Here we provide 4 topics for your final projects. Please take a look, and ideally, we encourage you to make your choice by Dec. 2. If you need more time, it should be no later than Dec. 9th to ensure that you have enough time to work on it.

If you choose one of the following projects, and work on your own, please directly inform us (email).

There is also the option that two students work on one project together. In this case, you need to discuss with one of us, and in this case, we will anticipate higher standards for grading.

You are also welcome to design your own final project. In this case, again, you need to talk to one of us for approval.

The final project is due on Jan. 10, and please submit your work online.

Final project: Thermochemical Equilibrium (Bai)

As mentioned in class, equilibrium chemistry at given pressure and temperature is often solved by minimizing the Gibbs free energy. This approach is often applied in atmospheric science, and more recently to model exoplanetary atmospheres.

While we have discussed various minimization algorithms in class, most of which are relatively straightforward, we understand that writing your own code for this task can be too demanding for a final project. There are two recent open-source codes for calculating thermochemical equilibrium:

The "TEA" code: `https://github.com/dzesmin/TEA` , and the corresponding paper can be found here: `https://ui.adsabs.harvard.edu/abs/2016ApJS..225....4B/abstract`

The GGchem code: `https://github.com/pw31/GGchem` , and the corresponding paper can be found here: `https://ui.adsabs.harvard.edu/abs/2018A%26A...614A...1W/abstract`

Both codes have sufficient documentation for usage, accompanied by a method paper that describes the relevant context, equations solved, input data, algorithms, and test problems.

This project is open-ended. You are encouraged to explore either or both of the codes, run some sample problems, discuss/compare the numerical algorithms used, reproduce some of the results in the papers, and/or run some calculations of your own interest. Grading will be mainly based on the following criteria:

- Understanding of the physical problem (5 pts)
- Understanding of the algorithm (5+2 pts)
- Comprehensiveness of the exploration (5+2 pts)
- Depth of the exploration (5+2 pts)
- Creativity (5+2 pts)

Note that each item has 5 nominal points, whereas four of them can be awarded 2 bonus points in exceptional cases. The total number of points caps at 25.

Final project: Demographics of exoplanets (Bai)

To date, more than 4000 exoplanets discovered. The sample is sufficiently large to allow us to study exoplanets as a population. In fact, there have been extensive studies on the demographics of exoplanets. One recent review paper can be found here:

`https://ui.adsabs.harvard.edu/abs/2015ARA%26A..53..409W/abstract`

The list of confirmed exoplanets is well documented in this archive:

`https://exoplanetarchive.ipac.caltech.edu/docs/data.html`

Note that the dataset is highly heterogeneous (quality and completeness of individual data vary depending on the detection method and sensitivity), and is of higher dimensions. This is an intrinsic difficulty for analyzing planet population data, and many studies select a more homogeneous subsample from the full data to avoid such complications.

One important subsample is from the California-Kepler Survey (CKS), where the accuracy of host star parameters (and hence the inferred planet radii) is substantially improved. This allows one to more more accurately discover features that are otherwise hidden in noisy data. More information about the CKS survey, and the associated publications can be found in their website: `https://california-planet-search.github.io/cks-website/` .

In this project, you are encouraged to explore either (or both) of the two datasets, and extract useful information/features out of them. Should you choose the first dataset, please be advised to make your own criteria for sample selection. Also, please note that different detection methods have different sensitivity limits, and the most fruitful methods such as radial velocity and transit methods becomes incapable of detecting planets with sufficiently small mass/radius and long orbital period.

This project is open-ended. You are free to explore the paper suggested here as well as other relevant literature, trying to reproduce and discuss some of the results, or analyze the data with your own idea. Given the diversity of possibilities, we suggest you choose one or two reasonably well-defined subjects (e.g., how do various planet populations depend on stellar metallicity?), and exploit the various methods you learned in this course to address the question(s). Grading will be mainly based on the following criteria:

- Choice of subjects (5 pts)
- Understanding of the statistical methods used (5+2 pts)
- Comprehensiveness of the exploration (5+2 pts)
- Depth of the exploration (5+2 pts)
- Creativity (5+2 pts)

Note that each item has 5 nominal points, whereas four of them can be awarded 2 bonus points in exceptional cases. The total number of points caps at 25.

Also, if you your research is already related to data analysis with exoplanets, I will have somewhat higher expectations on your work.

