An Optimal Transport Approach to Crop Root Systems and related Multiscale Structures

Young-Heon Kim (University of British Columbia),
Brendan Pass (University of Alberta),
David Schneider (University of Saskatchewan)

January 20 – January 27, 2019

1   Overview and background of the project

Given the root systems of several plants with identical genotypes (genome sequence) and grown in the same environment, how does one construct a "typical" representative of the set? What about variation within the set? How might one assess whether root systems of different genotypes are similar or different?

These are all important questions arising in the development of crops with improved traits, for example tolerance to drought or nutrient limitations. Root systems are notoriously difficult to study because, they exhibit biologically important structures with characteristic length scales varying over more than five orders of magnitude — primary roots have lengths on the order of one meter while root hairs (extensions of single cells) have diameters of 10^{-5} meters — but are not self-similar. While sophisticated mathematical techniques such as persistent homology have been applied to roots, most researchers rely an ad hoc collection of phenotypes or traits (e.g., rooting depth, total root length, total root mass, volume of convex hull) [2, 4, 5, 3, 7]. Extensive experience has shown that this ad hoc approach has critical limitations. Figures 1 and 2 represent two distinct root systems arising from the same species and illustrate the issues of using phenotypes to differentiate between them (their structures are clearly different, but it is not at all clear which traits should be used to quantify the difference).

This research project aims to develop metrics that can be used to quantify the differences between root systems and as a framework for interpolating or averaging over a set of biologically equivalent systems (same genotype and environment). The issue of averaging is ubiquitous in nature. Though most often thought of in the linear context of numbers or vectors, there is also a fairly well developed theory of averaging on metric spaces (collections of objects with a canonical distance, or measure of difference, between any two objects), known as the theory of metric barycenters.

Our motivation comes from the relationship of particular phenotypes exhibited by plants to their genotypes and the environmental conditions under which they were grown. Typically, fixed genotypes and environments result in plants with a diverse collection of different root systems, and it is desirable to choose a single system to associate with the genotype/environment pair (how does one "average" in a meaningful way among the various systems with different shapes?) and quantify the variation from that system exhibited by the other samples sharing the same genotype and environment.

The Wasserstein distance, or earth mover's, from optimal transport theory, is commonly used to compare distributions of mass: visualize two mass distribution as piles of dirt with different shapes but the same volume and transform the first pile into the second in a way that minimizes the average distance moved by the
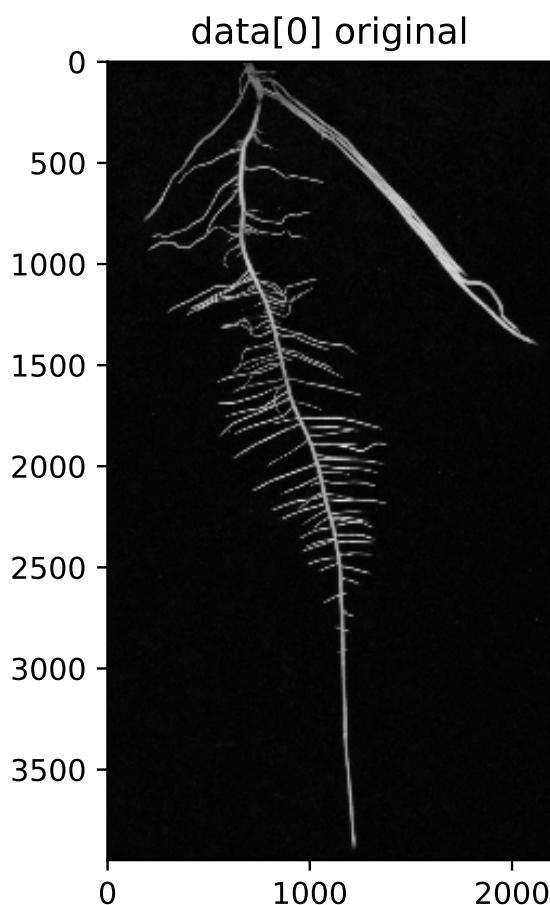
Figure 1: Example of root system.

dirt. This minimal average distance is then the earth mover's distance, and the theory of its metric barycenters (Wasserstein barycenters) is fairly well developed; see [1] for a seminal paper on Wasserstein barycenters and [6] for an overview of numerical methods in optimal transportation. Since one can interpret a root system as a mass distribution, this is in some ways a reasonable candidate for a metric between root systems. However, it fails to account for the topological, dendritic structure exhibited by root systems (for example, a single thick branch will be very close in Wasserstein distance to two thinner branches, emerging from a common node, sitting very close together, but these have very different biological properties). In addition, if we compute the Wasserstein barycenter, or interpolation, of two or more root systems, we generally obtain a distribution of mass, but not necessarily one that resembles a root system (that is, the dendritic structure may be lost). Classical Wasserstein distance is thus not well suited to comparing root systems.

