Internship in data science in cosmology

Encadrant: Laura Grigori (Director of Research INRIA), Laura.Grigori@inria.fr

Laboratoire d’accueil: INRIA Paris

Equipe d’accueil: In the context of the Alpines group, joint between Inria and Laboratoire J.L. Lions.

Funded internship, to start on March/April, for 5/6 months. Can be pursued with a Phd thesis.

Title: Data science in cosmology

Context
This internship is in the area of data analysis of cosmological data sets as collected by contemporary and forthcoming observatories. This is one of the most dynamic areas of modern cosmology. Our specific target are data sets of Cosmic Microwave Background (CMB) anisotropies, measurements of which have been one of the most fruitful of cosmological probes. CMB photons are remnants of the very early evolution of the Universe and carry information about its physical state at the time when the Universe was much younger, hotter and denser.

This internship will be performed within the B3DCMB project, Big Bang from Big Data (of the cosmic microwave background) funded by ANR. The main objective of this project is to empower the CMB data analysis with novel high performance tools and algorithms superior to those available today and which are capable of overcoming the existing performance gap. The challenge is multi-faceted as it involves at the same time need for computationally efficient numerical algorithms capable of exploiting power of the the largest current and forthcoming supercomputers, their advanced implementations and sophisticated statistical methods. This is a multi-disciplinary challenge which will be addressed with help of the interdisciplinary B3DCMB team.

Description of the project
The majority of the current CMB observatories perform observations by scanning the sky with a telescope, repeatedly revisiting the same sky areas and progressively covering a bigger and bigger part of the sky. The signal measured by arrays of thousands of detectors is registered at the constant rate of a few hundreds times per second over the periods of many years. CMB data analysis as collected by CMB instruments, known as the map making problem, can be modeled as a generalized least squares problem of very large scale. This internship will focus on the robust and accurate resolution of this generalized least squares problem and its suitability for massive parallelism. Our goal is to develop communication
avoiding preconditioned Krylov subspace iterative method for solving it, starting from the work in [1] and using low rank approximation techniques.

CMB data sets have volumes in excess of many Petabytes and so massively parallel supercomputers need to be employed to allow processing an entire data set at once. In this research we also address the important problem of increasing communication cost, and this is one of the main challenges that the high performance computing community faces today. Our research focuses on a novel approach to linear algebra algorithms, which aims at minimizing the communication, where communication refers to both its volume and the number of messages exchanged between parallel computers or between different levels of memory hierarchy. The algorithm that we will develop will be communication avoiding in the sense of [2, 3].

References

[1] M. Szydlarski, L. Grigori, and R. Stompor, Accelerating the Cosmic Microwave Background map-making problem through preconditioning, Astronomy and Astrophysics Journal, Section Numerical methods and codes, Volume 572, Article A39
