PHYS 391 – Lab 3: Astronomical Image Data Processing

Goals

This lab is designed to demonstrate basic astronomy data analysis including how stellar population information is extracted from images. With image analysis, we measure flux and determine the strength of signal through a concept called signal to noise (SNR). Using SNR we can determine the limits of an extended object, such as a galaxy or distinguish between random noise and a faint star.

After preliminary tasks that acquaint you with the software, you will measure the diameter of a galaxy and then calculate the distance and age of an open cluster of stars using image data from Sloan Digital Sky Survey and Pine Mountain Observatory, respectively. If you are new to programming, you will be introduced to some basic python scripts as well as working in a linux environment.

WARNING!!: PyRAF is potentially buggy. Write down measurements or save files/scripts in case you have to restart pyRAF.

I. Getting Familiar with terminal, data and software environments. Part I is an introduction to the software and data formats. There isn't any work to report in your lab notebook for part I, but I suggest making notes where necessary for your own benefit.

Everything in *italics* is something you type at the prompt and anything following it in **bold face** is something you put in based on your filenames, directory name, etc. ...

Setup and startup

Click on the virtual machine icon and in the terminal window do the following:

- In the lab391 directory (you start in this directory), make your own directory using your email name or whatever as long as it's not confused for someone else. Use the command 'mkdir'

  e.g. *mkdir* **elsa**

- Copy the files from the directory ~/labfiles/ to your directory using command 'cp' using the exact syntax except using your directory name:

  e.g. *cp ~/labfiles/** **~/elsa/**

- Type *ds9 &* at the prompt. This will bring up a window that will help you view the astronomical images.

- In the main directory, (not your directory) type *pyraf*.

- Once you are in the pyraf environment, change to your directory by using command 'cd'

  e.g. *cd* **~/elsa/**

Gedit

Before we go any further, bring up another terminal window by clicking on the terminal icon. Change to your newly made directory and type:

ls

These are all the files you will need for the lab. To view and edit myscript.py, type at the prompt:

gedit myscript.py

This should bring up a text editor window with the file myscript.py. If you want to create a new file, just type gedit at the prompt.

Okay now go back to your pyRAF window…

FITS files

FITS files are binary image files and thus can't be viewed with a text editor like Gedit or Word. All astronomical images and spectra are stored in this format. They consist of two parts: (1) the 'header' gives information about the observation and the image and (2) the image itself. Type

imheader m101u10.fits

(this is a UV filtered 10 minute exposure). Scroll up a bit and you can see all the information that it dumps out such as the date of the image and exposure time. Information varies depending on the observatory. This information would include variables that go into processing the data such as sky magnitude, the sky coordinates (ra and dec) etc. Now lets check out the image.

DS9

At the prompt, type: *display m101u10.fits*

After you press return, press return again (you can display up to 16 images at once in 'frames' on ds9 by typing in a frame number from 1-16). Bring up another image (select any fits file from your directory) and try different ways of displaying the frames. Click on an image, you should get a little green circle appearing on the image. Do the following to get familiar with the functionality of ds9. All of this is done through the ds9 window (no typing needed):

- Change the brightness and contrast (hint: try holding the mouse button down and dragging the cursor)
- change the scale
- zoom in and out
- click on a circle and change the radius, color, shape
- change the color of the image - for my own research, I find the rainbow color map most effective or inverted grayscale
- display both frames at once, blink them etc..

Note: to stop generating circles go under edit menu and select 'none', otherwise leave on 'pointer'.

As you move the cursor over the image you see numbers changing for 'X', 'Y' and 'Value'. X and Y - are simply pixel coordinates and 'Value' corresponds to the flux of photons in each

pixel. Scroll over to an empty patch of sky - notice it isn't zero. The sky has photons and this must be subtracted from the starlight when we measure the flux of stars.

Imexamine

The main software task that you will use in this lab is called imexamine. It does many, many things as well as measure stellar fluxes. At the pyRAF prompt type:

imexamine

When you press return, it won't return you to the command prompt. Imexamine is an interactive tool to use with ds9. Notice that the cursor is round and blinking when hovered over the image. Try the following imexamine functions 'r' and 's' by simply typing *r* or *s* when hovering the blinking cursor over a star. You should see a plot pop up. The function 'r' plots the radial profile of the star. It automatically finds the center of the star and fits the profile to a Gaussian. You will use this function later. The function 's' produces a surface plot where the z axis corresponds to counts. There are other functions you will use in imexamine that will be explained later in the lab. To quit imexamine, type *q* while the cursor is in the image.

pyRAF Math

The pyRAF environment is like other software packages (e.g. R, Matlab) such that mathematical expressions are calculated right on the command line. If you haven't done so, quit the task imexamine now (see previous paragraph). You should get the --> prompt back. If not press return again.

