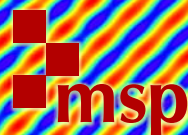


PURE and APPLIED ANALYSIS

PAM

YOUNG-HEON KIM, BRENDAN PASS AND DAVID J. SCHNEIDER

**OPTIMAL TRANSPORT AND BARYCENTERS FOR DENDRITIC
MEASURES**



vol. 2 no. 3 2020

OPTIMAL TRANSPORT AND BARYCENTERS FOR DENDRITIC MEASURES

YOUNG-HEON KIM, BRENDAN PASS AND DAVID J. SCHNEIDER

We introduce and study a variant of the Wasserstein distance on the space of probability measures, specially designed to deal with measures whose support has a dendritic, or tree-like structure with a particular direction of orientation. Our motivation is the comparison of and interpolation between plants' root systems. We characterize barycenters with respect to this metric, and establish that the interpolations of root-like measures, using this new metric, are also root-like, in a certain sense; this property fails for conventional Wasserstein barycenters. We also establish geodesic convexity with respect to this metric for a variety of functionals, some of which we expect to have biological importance.

1. Introduction

We introduce a new metric on the space of probability measures, the layerwise-Wasserstein distance. The motivation for this work is the need for a sound mathematical framework for describing the structure and diversity of dendritic structures in anisotropic environments. In particular, we are interested in the macroscopic structure of plant root systems developing under the influence of gravity and the stratification of chemical constituents, texture and microbial activity characteristic of soils. This biophysical context can be readily translated into mathematical terms. Plant tissues are composed of cells that physically partition $\mathbb{R}^3$ into two connected components – the “inside” and “outside”. The resulting structure roughly corresponds to a CW complex (see, e.g., [Hatcher 2002]) describing the topology of the plant. Ignoring complex features present at microscopic scales, the external surface can be viewed as a smooth, connected two-dimensional manifold with genus zero embedded in $\mathbb{R}^3$. Computational representations of these external surfaces can be reconstructed using standard methods of optical, X-ray and neutron tomography.

This idealization misses two essential points: the above- and below-ground portions of plants display intricate structural forms that are remarkably resistant to quantitative analysis, and the form and function of these complicated structures are intimately related to the anisotropic environment in which they develop. The first condition implies the need to handle arbitrarily complicated distributions of mass in space subject to very modest restrictions on the behavior of the surface, while the second suggests the need to handle preferred directions in space.

Natural challenges include quantifying the difference between two or more roots, summarizing or describing the typical structure of a family of root systems (for instance, the roots of several genetically identical plants, grown in nearly identical environments, which often exhibit considerable variation in their structure) in a succinct way, quantifying the variation within that family and comparing the

MSC2010: primary 49K99; secondary 92C80.

Keywords: optimal transport, dendritic measures, layerwise-Wasserstein distance, plant root architecture.

structure exhibited by one family to another. A typical approach to these problems is to compute a family of *phenotypes* for each system (including, for example, total root length, rooting depth, and various topological invariants, such as the Horton–Strahler index) and compare and average among them; see, for instance, [Clark et al. 2011; 2013; Famoso et al. 2010; 2011; Piñeros et al. 2016]. Though this has met with some success in distinguishing between particular choices of root systems, it is not generally clear which phenotypes are most useful for this purpose, and the choice in different applications is often done in an ad hoc way. For virtually any collection of phenotypes, it is not hard to come up with drastically different root shapes sharing the same phenotypes.

Our approach focuses on roots as mass distributions in $\mathbb{R}^3$ where the vertical and horizontal directions have distinct roles (roots which are related by a rotation about the vertical axis are considered identical). Natural mathematical goals include constructing a metric between root shapes reflecting both their downward-pointing dendritic topology as well as the distances and sizes in the underlying space,¹ and producing a representative of a family of root systems which captures the average or typical structure of the family. After normalization for overall mass, root systems can be modeled as probability distributions; the Wasserstein distance from optimal transport [Villani 2003; 2009; Santambrogio 2015] is then one candidate for such a metric, and Wasserstein barycenters (Fréchet means with respect to this metric; see [Agueh and Carlier 2011]) are a corresponding candidate for a representative of a family. While this metric has proved fruitful in related problems involving comparing and averaging among shapes (image processing, for instance), we demonstrate in this paper that it is not ideally adapted to the downward dendritic structure prominent among root systems, in large part because optimal matchings don’t generally exhibit monotonicity in the distinguished, vertical direction. While it is possible to incorporate vertical stratification in the usual definition of Wasserstein distance by penalizing transport in the vertical direction, the practical application of this formalism is limited by computational requirements. We propose a simple alternative based on a related metric, the layerwise-Wasserstein metric, derived from a variant of optimal transport in which monotonicity in the distinguished vertical direction is guaranteed; see Definition 2.2. The metric barycenter arising from this new metric is a natural candidate for a representative of a family of root systems. Furthermore, we suspect this distance may play a role in other applied problems featuring both tree-like and geometric structures (blood vessels in biology, river systems in topography, etc.). Our primary present goal is to develop the mathematical properties of the layerwise-Wasserstein distance and its interpolants, while the biological and methodological applications will be developed in subsequent work. However, we keep the motivating applications in mind as we go, and focus on properties of root systems that have potential biological relevance.

It is common in biology to model root systems by their *skeletons*, in which three-dimensional limbs are replaced by approximating one-dimensional curves [Bucksch 2014]; these skeletons retain the dendritic, or tree-like, structure of the root, but strip away its thickness (which is less crucial in some applications). As a corresponding mathematical object we introduce *skeletal measures*, which are essentially mass distributions supported on these skeletal structures; see Section 3. This gives a useful framework for

¹Ideally, the metric should detect geometric differences, between, for instance, a short limb and a long one, as well as topological differences, between say, a forked limb and a straight one.

studying the topological properties of roots and their interpolations, while avoiding difficulties that arise when dealing with their (more realistic) three-dimensional structure.

When building interpolants to use as representatives of families of roots, a desirable property is that the dendritic structure is preserved: given several root systems, does their metric barycenter look like a root? We are able to give a fairly satisfactory affirmative answer to this question for skeletal root systems, using our layerwise-Wasserstein distance as the metric; see Theorem 3.6. On the other hand, we exhibit examples illustrating that when the conventional Wasserstein distance is used, interpolants of root systems may not resemble root systems at all; more precisely, we show that the Wasserstein barycenter of several skeletal roots can have high-dimensional support, so that the dendritic structure is broken; see Section 3B. We also establish comparisons between the total root length (essentially the one-dimensional Hausdorff measure of the support) of several skeletal root measures and their layerwise-Wasserstein barycenter; see Proposition 3.14. This type of result is impossible in general with the Wasserstein barycenter, as the support may be more than one-dimensional.

Aside from being natural for certain applications, the layerwise-Wasserstein distance also has computational advantages over its classical Wasserstein counterpart in certain situations, as the sorting in the distinguished direction is monotone, and so optimization problems arise only in spaces of codimension 1. In $\mathbb{R}^2$, for instance, the layerwise-Wasserstein distance essentially corresponds to the Knothe–Rosenblatt rearrangement [Knothe 1957; Rosenblatt 1952], which can be computed much more easily than the two-dimensional Wasserstein distance; previous connections between the Knothe–Rosenblatt rearrangement and optimal transport have been established in [Carlier et al. 2010] (where the Knothe–Rosenblatt rearrangement was shown to arise as a limit of optimal transport maps with anisotropic costs) and in [Backhoff et al. 2017] (where an interpolation of probability measures using the Knothe–Rosenblatt rearrangement was introduced).² More generally, the layerwise-Wasserstein distance is a special instance of the Monge–Knothe maps recently introduced in [Muzellec and Cuturi 2019]; in that work, properties of the corresponding metric, including interpolation between measures and convexity were not studied.

We also note that our layerwise-Wasserstein distance is similar in spirit to the Radon–Wasserstein distance found in [Bonneel et al. 2015], as both approaches involve disintegrating the measures and transporting their fibers. The difference lies in how the measures are disintegrated; we disintegrate with respect to a distinguished, vertical variable on the underlying space (which is natural in the applications we have in mind), whereas the disintegration in [Bonneel et al. 2015] is done with respect to Radon transformed variables.

In addition, it is worth commenting briefly on the relationship between this work and another recent series of papers relating optimal transport to plant root shapes [Bressan and Sun 2018; Bressan et al. 2020]. In those works, the objective is to identify and characterize root (and tree) shapes which optimize certain functionals, modeling absorption of nutrients and sunlight and the cost (via ramified optimal transport) of returning those nutrients to the base of plant, whereas our goal is to differentiate and interpolate between various root systems.

²Interpolating between two-dimensional measures is in fact not merely a mathematical simplification or toy model, but has actual agricultural applications, since experiments are sometimes done growing plants between two panes of glass, placed very close together, resulting in essentially two-dimensional root shapes.

The manuscript is organized as follows. In Section 2 we introduce the layerwise Wasserstein distance and barycenters, and establish some basic properties. Section 3 focuses on skeletal measures, while Section 4 is devoted to layerwise displacement interpolation and convexity.

2. Layerwise-Wasserstein distance

Let $M^+(X)$, denote the space of finite, nonnegative Borel measures with positive total mass (i.e., $\mu(X) > 0$ for each $\mu \in M^+(X)$) on a metric space X equipped with the weak-* topology. Let $P(X)$ denote the space of finite, nonnegative Borel probability measures on a metric space X equipped with the weak-* topology. Consider $M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ and let $M_{ac}^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ be its subset consisting of absolutely continuous measures (with respect to Lebesgue); let $P_{ac}(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ be the corresponding subset of probability measures. For $\mu \in M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, let μ^V be its vertical marginal, defined by

$$\int_{\mathbb{R}_{\geq 0}} f(z) \mu^V(dz) = \int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} f(z) \mu(dx, dy) \quad \text{for all } f \in C(\mathbb{R}_{\geq 0}).$$

Note that $|\mu^V| = |\mu|$, where $|\mu|$ denotes the total mass of μ . The following vertical rescaling of the measures in $M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ is a key step in our construction of the Wasserstein-type distance that uses the distinguished coordinate $\mathbb{R}_{\geq 0}$. Note also that measures may not necessarily have the same mass, so we also normalize them to be probability measures.

