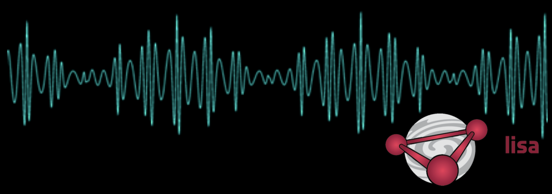
Internship on the topic of novel approaches for the gravitational wave data analysis

SYRTE l'Observatoire de Paris | PSL

SYstèmes de Référence Temps-Espace

Optimal representation of Extreme Mass Ratio Insiral waveforms using ML techniques



PROJECT DESCRIPTION

Extreme Mass Ratio Inspirals (EMRI) are one of the most interesting signals which will be detected by Laser Interferometer Space Antenna (LISA). They will sample highly curved space-time around Massive Black Holes (MBH) and allow for insight into the strong regime of gravity. EMRIs are signals which come from the compact objects falling onto MBH, they typically last for 10^5 number of cycles which translates into an order of one year length. They have relatively low signal-to-noise ratios (SNR) and typically it is required to observe them for a long time to be able to accumulate enough SNR.

The signal has a very complicated structure as it contains the evolution of the harmonics of three fundamental frequencies: orbital frequency, perihelion precession frequency and orbital plane precession frequency. At the moment the best up-to-date approaches for the EMRI parameters estimation involve semi-coherent searches, which are based on the combination of analyses for relatively short segments of data.

To facilitate attempts of applying other machine learning driven approaches to parameter estimation, such as Normalising Flows, we have to find the best way to represent a signal in the compressed way. Therefore this project will be focused on investigating different ways to accumulate information from long lasting signals. We will look at different types of Autoencoders, starting from linear Principle Component decomposition and simple fully connected networks, gradually increasing it to more complicated methods. The goal in the end of the project is to find the most compact representation of the data which will preserve sufficient amount of information.

CONTEXT

Since the first detection of gravitational waves by LIGO/Virgo the field of the gravitational wave astronomy has been developing very rapidly. The space based gravitational wave observatory – Laser Interferometer Space Antenna (LISA) is planned to be launched in 2034 and is preparing for mission adoption next year. LISA will observe gravitational wave signals with the peak sensitivity at 20 mHz. At this frequency range the signals will be long lasting. Moreover there will be thousands of signal appearing in the data at the same time. Among the signals that we expect to see are Galactic Binaries, Massive Black Hole Binaries, Extreme Mass Ratio Inspirals, Stochastic Background and maybe other unexpected sources. The extraction of the signals is done assuming that we model the data as a sum of the signal model and the noise of the observation. The signal models are called *waveforms*. They are defined by a set of astrophysical parameters and can be viewed as a time series. The difficulty for the data analysis arrises from the length of these time series which for some signals will last for years. Therefore the challenge for the data analysis, especially Machine Learning approaches rises in the ways the data can be represented and compressed. The investigation of the methods for the data compression for one type of signal, namely Extreme Mass Ratio Inspirals, will be the goal of this internship.

TEAM

The internship will be carried out at APC (Astroparticule et Cosmologie) lab – one of the main labs in France and worldwide working on the LISA data analysis. We will be collaborating with the SYRTE lab which among others specialises on the preprocessing of LISA data and on the observations of galactic centre.

PREREQUISITES

– Experience with Python programming language;
– Basic knowledge of the signal processing, such as Fourier Analysis;
– Basics knowledge of Probability Theory;
– Basic knowledge of Machine Learning.

ACCUIRED COMPETENCES

The student will learn modern methods of data analysis and signal processing. We will investigate different techniques which can be used for data compression, starting from linear methods such as Principle Component Analysis to Autoencoders with different architectures: CNNs, LSTMs, Transformers, Residual Networks and so on. The outcome of the project will be to find the method which will provide the optimal compression of the EMRI signal.

CONTACT

✉ korsakova@apc.in2p3.fr
✉ aurelien.hees@obspm.fr