To do math calculations in pyRAF type:

import math

Now try these calculations by typing in the following:

```
x=100
sqrt(x)
log10(x)
x**2
```

If you wish to use pi (fyi you will), you need to use math.pi
e.g.

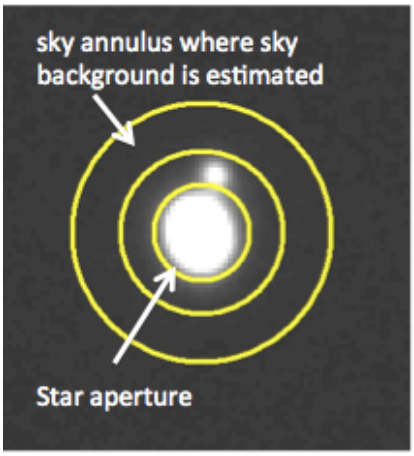
```
r = 4.5
2*r*math.pi
```

(Ans: 28.274333882308138)

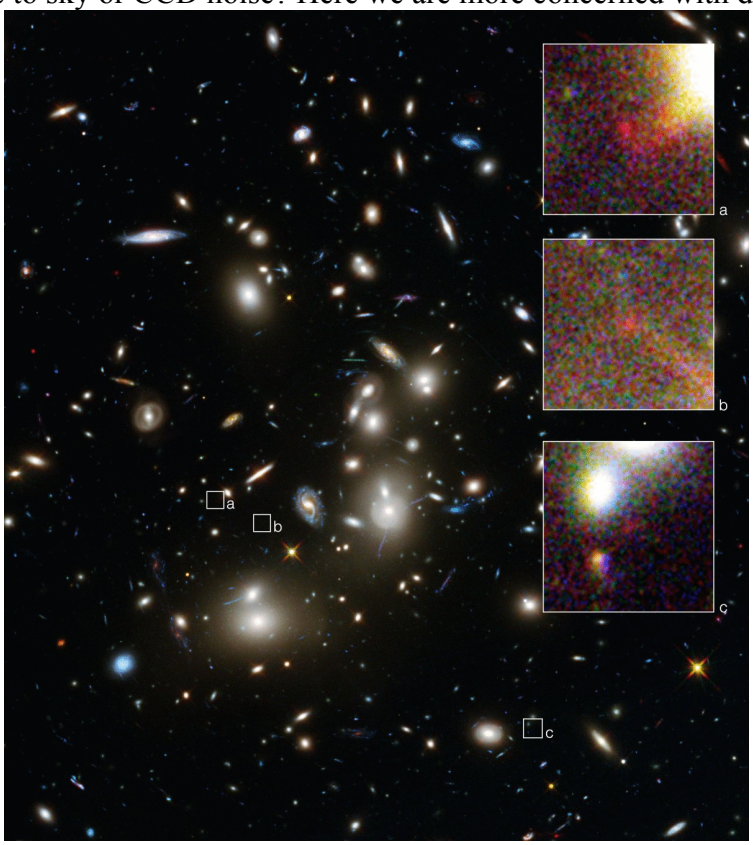
II. Photometry and Signal to Noise

Introduction:

Photometry is the term for measuring flux from any astrophysical object. The algorithm for stellar photometry goes as follows: An 'aperture' of a determined radius (or any shape for non spherically symmetric objects) is centered about a star and all of the photons (including the sky background and other sources of noise) are summed up in this aperture. The sky background is modeled and averaged based on a small piece of sky near the star (usually an annulus of about 3-10 pixels in width centered around the star and outside the aperture. The background is then subtracted leaving the flux. Below is a picture of this process.



In the image above, we have no doubt that there is a star in the middle of the aperture. In fact we would be just concerned about measuring the flux. But what about distant faint galaxies detected in the image below? How does one know that there is a galaxy and not a random fluctuation in the background due to sky or CCD noise? Here we are more concerned with detection than flux.