Definition 2.1 (vertical rescaling). Given $\mu \in M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, we define its *vertically rescaled version*, namely,

$$\tilde{\mu} \in P(\mathbb{R}^d \times [0, 1]),$$

as follows: Let $F_\mu : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be the cumulative function given by

$$F_\mu(y) = \frac{1}{|\mu|} \mu^V([0, y]).$$

Note that $(F_\mu)_\# \mu^V = |\mu| \mathcal{L}^1$, and F_μ is continuous for absolutely continuous μ^V .

Then, define

$$\tilde{\mu} = \frac{1}{|\mu|} (\text{id} \times F_\mu)_\# \mu,$$

where $\text{id} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the identity map. Notice that the map $\mu \mapsto \tilde{\mu}$ from $M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ to $P(\mathbb{R}^d \times [0, 1])$ is continuous with respect to the weak-* topology. In particular, this map pushes forward a given $\Omega \in P(M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$ to its vertically rescaled version

$$\tilde{\Omega} \in P(P(\mathbb{R}^d \times [0, 1])).$$

Note that the mapping F_μ depends on μ only through its vertical marginal, μ^V ; we will sometimes abuse notation and write F_{μ^V} instead of F_μ .

This normalization allows us to define a Wasserstein-type distance that uses the disintegration along the vertical line. In the following, W_2^2 denotes the quadratic Wasserstein distance.

Definition 2.2 (layerwise-Wasserstein distance). Given $\mu, \nu \in M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, define

$$d_{\text{LW}}^2(\mu, \nu) = W_2^2\left(\frac{1}{|\mu^V|}\mu^V, \frac{1}{|\nu^V|}\nu^V\right) + \int_0^1 W_2^2(\tilde{\mu}_l, \tilde{\nu}_l) dl, \quad (2-1)$$

where $\tilde{\mu}$ and $\tilde{\nu}$ have disintegrations $\tilde{\mu}(dx, dl) = \tilde{\mu}_l(dx) dl$, $\tilde{\nu}(dx, dl) = \tilde{\nu}_l(dx) dl$ with respect to the Lebesgue measure dl on $[0, 1]$.

Remark 2.3. We note that strictly speaking d_{LW}^2 does not give a distance on $M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, unless restricted to $P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, as different measures may have the same vertical rescaling $((1/|\mu^V|)\mu^V, \tilde{\mu})$; instead, it gives a metric on the set of equivalence classes, under the equivalence relation $\mu \sim \nu$ if $\mu/|\mu| = \nu/|\nu|$. To get a distance on $M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ one may add $(|\mu| - |\nu|)^2$ and consider the metric

$$W_2^2\left(\frac{1}{|\mu^V|}\mu^V, \frac{1}{|\nu^V|}\nu^V\right) + \int_0^1 W_2^2(\tilde{\mu}_l, \tilde{\nu}_l) dl + (|\mu| - |\nu|)^2.$$

In the following, however, we stick to (2-1) for simplicity (in fact, in subsequent sections, we restrict our attention entirely to $P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$).

We now consider the metric barycenter corresponding to the layerwise-Wasserstein distance (2-1),³ which we define below, and call it the layerwise-Wasserstein barycenter.

Definition 2.4 (layerwise Wasserstein barycenter). For $\Omega \in P(M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$, a *layerwise Wasserstein barycenter* $\text{Bar}^{\text{LW}}(\Omega) \in P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ is defined as an element of

$$\operatorname{argmin}_{\mu \in P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} d_{\text{LW}}^2(\mu, \nu) d\Omega(\nu).$$

To characterize layerwise-Wasserstein barycenters, we need a little more terminology. Define $\tilde{\Omega}_l := (\nu \mapsto \tilde{\nu}_l)_{\#}\Omega$. A Wasserstein barycenter of $\tilde{\Omega}_l$ is then a minimizer over $P(\mathbb{R}^d)$ of

$$\eta \mapsto \int_{P(\mathbb{R}^d)} W_2^2(\eta, \alpha) d\tilde{\Omega}_l(\alpha) = \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} W_2^2(\eta, \tilde{\nu}_l) d\Omega(\nu). \quad (2-2)$$

Similarly, defining $\Omega^V := (\nu \mapsto \nu^V/|\nu^V|)_{\#}\Omega$, a Wasserstein barycenter of Ω^V is a minimizer over $P(\mathbb{R}_{\geq 0})$ of

$$\eta \mapsto \int_{P(\mathbb{R}_{\geq 0})} W_2^2(\eta, \alpha) d\Omega^V(\alpha) = \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} W_2^2\left(\eta, \frac{\nu^V}{|\nu^V|}\right) d\Omega(\nu).$$

We then have the following:

Proposition 2.5. A measure $\mu \in P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ is a layerwise-Wasserstein barycenter of $\Omega \in P(M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$ if and only if its vertical marginal μ^V is a Wasserstein barycenter of Ω^V and for almost every layer l , $\tilde{\mu}_l$ is a Wasserstein barycenter of $\tilde{\Omega}_l$.

³Strictly speaking, given the remark above, the metric barycenter is an equivalence class of measures; we choose as a representative the unique probability measure in a given class.

Proof. By definition, a layerwise Wasserstein barycenter μ must minimize

$$\begin{aligned} \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} \left[W_2^2 \left(\mu^V, \frac{1}{|v^V|} v^V \right) + \int_0^1 W_2^2(\tilde{\mu}_l, \tilde{v}_l) dl \right] d\Omega(v) \\ = \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} W_2^2 \left(\mu^V, \frac{1}{|v^V|} v^V \right) d\Omega(v) + \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} \int_0^1 W_2^2(\tilde{\mu}_l, \tilde{v}_l) dl d\Omega(v) \\ = \int_{P(\mathbb{R}_{\geq 0})} W_2^2 \left(\mu^V, \frac{1}{|v^V|} v^V \right) d\Omega^V(v^V) + \int_0^1 \int_{P(\mathbb{R}^d)} W_2^2(\tilde{\mu}_l, \tilde{v}_l) d\tilde{\Omega}_l(\tilde{v}_l) dl. \end{aligned}$$

By changing μ^V and $\tilde{\mu}_l$ independently, we see that μ minimizes the last line if and only if its vertical marginal μ^V minimizes the first term and, for almost every l , $\tilde{\mu}_l$ minimizes $\int_{P(\mathbb{R}^d)} W_2^2(\tilde{\mu}_l, \tilde{v}_l) d\tilde{\Omega}_l(\tilde{v}_l)$; that is, μ^V is a Wasserstein barycenter of Ω^V and $\tilde{\mu}_l$ is a Wasserstein barycenter of $\tilde{\Omega}_l$. $\square$

The proposition gives a straightforward way to construct layerwise-Wasserstein barycenters; first construct the layers $\tilde{\mu}_l = \text{Bar}^W(\tilde{\Omega}_l)$ as Wasserstein barycenters of the $\tilde{\Omega}_l$. Then letting $\mu^V = \text{Bar}^W(\Omega^V)$ be the Wasserstein barycenter of Ω^V the layerwise Wasserstein barycenter $\mu = \text{Bar}^{\text{LW}}(\Omega)$ is defined by

$$d\mu(x, y) = d\tilde{\mu}_{F_{\mu^V}(y)}(x) d\mu^V(y).$$

Note that any $\text{Bar}^{\text{LW}}(\Omega)$ is written this way, and is uniquely determined if $\tilde{\mu}_l$ is uniquely determined for a.e. l . In particular, we have:

Corollary 2.6. *For $\Omega \in P(M_{\text{ac}}^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$, there is a unique $\mu = \text{Bar}^{\text{LW}}(\Omega)$.*

Proof. As $\Omega \in P(M_{\text{ac}}^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$, it also holds that $\tilde{\Omega}_l \in P(P_{\text{ac}}(\mathbb{R}^d))$ for a.e. l . Then uniqueness of $\tilde{\mu}_l$ follows from [Kim and Pass 2017]. $\square$

The rescaled version $\widetilde{\text{Bar}}^{\text{LW}}(\Omega) \in P(\mathbb{R}^d \times [0, 1])$ of the layerwise-Wasserstein barycenter $\text{Bar}^{\text{LW}}(\Omega)$ has the disintegration $d\widetilde{\text{Bar}}^{\text{LW}}(\Omega)(x, l) = d\widetilde{\text{Bar}}_l^{\text{LW}}(\Omega) dl$, where each $\widetilde{\text{Bar}}_l^{\text{LW}}(\Omega)$ is a Wasserstein barycenter of the $\tilde{\Omega}_l$. The rescaling mapping $F_{\text{Bar}^{\text{LW}}(\Omega)}$ satisfies

$$F_{\text{Bar}^{\text{LW}}(\Omega)}(y) = \left[\int F_v^{-1} d\Omega(v) \right]^{-1}(y). \quad (2-3)$$

We note here associativity, in the two-dimensional case, i.e., on $\mathbb{R} \times \mathbb{R}_{\geq 0}$, of the layerwise-Wasserstein barycenter of probability measures $\mu_1, \dots, \mu_N$ with weights $\lambda_1, \dots, \lambda_N$, where $\sum_{i=1}^N \lambda_i = 1$ and each $\lambda_i \geq 0$.

