., and Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli*, 20(3):1260–1291.

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

Müller, P., Quintana, F., and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):735–749.

Nguyen, X. (2016). Borrowing strengh in hierarchical bayes: Posterior concentration of the dirichlet base measure. *Bernoulli*, 22(3):1535–1571.

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

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2021). The dependent Dirichlet process and related models. *Statistical Science, to appear*.

Regazzini, E., Lijoi, A., and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31(2):560–585.

Riva-Palacio, A. and Leisen, F. (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*, 55(58-63):94.

Villani, C. (2003). *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Society.

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

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.

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 S1. *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$. We define the extended Wasserstein distance as*

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

A coupling γ^* that realizes the infimum in (1) is termed an optimal transport coupling.

Proposition S1 (Theorem A.5 in Guillen et al. (2019))**.** *If $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_d)$, there exists at least on 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.

Definition S2. *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, $\mathrm{supp}(\nu) = \cap\{C \subset [0, +\infty)^d \text{ closed set such that } \nu(C^c \cap \Omega_d) = 0\}$.

Proposition S2. *A coupling $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ is optimal if and only if* $\mathrm{supp}(\gamma) \cup \{\mathbf{0}\}$ *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. □

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 S4 in the next section.

Lemma S1. *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 the support of γ_n is c-cyclically monotone, then so is* $\mathrm{supp}(\gamma)$.

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 $\gamma|_K$, $\mathrm{supp}(\gamma_n|_K)$ converges in the Haussdorff topology to $\mathrm{supp}(\gamma|_K)$. Since c-cyclically monotonicity is preserved under Hausdorff limit, we deduce that $\mathrm{supp}(\gamma|_K)$ is c-cyclically monotone. As K is arbitrary, this concludes the proof. □

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

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

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

Theorem S1 (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 S1 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 S3. *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

$$\iint_{\Omega_{2d}} \|\boldsymbol{s} - \boldsymbol{s}'\|^2 \,\mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \varphi(\mathbf{s}) \,\mathrm{d}\nu^1(\mathbf{s}) + \int_{\Omega_d} \psi(\mathbf{s}') \,\mathrm{d}\nu^2(\mathbf{s}'). \tag{3}$$

(iii) For every $(\mathbf{s}, \mathbf{s}') \in \mathrm{supp}(\gamma)$ there is equality in (2)*, that is,*

$$\varphi(\mathbf{s}) + \psi(\mathbf{s}') = \frac{|\mathbf{s} - \mathbf{s}'|^2}{2}. \tag{4}$$

Proof. That (i) and (ii) are equivalent is a straightforward consequence of Theorem S1: indeed the left hand side of (3) 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. □

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 S2. *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*

$$\varphi(\mathbf{s}) = \frac{1}{2}\sum_{i=1}^d s_i^2 - u(\mathbf{s}), \qquad \psi(\mathbf{s}) = \frac{1}{2}\sum_{i=1}^d s_i^2 - u^*(\mathbf{s}), \tag{5}$$

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

$$\varphi(\mathbf{s}) = \inf_{\mathbf{s}'\in\Omega_d}\frac{\|\mathbf{s}-\mathbf{s}'\|^2}{2} - \tilde\psi(\mathbf{s}'), \qquad \psi(\mathbf{s}') = \inf_{\mathbf{s}\in\Omega_d}\frac{\|\mathbf{s}-\mathbf{s}'\|^2}{2} - \varphi(\mathbf{s}). \tag{6}$$

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 (5), 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 (6) 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. □

Remark S1. In particular, assuming that (φ,ψ) is given as above, the constraint (2) simply amounts to $u(\mathbf{s}) + u^*(\mathbf{s}')\geq\mathbf{s}\cdot\mathbf{s}'$, which holds by Young's inequality. Thus (4) 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 (4) 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 S3. *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 (5). Since $u(\mathbf{0}) = u^*(\mathbf{0}) = 0$, $(\varphi, \psi) \in \mathcal{K}_D$. In addition, thanks to Remark S1 we know that the triple $(\gamma_T, \varphi, \psi)$ satisfies (iii) of Proposition S3. This is enough to ensure optimality of γ_T. □

