CFHT/ESPaDOnS challenge: find the lines!

One of the main ways astronomers understand the objects they study is by looking at their spectra: what kind of light the objects emit, and at what intensities. We can learn a lot from dips and bumps in these spectra in particular. Can you write a program to identify these features in a set of real astrophysical spectra?

1. Introduction and Background Information

The **Canada-France-Hawaii Telescope (CFHT)** is an observatory that hosts a world-class, 3.6 meter optical and infrared telescope. The observatory is located atop Maunakea, a 4200 meter dormant volcano located on the island of Hawaii. The CFHT became operational in 1979. Its mission is to provide for its user community a versatile and state-of-the-art astronomical observing facility which is well matched to the scientific goals of that community and which fully utilizes the potential of the Maunakea site.

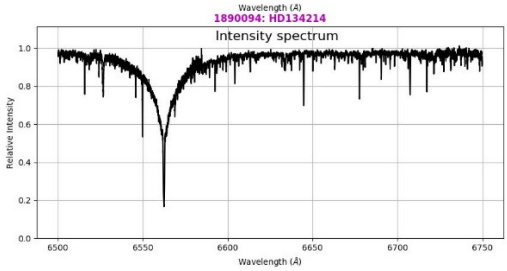
Among its suite of five instruments, CFHT has an **optical spectrograph and spectropolarimeter called ESPaDOnS** (Echelle SpectroPolarimetric Device for the Observation of Stars at CFHT). This instrument observes visible light, at similar wavelengths as those seen by our eyes, to record the spectra of stars. It can identify stars which are very hot and blue, or very cold and red, stars surrounded by planetary disks or nebulae, stars with weak or strong magnetic fields, stars with exoplanets, or stars with peculiar chemical compositions.

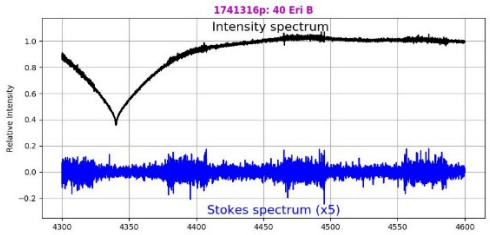
Each **spectrum** presents the intensity of the light as a function of wavelength. Depending on the physical processes present in the object observed, the spectrum will show **absorption lines** (a deficit of light at a very specific wavelength) or **emission lines** (an excess of light at a very specific wavelength). The location, width, depth, and shape of a line provide astronomers with the star's properties, such as chemical composition, temperature, presence of starspots, strength of its magnetic field, and even age, mass, or indication of the presence of an exoplanet in orbit around the star.

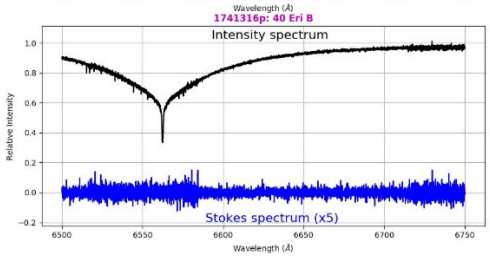
Raw images recorded on ESPaDOnS' detector often show absorption lines, which are the almost horizontal dark lines in the raw image below.

Raw images cannot be used to do analysis. They need to be reduced by a pipeline (code) that will provide astronomers with the numerical value of the intensity of light as a function of the wavelength. For ESPaDOnS, those spectra are plain ASCII files. One can plot the intensity of light as a function of wavelength, and see in a more quantitative way the absorption and emission lines.

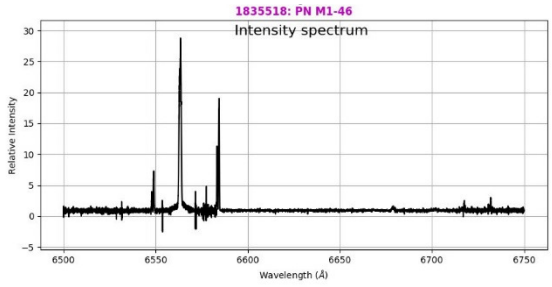
The following 3 sections of ESPaDOnS spectra (in black, for a restricted range of wavelengths, expressed in the plot in Angstroms Å[1]) show what absorption lines can look like. They can be narrow and shallow, deep and narrow, or deep and broad. Broad lines can even have different shapes.