Proposition 2.7 (associativity of two-dimensional layerwise-Wasserstein barycenters). *Assume that $d = 1$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Then*

$$\text{Bar}^{\text{LW}}(\lambda_1 \delta_{\mu_1} + \lambda_2 \delta_{\mu_2} + \lambda_3 \delta_{\mu_3}) = \text{Bar}^{\text{LW}}((\lambda_1 + \lambda_2) \delta_{\text{Bar}^{\text{LW}}(\delta_{\mu_1} \lambda_1 / (\lambda_1 + \lambda_2) + \delta_{\mu_2} \lambda_2 / (\lambda_1 + \lambda_2))} + \lambda_3 \delta_{\mu_3}).$$

This proposition is potentially useful in certain computations. For example, when one adds a new sample μ_{N+1} root system to a family of N root systems with a (previously computed) barycenter $\bar{\mu}$, one can find the barycenter of the augmented family by computing the appropriately weighted barycenter of μ_{N+1} and $\bar{\mu}$, rather than the more difficult computation of the barycenter of the new family of $N + 1$ systems.

Proof. The result follows immediately from the corresponding result in one dimension for Wasserstein barycenters. $\square$

Remark 2.8. In our motivating application, we only distinguish between root systems up to rotation about the vertical axis; that is, we wish to identify two systems whenever we can transform one system to the other via a rotation fixing y . For actual root systems then, the following distance is relevant:

Definition 2.9 (Horizontally symmetrized layerwise-Wasserstein distance). We define the horizontally symmetrized layerwise-Wasserstein distance $d_{\text{LW,symm}}^2(\mu, \nu)$ between μ and ν by

$$d_{\text{LW,symm}}^2(\mu, \nu) = \min_{R \in \text{SO}(d)} d_{\text{LW}}^2(R_{\#}\mu, \nu),$$

where $\text{SO}(d)$ denotes the special orthogonal group on the horizontal directions $\mathbb{R}^d$.

Note that $d_{\text{LW,symm}}$ is a metric on the set of equivalence classes of probability measures under horizontal rotational equivalence (that is, $\nu \sim \mu$ if $\nu = R_{\#}\mu$ for some rotation $R \in \text{SO}(d)$). A horizontally symmetrized Wasserstein barycenter $\text{Bar}_{\text{symm}}^{\text{LW}}(\Omega)$ of a measure $\Omega \in P(M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$ is then a metric barycenter with respect to this distance; that is, it is a minimizer of

$$\nu \mapsto \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} d_{\text{LW,symm}}^2(\nu, \mu) d\Omega(\mu).$$

Equivalently, $\text{Bar}_{\text{symm}}^{\text{LW}}(\Omega)$ minimizes

$$\nu \mapsto \min_{\substack{R_{\mu} \in \text{SO}(d) \\ \text{for all } \mu \in M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})}} \int_{M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} d_{\text{LW}}^2(\nu, (R_{\mu})_{\#}\mu) d\Omega(\mu).$$

Analogously, one could also consider rotationally symmetrized versions of the classical Wasserstein distance,

$$W_{2,\text{symm}}^2(\mu, \nu) := \min_{R \in \text{SO}(d)} W_2^2(\mu, R_{\#}\nu),$$

and corresponding barycenters, which are minimizers of

$$\nu \mapsto \min_{\substack{R_{\mu} \in \text{SO}(d) \\ \text{for all } \mu \in M^+(\mathbb{R}^d \times \mathbb{R}_{\geq 0})}} \int_{P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})} W_2^2(\nu, (R_{\mu})_{\#}\mu) d\Omega(\mu). \quad (2-4)$$

Symmetrized Wasserstein barycenters are more natural for the root interpolation problem than classical Wasserstein barycenters. One of our goals in this paper is to demonstrate that symmetrized layerwise-Wasserstein barycenters are better suited for this problem than classical (symmetrized or unsymmetrized) Wasserstein barycenters; to this end, we provide examples in Section 3B of measures $\mu_1, \dots, \mu_m$ which are root-like in a certain sense (skeletons in the nomenclature of the next section), for which the symmetrized Wasserstein barycenter of $(1/m) \sum_{i=1}^m \delta_{\mu_i}$ does not resemble a root (that is, is not a skeleton). Their layerwise-Wasserstein barycenter, on the other hand, has a much more root-like structure (see Theorem 3.6 below).

3. Skeletal measures

Real plant root systems consist of limbs with thickness. However, biologists often approximate roots by their “skeletons”, in which each limb is replaced by a one-dimensional curve, thus retaining the topological, or dendritic structure of the root, but losing its thickness. Below, we provide a formal mathematical definition of skeletons, and introduce skeletal measures, which are essentially distributions of mass supported on them.

Definition 3.1. Let $Y = [0, \bar{y}] \subset \mathbb{R}$, be an interval whose length $\bar{y}$ represents the vertical depth of the root. A *weak skeletal root* consists of the graphs of a finite union of curves,

$$\bigcup_{i=1}^N \text{graph}(g_i),$$

where each $g_i : [\underline{y}_i, \bar{y}_i] \rightarrow \mathbb{R}^d$ is a Lipschitz function defined on a subinterval $[\underline{y}_i, \bar{y}_i] \subseteq Y$, satisfying the following properties:

- (S1) (roots start from a common stem) $\underline{y}_1 = 0$ and $\underline{y}_i > 0$ for each $i = 2, \dots, N$.
- (S2) (limbs emerge from older limbs) For each $i = 2, \dots, N$, there is some $j < i$ such that $\underline{y}_i \in (\underline{y}_j, \bar{y}_j)$ and $g_i(\underline{y}_i) = g_j(\underline{y}_i)$.

A *strong skeletal root* is a weak skeletal root which satisfies the additional condition:

- (S3) (limbs never cross each other) For each $i \neq j$ and all $y \in (\underline{y}_i, \bar{y}_i] \cap (\underline{y}_j, \bar{y}_j]$, we have $g_i(y) \neq g_j(y)$.

We next define strong skeletal root measures.

Definition 3.2. A *strong skeletal root measure* is a probability measure whose support is an *entire* strong skeletal root, which is absolutely continuous with respect to the one-dimensional Hausdorff measure.

Strong skeletal root measures seem to be reasonable proxies for real roots. As we will see below, layerwise-Wasserstein barycenters of strong skeletal root measures preserve the one-dimensional structure of the support (this is an important distinction from conventional Wasserstein barycenters — see Example 3.12 below). Unfortunately, they are not always strong skeletal root measures for two reasons: the support may be disconnected, and the noncrossing property holds only in a weaker sense. This motivates the following definition:

Definition 3.3. A *weak skeletal root measure* is a probability measure supported on a weak skeletal root, which is absolutely continuous with respect to the one-dimensional Hausdorff measure, satisfying the following additional property:

- (W3) For each $i \neq j$ and all $y \in (\underline{y}_i, \bar{y}_i] \cap (\underline{y}_j, \bar{y}_j]$, such that $g_i(y) = g_j(y)$, we have either

$$\lim_{z \rightarrow y^-} \mu_z(\{g_i(z)\}) = 0 \quad \text{or} \quad \lim_{z \rightarrow y^-} \mu_z(\{g_j(z)\}) = 0,$$

where $\mu_y = \tilde{\mu}_{F_\mu(y)}$ is the conditional probability of $d\mu(x, y) = d\mu_y(x) d\mu^V(y)$.

Note that by construction, for each l , the layer $\tilde{\mu}_l$ of a weak skeletal root measure μ , is a convex combination of Dirac masses.

Obviously strong skeletal root measures are weak skeletal root measures; weak skeletal root measures are essentially “roots with missing parts” and have a weaker version (W3) of the noncrossing condition. While the layerwise-Wasserstein barycenter of several strong skeletal root measures may not be a strong skeletal root measure, we are able to show below that it is a weak skeletal root measure.

Remark 3.4. Interpreting each $\text{graph}(g_i)$ as a limb, condition (S3) expresses the natural expectation that limbs do not cross. We interpret (W3) as a weaker version of this: if $g_i(y) = g_j(y)$ and $\lim_{z \rightarrow y^-} \mu_z(\{g_i(z)\}) = 0$, we interpret g_i as consisting of two limbs: an upper limb g_i^1 , defined by restricting g_i to $[\underline{y}_i, y]$, and a lower limb, g_i^2 , obtained by restricting g_i to $[y, \bar{y}_i]$, emerging from the older limb y_j at the point y . This seems reasonable to us, since the hypothesis $\lim_{z \rightarrow y^-} \mu_z(\{g_i(z)\}) = 0$ means that there is no mass at y coming from the upper limb; the upper limb thus ends at the point y .

By interpreting a weak root as a tree in this sense, one can compute topological properties which are defined only for loop-free structures (including, for example, the Horton–Strahler index [Toroczkaï 2002], often used by biologists to measure the topological complexity of root systems).

Remark 3.5. Skeletons can be computationally useful in practice. Algorithms are available to construct skeletons from real root system data, essentially by tracing the center of mass of the cross sections of each limb [Bucksch 2014]. Computing layerwise-Wasserstein barycenters of these skeletons is then much less computationally intensive than computing the barycenters of the original roots, since each layer is discretized by many fewer points, but may still provide valuable biological insight about the “average” topological structure of the family of root systems.

Assuming μ is a (weak or strong) skeletal root measure, supported on the skeletal root $\bigcup_{i=1}^N \text{graph}(g_i)$, and the rescaling map F_μ is bi-Lipschitz, $\tilde{\mu}$ is also a (respectively weak or strong) skeletal root measure, supported on the skeletal root $\bigcup_{i=1}^N \text{graph}(\tilde{g}_i)$, where $\tilde{g}_i := g_i \circ F_\mu^{-1}$. Note that the domain $[\underline{l}_i, \bar{l}_i] := [F_\mu(\underline{y}_i), F_\mu(\bar{y}_i)]$ of each rescaled limb $\tilde{g}_i$ is contained in $[0, 1]$. We call $\bigcup_{i=1}^N \text{graph}(\tilde{g}_i)$ a *rescaled* skeletal root.

