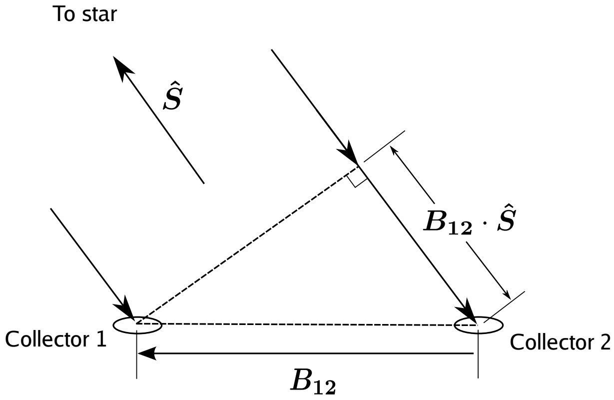
Project E: Surveying using stars



In this project you will do "data science" on a large dataset of optical path delay observations from the 6-telescope CHARA interferometer in California (see http://www.chara.gsu.edu for a general description). You will use this data to determine the 3-dimensional locations of the telescopes and try to infer something about their movement over long timescales (California is a seismically-active zone!).

The interferometer observes the interference between light arriving at different telescopes, separated by distances of up to 330 metres, in order to make images of stars at very high angular resolution. To generate interference fringes, the optical path delays from the star to the point of interference must be matched to a few microns. The optical path difference between two telescopes is given by $\hat{S} \cdot (\vec{x}_1 - \vec{x}_2) + d_1 - d_2$ where $\hat{S}$ is the unit vector pointing towards the star, $\vec{x}_1$ and $\vec{x}_2$ are the vector locations of the telescopes with respect to some origin and d_1 and d_2 are internal delays inside the interferometer associated with the optical paths from telescopes 1 and 2 respectively to the point of interference. Note that the vector between two telescopes is known as the "baseline" $\vec{B}_{12}$.

The "internal" delays consist of (a) a continuously-variable optical path length equaliser (OPLE) whose path delay is measured in real time by a laser interferometer to sub-micron accuracy (b) a set of discrete "Pipes of Pan" (POP) delays which can be switched in to access different parts of the sky (c) an quasi-fixed component dependent on optics in the beam path from each telescope to the beam combination point (known as the "constant term"). To the above can be added random errors due to atmospheric optical path delays above each telescope, which vary on timescales of less than a second but have a maximum amplitude of order 10–100 microns, and mechanical drifts in the POP and the "constant term", which can be at the level of hundreds of microns during a given night, but up to millimetres over the longer term.

The aim of this project is to determine the positions of the telescopes as precisely as possible from a sequence of delay measurements, made when the total delays have been equalised to such an extent that interference fringes have been observed.

The data for this project are available on Google Drive in the folder https://drive.google.com/drive/

folders/1Qgd51o0bbjOYv3ECVfdropuQRsMy7ReC?usp=sharing (links to the Google Drive folder can also be be found on the course Moodle site). You can access this data from Google Colab by following any of the options for accessing files given in https://colab.research.google.com/notebooks/io.ipynb. The delay data are in CSV (comma-separated-variable) files, contain tables of delay measurements and ancilliary data.

Each row in these tables represents the delay measured on a given star with a given pair of telescopes at a given time of night. The columns are:

utc : The date and time in UTC (Universal Coordinated Time, the successor to Greenwich Mean Time) when the observation was made.

star : Identifier for the star being observed. More detailed information about any given star can be found in the catalogue `starlist.zip` in the Google Drive folder.

elevation, azimuth : The elevation and azimuth of the star, in degrees, at the instant the delay was measured. Azimuth and elevation are angular coordinates defined relative to the local north and local horizontal plane. See https://en.wikipedia.org/wiki/Horizontal_coordinate_system for more information.

tel_1, tel_2 : Identifiers for the two telescopes making up the baseline being measured.

pop_1, pop_2 : Identifiers for the respective POP setups being used. Note that the P5B1 setup on the W1 telescope is unrelated to the P5B1 setup on the W2 telescope or any other telescope.

cart_1, cart_2 : delays measured in metres on the two OPLEs (one per telescope) at the same instant as fringes were observed.

You can start out by assuming that the delay introduced by any given POP setting is unknown but stable and that the “constant term” is stable on any given night but can vary by unknown amounts from night to night.