2 Proofs of Section 3: Wasserstein distance between Lévy measures

Proposition S4 (Proposition 1)**.** *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 Borel 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 S2 and on its preservation under weak* limits in Lemma S1. 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

$$\begin{aligned}&\sup_k \left| \int_{\Omega_{2d}} \phi(\mathbf{s}) g_m(\mathbf{s}') \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}') - \int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}') \right| \\ \leq &\sup_k \int_{\Omega_{2d}} |\phi(\mathbf{s})| \mathbb{1}_{A_m}(\mathbf{s}') \, \mathrm{d}\gamma_{n_k}(\mathbf{s}, \mathbf{s}'),\end{aligned}$$

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, $\int_{\Omega_{2d}} \phi(\mathbf{s}) \, \mathrm{d}\gamma(\mathbf{s}, \mathbf{s}') = \int_{\Omega_d} \phi(\mathbf{s}) \, \mathrm{d}\nu^1(\mathbf{s})$. 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 S1 the support of γ^* has a c-cyclically monotone. Next we claim that $\{\mathbf{0}\} \in \mathrm{supp}(\gamma^*)$, so that by Proposition S2 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, $+\infty = \gamma^*(\Omega_{2d}) \leq \gamma^*([\epsilon, +\infty)^d \times [0, +\infty)^d) + \gamma^*([0, +\infty)^d \times [\epsilon, +\infty)^d) = \nu^1([\epsilon, +\infty)^d) + \nu^1([\epsilon, +\infty)^d) < +\infty$. 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. □

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 pseudo inverse as $U_\nu^{-1}(t) = \inf_{x \geq 0}\{x \,:\, U_\nu(x) \geq t\}$. The function U_ν is non-increasing and right-continuous. Moreover, one can easily check that for any $a \geq 0$,

$$U_\nu^{-1}(t) > a \iff t < U_\nu(a). \tag{7}$$

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

Lemma S4. *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 $\{x > 0 \ : \ U_\nu(x) = r\}$ 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. □

We can now prove the Proposition 2 in the main manuscript.

Proposition S5 (Proposition 2)**.** *Let $\nu^1, \nu^2 \in \mathcal{M}_2(\Omega_1)$ and γ the restriction of $(U^{-1}_{\nu^1}, U^{-1}_{\nu^2})_\#\text{Leb}(\Omega_1)$ to Ω_2, where $\text{Leb}(\Omega_1)$ is the Lebesgue measure on Ω_1. Then $\gamma \in \overline{\Gamma}(\nu^1, \nu^2)$ is the unique optimal transport coupling. Moreover, if ν^1 is atomless and $\nu^1(\Omega_1) \geq \nu^2(\Omega_1)$, the mapping $T(x) = U^{-1}_{\nu^2}(U_{\nu^1}(x))$ is an optimal transport map between ν^1 and ν^2 and*

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

Proof. Thanks to Proposition S1 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 $(x', y') \in \text{supp}(\gamma^*)$ 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:

$$\begin{cases}
\gamma^*((a, +\infty) \times (b, +\infty)) = \min[U_{\nu^1}(a), U_{\nu^2}(b)] & \forall a, b > 0, \\
\gamma^*((a, +\infty) \times [0, +\infty)) = U_{\nu^1}(a) & \forall a > 0, \\
\gamma^*([0, +\infty) \times (b, +\infty)) = U_{\nu^2}(b) & \forall b > 0,
\end{cases} \tag{8}$$

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 (8) is in fact the optimal coupling.

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

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

From this we can use (7) to rewrite the quantity of interest as

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

In addition, as

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

and the second inequality is always satisfied, we deduce once again thanks to (7) that $\gamma((a,+\infty)\times[0,+\infty)) = U_{\nu^1}(a)$. The case when we exchange the marginals is similar. We conclude that γ satisfies (8) and thus it is the optimal coupling.

Eventually, let's turn to the case of where ν^1 is atomless and $\nu^1(\Omega_1) \geq \nu^2(\Omega_1)$, and let us define $T(x) = U^{-1}_{\nu^2}(U_{\nu^1}(x))$. 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 (8). Let's take $a, b \geq 0$ and we write:

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

where again we have use (7) 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 S4 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 S4. 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

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

where the second to last equality follow from Lemma S4 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 (8), 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. □

3 Proofs of Section 4: Evaluation of the index

Proposition 3 is implied by Proposition S6 and Corollary S1.