Final project: Using a particle-mesh code to calculate the gravitational acceleration field of a singular isothermal matter density distribution? (Xu)

- Using a Monte Carlo method to generate N_p particles that follow a singular isothermal sphere (SIS) density distribution $\rho(r) \propto r^{-2}$; compare the achieved radial density distribution with the analytical form. (3 pts)
- Assign the particles to a 3D mesh of N_m^3 dimension using: (1) Near-Grid-Point (NGP), (2) Cloud-in-Cell (CIC) and (3) Smooth Particle Hydrodynamics (SPH) algorithm; compare the density distribution in each case with the analytical form. (6 pts)
- Under an isolated boundary condition, using a convolution method achieved through FFT, calculate the gravitational potential $\frac{GM(\leqslant r)}{r}$ that each particle feels (a re-assigning process from mesh grids to particles is involved), compare the numerical results to the analytical solution. (6 pts)
- Using (1) Fourier derivation and (2) finite differencing method to calculate the gravitational accelaration fields that each particle feels; compare the numerical results in each case to the analytical solution. (6 pts)
- Change the number of Monte-Carlo particles N_p and mesh dimension N_m and re-run the exercises above, discuss the relevant performance, as well as comparisons among different methods. (4 pts)

Note the total number of points caps at 25. Your final score for this project will be given based on a comprehensively written report. Your source code shall be submitted along.

Final project: How do galaxy matter density profiles evolve throughout Cosmic time? (Xu)

Table 3 of Sonnenfeld et al. (2013, "The SL2S Galaxy-scale Lens Sample. IV. The Dependence of the Total Mass Density Profile of Early-type Galaxies on Redshift, Stellar Mass, and Size" published in *The Astrophysical Journal*, 777-98, `https://arxiv.org/pdf/1307.4759.pdf`) lists the observed lensing and dynamics properties of 25 early-type galaxies out to redshift $z \sim 0.7$. Use the first five columns in Table 3, i.e., galaxy ID, $z_{\rm d}$ (redshift), $R_{\rm eff}$ (effective radius), $R_{\rm Ein}$ (Einstein radius) and σ_{e2} (line-of-sight velocity dispersion measured within $R_{\rm eff}/2$), in combination with the source redshift $z_{\rm s}$ from Table 2 as our observational data, through a Bayesian approach find a model family of galaxy's matter density profiles that can reasonably well reproduce both the lensing and dynamics data. Can you further estimate how galaxy matter density profiles evolve throughout Cosmic time?

Hints: for example, the simplest matter density model $\rho(r) \propto r^{-\gamma}$ would be a spherical power-law model with a power-law slope of γ (as done in the paper). A best-fit γ is then sought for each galaxy according to its lensing and dynamics measurements. The Einstein radius $R_{\rm Ein}$ directly links to a projected mass $M_{\rm Ein}$ within $R_{\rm Ein}$, given the redshifts ($z_{\rm d}$ and $z_{\rm s}$, respectively) of the lens galaxy and of the source in the background (see references in the paper). The line-of-sight velocity dispersion σ_{e2} provides another dynamics estimate at a different radius from the galaxy centre. For a given matter density model, this can be obtained by solving the Jean equation under various assumption (e.g., a spherical symmetry, an isotropic velocity dispersion distribution etc, see references in the paper).

Of course, the model can be more sophisticated, e.g., a two-component model, one describing dark matter distribution and the other one describing stellar matter distribution. Of course, as the number of model parameters increases, more observational constraints are needed.

Now consider the simplest case where the total matter density profile is given by a simple power law. The strong lensing measurement, i.e., M_{Ein} essentially provides a normalization to this profile. Let us further assume that the surface brightness distribution of each galaxy well follows a de Vaucouleur's profile. Given this knowledge, you can work out the model expected line-of-sight velocity dispersion through the Jean equation under a number of reasonable assumptions. Now given observational data, how would you, using a Bayesian approach, infer the best power-law slope for each galaxy and further jointly infer the redshift evolution of the slope?

- A complete description of the matter density model and the assumptions involved in order to calculate the dynamics properties (through the Jeans equation) (5 pts)
- A reasonable model for the parameter prior distribution (5 pts)
- A reasonable model for the data likelihood function (5 pts)
- A reasonable parameter estimate procedure through a MCMC approach (5 pts)
- A reasonable inference of each galaxy's slope and overall redshift evolution (5 pts)

Note the total number of points caps at 25. Any novel methods reaching reasonable results can be awarded with up to 5 bonus points. Your final score for this project will be given based on a comprehensively written report. Your source code shall be submitted along.