3A. Layerwise-Wasserstein barycenters of skeletal root measures. We now prove that layerwise-Wasserstein barycenters of weak skeletal root measures are themselves weak skeletal root measures.

Theorem 3.6. *Let $\mu_1, \dots, \mu_m \in P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$ be compactly supported weak skeletal root measures such that $l \mapsto (\tilde{\mu}_i)_l$ is weak-* left continuous and F_{μ_i} is bi-Lipschitz for each i , and $\lambda_1, \dots, \lambda_m > 0$ with $\sum_{i=1}^m \lambda_i = 1$. Then any layerwise-Wasserstein barycenter $\text{Bar}^{\text{LW}}(\sum_{\alpha} \lambda_{\alpha} \mu_{\alpha})$ of $\mu_1, \dots, \mu_m$ with weights $\lambda_1, \dots, \lambda_m$ is also a weak skeletal root measure.*

Remark 3.7. We expect this result to play an important role in biological applications. As mentioned above, given a family of root systems, we will propose in future work interpreting the layerwise-Wasserstein barycenter as the best representative of that family. It is therefore desirable to compute certain biologically relevant traits of the barycenter, especially those traits that rely on its dendritic structure, for instance the total root length and the Horton–Strahler (HS) index [Toroczkaï 2002]. The HS index in particular relies on the noncrossing property, and can be defined for weak skeletal root measures, thanks to (W3), but not for more general unions of graphs such as weak skeletal roots.

The key tool in the proof of this theorem is the barycentric ghost, which we define now.

Definition 3.8. For $\alpha = 1, 2, \dots, m$, let $\tilde{S}_\alpha := \{\tilde{g}_{i_\alpha}^\alpha : i_\alpha = 1, 2, \dots, N_\alpha\}$ be a rescaled skeletal root, and let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ with $\lambda_1, \dots, \lambda_m > 0$ be a collection of weights with $\sum_{\alpha=1}^m \lambda_\alpha = 1$.

For fixed indices $i_1, \dots, i_m$, whenever the intersection $\bigcap_{\alpha=1}^m [l_{i_\alpha}^\alpha, \bar{l}_{i_\alpha}^\alpha]$ of domains $[l_{i_\alpha}^\alpha, \bar{l}_{i_\alpha}^\alpha]$ of the family $\{\tilde{g}_{i_\alpha}^\alpha\}$ is nonempty, we define the curve

$$\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda := \sum_{\alpha=1}^m \lambda_\alpha \tilde{g}_{i_\alpha}^\alpha.$$

The *ghost* of the family $\{\tilde{S}_\alpha\}$ with weights λ is then the collection of curves $\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda$.

At each slice $l \in [0, 1]$, the set $\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda(l)$ represents the Euclidean barycenters of all possible combinations of $\tilde{g}_{i_\alpha}^\alpha(l)$ in the supports of the discrete sliced layers. The Wasserstein barycenter of the layers is supported on these points; therefore, for skeletal root measures $\mu_1, \dots, \mu_m$, supported respectively on $\bigcup_{i_\alpha=1}^{N_\alpha} \text{graph}(g_{i_\alpha}^\alpha)$, we have

if $(x, y) \in \text{supp}(\text{Bar}^{\text{LW}}(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha}))$,

$$\text{then } x = \tilde{G}_{i_1, i_2, \dots, i_m}^\lambda \left(\left(\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1} \right)^{-1}(y) \right) \text{ for some choice of } i_1, \dots, i_m. \quad (3-1)$$

Given probability root measures, the ghost of their rescaled supports $\bigcup_{i_\alpha=1}^{N_\alpha} \text{graph}(\tilde{g}_{i_\alpha}^\alpha)$ can be unrescaled via the mapping $y \mapsto \left(\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1} \right)^{-1}(y)$; the unrescaled ghost is then the union of the graphs

$$G_{i_1, i_2, \dots, i_m}^\lambda(y) := \tilde{G}_{i_1, i_2, \dots, i_m}^\lambda \left(\left(\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1} \right)^{-1}(y) \right).$$

It is then easy to see that the layerwise-Wasserstein barycenter has support contained in the (unrescaled) ghost, though it typically won't fill it out. We think of the ghost sitting in the background; it is the largest possible potential support of the barycenter. We think of the actual support of the barycenter as sitting in the foreground on top of it.

Now, the ghost clearly satisfies (S1) (starting as stem) and (S2) (limbs emerge from older limbs) in the definition of skeletal roots. It does not generally satisfy (S3) (noncrossing). In order to verify that the layerwise-Wasserstein barycenter is a weak skeletal root measure, we must therefore show that it satisfies the weak noncrossing property (W3).

The following lemma essentially verifies (W3) for the rescaled barycenter; since it is clear that the bi-Lipschitz rescaling $\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1}$, which pushes $\widetilde{\text{Bar}}^{\text{LW}}(\sum_{\alpha} \lambda_\alpha \mu_\alpha)$ forward to $\text{Bar}^{\text{LW}}(\sum_{\alpha} \lambda_\alpha \mu_\alpha)$, preserves this property, the lemma implies Theorem 3.6.

Lemma 3.9. *Under the same assumptions as in Theorem 3.6, let $\mu = \text{Bar}^{\text{LW}}(\sum_{\alpha} \lambda_\alpha \mu_\alpha)$ be a layerwise-Wasserstein barycenter of the $\{\mu_\alpha\}$'s. Set $l = \left(\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1} \right)^{-1}(y)$, and suppose $x = \tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(l) = \tilde{G}_{i_1, i_2, \dots, i_m}^\lambda(l)$, where $j_\alpha \neq i_\alpha$ for at least one α and l is not the minimal point in the domain of $\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda$ or $\tilde{G}_{j_1, j_2, \dots, j_m}^\lambda$. Moreover, suppose that $\lim_{z \rightarrow l^-} \tilde{\mu}_z(\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda(z)) > 0$. Then,*

$$\lim_{z \rightarrow l^-} \tilde{\mu}_z(\{\tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(z)\} \setminus \{\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda(z)\}) = 0.$$

The proof of this lemma leverages a connection between the Wasserstein barycenter of $\sum_{\alpha} \lambda_{\alpha} \delta_{(\tilde{\mu}_{\alpha})_l}$ and the multimarginal extension of optimal transport, which is to minimize

$$\int_{(\mathbb{R}^d)^m} \sum_{\alpha, \beta} \lambda_{\alpha} \lambda_{\beta} |x_{\alpha} - x_{\beta}|^2 d\gamma(x_1, x_2, \dots, x_m) \quad (3-2)$$

among all probability measures γ on $(\mathbb{R}^d)^m$ whose marginals are the $(\tilde{\mu}_{\alpha})_l$. It is well known that the mapping

$$\Delta_{\lambda} : (x_1, x_2, \dots, x_m) \rightarrow \sum_{\alpha} \lambda_{\alpha} x_{\alpha}$$

pushes each solution $\tilde{\gamma}_l$ forward to a Wasserstein barycenter $\tilde{\mu}_l$ [Agueh and Carlier 2011], and this mapping is invertible with a Lipschitz inverse on the support of $\tilde{\mu}_l$; see, e.g., [Kim and Pass 2017].

Proof of Lemma 3.9. For each layer l , we will denote by $\tilde{\gamma}_l \in P(\mathbb{R}^d \times \dots \times \mathbb{R}^d)$ a solution to the multimarginal optimal transport problem (3-2). Assume that the conclusion of the lemma fails. Then there exists $\epsilon > 0$ and a sequence $l_k < l$ converging to l such that $\tilde{G}_{j_1, j_2, \dots, j_m}^{\lambda}(l_k) \neq \tilde{G}_{i_1, i_2, \dots, i_m}^{\lambda}(l_k)$ and $\tilde{\mu}_{l_k}(\tilde{G}_{j_1, j_2, \dots, j_m}^{\lambda}(l_k)) > \epsilon$ for large enough k .

We first prove the lemma under the simplifying assumption that $\tilde{G}_{j_1, j_2, \dots, j_m}^{\lambda}(l_k) \neq \tilde{G}_{j'_1, j'_2, \dots, j'_m}^{\lambda}(l_k)$ for all $(j'_1, j'_2, \dots, j'_m) \neq (j_1, j_2, \dots, j_m)$.

This immediately implies

$$\tilde{\gamma}_{l_k}(\tilde{g}_{j_1}^1(l_k), \tilde{g}_{j_2}^2(l_k), \dots, \tilde{g}_{j_m}^m(l_k)) > \epsilon \quad (3-3)$$

and in particular

$$(\tilde{\mu}_{\alpha})_{l_k}(g_{j_{\alpha}}^{\alpha}(l_k)) > \epsilon \quad (3-4)$$

for $\alpha = 1, 2, \dots, m$. After passing to a convergent subsequence, the $\tilde{\gamma}_{l_k}$ converge (in the weak-* sense) to a measure $\tilde{\gamma}_l$ which is optimal in the multimarginal problem for the $(\tilde{\mu}_{i_{\alpha}})_l$, and

$$\tilde{\gamma}_l(\tilde{g}_{j_1}^1(l), \tilde{g}_{j_2}^2(l), \dots, \tilde{g}_{j_m}^m(l)) \neq 0.$$

Exactly the same argument implies the existence of a second minimizer $\tilde{\gamma}'_l$ to the multimarginal problem such that $\tilde{\gamma}'_l(\tilde{g}_{i_1}^1(l), \tilde{g}_{i_2}^2(l), \dots, \tilde{g}_{i_m}^m(l)) \neq 0$.

