

平成 29 年 6 月 26 日現在

機関番号：14301

研究種目：若手研究(A)

研究期間：2014～2016

課題番号：26700002

研究課題名(和文) Fast Optimal Transport and Applications to Inference and Simulation in Large Scale Statistical Machine Learning

研究課題名(英文) Fast Optimal Transport and Applications to Inference and Simulation in Large Scale Statistical Machine Learning

研究代表者

Cuturi Marco (CUTURI, Marco)

京都大学・情報学研究科・准教授

研究者番号：80597344

交付決定額(研究期間全体)：(直接経費) 19,300,000円

研究成果の概要(和文)：We showed over the course of this project that optimal transport theory could have an impact on real world applications (machine learning, imaging, graphics) using a proper regularization of the well known optimal transport problem.

研究成果の概要(英文)：This funding was used to push forward the idea that optimal transport could be used numerically to solve real life problems using a regularization approach. We have demonstrated over the course of this project that these ideas were feasible, and have shown their applicability to a very wide range of applications, ranging from graphics and medical imaging to graphics and machine learning. These ideas were presented in top conferences and journals.

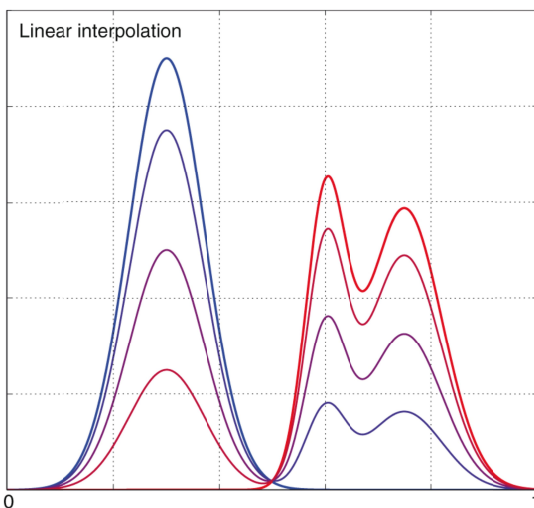
研究分野：統計学

キーワード：最適輸送理論 機械学習 最適化 グラフィックス

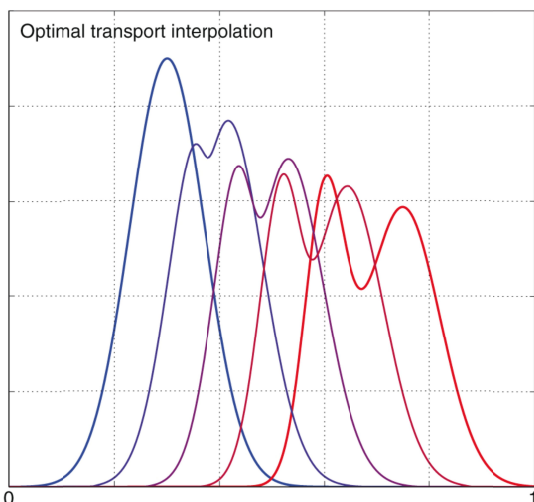
1. 研究開始当初の背景

The optimal transport (OT) geometry is a very useful tool to compare probability distributions. That geometry provides not only mathematicians with a useful distance between probability measures (the so-called optimal transport distances, also known as Wasserstein or earth mover's), but also with a very rich family of geometric tools to interpolate and handle probability distributions in an intuitive way.

As a first illustration, consider the two figures below, which illustrate how two mixtures of Gaussians can be interpolated. The first Gaussian, the higher curve on the left, can be gradually turned into the mixture on the right (the higher distribution on the right) by means of a standard (Euclidean) interpolation, which results essentially in purely “vertical” changes, as displayed below, where intermediary lines stand for interpolations.



The optimal transport interpolation (displayed in the second figure below) between these two measures has a more physical interpretation: the mass of each measure is transported “laterally”, by moving the first measure on the left towards the measure on the right in a spatially coherent and meaningful way. That transformation is called a displacement interpolation.