Proposition S6. *Let $\nu \in \mathcal{M}_2(\Omega_d)$ be a Lévy measure with equal marginals $\pi_{i\#}\nu = \overline{\nu}$ on Ω_1, for $i = 1, \dots, d$. Denote by $\nu^+ = \Sigma_\# \nu \in \mathcal{M}_2(\Omega_1)$, where $\Sigma(\mathbf{s}) = \sum_{i=1}^d s_i$. Then*

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = 2dM_2(\overline{\nu}) - 2\int_0^{+\infty} U^{-1}_{\nu^+}(s)\, U^{-1}_{\overline{\nu}}(s)\, \mathrm{d}s.$$

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 $\max(\mathbf{s}) = \max(s_1, \dots, s_d)$.

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 S5 there

exists an optimal transport map $T^+ = \nabla u$ between ν^+ and $\bar{\nu}$. We prove that $T(\mathbf{s}) = (T^+(\Sigma(\mathbf{s})), \dots, T^+(\Sigma(\mathbf{s})))$ is a transport map between ν and ν^{co}. We define $\Delta(\mathbf{s}) = (s, \dots, s)$ 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 $\overline{\nu}$, $T_{\#}\nu = \Delta_{\#}\overline{\nu} = \nu^{\text{co}}$ 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 S3, 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 Lemma S5 there exists a unique optimal transport plan $\gamma^+ = (U^{-1}_{\nu^+}, U^{-1}_{\overline{\nu}})_{\#}\text{Leb}(\Omega_1)$ 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 $\mathbf{s} \in \text{supp}(\nu^+)$. 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 $\gamma \in \overline{\Gamma}(\nu, \nu^{\text{co}})$, 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 $f \in C_c(\Omega_d)$,

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

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 $\Delta_{\#}\overline{\nu} = \nu^{\text{co}}$.

Consider the convex function u, u^* arising from the dual formulation of the transport problem between ν^+ and $\bar{\nu}$, as in Lemma S2. We claim that (φ, ψ) in the statement are an optimal pair of Kantorovich potentials and γ is the optimal transport coupling. We will use Proposition S3. We observe that $(\mathbf{s}, \mathbf{s}') \in \text{supp}(\gamma)$ if and only if the following conditions hold: (i) $\mathbf{s} \in \text{supp}(\nu)$; (ii) $\mathbf{s}' = (s', \cdots, s')$ for some $s' \in \Omega_1$; (iii) $(\Sigma(\mathbf{s}), \mathbf{s}') \in \text{supp}(\gamma^+)$. Thus in this case $\mathbf{s} \cdot \mathbf{s}' = \Sigma(\mathbf{s})s'$. Using the implication (i) $\Rightarrow$ (iii) from Proposition S3 for the transport from ν^+ onto $\overline{\nu}$ (combined with Remark S1), 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 (4) actually holds for the $(\mathbf{s}, \mathbf{s}')$ that we chose in $\text{supp}(\gamma)$. Using this time the implication (iii) $\Rightarrow$ (i) from Proposition S3 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

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

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 $\gamma^+ = (U^{-1}_{\nu^+}, U^{-1}_{\overline{\nu}})_{\#}\text{Leb}(\Omega_1)$. □

Corollary S1. *Let $\nu^\perp \in \mathcal{M}_2(\Omega_d)$ be an independent Lévy measure with equal marginals $\pi_{i\#}\nu = \overline{\nu}$ on Ω_1, for $i = 1, \dots, d$. Then,*

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

Proof. Let $\nu = \nu^\perp$ indicate the Lévy measure. By looking at the support of ν, we observe that $U_\nu^+(s) = d\, U_{\overline{\nu}}(s)$. By absolute continuity of $\overline{\nu}$, it follows that $\nu^+(s) = -\partial/\partial s\, U_\nu^+(s) = -d\, \partial/\partial s\, U_{\overline{\nu}}(s) = d\, \overline{\nu}(s)$ is atomless. Thus, with a change of variable $s = U_{\nu^+}^{-1}(t)$, the expression in Proposition S6 becomes

$$\mathcal{W}_*(\nu, \nu^{\mathrm{co}})^2 = 2dM_2(\overline{\nu}) - 2\int_0^{+\infty} s\, U_{\overline{\nu}}^{-1}(U_{\nu^+}(s))\nu^+(s)\, \mathrm{d}s. \tag{9}$$

We conclude by substituting the expression of ν^+ and U_ν^+ above. □

4 Proofs of Section 2: Main result

Given the expression of the index in (6) of the main manuscript, Theorem 1 and Remark 1 are easily implied by the following result.