Although it is possible that $\tilde{\gamma}'_l \neq \tilde{\gamma}_l$, their linear average $\frac{1}{2}\tilde{\gamma}'_l + \frac{1}{2}\tilde{\gamma}_l$ is also optimal for the multimarginal problem and has both

$$(\tilde{g}_{i_1}^1(l), \tilde{g}_{i_2}^2(l), \dots, \tilde{g}_{i_m}^m(l)) \quad \text{and} \quad (\tilde{g}_{j_1}^1(l), \tilde{g}_{j_2}^2(l), \dots, \tilde{g}_{j_m}^m(l))$$

in its support, with the corresponding Wasserstein barycenter $\hat{\mu}_l = \frac{1}{2}\tilde{\mu}_l + \frac{1}{2}\tilde{\mu}'_l$. Notice that

$$\Delta_{\lambda}((\tilde{g}_{i_1}^1(l), \tilde{g}_{i_2}^2(l), \dots, \tilde{g}_{i_m}^m(l))) = x = \Delta_{\lambda}((\tilde{g}_{j_1}^1(l), \tilde{g}_{j_2}^2(l), \dots, \tilde{g}_{j_m}^m(l))).$$

Because Δ_{λ} has a Lipschitz inverse, we have

$$|(\tilde{g}_{i_1}^1(l), \tilde{g}_{i_2}^2(l), \dots, \tilde{g}_{i_m}^m(l)) - (\tilde{g}_{j_1}^1(l), \tilde{g}_{j_2}^2(l), \dots, \tilde{g}_{j_m}^m(l))| \leq C|x - x| = 0.$$

Now, letting α be such that $j_\alpha \neq i_\alpha$, the above implies that $\tilde{g}_{i_\alpha}^\alpha(l) = \tilde{g}_{j_\alpha}^\alpha(l)$. Since μ_α is a weak root measure, this means that, without loss of generality,

$$\lim_{z \rightarrow l^-} (\tilde{\mu}_\alpha)_z(\{\tilde{g}_{j_\alpha}^\alpha(z)\}) = 0.$$

This contradicts (3-4) and completes the proof under the additional assumption.

Now, if the assumption fails, instead of (3-3), we can conclude only that

$$\tilde{\gamma}_k(\tilde{g}_{j'_1}^1(l_k), \tilde{g}_{j'_2}^2(l'_k), \dots, \tilde{g}_{j'_m}^m(l'_k)) > \epsilon$$

for some $(j'_1, \dots, j'_m)$ with $G_{j'_1, j'_2, \dots, j'_m}^\lambda(l_k) = \tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(l_k)$, and by passing to a subsequence if necessary, we can take it to be the same $(j'_1, \dots, j'_m)$ for each k . As above, this implies

$$\tilde{\gamma}_l(\tilde{g}_{j'_1}^1(l), \tilde{g}_{j'_2}^2(l), \dots, \tilde{g}_{j'_m}^m(l)) \neq 0,$$

and since $\tilde{G}_{j'_1, j'_2, \dots, j'_m}^\lambda(l_k) = \tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(l_k)$, passing to the limit implies

$$\tilde{G}_{j'_1, j'_2, \dots, j'_m}^\lambda(l) = \tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(l) = \tilde{G}_{i_1, i_2, \dots, i_m}^\lambda(l).$$

The rest of the proof follows exactly as in the special case above. $\square$

Remark 3.10. If the solution $\tilde{\gamma}_l$ to the multimarginal problem (3-2) in the proof above is *unique*, and the μ_α are all strong root measures, then more is true: if $\lim_{z \rightarrow l^-} \tilde{\mu}_z(\{\tilde{G}_{j_1, j_2, \dots, j_m}^\lambda(z)\}) \neq 0$ we actually have $\tilde{\mu}_l(G_{i_1, i_2, \dots, i_m}^\lambda(z)) = 0$ for $z < l$ sufficiently close to l .

To see this, note that as above, $\tilde{\gamma}_l(\tilde{g}_{j'_1}^1(l), \tilde{g}_{j'_2}^2(l), \dots, \tilde{g}_{j'_m}^m(l)) \neq 0$. Since the root measures are strong, we must have $\tilde{g}_{i_\alpha}^\alpha(l) \neq \tilde{g}_{j_\alpha}^\alpha(l)$ for the α such that $i_\alpha \neq j_\alpha$. For $z < l$ with z close to l , any solution $\tilde{\gamma}_z$ to the multimarginal plan (3-2) must be weak-* close to $\tilde{\gamma}_l$ (by uniqueness) and so must satisfy $\tilde{\gamma}_z(\tilde{g}_{j'_1}^1(z), \tilde{g}_{j'_2}^2(z), \dots, \tilde{g}_{j'_m}^m(z)) \neq 0$. If such a solution satisfied $\tilde{\gamma}_z(\tilde{g}_{i_1}^1(z), \tilde{g}_{i_2}^2(z), \dots, \tilde{g}_{i_m}^m(z)) \neq 0$ as well, we would then have, by the Lipschitz property of Δ_λ^{-1} ,

$$|(g_{i_1}^1(z), g_{i_2}^2(z), \dots, g_{i_m}^m(z)) - (g_{j_1}^1(z), g_{j_2}^2(z), \dots, g_{j_m}^m(z))| \leq C |G_{i_1, i_2, \dots, i_m}^\lambda(z) - G_{j_1, j_2, \dots, j_m}^\lambda(z)|.$$

However, this is impossible since the right-hand side tends to 0 as z tends to l , but the left-hand side does not (as $\tilde{g}_{i_\alpha}^\alpha(l) \neq \tilde{g}_{j_\alpha}^\alpha(l)$ for at least one α , as described above). We conclude that we must have $\tilde{\gamma}_z(\tilde{g}_{i_1}^1(z), \tilde{g}_{i_2}^2(z), \dots, \tilde{g}_{i_m}^m(z)) = 0$ for any solution to the multimarginal problem and all $z < l$ sufficiently close to l ; therefore, $\tilde{\mu}_z(G_{i_1, i_2, \dots, i_m}^\lambda(z)) = 0$ for any Wasserstein barycenter $\tilde{\mu}_z$ of the $\tilde{\mu}_1, \dots, \tilde{\mu}_m$.

This applies, for instance, when $d = 1$, in which case Wasserstein barycenters are always unique.

3B. Comparison with the Wasserstein barycenter. If we instead use the standard notion of the Wasserstein barycenter to interpolate between several root measures, the barycenter may not be a weak root measure, as the following examples show.

Example 3.11. Several constructions of [Santambrogio and Wang 2016] show that displacement interpolation does not generally preserve convexity of sets. In one of these, the two marginals measures are concentrated on line segments embedded in $\mathbb{R}^2$, while their displacement interpolant (or Wasserstein barycenter) is supported on a curve $y = f(x)$ with a strict local minimum, where y is the vertical direction;

see $\mu_{1/2}$ in Section 2 in [Santambrogio and Wang 2016]. In our context, the two line segments constitute simple strong skeletal root measures, whereas the displacement interpolant is not even a weak skeletal root measure (as the two limbs meeting at the minimum point x_0 of f violate (W3)). Note that this is precisely because the angle between the two limbs is greater than $\frac{\pi}{2}$, and so the optimal map is not monotone in the vertical direction.

Our second example is even less well-behaved; here we take three strong root measures for which the Wasserstein barycenter has three-dimensional support.

Example 3.12. Consider uniform measure on the mutually orthogonal segments

$$\begin{aligned} T &:= \{(t, t, t) : t \in [0, 1]\}, \\ R &:= \left\{ \left(r, \left(\frac{-1+\sqrt{3}}{2} \right) r, \left(\frac{-1-\sqrt{3}}{2} \right) r \right) : r \in [0, 1] \right\}, \\ S &:= \left\{ \left(s, \left(\frac{-1-\sqrt{3}}{2} \right) s, \left(\frac{-1+\sqrt{3}}{2} \right) s \right) : s \in [0, 1] \right\} \end{aligned}$$

in $\mathbb{R}^3$.

Since the segments are orthogonal, the interaction terms $x_\alpha \cdot x_\beta = 0$ in the Gangbo–Swiech cost $(\mathbb{R}^3)^3$ (with, say, $\lambda_\alpha = 1/m$, $m = 3$)

$$\sum_{\alpha, \beta} \frac{1}{9} |x_\alpha - x_\beta|^2 = - \sum_{\alpha, \beta=1}^3 \frac{4}{9} x_\alpha \cdot x_\beta + \sum_{\alpha=1}^3 \frac{4}{9} |x_\alpha|^2$$

vanish.

Therefore, any measure with Lebesgue marginals supported on the product space $T \times R \times S$ is optimal in the multimarginal problem (3-2). The pushforward of any such measure γ by the mapping

$$\begin{aligned} & \left((t, t, t), \left(r, \left(\frac{-1+\sqrt{3}}{2} \right) r, \left(\frac{-1-\sqrt{3}}{2} \right) r \right), \left(s, \left(\frac{-1-\sqrt{3}}{2} \right) s, \left(\frac{-1+\sqrt{3}}{2} \right) s \right) \right) \\ & \mapsto \frac{(t, t, t) + (r, r, -2r) + \left(r, \left(\frac{-1+\sqrt{3}}{2} \right) r, \left(\frac{-1-\sqrt{3}}{2} \right) r \right) + \left(s, \left(\frac{-1-\sqrt{3}}{2} \right) s, \left(\frac{-1+\sqrt{3}}{2} \right) s \right)}{3} \end{aligned}$$

is a Wasserstein barycenter. If γ is, for example, a product measure, this push forward is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^3$.

This is certainly not a skeletal measure, and cannot be interpreted as a root in any reasonable way.

As with the layerwise-Wasserstein distance, one might suggest that horizontal symmetrization of the classical Wasserstein distance is more appropriate for comparing root shapes. That is, one should consider minimizers of (2-4). In the preceding example, although the Wasserstein barycenter has three-dimensional support, the biologically more relevant horizontally symmetrized version is concentrated on a line segment (since after appropriate rotations, the three sample measures are the same). Below, we augment the sample measures to produce a horizontally symmetrized Wasserstein barycenter with three-dimensional support.

