Titouan Vayer (titouan.vayer@inria.fr)
Rémi Gribonval (remi.gribonval@inria.fr)   **October 18, 2023**

Internship proposal: Butterfly Sparse Matrix Factorization for Regularized Optimal Transport

Keywords: Optimal transport, matrix factorization, sparsity, machine learning.

Topic

Optimal transport (OT) is a tool that plays a central role in various machine learning applications today, whether it's for generative models, domain adaptation, analysis of cellular dynamics, neural networks, or graph models (for more details, see [PC+19]). The essence of OT can be distilled into the following question: "how can one move masses from one place to another in a way that minimizes the overall effort of transportation?" This simple idea has evolved over the years into an elegant theory at the intersection of mathematics, optimization, and computer science.

In its original formulation, OT faces a scalability issue: solving the underlying optimization problem has *cubic complexity with respect to the number of points*. The introduction of *regularized transport* [Cut13] marked a groundbreaking advancement in the field by achieving a mere *quadratic complexity*. This breakthrough has, in turn, facilitated the deployment of OT in medium to large-scale applications. This regularization leads to a matrix scaling problem where a specific kernel matrix $\mathbf{K}$ is normalized to have prescribed marginals, achieved through the well-known Sinkhorn-Knopp matrix scaling algorithm [SK67; Ide16].

However, for truly large-scale applications, this quadratic complexity remains prohibitive. Hence, alternative approaches have been proposed, e.g. by leveraging the low-rank structure of the kernel $\mathbf{K}$ [SC20], exploring factorization techniques like Nyström-based low-rank approaches [Alt+19], or employing sketching methods [Cha+23] to achieve nearly linear complexity.

On the other hand, *sparse matrix factorization* approaches aim to approximate a given matrix $\mathbf{K}$ with a product of sparse factors $\mathbf{X}^{(1)} \cdots \mathbf{X}^{(J)}$. Such factorization is desired to reduce time and memory complexity in numerical methods involving $\mathbf{y} \to \mathbf{Ky}$, as each matrix multiplication with a sparse factor involves only a few non-zero coefficients. Recently, a specific structure of sparse factors, known as the *butterfly structure*, has shown promising results [Dao+19; ZRG23]. It enables, among other advantages, a $O(n\log(n))$ matrix multiplication by employing fast transforms akin to the discrete Fourier transform.

With the framework established, **the goal of this internship is to explore the theoretical and practical potential of butterfly matrix factorization for regularized optimal transport**. A first avenue of investigation will be to assess whether such a factorization of the kernel $\mathbf{K}$ can reasonably approximate the transport problem. A secondary aim is to show that this factorization can lead to the design of an algorithm for OT with an almost linear time complexity. Lastly, a more theoretically-oriented task will encompass the quantification of the error resulting from this approximation.

The internship will also provide an opportunity to contribute to the Python library FaμST developed by the team.

Practical aspects

We are looking for a highly motivated student, willing to continue with a PhD thesis, with a background in mathematics (optimization, probability and statistics) and/or electrical engineering (signal/image processing, harmonic analysis). Strong abilities in computer sciences will be appreciated.

The intern will be granted the usual stipend of ≈ 600 euros/month. If the candidate is successful, this internship could be pursued by a PhD.

This internship will take place at ENS Lyon in the OCKHAM Inria team. It will be co-supervised by Titouan Vayer (CR Inria) and Rémi Gribonval (DR Inria).

Do not hesitate to contact us for more information.

References

[PC+19]   Gabriel Peyré, Marco Cuturi, et al. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[Cut13]   Marco Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems* 26 (2013).

[SK67]   Richard Sinkhorn and Paul Knopp. "Concerning nonnegative matrices and doubly stochastic matrices". In: *Pacific Journal of Mathematics* 21.2 (1967), pp. 343–348.

[Ide16]   Martin Idel. *A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps.* 2016.

[SC20]   Meyer Scetbon and Marco Cuturi. "Linear time Sinkhorn divergences using positive features". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 13468–13480.

[Alt+19] Jason Altschuler et al. "Massively scalable Sinkhorn distances via the Nyström method". In: *Advances in neural information processing systems* 32 (2019).

[Cha+23] Moses Charikar et al. "Fast Algorithms for a New Relaxation of Optimal Transport". In: *The Thirty Sixth Annual Conference on Learning Theory.* PMLR. 2023, pp. 4831–4862.

[Dao+19] Tri Dao et al. "Learning fast algorithms for linear transforms using butterfly factorizations". In: *International conference on machine learning.* PMLR. 2019, pp. 1517–1527.

[ZRG23]   Léon Zheng, Elisa Riccietti, and Rémi Gribonval. "Efficient identification of butterfly sparse matrix factorizations". In: *SIAM Journal on Mathematics of Data Science* 5.1 (2023), pp. 22–49.