Estimation of smooth densities in Wasserstein distance

Jonathan Weed JWEED@MIT.EDU
Department of Mathematics, Massachusetts Institute of Technology

Quentin Berthet Q.BERTHET@STATSLAB.CAM.AC.UK
Statistical Laboratory, DPMMS, University of Cambridge

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

The Wasserstein distances are a set of metrics on probability distributions supported on $\mathbb{R}^d$ with applications throughout statistics and machine learning. Often, such distances are used in the context of variational problems, in which the statistician employs in place of an unknown measure a proxy constructed on the basis of independent samples. This raises the basic question of how well measures can be approximated in Wasserstein distance. While it is known that an empirical measure comprising i.i.d. samples is rate-optimal for general measures, no improved results were known for measures possessing smooth densities. We prove the first minimax rates for estimation of smooth densities for general Wasserstein distances, thereby showing how the curse of dimensionality can be alleviated for sufficiently regular measures. We also show how to construct discretely supported measures, suitable for computational purposes, which enjoy improved rates. Our approach is based on novel bounds between the Wasserstein distances and suitable Besov norms, which may be of independent interest.[1]

Keywords: Wasserstein distance, nonparametric density estimation, optimal transport

1. Introduction

Wasserstein distances are an increasingly common tool in statistics and machine learning. These distances are a special case of the problem of *optimal transport*, one of the foundational problems of optimization (Kantorovitch, 1942; Monge, 1781), and a very important topic in analysis (Villani, 2008). Their recent popularity can be traced back to their empirical success on a wide range of practical problems (see, e.g., Peyré and Cuturi, 2017, for a survey).

In many modern applications, a Wasserstein distance is used as a loss function in an optimization problem over measures. Solving such problems involves optimizing functionals of the form $\nu \mapsto W_p(\nu, \mu)$ where μ is unknown. Given n i.i.d. samples from μ, one therefore seeks an estimator $\tilde{\mu}_n$ such that $W_p(\tilde{\mu}_n, \mu)$ is as small as possible. Much of the statistics literature adopts the plug-in approach and focuses on using the empirical distribution $\hat{\mu}_n$ as an estimator. In this case, the rates of convergence are of order $n^{-1/d}$, and the sample size required for a particular precision is exponential in the dimension, a phenomenon known as the *curse of dimensionality*.

Our work shows that a wavelet estimator achieves substantially better rates of convergence when μ possesses a smooth density. We give minimax-optimal rates over Besov classes $\mathcal{B}^s_{p,q}([0,1]^d)$, and we show that the optimal rates depend strongly on whether the density in question is bounded

1. Extended abstract. Full version appears as arXiv:1902.01778v3.

away from 0. Indeed, we show that the optimal rate for general densities is strictly worse than the corresponding rate for densities bounded below, no matter the smoothness. Our bounds are obtained via a new technical result, which shows that if two probability measures on $[0, 1]^d$ have densities bounded away from zero, then the Wasserstein distance between them can be controlled by a Besov norm of negative smoothness.

Algorithmic aspects are an important part of optimal transport problems. For practical applications, the proposed estimators must therefore also be computationally tractable. We describe a method to produce computationally tractable atomic estimators via resampling from any estimator that outperforms the empirical distribution, under minimal assumptions. We study the computational cost of this method, compared to the cost of using the empirical distribution with n atoms, and exhibit a trade-off between computational cost and statistical precision.

References

L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37: 199–201, 1942.

Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie royale des sciences*, 1:666–704, 1781.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. Technical report, 2017.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.