Example 3.13. Let μ_1, μ_2 and μ_3 be uniform measures on the respective domains S_i defined by

$$\begin{aligned} S_1 &:= \{(t, t, t) : t \in [0, 1 + \epsilon]\}, \\ S_2 &:= \{(t, t, t) : t \in [0, 1]\} \cup \left\{ \left(1 + t, 1 + \left(\frac{-1+\sqrt{3}}{2} \right) t, 1 + \left(\frac{-1-\sqrt{3}}{2} \right) t \right) : t \in [0, \epsilon] \right\}, \\ S_3 &:= \{(t, t, t) : t \in [0, 1]\} \cup \left\{ \left(1 + t, 1 + \left(\frac{-1-\sqrt{3}}{2} \right) t, 1 + \left(\frac{-1+\sqrt{3}}{2} \right) t \right) : t \in [0, \epsilon] \right\}. \end{aligned}$$

It is not hard to show that the identity rotation minimizes the Wasserstein distance between μ_i and $R_{\#}\mu_j$ among horizontal rotations R for sufficiently small ϵ .

Furthermore, the optimal plans between μ_i and μ_j couple the top limbs via the identify mappings and the bottom limbs via product measure (or any other coupling between the bottom limbs—the solution is nonunique). Therefore, the measure

$$\gamma = (\text{id} \times \text{id} \times \text{id})_{\#}(\mu_1|_{\{(t,t,t):t \in [0,1]\}} + (\mu_1|_{\{(t,t,t):t \in [1,1+\epsilon]\}})(\mu_2|_{\{(1+t,1+(-\frac{1+\sqrt{3}}{2})t,1+(-\frac{1-\sqrt{3}}{2})t):t \in [0,\epsilon]\}})(\mu_3|_{\{(1+t,1+(-\frac{1-\sqrt{3}}{2})t,1+(-\frac{1+\sqrt{3}}{2})t):t \in [0,\epsilon]\}}))$$

is optimal in the multimarginal problem (3-2), and this plan has minimal cost among all multimarginal problems with marginals $(\mu_1, R_{2\#}\mu_2, R_{3\#}\mu_3)$ for horizontal rotations R_2 and R_3 . Consequently the symmetrized Wasserstein barycenter from (2-4) is then the pushforward of this measure under the mapping $(x_1, x_2, x_3) \mapsto \frac{1}{3}(x_1 + x_2 + x_3)$; this consists of the uniform measure on $\{(t, t, t) : t \in [0, 1]\}$ and a measure constructed as in the previous example, with three-dimensional support, arising from coupling the three orthogonal lower limbs.

3C. Total root length. An important phenotype used by biologists to compare root systems is the total root length, which is well-defined for skeletal root systems.

Given a strong skeletal root measure μ supported on $\bigcup_{i=1}^N \text{graph}(g_i)$ on $[0, \bar{y}]$, the total root length of μ is simply the one-dimensional Hausdorff measure of its support. Letting χ_i be the indicator function of the domain $[y_i, \bar{y}_i] \subseteq [0, \bar{y}]$ of g_i , we note that the root length is

$$R(\mu) = \sum_{i=1}^N \int_0^{\bar{y}} \sqrt{1 + |(g_i)'(y)|^2} \chi_i(y) dy. \quad (3-5)$$

Here we establish a result comparing the total root lengths of several skeletal root systems and their layerwise-Wasserstein barycenter. Given strong skeletal root measures

$$\mu_{\alpha} \text{ supported on } \bigcup_{i=1}^{N_{\alpha}} \text{graph}(g_{i_{\alpha}}^{\alpha}) \text{ on } [0, \bar{y}^{\alpha}] \text{ for } \alpha = 1, 2, \dots, m,$$

we compare their total root lengths to that of (a selected) layerwise-Wasserstein barycenter, with weights $\lambda_1, \dots, \lambda_m$. As above, we will also assume two sided bound

$$0 < L \leq f_{\alpha}^V(y) \leq U < \infty$$

on each μ_{α} , where f_{α}^V is the density of the vertical marginal μ_{α}^V . We have that $F'_{\mu_{\alpha}}(y) = f_{\alpha}^V(y)$, which implies that each rescaling change of variables is bi-Lipschitz.

Let $\bar{y} = \sum_{\alpha=1}^m \lambda_{\alpha} \bar{y}^{\alpha}$, so that any layerwise-Wasserstein barycenter of the μ_{α} is supported on $\mathbb{R}^d \times [0, \bar{y}]$. Assume that each $g_{i_{\alpha}}^{\alpha}$ is in $C^1([y_{i_{\alpha}}^{\alpha}, \bar{y}_{i_{\alpha}}^{\alpha}])$ and let C be an upper bound on each $|(g_{i_{\alpha}}^{\alpha})'|$. We define the total root length of a layerwise-Wasserstein barycenter $\text{Bar}^{\text{LW}}(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}})$ as the one-dimensional Hausdorff measure of its support, namely, the set

$$\left\{ (x, y) : x \in \text{spt} \left(\widetilde{\text{Bar}}_l^{\text{LW}} \left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}} \right) \right), l = \left(\sum_{\alpha=1}^m \lambda_{\alpha} F_{\mu_{\alpha}}^{-1} \right)^{-1}(y) \right\},$$

where, as before $\widetilde{\text{Bar}}_l^{\text{LW}}(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha})$ is the horizontal slice of the layerwise-Wasserstein barycenter at level l (that is, the Wasserstein barycenter of the $(\tilde{\mu}_\alpha)_l$).

Letting $G_{i_1, i_2, \dots, i_m}^\lambda$ be one of the graphs in the ghost, we let $\chi_{i_1, i_2, \dots, i_m}^\lambda$ be the indicator function of its active set; that is,

$$\chi_{i_1, i_2, \dots, i_m}^\lambda(y) = \begin{cases} 1 & \text{if } G_{i_1, i_2, \dots, i_m}^\lambda(y) \text{ is well-defined and in the support of } \widetilde{\text{Bar}}_{(\sum_{\alpha=1}^m \lambda_\alpha F_{\mu_\alpha}^{-1})^{-1}(y)}^{\text{LW}}(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha}), \\ 0 & \text{otherwise.} \end{cases}$$

The root length is then

$$\sum_{i_1, \dots, i_m} \int_0^{\bar{y}} \sqrt{1 + |(G_{i_1, i_2, \dots, i_m}^\lambda)'(y)|^2} \chi_{i_1, i_2, \dots, i_m}^\lambda(y) dy.$$

Proposition 3.14. *Letting μ_α be skeletal roots for $\alpha = 1, 2, \dots, m$, there is a layerwise-Wasserstein barycenter $\text{Bar}^{\text{LW}}(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha})$ of $\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha}$ for which*

$$C_0 R(\mu_\beta) \leq R\left(\text{Bar}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha}\right)\right) \leq C_1 \left[C_2 \sum_{\alpha=1}^m R(\mu_\alpha) - (m-1) \right]$$

for any $\beta = 1, 2, \dots, m$. The constants C_0 , C_1 and C_2 depend only on $C = \sup_{\alpha, i_\alpha} \|(g_{i_\alpha}^\alpha)'\|_{L^\infty}$, L and U .

The proof essentially consists of two steps: first, we establish a similar result for the rescaled versions. We then use bounds on the rescaling change of variables to translate the rescaled inequalities back to the original coordinates. We isolate the first step as a separate lemma.

Lemma 3.15. *Using the notation in the proposition above, there exists a layerwise Wasserstein barycenter such that*

$$\tilde{C}_0 R(\tilde{\mu}_\beta) \leq R\left(\widetilde{\text{Bar}}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_\alpha \delta_{\mu_\alpha}\right)\right) \leq \tilde{C}_1 \left[\sum_{\alpha=1}^m R(\tilde{\mu}_\alpha) - (m-1) \right]$$

for any $\beta = 1, 2, \dots, m$.

Proof. Note that $\tilde{\mu}_\alpha$ is supported on the skeletal set $\bigcup_{i=1}^{N_\alpha} (\tilde{g}_{i_\alpha}^\alpha)$, where $\tilde{g}_{i_\alpha}^\alpha := g_{i_\alpha}^\alpha \circ F_{\mu_\alpha}^{-1}$. The $\tilde{g}_{i_\alpha}^\alpha$ then have derivatives bounded by $\tilde{C} = C/L$.

The ghost of the rescaled system consists of the limbs

$$\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda = \sum_{\alpha=1}^m \lambda_\alpha \tilde{g}_{i_\alpha}^\alpha,$$

which inherit the same derivative bounds as the $\tilde{g}_{i_\alpha}^\alpha$,

$$|(\tilde{G}_{i_1, i_2, \dots, i_m}^\lambda)'| \leq \tilde{C},$$

and at each $l \in [0, 1]$ it is shown in [Anderes et al. 2016] that there is a Wasserstein barycenter $\text{Bar}^W(\sum_{\alpha} \lambda_\alpha \delta_{(\tilde{\mu}_\alpha)_l})$ of the discrete measures $(\tilde{\mu}_\alpha)_l$ such that the number $S(l)$ of points in its support is at most $\sum_{\alpha=1}^m S_\alpha(y) - m + 1$, where $S_\alpha(l)$ is the number of points in the support of $(\tilde{\mu}_\alpha)_l$. We use this

Wasserstein barycenter in our construction of $\widetilde{\text{Bar}}^{\text{LW}}(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}})$. Therefore,

$$\begin{aligned} R\left(\widetilde{\text{Bar}}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}}\right)\right) &= \sum_{i_1, \dots, i_m} \int_0^1 \sqrt{1 + |(\widetilde{G}_{i_1, i_2, \dots, i_m}^{\lambda})'(l)|^2} \chi_{i_1, i_2, \dots, i_m}^{\lambda}(l) dl \\ &\leq \int_0^1 \sqrt{1 + \widetilde{C}^2} \left[\sum_{\alpha=1}^m S_{\alpha}(l) - m + 1 \right] dl \\ &\leq \int_0^1 \sum_{\alpha=1}^m \sum_{i_{\alpha}=1}^{N_{\alpha}} \sqrt{1 + |(\widetilde{g}_{i_{\alpha}}^{\alpha})'(l)|^2} \sqrt{1 + \widetilde{C}^2} \chi_{i_{\alpha}}^{\alpha}(l) dl - \sqrt{1 + \widetilde{C}^2} (m - 1) \\ &= \sqrt{1 + \widetilde{C}^2} \sum_{\alpha=1}^m R(\tilde{\mu}_{\alpha}) - \sqrt{1 + \widetilde{C}^2} (m - 1). \end{aligned}$$

