Large-scale Optimal Transport Map Estimation using Projection Pursuit

Cheng Meng, Yuan Ke, Jingyi Zhang, Mengrui Zhang, Wenxuan Zhong, Ping Ma
The University of Georgia

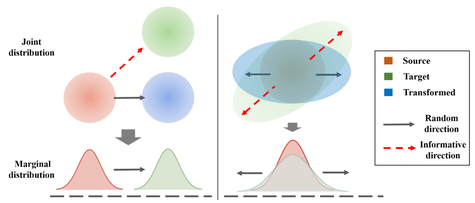
Abstract

This paper studies the estimation of large-scale optimal transport map (OTM), which is a well-known challenging problem owing to the curse of dimensionality. Existing literature approximates the large-scale OTM by a series of one-dimensional OTM problems through iterative random projection. Such methods, however, suffer from slow or none convergence in practice due to the nature of randomly selected projection directions. Instead, we propose an estimation method of large-scale OTM by combining the idea of projection pursuit regression and sufficient dimension reduction. The proposed method, named projection pursuit Monge map (PPMM), adaptively selects the most "informative" projection direction in each iteration. We theoretically show the proposed dimension reduction method can consistently estimate the most "informative" projection direction in each iteration. Furthermore, the PPMM algorithm weakly convergences to the target large-scale OTM in a reasonable number of steps. Empirically, PPMM is computationally easy and converges fast. We assess its finite sample performance through the applications of Wasserstein distance estimation and generative models.

Motivation

Recently, optimal transport map (OTM) draws great attention in machine learning, statistics, and computer science. Nowadays, generative models have been widely-used for generating realistic images, songs and videos. OTM also plays essential roles in various machine learning applications, say color transfer, shape match, transfer learning and natural language processing.

Our contributions. To address the issues mentioned above, this paper introduces a novel statistical approach to estimate large-scale OTMs. The proposed method, improves the existing projection-based approaches from two aspects.



First, PPMM uses sufficient dimension reduction technique to estimate the most "informative" projection direction in each iteration. Second, PPMM is based on projection pursuit. The idea is similar to boosting that search the next optimal direction based on the residual of previous ones.

Problem setup and methodology

Optimal transport map and Wasserstein distance. Denote $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}^d$ as two continuous random variables with probability distribution functions p_X and p_Y, respectively. The problem is to find a transport map $\phi: \mathbb{R}^d \to \mathbb{R}^d$ such that $\phi(X)$ and Y have the same distribution. A standard approach is to find the optimal transport map ϕ^* that satisfies:

$$\phi^* = \inf_{\phi \in \Phi} \int_{\mathbb{R}^d} || X - \phi(X)||^p \, dp_X$$

where Φ is the set of all transport maps, $||\cdot||$ is the vector norm and p is a positive integer. The Wasserstein distance (of order p) between p_X and p_Y is then define as:

$$W_p(p_X, p_Y) = \left(\inf_{J \in J(X,Y)} \int_{\mathbb{R}^d \times \mathbb{R}^d} ||X - Y||^p \, d\, J(X,Y) \right)^{\frac{1}{p}} = \left(\int_{\mathbb{R}^d} ||X - \phi^*(X)||^p \, d\, p_X \right)^{1/p}$$

where $J(X,Y)$ contains all joint distributions J for (X,Y) that have marginals p_X and p_Y.

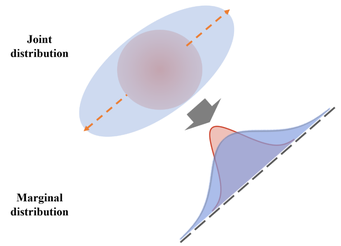
Problem setup and methodology

Denote $\hat{\phi}$ as an estimator of ϕ^*. Suppose one observe $\boldsymbol{X} = (x_1, \dots, x_n)^T \in \mathbb{R}^{n \times d}$ and $\boldsymbol{Y} = (y_1, \dots, y_n)^T \in \mathbb{R}^{n \times d}$ from p_X and p_Y, respectively. The Wasserstein distance $W_{p(p_X, p_Y)}$ thus can be estimated by:

$$\widehat{W_p}(X,Y) = \left(\frac{1}{n} \sum_{i=1}^{n} || \boldsymbol{x}_i - \hat{\phi}(\boldsymbol{x}_i)||^p \right)^{\frac{1}{p}}$$

Projection pursuit method. Projection pursuit regression is widely-used for high-dimensional nonparametric regression models.

Sufficient dimension reduction. Sufficient dimension reduction for regression aims to reduce the dimension of X while preserving its regression relation with Z.



Estimation of the most "informative" projection direction. Consider the problem of estimating an OTM. We regard the input as a binary-response sample, and we utilize the sufficient dimension reduction technique to select the most "informative" projection direction. The metric to quantify the "discrepancy" depends on the choice of sufficient dimension reduction technique.