In astronomy we measure a detection of an astrophysical object in terms of signal to noise ratio (SNR). Noise takes many forms (light pollution, sky background, detector electronics, cosmic rays, bad pixels, etc.). SNR is a way to quantify the strength of a signal above random fluctuations in the noise rather than qualitatively determining whether an object is visible in an image. In fact our eyes aren't sensitive enough to see faint significant signals on an image or in some cases detect bright signals against even brighter backgrounds.

SNR comes from counting statistics, which basically means that standard deviation or noise goes as the square root of the counts. SNR is the signal count divided by the standard deviation:

N = number of counts
$(N)^{1/2}$ = standard deviation
SNR ~ $N/(N)^{1/2}$
~ $(N)^{1/2}$

In photometry, SNR is the number of photons from the signal alone (i.e. star) divided by the standard deviation of the photons from the signal plus the background noise. For instance if you collect 10,000 photons in an aperture where 100 are from the star and 9900 are from noise, the SNR is:

SNR = $100/(100+9900)^{1/2}$
SNR = 1

Conceptually, this means that the amplitude of your signal is as big as a random fluctuation of noise – the signal is indistinguishable from the noise. An SNR ~ 3 is the bare minimum for detecting object. A bright star will have an SNR of at least 100 or have a signal that is 100 times greater than a random fluctuation of noise. For this lab you will only be concerned with sky noise.

Astronomy Terminology:

- Flux (**F**) is the amount of energy/sec/area from a star. More specifically, it is the number of photons from a star or galaxy that intersects a detector in an interval of time.
- Apparent brightness is also flux, but measured in a $\log_{10}$ scale based on how our eyes perceive different levels of brightness. It is denoted by **m** and measured in magnitudes.
- Absolute brightness is similar to apparent brightness, but it's the brightness measure of any object at a distance of 10 parsecs (32.6 ly or $3x10^{19}$ cm) away. It is denoted by **M** and measured in magnitudes.
- Background (**S**) refers to the sky photons contaminating the star signal. It has the same units as flux.
- Gain is a scale factor that compares the counts registered on the CCD to actual photoelectrons. It is always a value greater than or equal to 1. For instance, if it takes 10 photoelectrons to register 1 count on a CCD, that CCD has a gain of 10. You measure counts in the image and will convert them to photons (actually photo-electrons). For this lab, we will deal only with counts for measuring flux and photo-electrons for calculating SNR.
- Zero point (**zp**) refers to a magnitude calibration that is specific to the observing conditions, filters and telescope. Strictly speaking, it represents the magnitude of an object with a flux of one count per second, but it also serves as a term that encompasses all calibrations (e.g. gain, filter system) and wraps them up into one neat term. It is estimated by comparing measured magnitudes to catalogued values of known stars. You will derive it for your data set when you analyze M39.

These quantities are related:

i. $m = -2.5\log_{10}(F) + zp$
ii. $SNR = Ft\sqrt{Gain/(F+S)t}$

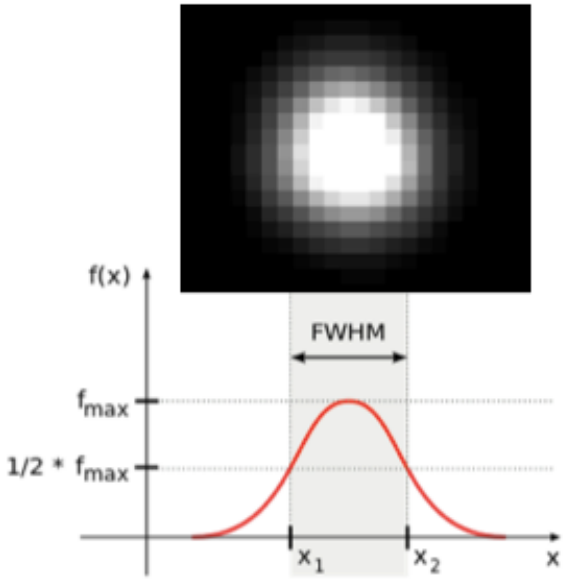
iii. $m_{error} = 1.0875/SNR$
iv. $F = L/d^2$
v. $m - M = 5log_{10}(d) - 5$
vi. $m_1 - m_2 = -2.5log_{10}(F_1/F_2)$

where d is the distance to the star and t is exposure time.