Similarly, the $S(l)$ is bounded below by the support of each marginal, $S(l) \geq S_{\beta}(l)$, and so, for each β

$$\begin{aligned} R\left(\widetilde{\text{Bar}}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}}\right)\right) &= \sum_{i_1, \dots, i_m} \int_0^1 \sqrt{1 + |(\widetilde{G}_{i_1, i_2, \dots, i_m}^{\lambda})'(l)|^2} \chi_{i_1, i_2, \dots, i_m}^{\lambda}(l) dl \\ &\geq \int_0^1 S_{\beta}(l) dl \\ &\geq \int_0^1 \sum_{i=1}^{N_{\beta}} \frac{\sqrt{1 + |(\widetilde{g}_i^{\beta})'(l)|^2}}{\sqrt{1 + \widetilde{C}^2}} \chi_i^{\beta}(l) dl \\ &= \frac{1}{\sqrt{1 + \widetilde{C}^2}} R(\tilde{\mu}_{\beta}). \quad \square \end{aligned}$$

The proof of the proposition combines the lemma with straightforward estimates on the change of variables $F_{\mu_{\alpha}}$.

Proof of Proposition 3.14. The root length of each limb satisfies:

$$\begin{aligned} \int_{\underline{y}_i}^{\bar{y}_i} \sqrt{1 + [(g_{i_{\alpha}}^{\alpha})'(y)]^2} dy &= \int_{l_i}^{\bar{l}_i} \sqrt{1 + [(g_{i_{\alpha}}^{\alpha})'(F_{\mu_{\alpha}}^{-1}(l))]^2 [F_{\mu_{\alpha}}^{-1}]'(l)} dl \\ &= \int_{l_i}^{\bar{l}_i} \sqrt{[F_{\mu_{\alpha}}^{-1}]'(l)^2 + [(g_{i_{\alpha}}^{\alpha})'(F_{\mu_{\alpha}}^{-1}(l))]^2 [F_{\mu_{\alpha}}^{-1}]'(l)^2} dl \\ &\leq K \int_{l_i}^{\bar{l}_i} \sqrt{1 + [(g_{i_{\alpha}}^{\alpha})'(F_{\mu_{\alpha}}^{-1}(l))]^2 [F_{\mu_{\alpha}}^{-1}]'(l)^2} dl, \end{aligned}$$

where $K = \max(1/L, 1)$. The last term corresponds to the root length of the corresponding limb of $\tilde{\mu}_{\alpha}$. Adding over all limbs we get

$$R(\mu_{\alpha}) \leq K R(\tilde{\mu}_{\alpha}),$$

while a symmetric argument yields

$$R(\mu_{\alpha}) \geq k R(\tilde{\mu}_{\alpha}),$$

with $k = \min(1/U, 1)$.

Similarly, since the vertical rescaling for the barycenter

$$F_{\text{Bar}^{\text{LW}}}(\sum_{\alpha} \lambda_{\alpha} \delta_{\mu_{\alpha}}) = \left(\sum_{\alpha=1}^m \lambda_{\alpha} F_{\mu_{\alpha}}^{-1} \right)^{-1}$$

inherits first derivative bounds from the μ_i , we also get

$$kR\left(\widetilde{\text{Bar}}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}}\right)\right) \leq R\left(\text{Bar}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}}\right)\right) \leq KR\left(\widetilde{\text{Bar}}^{\text{LW}}\left(\sum_{\alpha=1}^m \lambda_{\alpha} \delta_{\mu_{\alpha}}\right)\right).$$

Combined with the Lemma 3.15, these estimates yield the desired result. $\square$

Remark 3.16. The result also holds, with essentially the same proof, for weak skeletal root measures, provided we take χ_i^{α} in (3-5) to be the indicator function of the subset of the domain $[\underline{y}_i^{\alpha}, \bar{y}_i^{\alpha}]$ where $(\tilde{\mu}_{\alpha})_l(g_{i_{\alpha}}^{\alpha}(y)) > 0$ for $l = F_{\mu}(y)$.

Remark 3.17. It is unfortunately not possible to establish an upper bound on the root length of the layerwise Wasserstein barycenter which is *independent* of the number of samples m .

To see this, consider the skeletal root measures μ_{α} , each concentrated on two curves $g_{\alpha_1}, g_{\alpha_2} : [0, 1] \rightarrow \mathbb{R}$, with $g_{\alpha_1}(y) = 0$ and $g_{\alpha_2}(y) = y$. We let the one-dimensional density of each μ_i be constant on each of the two limbs, with densities $1/\alpha$ on g_{α_1} and $1 - 1/\alpha$ on g_{α_2} (normalized to have total mass 1). The vertical marginals of each μ_{α} are then uniform, so $F_{\mu_{\alpha}}(y) = y$ and each $(\tilde{\mu}_{\alpha})_l = (1/\alpha)\delta_0 + (1 - 1/\alpha)\delta_l$.

It is then not hard to see that the Wasserstein barycenter of $\sum_{\alpha=1}^m (1/m)(\tilde{\mu}_{\alpha})_l$ is then concentrated on the $m + 1$ points $0, l/m, 2l/m, \dots, l$, and so the support of the layerwise-Wasserstein barycenter of $\sum_{\alpha=1}^m (1/m)\mu_{\alpha}$ consists of the m curves $g_1, \dots, g_m : [0, 1] \rightarrow \mathbb{R}$, with $g_{\alpha}(y) = \alpha y/m$; the total root length clearly grows with m .

Interpolating between a large number of marginals, or samples, m , can therefore result in weak skeletal measures with very large total root length,

4. Layerwise Wasserstein convexity

We will call a function $\mathcal{F} : P(\mathbb{R}^d \times \mathbb{R}_{\geq 0}) \rightarrow \mathbb{R}$ *layerwise-Wasserstein convex* if for any $\Omega \in P(P(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$,

$$\mathcal{F}(\text{Bar}^{\text{LW}}(\Omega)) \leq \int \mathcal{F}(\mu) d\Omega(\mu).$$

This notion of convexity may potentially play an important role in applications. Given a family of root systems, corresponding to a family of genetically identical plants, grown under identical environmental conditions, we will in forthcoming work propose interpreting the layerwise-Wasserstein barycenter of the systems as the single root system which best represents the family. It is natural to compare phenotypes (for instance, center of mass, variance, entropy, total root length, etc.) of that barycenter with the phenotypes of the actual observed roots in the original family. If these phenotypes (interpreted as functionals on the space of measures) are layerwise-Wasserstein convex, the phenotype of the barycenter is always less than the average of the phenotypes of the samples.

The theory of layerwise convexity, which we begin to develop below, has a strong connection to the theory of displacement convexity, or convexity along geodesics with respect to the Wasserstein metric,

introduced in [McCann 1997], and its extension to convexity over Wasserstein barycenters, introduced in [Agueh and Carlier 2011].

We begin with the Shannon entropy, perhaps the best known displacement convex functional. As we show below, it is also layerwise-Wasserstein convex. For roots, it can be regarded as a measure of the concentration of mass and therefore has potential biological interest. Given $\mu \in P(\mathbb{R}^d \times \mathbb{R}_{\geq 0})$, with

$$\mu(x, y) = f(x, y) dx dy, \quad \text{where } x \in \mathbb{R}^d, y \in \mathbb{R}_{\geq 0},$$

the vertical marginal μ^V has density

$$f^V(y) = \int_{\mathbb{R}^d} f(x, y) dx.$$

Note that for fixed y , the probability measure

$$d\mu_y(x) = \frac{f(x, y)}{f^V(y)} dx$$

coincides with $\tilde{\mu}_l$ for $l = F_\mu(y)$. Recall that the Shannon entropy of μ is defined as

$$S(\mu) = \int_{\mathbb{R}^d} \int_{\mathbb{R}_{\geq 0}} f(x, y) \log f(x, y) dx dy.$$

This formula allows one to rewrite $S(\mu)$ using the layerwise decomposition.

Proposition 4.1. *The Shannon entropy $S(\mu)$ satisfies*

$$S(\mu) = \int_{\mathbb{R}_{\geq 0}} S(\mu_y) d\mu^V(y) + S(\mu^V) = \int_0^1 S(\tilde{\mu}_l) dl + S(\mu^V).$$

Proof. The proof is a calculation. In the following we use the standard convention that $0 \log(0) = 0$. We have

$$\begin{aligned} S(\mu) &= \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^d} f(x, y) \log \left[\frac{f(x, y) f^V(y)}{f^V(y)} \right] dx dy \\ &= \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^d} f(x, y) \left[\log \left[\frac{f(x, y)}{f^V(y)} \right] + \log f^V(y) \right] dx dy \\ &= \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^d} f(x, y) \log \left[\frac{f(x, y)}{f^V(y)} \right] dx dy + \int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} f(x, y) dx \right) \log f^V(y) dy \\ &= \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^d} f^V(y) \frac{f(x, y)}{f^V(y)} \log \left[\frac{f(x, y)}{f^V(y)} \right] dx dy + \int_{\mathbb{R}} f^V(y) \log f^V(y) dy \\ &= \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^d} \left(\frac{f(x, y)}{f^V(y)} \log \left[\frac{f(x, y)}{f^V(y)} \right] dx \right) f^V(y) dy + S(\mu^V) \\ &= \int_{\mathbb{R}_{\geq 0}} S(\mu_y) d\mu^V(y) + S(\mu^V). \end{aligned}$$

The final equality follows by noting that the cumulative distribution function F_μ satisfies $F'_\mu(y) = f^V(y)$ and by changing variables from y to $l = F_\mu(y)$. $\square$

Corollary 4.2. *The Shannon entropy is layerwise-Wasserstein convex.*

Proof. Recall that from Proposition 2.5 the layerwise-Wasserstein interpolation of $\Omega \in P(P(\mathbb{R}^d \times \mathbb{R}_{\geq 0}))$ amounts to constructing the probability measure η whose vertical marginal η^V is the Wasserstein barycenter of Ω^V , and whose conditional probabilities $\tilde{\eta}_l$ are the Wasserstein barycenters of the $\tilde{\Omega}_l$. Since the entropy $S(\mu)$ depends additively on μ^V and the $\tilde{\mu}_l$, Wasserstein convexity of the entropy (see [Kim and Pass 2017] for convexity with respect to general barycenters) yields the result. $\square$

Many phenotypes of interest concern only the depth of the root, and not its horizontal distribution of mass (since, for instance, nutrient concentration in soil is largely determined by depth). Therefore the following simple observation is relevant.

