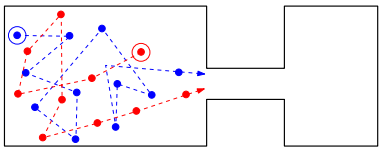
Supervisor: Frédéric Cazals
Centre Inria at Université Côte d'Azur, France
3IA Côte d'Azur
Email: Frederic.Cazals@inria.fr
Web: `http://team.inria.fr/abs`

Master internship proposal



On the convergence of iterative sampling methods and the detection of meta-stable states in dynamical systems

Keywords: dynamical systems, meta-stable states, Monte Carlo Markov Chain, two-sample tests, high dimensional geometry, molecular conformations.

Context. In the theory of dynamical systems, loosely speaking, a *meta-stable state* is a region in phase or conformational space where the system remains sufficiently long before jumping to another such state, via some *transition* which is in general a *rare* event. Equivalently, such a state may be characterized by *local* ergodicity, meaning that for such a region and at the relevant time scale, spatial averages equal time averages [1]. A key difficulty for complex systems, for example a molecule undergoing conformational changes, is to understand the multiple scales at which the system is meta stable.

In statistics, the convergence of iterative methods in general and Monte Carlo Markov chains in particular relies on techniques related to r-hat ($\hat{R}$) and effective sample sizes [2]. In the theory of statistical hypothesis testing [3], a two-sample test is a statistical test aiming at detecting whether two collections of samples (*e.g.* in a high dimensional space, on a manifold, etc) have the same underlying distribution [4].

Goals. The goal of this master internship is to develop a novel approach for the detection of meta-stable states in dynamical systems, using ideas from geometry [5], information theory [6, 7], and statistical hypothesis testing [4]. Tests will be conducted on classical test systems [8], as well as on proteins undergoing conformational changes [9].

The work envisioned encompasses the design and mathematical analysis of algorithms, their coding (C++ and python), as well their experimental evaluation.

Training. Master 2 or equivalent degree in Computer science (algorithms) or machine learning or statistics or statistical physics.

Misc. Internship with *gratification*. Possibility to follow-up with a **PhD thesis**.

References

[1] C. Schön and M. Jansen. Prediction, determination and validation of phase diagrams via the global study of energy landscapes. *Int. J. of Materials Research*, 100(2):135, 2009.

[2] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of mcmc (with discussion). *Bayesian analysis*, 16(2):667–718, 2021.

[3] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[4] A. Gretton, K. M. Borgwardt, J.R. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[5] N. Verma, S. Kpotufe, and S. Dasgupta. Which spatial partition trees are adaptive to intrinsic dimension? In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 565–574. AUAI Press, 2009.

[6] A. Lhéritier and F. Cazals. A sequential non-parametric multivariate two-sample test. *IEEE Transactions on Information Theory*, 64(5):3361–3370, 2018.

[7] F. Cazals and A. Lhéritier. Beyond two-sample-tests: Localizing data discrepancies in high-dimensional spaces. In P. Gallinari, J. Kwok, G. Pasi, and O. Zaiane, editors, *IEEE/ACM International Conference on Data Science and Advanced Analytics*, Paris, 2015.

[8] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[9] T. O'Donnell and F. Cazals. Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry. *J. Comp. Chem.*, 2023.