Onlicorne: optimal transportation distances from sample streams

Anonymous Authors[1]

Abstract

This paper introduces a new online estimator of optimal transport distances between two continuous probability distributions. Our algorithm handles sample streams from two distributions, that it uses to iteratively enrich a non-parametric representation of the transportation plan. This plan is naturally represented using a mixture of simple (e.g. Gaussian) functions, of increasing memory complexity. Compared to the classic Sinkhorn algorithm, our method handles new samples at each iteration, which permits a consistent estimation of the true regularized OT distance. We show how our algorithm can be casted as a block-convex mirror descent in the space of positive distributions; from this point of view, we analyse its convergence. We illustrate the performance of our method compared to concurrent approaches.

1. Introduction

Optimal transport (OT) distances are fundamental in statistical learning, both as a tool for analyzing the convergence of various algorithms, and as a data-dependant term for estimating data density, e.g. using generative models. OT lifts a given distance over data points living in space $\mathcal{X}$ into a distance between probability distributions over the data space $\mathcal{X}$ $\mathcal{P}(\mathcal{X})$; as such, it allows to compare distributions with disjoint support. To alleviate the computational burden of optimal transport, that is cubic in the number of points, it is common to regularize the linear problem that defines it, using an entropic barrier term. This approach, that has been rediscovered many times in the previous thirty years, allows to approximate OT distances using a matrix balancing algorithm, amenable to GPU computations.

The Sinkhorn algorithm was introduced in a discrete setting, i.e. when both distributions to compare are a set of realizations. The so-called Sinkhorn distances between empirical distributions indeed form an estimate of the OT distance between the true distributions from which the samples are drawn. This approach estimates the OT distance in two distinct phases: we draw samples and evaluate a pairwise distance matrix in the first phase; we balance this distance matrix using Sinkhorn-Knopp iterations in the second phase, thereby obtaining a transportation plan and distance.

In this paper, we show how mingling together these two phases can be beneficial to quickly estimate OT distances. Our approach relies on three observations. First, Sinkhorn iterations can be rewritten as a block convex mirror descent on the space of positive distributions. This formulation is valid in the discrete and continuous setting. Second, we can modify these iterations to rely on realizations $\hat\alpha_t$, $\hat\beta_t$ of the two distributions α and β, renewed at each iteration t. Finally, we can represent the iterates produced by such approximations in a space of mixtures of simple functions. Those iterates are a simple transformation of the potentials in the Sinkhorn optimization problem.

Contribution. These observations allows us to propose the following material.

- We introduce a new *online Sinkhorn* algorithm. It produces a sequence of estimates $(\hat w_t)_t \in \mathbb{R}$ and of transportation plans $\hat\pi_t \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$, using two stream of renewed samples $\hat\alpha_t = \sum_{i=1}^n \delta_{x_i^t}$, $\hat\beta_t = \sum_{i=1}^n \delta_{y_i^t}$, where x_i^t and y_i^t are sampled from α and β.

- We show that those estimations are consistent, in the sense that $\hat w_t \to \mathcal{W}_{C,\varepsilon}(\alpha,\beta)$, and $\hat\pi_t \to \pi^\star$.

- We empirically demonstrate that our algorithm permits a faster estimation of optimal transportation distances for discrete distributions, and a convincing estimation of OT distances between *continous* distributions.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

References