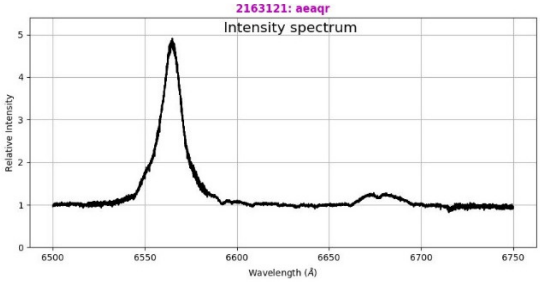






Emission lines can also be narrow or broad, as shown below:





Spectra can have a mix of absorption and emission lines. Often, one of an astronomer's first tasks is to identify these lines and find their central wavelength!

1 Where 10 Å = 1 nm = 10^{-9} m.

2. Goal of this Challenge

The goal of this coding challenge is to produce a file containing a list of absorption or emission lines found in a spectrum taken with the ESPaDOnS instrument.

The **input** to your code should be the name of the file (which is in an ascii format). The **output** should be a list that should have **(1)** the central wavelength of the emission or absorption line in nanometers (nm) and **(2)** the depth of the line.

By central wavelength, we mean the wavelength of the bottom of the line (for an absorption line) or the peak of the line (for an emission line), regardless of the shape of the line.

3. Input Files

The data to use as input consist of 2 ASCII files (`1835518in.s` and `2107396pn.s`). Each file has a complete spectrum, from ~400 nm to ~10,000 nm, for 2 different types of astronomical objects. Those spectra have absorption and/or emission lines.

Each file starts with a header of 2 lines. The first line has the name of the object. The second line has the number of lines in the file (around a quarter of a million lines!), then the number of columns, excluding the first column. Here are 2 examples:

```
***Reduced spectrum of 'FS CMa  '
 213722 2
  369.3997 -3.0389e-01  3.1139e-01
  369.4020  4.3935e-01  3.6575e-01
  369.4044  1.4792e-01  3.2914e-01
  369.4067  8.2091e-03  3.1969e-01

***Reduced spectrum of 'gam CrA A'
 214823 5
  366.9261  3.1077e+00 -3.7206e-02  2.4856e-02  3.5294e-03  1.8397e+00
  366.9286 -7.6380e+00  9.1527e-02 -6.0928e-02 -8.6016e-03  2.0184e+00
  366.9310  1.2284e+00 -1.4734e-02  9.7726e-03  1.3715e-03  1.9371e+00
  366.9335 -1.6343e+00  1.9620e-02 -1.2967e-02 -1.8091e-03  1.9903e+00
  366.9359  6.0785e-01 -7.3041e-03  4.8100e-03  6.6714e-04  2.0776e+00
  366.9384 -4.7063e-01  5.6605e-03 -3.7141e-03 -5.1201e-04  2.0594e+00
```

In the first column of each file, you have the wavelength in nanometers. The interval between each wavelength is not constant throughout the file. The wavelengths are not strictly listed in increasing order.

The second column has the intensity of the light. Some of the columns have zero, and that's normal.

If there are other columns, they are to be ignored for this challenge. Only the first 2 columns are useful here.

The spectra are normalized: the continuum is close to the value 1. (A spectrum without any absorption or emission lines would be a straight line centered around 1, with some random noise added.) Features that are roughly below 1 are absorption lines. Features that are roughly above 1 are emission lines.

Some parts of the spectra are very noisy, especially at blue (short) wavelengths. Therefore, some features that might look like small absorption or emission lines are just noise and not real.

4. Outputs

Your program should either find features that have a predetermined (hard-coded) depth (e.g. all absorption lines deeper than 0.9 times the continuum, or emission lines that are brighter than 1.1 times the continuum) or should ask the user for thresholds (either as a command line argument or interactively).

The program should either find all features for the whole spectrum or find features in a user-specified range of wavelengths.

The output should be a plain ASCII file with one feature per line. Each line (feature) should indicate the central wavelength and depth of the feature.

The program will preferably be written in Python. The least preferred language is IDL.

The code should have enough comments so that a high school student who wants to understand how the program works or modify it can do that by reading the comments.

Bonus: plot the spectrum provided as input and indicate (from the content of the output file) where lines (features) were found.

5. Submission

Please submit your work by email to manset@cfht.hawaii.edu (Nadine Manset).

Include the following:

- The file (or files) for your code,
- At least one example of an output file (ascii format) with results (i.e. the list of lines with their central wavelength and depth).
- A README file that explains how to run the code (this will depend on the programming language you used, does it need to be compiled?) and how to provide inputs.

Optional: you may include screenshots of what you see on your screen when you run your code (if there is a plot for example).

Good luck!