Instructions: Part A & B will acquaint you will detection and signal to noise. Part C will demonstrate astronomy research techniques. It's expected that you write all your work, data, equations and answers to questions down in your lab book. Keep a text file going of python math and plotting scripts used in pyRAF along with the output (unless it's a plot). You will email this file of compiled scripts to me when you turn in your lab. For your convenience, I've denoted any instructions that use python scripts with a symbol ★. You will use imexamine to measure the flux for various objects.

Note: Imexamine does all of the basic photometry tasks automatically including:

- Finding center of the star based on the stellar profile
- Deriving an aperture radius based on the full width half maximum (FWHM) of the stellar profile (radius = 3xFWHM, see figure below)
- Calculates and subtracts the sky background.



Light from a star is fit to a Gaussian profile and FWHM is estimated. f is photon counts, the x axis a spatial dimension

A. Measure SNR of Faint and Bright Stars in single and combined images of different exposure times.

Depending on your science goals, it is better to take several shorter images and combine them in various ways rather than take one long exposure. In this quick exercise you will compare the SNR and other attributes of four images of the same objects

1. Display these four UV images in four different frames and adjust the brightness/ contrast as needed.
   - m101u10.fits (1-10 minute exposure)
   - m101u15.fits (1-15 minute exposure)
   - m101ucombmed.fits (median of 4 – 15 minute exposures -15 total minutes)
   - m101ucombadd.fits (4 – 15 minute exposures added together – 60 total minutes)

The median exposure takes the median value of each pixel from all four images where as the added exposures is simply the addition of values in each pixel leading to a larger count per pixel.

2. Type *imexamine* at pyraf prompt. Pick out a star –find one isn't saturated which appears as spikes coming from the center. Hover the cursor over a star and type *a*

You will get the following two-line output (highlighted values are the important ones):
COL (pixel column) LINE (pixel row) COORDINATES (a repeat of COL and LINE)
R (aperture radius derived from 3*FWHM) **MAG** (apparent magnitude) **FLUX** (total number of counts) **SKY** (sky value per pixel). The middle 3 terms are not important and the last 3 are derived values of the FWHM of the star. The apparent magnitude is calculated via equation (i) with zeropoint = 0.

SIDENOTE The term FLUX here is misleading. Imexamine assumes that the pixels are in units of counts/second. These are actually just counts. To convert to proper flux units divide the FLUX value by exposure time (in seconds) or add factor: 2.5*log10(time) to the MAG value.

3. Record R, MAG, FLUX, SKY and exposure time in your lab book.
4. Do this entire process again for the faintest star-like object that you can find on the image.
5. ★ Recalculate the flux and magnitude as per the note above. Do this in pyRAF (you will have to quit imexamine). Recall from part I that you need to type import math in order to do math calculations.
6. ★ Calculate the SNR. When combining images, the SNR changes slightly. Use this equation for added and median images where n=4, the number of images:

$$SNR = Ft\sqrt{Gain/(F/n + S)t}$$

Answer the following questions:
Q1. By what factor is the SNR greater for the 15 min exposure to the 10 min exposure. Based on the formula, is this what you expect? What about the added exposure vs the 15 min exposure? Did the SNR improve according to what you expect?

Q2. Compare the median SNR to the 15 minute exposure SNR. What advantage does median have over the 15 minute exposure? Over the added exposure?

Q3. If there is a neighboring star in this annulus, how will that effect the flux estimated from the star?

B. Determining the size of a Galaxy

Signal to noise ratio is also convenient for finding the observable limits of a galaxy. From the previous exercise, it should be apparent that finding a SNR~3 star visually is difficult, this is also true with finding the extent of galaxies, that is our eyes can't detect the outer regions of galaxies like CCDs. Use the fact that a SNR =3 is the limit for detection and find the radius of a galaxy. You will be using Sloan Digital Sky Survey (SDSS) data, which comes in a different filter system.

1. View the header of file ngc4636g.fit. As you can see there is more information on SDSS images than Pine Mountain images. SDSS adds a DC offset to each pixel. This is called SOFTBIAS; find this value in the header, you will need it for calculations
2. Display the ngc4636g.fit file and adjust brightness if necessary.
3. Use the key 'm' in imexamine to find the average value of a 10x10 box at various points in the image. Try using it at any point to see how it works. It will bring up the following on the pyRAF command line:

SECTION (Range of x and y coordinates defining the box used for statistics) NPIX (number of pixels in box used for statistics), MEAN, MEDIAN, STANDARD DEVIATION, MIN AND MAX.

