PhD thesis proposal: On the convergence of iterative sampling methods and the detection of meta-stable states in protein science

Keywords: dynamical systems, meta-stable states, Monte Carlo Markov Chain, two-sample tests, high dimensional geometry, tree-like decompositions, random forests, 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 samples sizes [2,3]. In the theory of statistical hypothesis testing [4], 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 [5].

Goals. The goal of this thesis is to develop a novel approach for the detection of meta-stable states in dynamical systems, using ideas from geometry [6], information theory [7,8], and statistical hypothesis testing [5]. The line pursued will be to use properties from spatial decompositions yielded by trees (cf random projections trees and random forests), to exploit local properties in the sample space, and the ability to update these data structures dynamically. Tests will be conducted on classical test systems and proteins undergoing conformational changes [9].

The work envisioned encompasses the design and the 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.

Conditions. Position at Centre Inria at Université Côte d’Azur, France.

References


On the convergence of iterative sampling methods and the detection of meta-stable states in protein science

Etude de la convergence de méthodes de sampling itératives, et detection d’états meta-stables pour les protéines

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, a state is characterized by local ergodicity, meaning that for such a region and at the relevant time scale, spatial averages equal time averages. In the context of biophysics, meta-stable states are especially important for proteins, as they provide insights on the stable conformations and important intermediates on reaction pathways accounting for the function of proteins. Alas, while the prediction of folded protein structures is now possible thanks to Alphafold, the study of dynamics remains a major open problem.

The work envisioned is to develop novel multi-scale methods to detected meta-stable state in protein simulations. The line pursued will be to use spatial decompositions yielded by (random) trees and forests, to exploits local properties and also make the detection methods dynamic.