The main observation behind this simple example is that optimal transport provides very different tools than the standard Euclidean geometry.

Because of its origins in fluid mechanics and the movement of gases and masses, the optimal transport geometry has been widely studied in the mathematics literature, notably partial differential equations. That field has received ample attention thanks to Cédric Villani's fields medal in 2010.

OT has, however, an important drawback. As is often the case with tools that have an intuitive physical interpretation, the computational cost associated with recreating that physically plausible displacement is usually very high. Although the simple example above is easy to create, higher-dimensional problems (both in terms of the dimension of the ambient space and in terms of the number of points that are tracked when considering empirical measures) such computations are usually intractable. It comes therefore as no surprise that OT was barely used at all in practice for data analysis until recent developments.

Within that context, I have proposed in 2013 a novel numerical method that can be used to approximate very efficiently optimal transport between probability measures. The main tool of that approach is regularization. The main engine of that method is an existing iterative algorithm known as the Sinkhorn fixed-point iteration. After showing that this algorithm results in an approximation of optimal transport that has favorable mathematical and computational properties (quadratic complexity, differentiability, parallelism) I seized that opportunity to use the Sinkhorn algorithm for more ambitious tasks in machine learning and computational statistics. This opportunity was the main context behind my kakenhi Wakate-A project.

2. 研究の目的

The goal of this project was to build upon my early proposal to regularize the optimal transport with an entropic term. In that paper my contribution was mostly limited to that of re-discovering a known algorithm. I believed very early in the project that there were several worthwhile research opportunities and low-hanging fruits that could be solved using the Sinkhorn algorithm. Therefore, I focused during this kakenhi projects on three main contributions: *algorithms*, *methodologies* and *applications*. My goals were therefore to:

1. Explore further the computational aspects of that regularization. Study whether it was

possible to use other standard *algorithms* and theoretical tools from convex optimization (Fenchel duality, Bregman divergences, acceleration and quasi Newton methods) to obtain better numerical performance.

2. Create new flexible *methodologies* that could incorporate a priori (geometric) information on the space of observations to compare probability distributions on such observations. These methodologies were to be used with different usages: averaging (the problem of computing barycenters and centroids), dimensionality reduction, or supervised learning with structured output spaces (when the output of a classifier is itself a probability distribution), interpolations in graphics or between texts seen as clouds of points in word-embedding spaces, *etc.*
3. Test these new algorithms and methodologies on real world data, such as that coming from natural language processing, neuroscience, imaging or graphics, and see whether they could result in substantially different or better results.

3 . 研究の方法

Our research was goal driven, and we tried to fulfill all of the three goals outlined above in parallel.

To realize the first goal, I have worked extensively with a team from INRIA (France) that was specialized in optimization (G. Carlier, J.D. Benamou, G. Peyré, F. Bach). In terms of methodologies, I worked with two students at Kyoto University (V. Seguy, A. Rolet) as well as a postdoctoral researcher from U. Berkeley (A. Ramdas). To obtain convincing results in applicative fields I teamed up with experts in graphics (J. Solomon, MIT; N. Bonneel, INRIA) and an expert in neuroscience (A. Gramfort, Telecom).

From a computational perspective, I detected very early the potential benefits of using general purpose graphical processing units (GPGPU). I have therefore invested a substantial amount of money to equip myself with an efficient computational server with several NVIDIA cards, and taught myself as well as students working with me how to use such cards.

4 . 研究成果

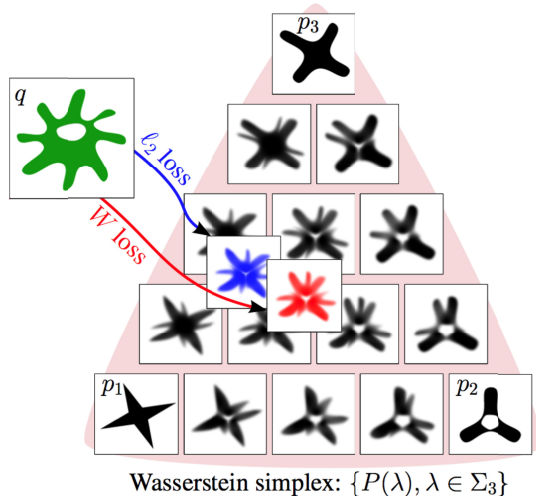
This Wakate-A project has directly funded a total of 13 publications, among which 2 ACM SIGGRAPH papers, 3 NIPS conference proceedings, 2 ICML conference proceedings, 1 JMLR and 2 SIAM journal articles. All of these conferences and journals are top tier, if not the most respected venues in their respective fields. Several of these papers have now more than 50 citations, which reflects their immediate impact in these fields.

