

**Geert-Jan Huizing**

ENS PSL, France

geert-jan.huizing@ens.fr

Defining meaningful distances between samples in a dataset is a fundamental problem in machine learning. Optimal Transport (OT) lifts a distance between features (the “ground metric”) to a geometrically meaningful distance between samples. However, there is usually no straightforward choice of ground metric. Supervised ground metric learning approaches exist but require labeled data. In absence of labels, only ad-hoc ground metrics remain. Unsupervised ground metric learning is thus a fundamental problem to enable data-driven applications of OT. In this paper, we propose for the first time a canonical answer by simultaneously computing an OT distance between samples and between features of a dataset. These distance matrices emerge naturally as positive singular vectors of the function mapping ground metrics to OT distances. We provide criteria to ensure the existence and uniqueness of these singular vectors. We then introduce scalable computational methods to approximate them in high-dimensional settings, using stochastic approximation and entropic regularization. Finally, we showcase Wasserstein Singular Vectors on a single-cell RNA-sequencing dataset.

*Joint work with Laura Cantini (ENS PSL) and Gabriel Peyré (ENS PSL).*