

Astrophysics with the Computer: Interpretation of Emission Line Profiles: Radial Velocity and Internal Kinematics

Joachim Köppen Strasbourg 2001/02

1 Astrophysics

Why does one do observations? To find out about an object. This may be either something new, previously unknown, or we wish to determine some physical quantity. In either case, we compare the observations with our present concepts or models for the object. If the observations contradict our expectations, we may have discovered something new; if they agree with the predictions, we are happy that our model is confirmed and we may be able to derive from the observational data some physical property of the object. This process is the base of science, and is present in some form whenever we look at the world around us.

How does one decide whether observations agree with a model? How do we derive a parameter of the model? How certain are the values we derive? Is it necessary to introduce another parameter? All these questions we have to deal with in the interpretation of (astrophysical) data. In this exercise we shall go through these problems. We shall take as a simple example the determination of the radial velocity from the spectrum of an emission line nebula, and we shall use a general statistical approach which can easily be formulated and adapted to other problems.

You may remember the χ^2 -method to find the best fit of data to a model, that one calculates from the difference between observed and predicted data a quantity and one searches for that value of the model parameter whose χ^2 is minimal. Then one can also give a confidence interval for the parameter, and the absolute value of χ^2 gives a measure of how acceptable this best-fit model is. However, all this is only correct if the errors are distributed like a Gaussian, and the error amplitude σ is known. In this exercise, we shall use a more general approach.

The aim of the project is to write a program to make this comparison of data and model, to create artificial data to test the program, and to apply the program to some real spectra of a planetary nebula.

2 The Method: Bayesian and Maximum Likelihood Fits

The principle of the method is this: Suppose we have some observational data D_i given at a number of (wavelength) points, and we wish to explain these by a model which depends

on some parameter p . This means we wish to determine that value of p for which the model gives a prediction for the observations that is closest to the data.

To compare the predictions of the model – call them $M_i(p)$ – with the data we also need to know or assume the source of errors that could lead to deviations between model and observation. As we shall assume that we have corrected the data for all systematic errors, we are left with random errors, such as noise in the detector etc. Let us denote by $\psi(x)$ the distribution function for the probability of an error x . For example, ψ would be a Gaussian function in the case of white noise in the data (see Section 5). Then, the likelihood that at point i the deviation between observation and the model with parameter p is $\psi(O_i - M_i(p))$. If the noise amplitudes in the various points are not correlated with each other, the likelihood that the deviation in point 1 is $O_1 - M_1(p)$ **and** in point 2 is $O_2 - M_2(p)$... etc ... is the product

$$\Lambda(p) = \prod_{i=1}^n \psi(O_i - M_i(p)) \quad (1)$$

This is the likelihood for the model with parameter p . It is a measure for the ‘distance’ between model and observation, and also for the probability that the model parameter is p .

To find the ”best” value for p , one calculates the likelihood as a function of p and searches for its maximum. This is the principle of a Maximum Likelihood fit of the data.

Such a fit makes use only of the position of the peak. But one can go further: Bayes’ Theorem from probability theory tells us that the probability for the parameter p **after** taking into account the (new) observation is computed from the Likelihood and the probability distribution for p **before** this observation:

$$P(p|O) = (P(p) \cdot \Lambda(O|p)) / \int P(p)\Lambda(O|p)dp \quad (2)$$

Here $P(p)$ is the prior distribution for p , i.e. what had been known before we use this present observation. If we want to suppose that we knew nothing about p , we set $P(p) = const$. The good thing about this formula is that it tells us how to combine our previous knowledge about p with the information from the new information, and also from different observations.

Note that the prior distribution is a normalized probability distribution, so that the integral over the parameter space

$$\int_{p_a}^{p_b} P(p)dp = 1 \quad (3)$$

We also note that the integral in the denominator of Eqn. 2 is nothing but a normalization over the whole parameter space.

The merit of the posterior distribution $P(p|O)$ is that it tells us which values for p are more probable than others, and by how much. In our simple exercise, the posterior distribution will be a simple Gaussian-shape function, symmetrical about its single peak;

but in other problems this is no longer so simple! Of particular value is the computation of the median value, confidence intervals and error bars. The median value for p is found by the condition

$$P(p_{50}|O) = 0.5 \quad (4)$$

In practice, one computes $P(p_k|O)$ for a number of evenly-spaced p_k , then searches for the point closest to $P = 0.5$, and then makes a linear interpolation in p to find the exact value for p_{50} . Analogously, the central 90 percent confidence region is found by the two points p_5 and p_{95} .

It is worth pointing out that the form of the error distribution function, and thus the amplitude of the errors or noise in the data, enters into the likelihood, and may well be taken as additional parameters! They would not be parameters of the model proper, but of the instrument. So we also model the instrument.

The formulation in terms of probabilities makes it straightforward to formulate the demand that a model should fit both observations A **AND** B:

$$P(p|A\&B) = P(p|A) \cdot P(p|B) \quad (5)$$

$$\propto P(p) \cdot \Lambda(A|p) \cdot \Lambda(B|p) \quad (6)$$

The prior and the two likelihoods, representing our initial knowledge and the information gained by the two observations, are entering equally.