From an algorithmic perspective, we proposed in papers [3] (see bibliographic section) novel convex optimization tools to compute barycenters. These tools were natural generalizations of the Sinkhorn algorithm. That paper has already been cited more than 90 times in 2 years. In [2] we proposed to solve more advanced and general variational problems involving Wasserstein distances using Fenchel duality. We believe that this paper holds several ideas for future applied work. In [5] we proposed the first stochastic optimization approach to approximate optimal transport. That paper is being used now by the deep learning community to train generative models for natural images.

In the field of computer graphics, our methods gave for the very first time the ability to compute automatically, and in very reasonable time, shape interpolations in 3 dimensions with very natural results [11]. That feat was made possible through the important observation that the Sinkhorn algorithm can be implemented in linear time when using shapes in regular grids with separable metrics. We propose such interpolations in the figure below, in which the three 3D shapes at the corners of the triangle are handled as 3D uniform volumetric distributions.



We were able to obtain shortly after, in our second SIGGRAPH paper [8], an automatic differentiation algorithm to solve the inverse problem related to that interpolation goal.



Some of the applications in [8] including analysis of MRI data. We have also explored such applications in [12].

In machine learning, our goal was to analyze histograms of features, such as bag-of-words, and we were able to propose dictionary learning methods [9] as well as PCA methods [10] to carry out dimensionality reduction.

Finally, two of our recent papers provide several research opportunities. We have explored more advanced ideas that generalize the Wasserstein distance and give it additional, desirable invariance properties in [7]. We were also the first to provide an algorithm to estimate the density of a probability distribution under a Wasserstein loss between data and the probability model [6]. These ideas have now reached the ML community, where algorithms such as Wasserstein GAN have received now wide attention.

We can thus safely claim that this Wakate-A project was a success, since we were able to introduce these results both to audiences interested in the mathematical, statistical, applied, hardware, and machine learning aspects of optimal transport.

5 . 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 13 件)

1. Ramdas, N. Garcia Trillos, M. Cuturi, On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests, *Entropy*, 19(2), 2017
2. M. Cuturi, G. Peyré. A Smoothed Dual Formulation for Variational Wasserstein Problems, *SIAM Journal on Imaging Sciences*, 9(1), 320–343, 2016. 28 citations.