1. You will need MEAN and SECTION. MEAN gives the average count per pixel. First find the mean sky background of an area far from the galaxy light and stars. This is the nominal sky background value. Do this a few times around the image to be sure you found a low sky brightness area. Once you have, you will use the mean value to as your sky background value for SNR ratio.
2. Starting from the center of the galaxy, work your way out measuring the mean value of the region. Measure until the value is close if not the same as your sky background value. Note, you can find the galaxy center by using the 'r' function on imexamine. Get the cursor close to the center and press r. Use the coordinate given on the graph title where it says 'Radial profile at …'.
3. ★ Use equation (ii) to find the SNR. Gain = 4 for SDSS data. Pixel value units are in counts, not in counts/sec so you effectively have F*t and S*t for SNR calculation. Be sure to subtract the softbias value from your mean values, multiply each mean value by NPIX, and subtract the sky background from your measurements to get flux. For example:

   skybackground measurement = 402
   softbias = 100
   galaxy measurement = 410
   npix = 25

   Then:
   skybackground = (402-100)*25 = 7550
   flux = (410-100-302)*25 = 200

4. ★ Once you find the approximate location of the galaxy edge, find this distance in pixel units and convert to arcminutes ('), the proper unit for apparent radius. There are 0.4 arcseconds (") per pixel. You should do this exercise in at least two different directions away from the galaxy.

Q1 : Is this measurement the true extent of the galaxy? Why or why not?

C. Calculating the distance and age of M39

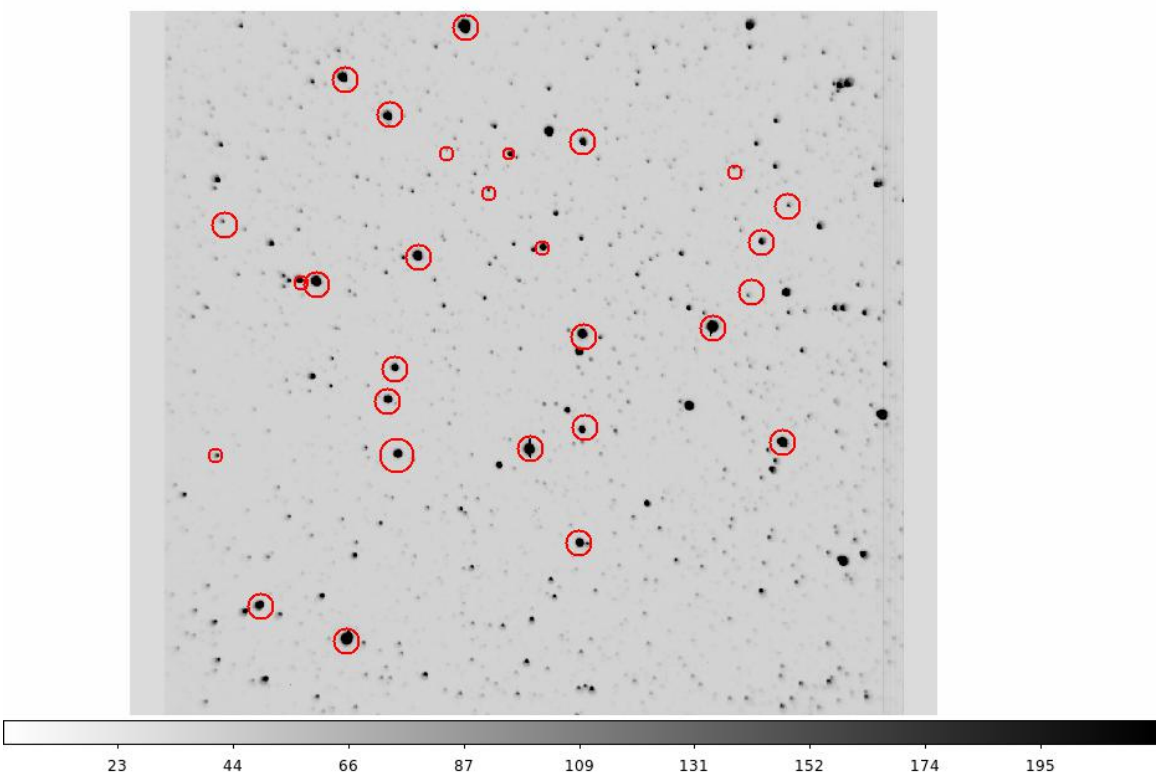
You will create a Hertzsprung-Russell diagram of an open cluster of stars by measuring B and V fluxes with imexamine. From the diagram you will determine the distance and age of this open cluster of stars. I suggest that you refer to the lecture notes if you don't understand the astronomy portion. It won't be necessary but it will enhance your lab experience.