Our proposal is to adapt the earth mover's distance to build a metric better suited for this purpose, a root matcher's metric. When comparing to root systems, one tries to move the mass from one onto the other in way that minimizes as much as possible the overall movement of the mass (as with the earth mover's metric), but also minimizes the distortion of the topology. This will yield a more biologically meaningful way to quantify the difference between two root systems, and the corresponding metric barycenters (defined using the root matcher's metric) will be meaningful representatives of a family of such systems
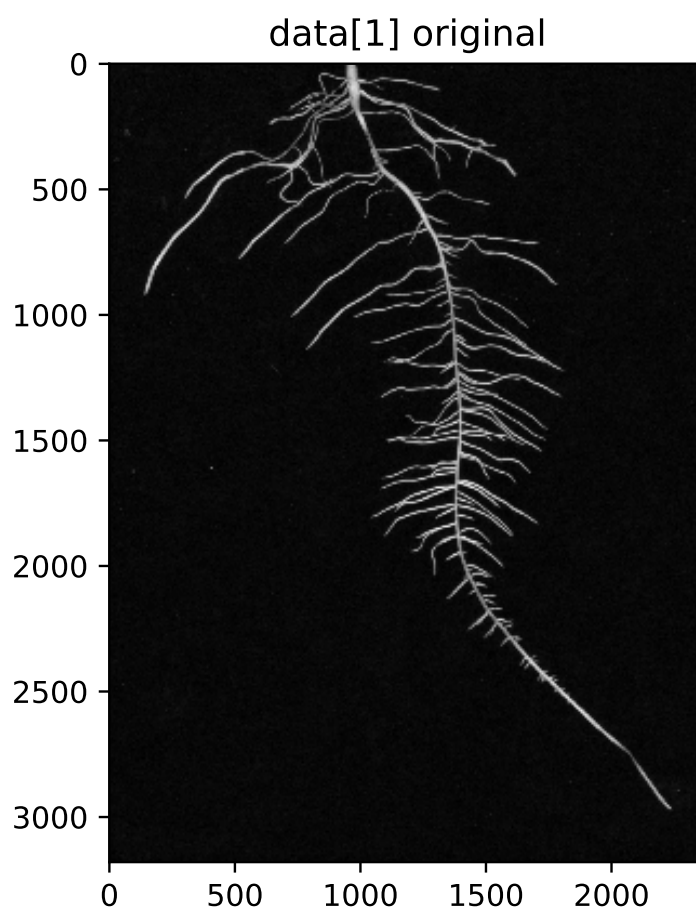
Figure 2: Example of a root system.

2 Objectives of the meeting

The goal of this team meeting is to boost up progress in the very early stage of the proposed research program we have recently initiated; two of us (Kim and Pass) are mathematicians who have been focusing on optimal transport theory, and the third member (Schneider) is a computational biologist who have been working on complex biological systems including analyzing crop root systems. Though we have been having regular research meetings over the internet to discuss issues and exchange ideas, it is desirable for the three of us to meet together in person for a week, free from other distractions, to concentrate on the research problems. This first in-person meeting focused on developing new versions of Wasserstein distance and barycenters which are well suited for analyzing root structures.

3 Scientific Progress Made

We have developed a new metric on the space of the probability measures that is naturally suited to dentritic structures, including plant roots and trees. Unlike the usual Wasserstein barycenters, interpolants of several root systems (viewed as mass distributions) under this new interpolation seem to largely preserve root like structure (we have already proven the preservation of several desirable properties characteristic of roots sys-

tems, while others remain conjectures.) The approach also seems advantageous computationally, and we have almost completed writing code for a preliminary algorithm.

4 Anticipated Outcome of the Meeting

We are working on three initial papers, one developing the theoretical framework for root matching, targeted towards mathematicians, another focusing on methodological issues (explaining how this framework can be used to distinguish between collections of root systems) and a third targeted towards biologists, in which we apply this methodology to real data and establish distinctions between collections of root systems that previous, phenotype-based approaches were unable to detect.

We are also optimistic that this approach and variations of it will lead to much more fruitful research in the future as well. We plan to hire at least on postdoctoral researcher, working closely with all three us, to investigate future directions.

Our work completed at BIRS represents a first step towards a mathematical foundation of "optimal transport of graphs with spatial embedding and mass," and seems likely to lead to fascinating new questions at the intersection of graph theory and analysis. It is reasonable to expect the same underlying mathematical structure (mass distributions on dendritic structures) to occur in other areas such as vascular structures in leaves, blood vessels and nerves in animals, and trees in decision science. Therefore it is reasonable to expect that this work will be broadly applicable.

References

[1] M. Agueh and G. Carlier, Barycenters in the Wasserstein space, *SIAM J. Math. Anal.* **43** (2011), 904–924.

[2] Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J., and Kochian, L. V. (2011). Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol., 156(2), 455465. doi: http://dx.doi.org/10.1104/pp.110.169102.

[3] Randy T. Clark, Adam N. Famoso, Keyan Zhao, Jon E. Shaff, Eric J. Craft, Carlos D. Bustamante, Susan R. McCouch, Daniel J. Aneshansley and Leon V. Kochian (2013). High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant, Cell and Environment, 36, 454466. doi:10.1111/j.1365-3040.2012.02587.x.

[4] Famoso, A. N., Clark, R. T., Shaff, J. E., Craft, E., McCouch, S. R., and Kochian L. V. (2010). 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), 16781691. doi: http://dx.doi.org/10.1104/pp.110.156794.

[5] Famoso, A. N., Zhao, K., Clark, R. T., Tung, C.-W., Wright, M. H., Bustamante, C., Kochian, L. V., and McCouch, S. R. (2011). Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-Wide Association Analysis and QTL Mapping. PLoS Genetics, 7, e1002221. doi: https://doi.org/10.1371/journal.pgen.1002221.

[6] G. Peyre and M. Cuturi, Computational optimal transport. Preprint available at: https://arxiv.org/abs/1803.00567.

[7] Pieros MA, Larson BG, Shaff JE, Schneider DJ, Falco AX, Yuan L, Clark RT, Craft EJ, Davis TW, Pradier PL, Shaw NM, Assaranurak I, McCouch SR, Sturrock C, Bennett M, Kochian LV (2016) Evolving technologies for growing, imaging and analyzing 3D root system architecture of crop plants. J Integr Plant Biol 58: 230241