Proposition 4.3. *Any Wasserstein convex function of the vertical marginal is layerwise Wasserstein convex.*

Proof. This follows immediately from the structure of the layerwise-Wasserstein distance, given in Proposition 2.5. $\square$

Example 4.4. Let us list a few examples of functionals with possible biological applications, which are layerwise-Wasserstein convex by Proposition 4.3:

- The vertical mean

$$\mu \mapsto \bar{y} := \int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} y \, d\mu(x, y) = \int_{\mathbb{R}_{\geq 0}} y \, d\mu^V(y);$$

one can verify easily that this is in fact affine along displacement (and hence layerwise-Wasserstein) interpolations.

- The vertical variance

$$\int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} |y - \bar{y}|^2 \, d\mu(x, y) = \int_{\mathbb{R}_{\geq 0}} |y - \bar{y}|^2 \, d\mu^V(y),$$

a measure the spread of the mass in the vertical direction [Kim and Pass 2020].

- The vertical internal energy

$$\int_{\mathbb{R}_{\geq 0}} (f^V(y))^r \, dy \quad \text{for } r \geq 1.$$

- Vertical quantiles $F_\mu^{-1}(l)$ for each fixed $l \in (0, 1)$. For instance, the vertical median ($l = \frac{1}{2}$) is the depth above which half the mass of the root lies. The 100th quantile (the maximal depth of the root) is often called the rooting depth, while the 87th quantile ($l = \frac{87}{100}$) is a conventional phenotype often used as a measure of the root depth. The displacement convexity (and layerwise-Wasserstein convexity) of these follows immediately from the monotone structure of one-dimensional optimal couplings with respect to the distance-squared cost; in fact, it is layerwise-Wasserstein affine.

Although, unlike the examples above, it is not a functional of the vertical marginal, the structure of the layerwise-Wasserstein distance easily implies that the class of functionals in the following example are layerwise-Wasserstein convex as well.

Example 4.5. Any functional of the form

$$\bar{\mathcal{F}}(\mu) = \int_0^1 \bar{\mathcal{F}}_l(\tilde{\mu}_l) dl,$$

where each $\bar{\mathcal{F}}_l$ is Wasserstein convex, is layerwise-Wasserstein convex.

Acknowledgements

Kim and Pass are partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant program. Schneider is supported by the Canada Excellence in Research Chair in Food Systems and Security and the Global Institute for Food Security.

The authors gratefully acknowledge the support of the Pacific Institute of Mathematical Sciences (PIMS); in particular, they would like to thank PIMS director James Colliander for initiating this project by bringing the three of them together. Part of this work was done while the authors participated in a research-in-teams program at the Banff International Research Station (BIRS), and we thank BIRS for the excellent working environment provided there.

References

- [Agueh and Carlier 2011] M. Agueh and G. Carlier, “Barycenters in the Wasserstein space”, *SIAM J. Math. Anal.* **43**:2 (2011), 904–924. MR Zbl
- [Anderes et al. 2016] E. Anderes, S. Borgwardt, and J. Miller, “Discrete Wasserstein barycenters: optimal transport for discrete data”, *Math. Methods Oper. Res.* **84**:2 (2016), 389–409. MR Zbl
- [Backhoff et al. 2017] J. Backhoff, M. Beiglböck, Y. Lin, and A. Zalashko, “Causal transport in discrete time and applications”, *SIAM J. Optim.* **27**:4 (2017), 2528–2562. MR Zbl
- [Bonneel et al. 2015] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister, “Sliced and Radon Wasserstein barycenters of measures”, *J. Math. Imaging Vision* **51**:1 (2015), 22–45. MR Zbl
- [Bressan and Sun 2018] A. Bressan and Q. Sun, “On the optimal shape of tree roots and branches”, *Math. Models Methods Appl. Sci.* **28**:14 (2018), 2763–2801. MR Zbl
- [Bressan et al. 2020] A. Bressan, M. Palladino, and Q. Sun, “Variational problems for tree roots and branches”, *Calc. Var. Partial Differential Equations* **59**:1 (2020), art. id. 7. MR Zbl
- [Bucksch 2014] A. Bucksch, “A practical introduction to skeletons for the plant sciences”, *Appl. Plant Sci.* **2**:8 (2014), art. id. 1400005.
- [Carlier et al. 2010] G. Carlier, A. Galichon, and F. Santambrogio, “From Knothe’s transport to Brenier’s map and a continuation method for optimal transport”, *SIAM J. Math. Anal.* **41**:6 (2010), 2554–2576. MR Zbl
- [Clark et al. 2011] R. T. Clark, R. B. MacCurdy, J. K. Jung, J. E. Shaff, S. R. McCouch, D. J. Aneshansley, and L. V. Kochian, “Three-dimensional root phenotyping with a novel imaging and software platform”, *Plant Physiol.* **156**:2 (2011), 455–465.
- [Clark et al. 2013] R. T. Clark, A. N. Famoso, K. Zhao, J. E. Shaff, E. J. Craft, C. D. Bustamente, S. R. McCouch, D. J. Aneshansley, and L. V. Kochian, “High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development”, *Plant Cell Environ.* **36**:2 (2013), 454–466.
- [Famoso et al. 2010] A. N. Famoso, R. T. Clark, J. E. Shaff, E. Craft, S. R. McCouch, and L. V. Kochian, “Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms”, *Plant Physiol.* **153**:4 (2010), 1678–1691.
- [Famoso et al. 2011] A. N. Famoso, K. Zhao, R. T. Clark, C.-W. Tung, M. H. Wright, C. Bustamente, L. V. Kochian, and S. R. McCouch, “Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping”, *PLoS Genetics* **7**:8 (2011), art. id. e1002221.

- [Hatcher 2002] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, 2002. MR Zbl
- [Kim and Pass 2017] Y.-H. Kim and B. Pass, “Wasserstein barycenters over Riemannian manifolds”, *Adv. Math.* **307** (2017), 640–683. MR Zbl
- [Kim and Pass 2020] Y.-H. Kim and B. Pass, “Nonpositive curvature, the variance functional, and the Wasserstein barycenter”, *Proc. Amer. Math. Soc.* **148**:4 (2020), 1745–1756. MR Zbl
- [Knothe 1957] H. Knothe, “Contributions to the theory of convex bodies”, *Michigan Math. J.* **4** (1957), 39–52. MR Zbl
- [McCann 1997] R. J. McCann, “A convexity principle for interacting gases”, *Adv. Math.* **128**:1 (1997), 153–179. MR Zbl
- [Muzellec and Cuturi 2019] B. Muzellec and M. Cuturi, “Subspace detours: building transport plans that are optimal on subspace projections”, preprint, 2019. arXiv
- [Piñeros et al. 2016] M. A. Piñeros, B. G. Larson, J. E. Shaff, D. J. Schneider, A. X. Falcão, L. Yuan, R. T. Clark, E. J. Craft, T. W. Davis, P.-L. Pradier, N. M. Shaw, I. Assaranurak, S. R. McCouch, C. Sturrock, M. Bennett, and L. V. Kochian, “Evolving technologies for growing, imaging and analyzing 3D root system architecture of crop plants”, *J. Integr. Plant Biol.* **58**:3 (2016), 230–241.
- [Rosenblatt 1952] M. Rosenblatt, “Remarks on a multivariate transformation”, *Ann. Math. Stat.* **23** (1952), 470–472. MR Zbl
- [Santambrogio 2015] F. Santambrogio, *Optimal transport for applied mathematicians*, Progr. Nonlin. Diff. Eq. Appl. **87**, Birkhäuser, Cham, 2015. MR Zbl
- [Santambrogio and Wang 2016] F. Santambrogio and X.-J. Wang, “Convexity of the support of the displacement interpolation: counterexamples”, *Appl. Math. Lett.* **58** (2016), 152–158. MR Zbl
- [Toroczkai 2002] Z. Toroczkai, “Topological classification of binary trees using the Horton–Strahler index”, *Phys. Rev. E* (3) **65**:1 (2002), art. id. 016130. MR
- [Villani 2003] C. Villani, *Topics in optimal transportation*, Grad. Studies in Math. **58**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl
- [Villani 2009] C. Villani, *Optimal transport: old and new*, Grundlehren der Math. Wissenschaften **338**, Springer, 2009. MR Zbl

Received 10 Oct 2019. Revised 21 Feb 2020. Accepted 28 Mar 2020.

YOUNG-HEON KIM: yhkim@math.ubc.ca

Department of Mathematics, University of British Columbia, Vancouver, BC, Canada

BRENDAN PASS: pass@ualberta.ca

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada

DAVID J. SCHNEIDER: dave.schneider@gifs.ca

Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK, Canada

