17 Chemo-dynamical analysis of Milky Way's stellar populations with unsupervised multi-dimensional clustering

Proposer Name	Dr GyuChul Myeong
Proposer Role and Affiliation	Research Fellow, Institute of Astronomy
Proposer Contact Email	gm564@cam.ac.uk
Key Publication	Milky Way's Eccentric Constitutents with Gaia, APOGEE, and GALAH

Project Description

Large galaxies, such as our Milky Way, go through complex evolutionary phases governed by various different mechanisms. As a result, the present day Galaxy is a composed product of stars from different origins and epochs, although it is not easy to distinguish one star's origin from another individually. Stars formed from the same process and the same source share comparable chemical and dynamical properties which reflects their formation environment. To understand the complex evolutionary history and underlying formation mechanisms of the Galaxy, it is essential to know the chemical and dynamical properties of the stars produced from each process. Using a combination of detailed chemical and dynamical information from Gaia, APOGEE, and GALAH, we attempt to classify the Milky Way stars into an unspecified number of sub-groups with distinguishable chemo-dynamical trends. Gaussian mixture modelling is adopted as an unsupervised clustering method. The chemo-dynamical property of each identified component can help us to trace the Galaxy's evolutionary history. In specific we focus on identifying and studying the Milky Way's known stellar populations, such as GS/E (ex-situ population from a major galactic merger event the Milky Way experienced), Aurora (ancient population from early epoch of Milky Way evolution), Splash (dynamically heated population due to the GS/E merger), and Eos (result of a star-formation as a consequence of GS/E merger).

Project goals

Main goals of the project:

1. Understand the high-dimensional dataset, implement/test unsupervised clustering models
2. Apply model selection criteria based on the goodness-of-fit and complexity
3. Analyse the output result and interpret the physics behind the features uncovered from data

Extension directions:

- Apply dimensionality reduction and visualisation for high-dimensional data
- Explore the effect of data uncertainty in unsupervised clustering method

Prerequisites

Required:

1. Some coding skills in Python or other suitable programming languages

Desirable:

- Some knowledge related to Galactic Archaeology (there will be a minor module course) would be helpful but is not required

Reading List

- From dawn till disc: Milky Way's turbulent youth revealed by the APOGEE+Gaia data, Belokurov et al., 2022, MNRAS, 514, 689B
- Co-formation of the disc and the stellar halo, Belokurov et al, 2018, MNRAS, 478, 611B
- The biggest splash, Belokurov et al, 2020, MNRAS, 494, 3880B
- Extreme deconvolution: Inferring complete distribution functions from noisy, hetrogeneous and incomplete observations, Bovy et al., 2011, AnApS, 3.1657B

Data Access

The main datasets are publicly accessible (Gaia DR3, APOGEE DR17, GALAH DR3), but some of the necessary information requires further computations (e.g., orbital integration, energy calculation) which is beyond the scope of the project. The student can contact the PI (GyuChul Myeong; gm564@cam.ac.uk) for the dataset that can be used for the project. The main software required for the project (Extreme Deconvolution) is publicly available (https://github.com/jobovy/extreme-deconvolution). Alternatively, the student can try scikit-learn's Gaussian Mixture Modelling (https://scikit-learn.org/stable/).