Junior researcher (Ph.D. or Postdoc) position

Institut 3IA Côte d'Azur
Université Côte d'Azur & INRIA

Optimal Transport of High-Dimensional Data for Unsupervised Domain Adaptation

Advisors and location:

- Advisor: Pr. Charles Bouveyron
- Co-advisor: Marco Corneli
- Team: Maasai project-team, LJAD, Université Côte d'Azur & Inria
- Localisation: Maasai project-team, Centre Inria d'Université Côte d'Azur, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France

Context and project: Due to its proven versatility, optimal transport (OT) is becoming more and more popular within the machine learning community. Basically, once the observed data is identified with a probability distribution (possibly the empirical mass function), optimal transport allows to consistently assess the similarity between complex instances such as point clouds, images or graphs. However, as the modern data are increasingly high-dimensional, OT is also now facing an old problem in optimization and statistical learning: the curse of dimensionality. Among the OT problems that have to face the high dimensionality of the data, we can mention as a popular example the calculation of the Frechet inception distance for comparing the distribution of generated images with the distribution of a set of ground-truth images, using the Wasserstein distance between two full Gaussian distributions. Current OT approaches for high-dimensional situations rely on projections of the data or measures onto low-dimensional spaces, which inevitably results in information loss.

In a recent work [1], we considered the case of high-dimensional Gaussian distributions with parsimonious covariance structures and lower intrinsic dimension. We exhibited a simplified closed-form expression of the Wasserstein distance with an efficient and robust calculation procedure based on a low-dimensional decomposition of empirical covariance matrices, without relying on data projections. Furthermore, we provided a closed-form expression for the Monge map, which involves the exact calculation of the square-root and inverse square-root of the source distribution covariance matrix. This approach offers analytical and computational advantages in comparison to existing methods. In addition to being able to compute both the Wasserstein distance and the transport map, our method outperforms model-free methods, in high dimension, even in the case of non-Gaussian distributions.

To go further in this context of model-based optimal transport (MBOT), the purpose of this Ph.D (or postdoc) position, within the Institut 3IA Côte d'Azur (Univ. Côte d'Azur & INRIA), will be focused on **extending this seminal work to other high-dimensional distributions, and optimal transport distances with applications to unsupervised domain adaptation.** In particular, we will consider the extensions to the family of exponential distributions and to mixture models. The proposed methodologies will be then applied to real-world situations in either Medicine (omics data, medical imaging, …) or Digital Humanities (History, Archeology, …).

Expected skills: The candidate should have a graduate degree (Master 2 degree). His/her scholar background should include:

- statistical/machine learning, statistical inference, clustering, classification
- deep learning, variational auto-encoding, back-propagation,
- excellent knowledge of R and Python. The knowledge of C++ would be a plus.

Application: Application files should contain a resumé, an application letter and grade records of the 2 last years (M1 & M2). Support letters from senior researchers will be also appreciated.

Applications should be sent by email to charles.bouveyron@inria.fr and marco.corneli@inria.fr.

References:

[1] Charles Bouveyron and Marco Corneli (2025), *Scaling Optimal Transport to High-Dimensional Gaussian Distributions,* Preprint HAL n°04930868, https://hal.science/hal-04930868.

[2] Peyré, G., Cuturi, M., et al. (2019). Computational optimal transport: With applications to data science. Foundations and Trends in Machine Learning, 11(5-6):355–607.

[3] Villani, C. et al. (2009). Optimal transport: old and new, volume 338. Springer.

[4] Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2016). Optimal transport for domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 39(9):1853–1865.