3. J.D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyre. Iterative Bregman projections for regularized transportation problems, *SIAM Journal on Scientific Computing* Vol. 37 (2), A1111-A1138, 2015. 93 citations.
4. M. Cuturi and D. Avis. Ground metric learning. *Journal of Machine Learning Research* (15) 533-564, 2014. 28 citations.
5. A. Genevay, M. Cuturi, G. Peyré, F. Bach. Stochastic Optimization for Large Scale Optimal Transport, In *Advances in Neural Information Processing Systems (NIPS)* 30, 2016. Acceptance: 568/2500=22%. 10 citations.
6. G. Montavon, K.-R. Muller and M. Cuturi. Wasserstein Training of Restricted Boltzmann Machines, In *Advances in Neural Information Processing Systems (NIPS)* 30, 2016. Acceptance: 568/2500=22%. 8 citations.
7. G. Peyre, M. Cuturi and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. *Proceedings of the 33rd International Conference on Machine Learning (ICML), JMLR W&CP*, volume 48, 2016. ACM acceptance: 322/1327=24%. 5 citations.
8. N. Bonneel, G. Peyre, M. Cuturi. Wasserstein Barycentric Coordinates: Histogram Regression using Optimal Transport, *ACM SIGGRAPH* 2016. Acceptance: 119/467=25%. 6 citations.
9. A. Rolet, M. Cuturi, G. Peyre. Fast Dictionary Learning with a Smoothed Wasserstein Loss, *Proc. of AISTATS* 2016. Acceptance: 165/537=30%. 8 citations.
10. V. Seguy and M. Cuturi. Principal Geodesic Analysis for Probability Measures under the Optimal Transport Metric, In *Advances in Neural Information Processing Systems (NIPS)* 29, 2015. Acceptance: 403/1838=22%. 11 citations.
11. J. Solomon, F. de Goes, G. Peyre, M. Cuturi, A. Butscher, A. Nguyen, T. Du, L. Guibas. Convolutional Wasserstein distances: efficient optimal transportation on geometric domains, *ACM SIGGRAPH* 2015. Acceptance: 118/650=18%. 49 citations.
12. A. Gramfort, G. Peyre, M. Cuturi, Fast optimal transport averaging of neuroimaging data, *Information Processing in Medical Imaging (IPMI)* 2015, 10 citations.
13. M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. *Proceedings of the 31st International Conference on Machine Learning (ICML), JMLR W&CP*, volume 32, 2014. Acceptance:

310/1238=25%. 87 citations

〔学会発表〕(計 19 件)

1. 26.09.16 M.Cuturi, *Regularized Optimal Transport*, IPAM Workshop, ML meets many particle problems, UCLA
2. 19.09.16 M.Cuturi, *Soft-DTW*, ECML/PKDD Workshop, Adv. Analytics and Learning on Temporal Data, Riva del Garda
3. 18.07.16 M.Cuturi, *Regularized Optimal Transport*, Computational Optimal Transport, U. Montréal
4. 03.06.16 M.Cuturi, *Regularized Optimal Transport*, Korea/Japan ML Symposium, Seoul
5. 07.04.16 M.Cuturi, *Regularized Optimal Transport*, Statlearn 2016, Vannes
6. 16.03.16 M.Cuturi, *Regularized Optimal Transport*, EECS - BLISS Seminar, UC Berkeley
7. 19.02.16 M.Cuturi, *Regularized Optimal Transport*, Microsoft Research Cambridge talk
8. 04.12.2015 M.Cuturi, *Regularized Optimal Transport*, ML Tea seminar, MIT
9. 01.12.2015 M.Cuturi, *Regularized Optimal Transport*, CILVR Seminar, NYU
10. 25.11.2015 M.Cuturi, *Regularized Optimal Transport*, IBIS Workshop, Tsukuba
11. 28.10.2015 M.Cuturi, *Regularized Optimal Transport*, Germany-Japan Adaptive BCI Workshop, Kyoto
12. 06.10.2015 M.Cuturi, *Regularized Optimal Transport*, Optimization in Machine Learning, Vision and Image Processing, Toulouse
13. 18.09.2015 M.Cuturi, *Regularized Optimal Transport*, GPU Technology Conference, Tokyo
14. 06.03.2015 M.Cuturi, *Regularized Optimal Transport*, New trends in Optimal Transport, Bonn
15. 15.02.2015 M.Cuturi, *Regularized Optimal Transport*, Advances in Numerical Optimal Transport, Banff, Canada
16. 11.01.2015 M.Cuturi, *Regularized Optimal Transport*, Optimization and Statistical Learning Workshop, Les Houches, France
17. 18.12.2014 M.Cuturi, *Regularized Optimal Transport*, Learning Theory Workshop, FoCM, Montevideo, Uruguay
18. 23.09.2014 M.Cuturi, *Regularized Optimal Transport*, Gatsby Neuroscience Unit, UCL, London
19. 22.05.2014 M.Cuturi, *Regularized Optimal Transport*, Deep Learning Workshop, Shonan village, Japan

6 . 研究組織

(1)研究代表者

クトゥリ マルコ (Cuturi Marco)

京都大学・情報学研究科・准教授

研究者番号 : 80597344