Files Used:

- m39b10.fits
- m39v10.fits
- myscript.py
- simpleplots.py

1. Display the two M39 files in your directory (you can close the frames from part A)
   m39b10.fits (10 second blue exposure)
   m39v10.fits (10 second visual exposure – effectively green)

2. Using the image below, measure the flux of at least 26 circled cluster stars in both the B and V images as you did in part A recording R, MAG, FLUX and SKY. Don't worry about dividing the flux by the exposure time and re-calculating the magnitude. It is absorbed into the zero point, which you will later calculate. Use the printed image to label (e.g. 1, 2,…) each star so that you associate fluxes in B and V with the same star. When you are done with the lab, cut and paste this image with your labels in your lab book.



3. Enter your values in your lab book. Here is a suggestion:

Star #	R (V)	MAG (V)	FLUX (V)	SKY (V)	R (B)	MAG (B)	FLUX (B)	SKY (B)

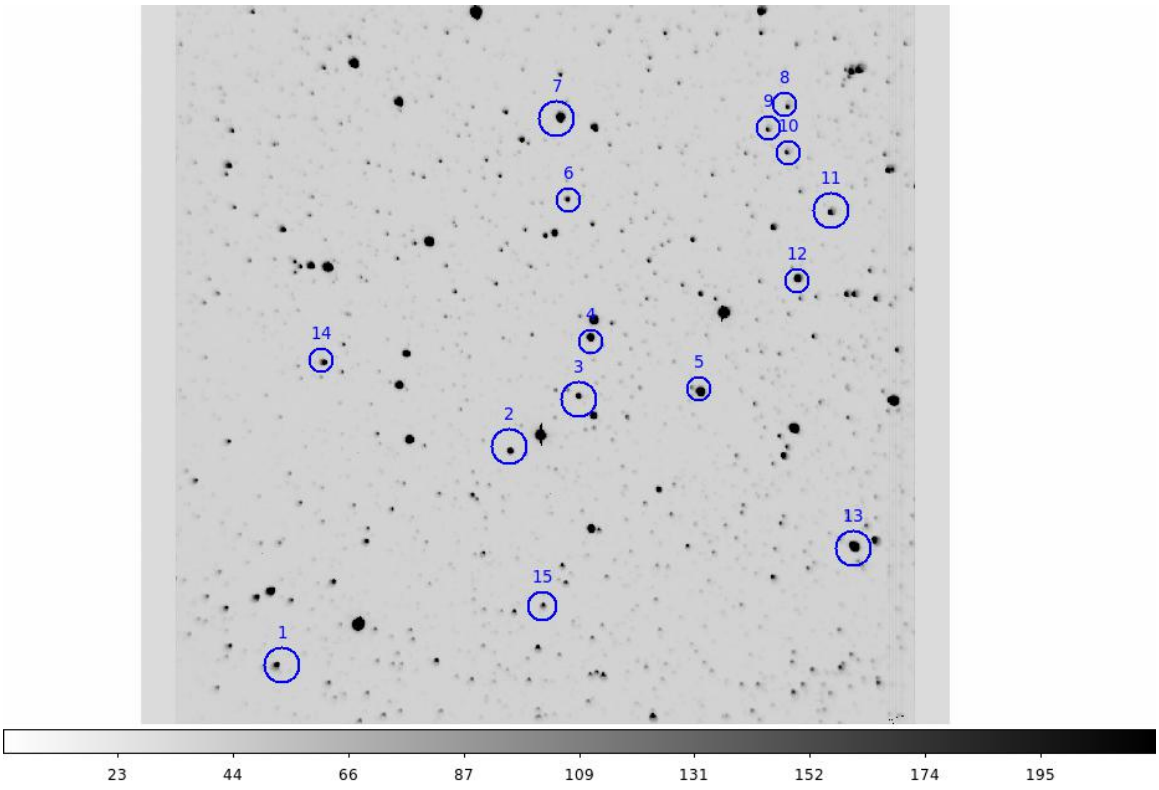
4. ★Calculate the magnitude error for each B and V.
5. Calculate zero point: In order to make sense of these fluxes, they need to be calibrated to a "standard".