Theorem S2. *Let $\nu \in \mathcal{M}_2(\Omega_d)$ with equal marginals $\pi_{i\#}\nu = \overline{\nu}$ on Ω_1, for $i = 1, \dots, d$. Then $\mathcal{W}_*(\nu, \nu^{\mathrm{co}}) \leq \mathcal{W}_*(\nu^\perp, \nu^{\mathrm{co}})$. If $\overline{\nu}(\Omega_1) = +\infty$, 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 Proposition S6. Then,

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

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 Proposition S6, $\varphi(\mathbf{s}) = \frac{1}{2}\sum_{i=1}^d s_i^2 - u(\Sigma(\mathbf{s}))$, 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,

$$\int_{\Omega_d} \varphi(\mathbf{s})\, \mathrm{d}(\nu^\perp - \nu)(\mathbf{s}) = \int_{\Omega_d} \bigg(u(\Sigma(\mathbf{s})) - \sum_{i=1}^d u(s_i)\bigg)\, \mathrm{d}\nu(\mathbf{s}), \tag{10}$$

which is non-negative thanks to superadditivity of convex functions, see Lemma S5 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 (10), 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 S5 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 S3 and Remark S1 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. □

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

Lemma S5. *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,

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

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. □

5 Proofs of Section 5: Examples

Lemma 1 easily follows from the following.

Lemma S6. *Let $\tilde{\boldsymbol{\mu}}^1$ and $\tilde{\boldsymbol{\mu}}^2$ be independent homogeneous completely random vectors with same base measure α and Lévy measures ν^1, ν^2, respectively. Then $\tilde{\boldsymbol{\mu}}^1 + \tilde{\boldsymbol{\mu}}^2$ is a completely random vector with base measure α and Lévy measure $\nu^1 + \nu^2$.*

Proof. The law of a completely random vector 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 Borel 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 $\mathbb{E}(e^{-\boldsymbol{\lambda}\,(\tilde{\boldsymbol{\mu}}^1+\tilde{\boldsymbol{\mu}}^2)(A)})$. □

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 S7 (Proposition 4)**.** *Let $\tilde{\boldsymbol{\mu}}$ be a d-dimensional additive completely random vector of parameter z. Then $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq z$ and*

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

Proof. Let $\nu = \nu_z$. The expression in the denominator follows by Corollary S6. 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 $I_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}) \geq 1 - (1-z) = z$. □

We define the following quantities.

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

Proposition S8 (Proposition 5)**.** *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 S6,

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

Let $\nu = \nu^{\phi}$. Since ν is diffuse, ν^+ is atomless and we may apply (9). 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

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

By definition of pushforward measure,

$$\begin{aligned} U_{\nu^+}(t) &= \int_{(0,+\infty)^d} \mathbb{1}_{(t,+\infty)}(s_1 + \cdots + s_d)\nu(\boldsymbol{s})\,\mathrm{d}s_1 \ldots \mathrm{d}s_d \\ &= \int_0^1 \frac{(1-u)^{\phi-1}}{u}\Big(\int_{(0,+\infty)^d} \mathbb{1}_{\left(\frac{t}{u},+\infty\right)}(v_1 + \cdots + v_d)\prod_{i=1}^d p_\phi(v_i)\,\mathrm{d}v_1 \ldots \mathrm{d}v_d\Big)\mathrm{d}u, \end{aligned}$$

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$(d\phi, 1)$,

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

The expression of v^+ easily derives by differentiating U_{ν^+}. □

References

Figalli, A. and Gigli, N. (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(2):107–130.

Guillen, N., Mou, C., and Święch, A. (2019). Coupling Lévy measures and comparison principles for viscosity solutions. *Transactions of the American Mathematical Society*, 372(10):7327–7370.

Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94.

Villani, C. (2003). *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Society.