Algorithm 1 Select the most "informative" projection direction using SAVE

Input: two standardized matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} \in \mathbb{R}^{n \times d}$
Step 1: calculate $\widehat{\Sigma}^{-1/2}$, where $\widehat{\Sigma}$ denotes the sample variance-covariance matrix of $\binom{\mathbf{X}}{\mathbf{Y}}$
Step 2: calculate the sample variance-covariance matrix of $\mathbf{X}\widehat{\Sigma}^{-1/2}$ and $\mathbf{Y}\widehat{\Sigma}^{-1/2}$, denoted as $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, respectively
Step 3: Calculate the eigenvector $\boldsymbol{\xi} \in \mathbb{R}^d$, which corresponding to the largest eigenvalue of the matrix $((\widehat{\Sigma}_1 - I_p)^2 + (\widehat{\Sigma}_2 - I_p)^2)/4$
Output: the final result is given by $\widehat{\Sigma}^{-1/2}\boldsymbol{\xi}/||\widehat{\Sigma}^{-1/2}\boldsymbol{\xi}||$, where $||\cdot||$ denotes the Euclidean norm

Projection pursuit Monge map Algorithm. Now, we are ready to present our estimation method for large-scale OTM. In each iteration, the PPMM applies a one-dimensional OTM following the most "informative" projection direction selected by the Algorithm 1.

Algorithm 2 Projection pursuit Monge map (PPMM)

Input: two matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} \in \mathbb{R}^{n \times d}$
$k \leftarrow 0$, $\mathbf{X}^{[0]} \leftarrow \mathbf{X}$
repeat
(a) calculate the projection direction $\boldsymbol{\xi}_k \in \mathbb{R}^d$ between $\mathbf{X}^{[k]}$ and $\mathbf{Y}$ (using Algorithm 1)
(b) find the one-dimensional OTM $\phi^{(k)}$ that matches $\mathbf{X}^{[k]}\boldsymbol{\xi}_k$ to $\mathbf{Y}\boldsymbol{\xi}_k$ (using look-up table)
(c) $\mathbf{X}^{[k+1]} \leftarrow \mathbf{X}^{[k]} + (\phi^{(k)}(\mathbf{X}^{[k]}\boldsymbol{\xi}_k) - \mathbf{X}^{[k]}\boldsymbol{\xi}_k)\boldsymbol{\xi}_k^{\mathsf{T}}$ and $k \leftarrow k + 1$
until converge
The final estimator is given by $\widehat{\phi}: \mathbf{X} \to \mathbf{X}^{[k]}$

Computational cost of PPMM. In Algorithm 2, the computational cost mainly resides in the first two steps within each iteration. The overall computational cost of Algorithm 2 is of order $O(K\,nd^2 + Kn\log(n))$.

	PPMM	RANDOM	SLICED(10)	SLICED(20)	SLICED(50)
$d = 10$	0.019 (0.008)	0.011 (0.008)	0.111 (0.019)	0.213 (0.024)	0.529 (0.031)
$d = 20$	0.027 (0.011)	0.014 (0.008)	0.125 (0.027)	0.247 (0.033)	0.605 (0.058)
$d = 50$	0.059 (0.036)	0.015 (0.008)	0.171 (0.037)	0.338 (0.049)	0.863 (0.117)

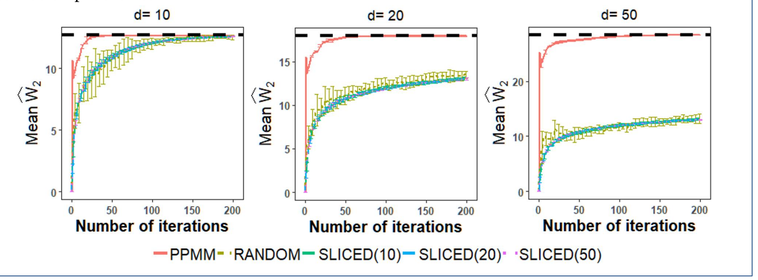
Table 1. The mean CPU time (sec) per iteration, with standard deviations presented in parentheses

	PPMM	RANDOM	SLICED(10)	SLICED(20)	SLICED(50)
MNIST	**0.17** (0.01)	4.62 (0.02)	2.98 (0.01)	3.04 (0.01)	3.12 (0.01)
Doodle (face)	**0.59** (0.09)	8.78 (0.04)	5.69 (0.01)	6.01 (0.01)	5.52 (0.01)
Doodle (cat)	**0.24** (0.03)	8.93 (0.03)	5.99 (0.01)	5.26 (0.01)	5.33 (0.01)
Doodle (bird)	**0.36** (0.03)	7.81 (0.03)	5.44 (0.01)	5.50 (0.01)	4.98 (0.01)

Table 2. The FID for the generated samples (lower the better), with standard deviations presented in parentheses

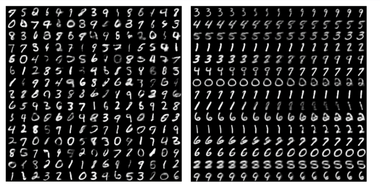
Estimation of optimal transport map

When $d = 10$, RANDOM and SLICED converge to the ground truth but in a much slower manner. When $d = 20$ and 50, neither RANDOM nor SLICED manages to converge within 200 iterations. PPMM is the only one among three that is adaptive to large-scale OTM estimation problems.



Application to generative models

MNIST. We first study the MNIST dataset. First, we visually examine the fake sample generated with PPMM. In the left-hand panel, we display some random images generated by PPMM. The right-hand panel shows that PPMM can predict the continuous shift from one digit to another.



The Google Doodle dataset 1. Predict the continuous shift between two categories. 2. Quantify the similarity between the generated fake samples by calculating the FID in the latent space. The results in justify the superior performance of PPMM over existing projection-based methods.

Acknowledgements

This work was partially supported by National Science Foundation grants DMS-1440037, DMS-1440038, DMS-1438957 and NIH grants R01GM113242, R01GM122080.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In International Conference on Machine Learning, 2017

[2] M. Blaauw and J. Bonada. Modeling and transforming speech using variational autoencoders. In Inter speech, 2016.

[3] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 2017.