The tasks are as follows:

1. Analyse the data from the `2019_04_07.csv` file to yield estimates for the locations of all the telescopes on that night. I recommend using the Moore-Penrose pseudo-inverse (based on Singular Value Decomposition) to solve for the positions (as well as all the other model parameters).
   You should construct a “design matrix” for the problem, noting that it is normally a good idea to use the telescope positions rather than the baseline vectors as model parameters. This will allow you to take advantage of the additional constraints arising from the fact there are 15 possible baseline vectors but we need only 5 independent vectors to represent the relative telescope locations.
   Compare your results with the results presented by ten Brummelaar et al. (2005, doi:10.1086/430729) who give the following table:

TABLE 1
AVAILABLE BASELINES

Telescopes	East (m)	North (m)	Height (m)	Baseline (m)
S2-S1	−5.748	33.581	0.644	34.076
E2-E1	−54.970	−36.246	3.077	65.917
W2-W1	105.990	−16.979	11.272	107.932
W2-E2	−139.481	−70.372	3.241	156.262
W2-S2	−63.331	165.764	−0.190	177.450
W2-S1	−69.080	199.345	0.454	210.976
W2-E1	−194.451	−106.618	6.318	221.853
E2-S2	76.149	236.135	−3.432	248.134
W1-S2	−169.322	182.743	−11.462	249.392
W1-E2	−245.471	−53.393	−8.031	251.340
W1-S1	−175.071	216.324	−10.818	278.501
E2-S1	70.401	269.717	−2.788	278.767
E1-S2	131.120	272.382	−6.508	302.368
W1-E1	−300.442	−89.639	−4.954	313.568
E1-S1	125.371	305.963	−5.865	330.705

NOTES.—These numbers are based on the results of a global baseline solution from mid-2004. The rms error of this fitting process was 1891 μm.

2. Analyse all the available nights of data in 2019 to see if there is any evidence for the telescopes moving due to seasonal effects between April and November of that year.
   You will need to have estimates for the uncertainties of your telescope position solution for each month. There are two possible to estimate the solution errors. The first is to compare solutions on two different nights in the same month and use the variation between solutions to estimate the solution uncertainty.
   The second way to derive uncertainty estimates is to use Singular Value Decomposition (SVD) of the design matrix to determine the U, w and V^T matrices. From this we can determine the uncertainty on a model parameter θ_j by using the singular values w_i and the columns of the V matrix (rows of V^T) as follows:

$$\sigma^2(\theta_j) = \sum_i \left(\frac{V_{ji}}{w_i} \right)^2 \sigma_d^2$$

   where σ_d is the RMS error on the delay measurements, assumed to be the same for all measurements in a given fit. Remember to check for particularly small values of the singular values w_i, corresponding to degeneracies, and replace $1/w_i$ with zero in these cases. Small is usually defined as around $N\epsilon$ times the maximum singular value where N is the number of data points and ϵ is the machine precision.
   You can estimate σ_d from the residuals in your data, assuming that χ^2 of the best-fit model is approximately equal to the number of degrees of freedom.
3. Now analyse the data from the `2012_all_v2.csv` file, which contains data from all the nights in 2012 to see if there is evidence for motion of the telescopes between 2012 and 2019.
   Again it is important to derive uncertainties in order to quantify whether any differences are real or just due to experimental uncertainty.

It may help in quantifying uncertainties to divide the 2012 dataset into subsets (perhaps one or two months' worth of data per subset) and compare the solutions from different subsets.
In the 2012 dataset there are typically fewer measurements per night, which leads to the possibilities in degeneracies in the data, i.e. singularities in the design matrix, particularly if you choose subsets which contain only a few nights of data.
You can find any degeneracies in a given dataset by using SVD of the design matrix and looking for particularly small singular values. The corresponding column of the V matrix can be used to understand the linear combination of model parameters which are involved in the degeneracy (see "Numerical Recipes" for more information on how to interpret and deal with singular values which are close to zero). Use this to interpret the form of the degeneracies and whether/how they affect the model parameters we are interested in, namely the telescope locations.

4. If you have time, you could also look for evidence seasonal variations in the telescope positions during 2012.

Remember to check and potentially "clean up" your data: for example you should plot residuals to look for outliers. Make sure to present visualisations of the data which back up your conclusions.