   a. Using the image below of standard stars with (catalogued) magnitudes, measure the flux and magnitude of each (R, MAG, FLUX and SKY) circled star as you've done before.

   b. ★ Compare these measured magnitudes to the standard values using equation (i) and standard.txt (contains all the catalogue star values) to find the zero point. Do this for both B and V images and write down in lab book. See myscript.py for calculations and arranging lists of data. (NOTE: If you feel confident about your python skills, write a little program and execute it in a terminal window by typing
   ./your_python_script.py
   at the prompt. Note the period and the backslash before the file name. I strongly suggest copying file myscript.py and altering it with your own expressions)

   c. ★Average these values to get the B and V zero points. Calculate the standard deviation for each zero point, this will be the error (don't worry about individual measurement errors).

   d. ★Apply these zero points to your cluster star data.



HR diagram: Now you are ready to create your HR diagram.

6. ★ Calculate B-V and errors (these add in quadrature)
7. ★ Plot V as a function of B-V with error bars. Use the scripts in simpleplot.py for help on plotting. Be sure that V goes in order of increasing brightness and B-V goes from smaller to bigger numbers. See lecture notes. Your plot should have data points from the bottom right corner to the top left corner of the plot along the main sequence.
8. Once you have labeled the axes save it to a file - you will need to print it out and paste it in your notebook (again ftp it or attach it to an email and send it to yourself)

Understanding the diagram:
Q1. Where are the most luminous, faintest, reddest, bluest stars located on a HR diagram?
Q2. Near the top of the main sequence, you will see the "turn-off" point. Physically, these stars are running out of hydrogen fuel in their cores and are now beginning the process of evolving off of the main sequence. Do they emit more red or blue light than stars of equivalent V magnitude on the main sequence?

Distance to Cluster
The V magnitudes (and B magnitudes) you just measured are fluxes measured at Earth. We know what their relative brightnesses to one another are but don't know how they compare to other stars with known luminosity such as the Sun, Procyon, Sirius, etc. Therefore we convert this apparent magnitude (ie flux measured at Earth) to absolute magnitude (the flux of any astrophysical object standing 10pc away).

★Using the fact that the sun has a B-V =+0.6 and absolute visual magnitude of Mv = +4.8 (just to reiterate, this is the magnitude the sun would be 10pc away from us), estimate the V band apparent magnitude for a star with a B-V=0.6 in the cluster. Using the equations given in the introduction of part II, find the distance to the cluster. Find the min and max distance based your errors.
→Google the actual value. How much bigger or smaller is your derived value (in percentage)?

Cluster Age:
★The age of the cluster is determined by the life expectancy of the most massive main sequence star. Massive stars burn through their fuel faster than smaller main sequence stars because they require more internal pressure generated in their core to support greater overlying mass. As a result, the most massive star in a cluster is the first to evolve off the main sequence or to "die" because the lifetime of a star is defined by its time fusing H into He in its core (main sequence). For instance, a 2 solar mass star has a lifetime of about 1.5 billion years, if the most massive star in a cluster was 2 solar masses, then the cluster is at most 1.5 billion years old.
Using your diagram and newly determined distance (don't use the published one), figure out the absolute magnitude of the most massive main sequence star. Find the corresponding luminosity L (in solar units) using the provided equations (hint, compare the apparent magnitude of a sun-like star in this cluster to the brightest cluster star to find the relative luminosity in solar units). Using the following equations, derive the age of the cluster:

$\mathcal{M}$ here is mass and in solar mass units.

$L \sim \mathcal{M}^{3.5}$

Age ~ $\mathcal{M}/L$ → Note: $1\mathcal{M}_{sun}/L_{sun}$ = 10 billion years.

Q3. Research the quantum efficiency of a CCD. Why is the zero point different for the B and V image? Each image has the same exposure time.

Q4. Conceptually, describe the gain of a CCD. You can use the values for PMO and SDSS to explain your answer. For instance, what does it mean that PMO has a gain of 35? Which is a better CCD, one with a gain of 4 or 35 and why?