Likewise, if we do not care whether the parameter has one value or the other (for $p_1 \neq p_2$), we have

$$P(p_1 \text{ or } p_2|O) = P(p_1|O) + P(p_2|O) \quad (7)$$

$$\propto P(p_1)\Lambda(O|p_1) + P(p_2)\Lambda(O|p_2) \quad (8)$$

Or, if we only care whether the parameter is within a certain interval $a < p < b$:

$$P(a < p < b|O) = \int_a^b P(p|O)dp \quad (9)$$

$$\propto \int_a^b P(p)\Lambda(O|p)dp \quad (10)$$

Note here a small inconsistency in the notation: $P(p|O)$ is a probability **density** i.e. per unit p interval, so $P(p|O)dp$ is a genuine probability, such as $P(a < p < b|O)$! Usually, one does not make a difference in the notation, because it could get quite clumsy. One may keep in mind that the density $P(p)$ is really $dP(p)/dp$.

3 The Model

To derive the radial velocity from emission lines we compare the observed line profile with the line profile "predicted" by a simple model. The lines are optically thin, so we do

not have to take into account any modifications of the line profile by radiative transfer. Furthermore, we shall assume that at any line-of-sight we look at a single thin shell of gas. We also assume that the motion of the emitting gas ions is dominated by random thermal motions, perhaps some small-scale turbulence. Under these circumstances the line profile will be a simple Gaussian, with the width being equal to the thermal Doppler width

$$\Delta\lambda_D = \frac{\lambda_0}{c} \sqrt{\frac{kT}{m_{\text{ion}}} + \xi^2} \quad (11)$$

where ξ is the velocity amplitude of microturbulent motions. We shall deal with a planetary nebula or H II region, and we use the [N II] lines and the H α line. The electron temperature $T = 10\,000\text{K}$ is a very good guess for such an ionized nebula.

Because our observations were done with a long slit across the nebula's face, we have some spatial resolution, and by using the spectra extracted from single pixel rows of the CCD image, we can measure the radial velocity of different parts of the nebula. Because these single-pixel spectra show only one peak of the line profile, we probably see only a single shell of plasma in each location. This is nice, as it makes life more easy. But in many other nebulae one does observe a doubly peaked emission line profile!

So, if there are no turbulent motions, the line width is fixed by the atomic mass and the temperature. Thus, our unknown parameter is the radial velocity which causes a shift of the wavelength of the line's centre. The model line profile would be:

$$\phi_{\text{mod}}(v_{\text{rad}}) = \frac{1}{\sqrt{2\pi}\Delta\lambda_D} \exp\left(-\frac{1}{2}\left(\frac{\lambda - \lambda_0 - v_{\text{rad}}\lambda_0/c}{\Delta\lambda_D}\right)^2\right) \quad (12)$$

But this means that the model predicts a certain **line shape and width**. So we better check whether this agrees with the observations! We take the line width as another model parameter and now find the peak likelihood in two dimensions.

The **shape** of the model line profile may not be the best one. We may try a Lorentz profile:

$$\phi_{\text{mod}}(v_{\text{rad}}) = \frac{\text{const.}}{\Delta\lambda_L^2 + (\lambda - \lambda_0 - v_{\text{rad}}\lambda_0/c)^2} \quad (13)$$

What do you get when you analyze data created with say a Gaussian shape with say a Lorentz profile? What line width do you get? And what do you get for the maximum of the likelihood? Since this is not the same line shape, one should expect that the fit becomes worse, and the maximum likelihood is smaller than if one uses the same line shape! What is the best line shape for the observations? Can you say anything significant?

The question whether the observed line profiles have this or that shape can be answered by this method by comparing the data with either type of profile ($k = \text{G}$ or L) and computing

$$P(k|O) = \int_{w_a}^{w_b} P(k, w|O)dw \quad (14)$$

$$\propto \int_{w_a}^{w_b} P(k, w)\Lambda(O|k, w)dw \quad (15)$$

We integrate over all line widths w to get the probability for either model irrespective of the line width. In comparing $P(G|O)$ and $P(L|O)$ we can decide which model is better.

4 Observational Data

Our observations is a spectrum, the measured flux as a function of wavelength. This is usually obtained by a CCD, so the data consist of a number (e.g. 1024) of values for the flux $F(i)$ and the wavelength for each pixel $\lambda(i)$. The wavelengths are obtained by calibrating the CCD image with the well known spectrum of a light source – usually a gas discharge lamp which produces many sharp emission lines.

The flux values are **not** the fluxes $F(\lambda(i))$ at the wavelength $\lambda(i)$ corresponding to the pixel's centre, but as each pixel has a finite size, each CCD element measures the integral of the true spectrum over the wavelength interval p corresponding to its size:

$$F(i) = \int_{\lambda(i)-p/2}^{\lambda(i)+p/2} F_{\text{CCD input}}(\lambda) d\lambda \quad (16)$$

Of course, we usually take a spectrograph configuration so that the pixel size is smaller than what we need for a wavelength resolution! But sometimes we have to push the instrument to its limits, or we have to squeeze out all information from the data available.

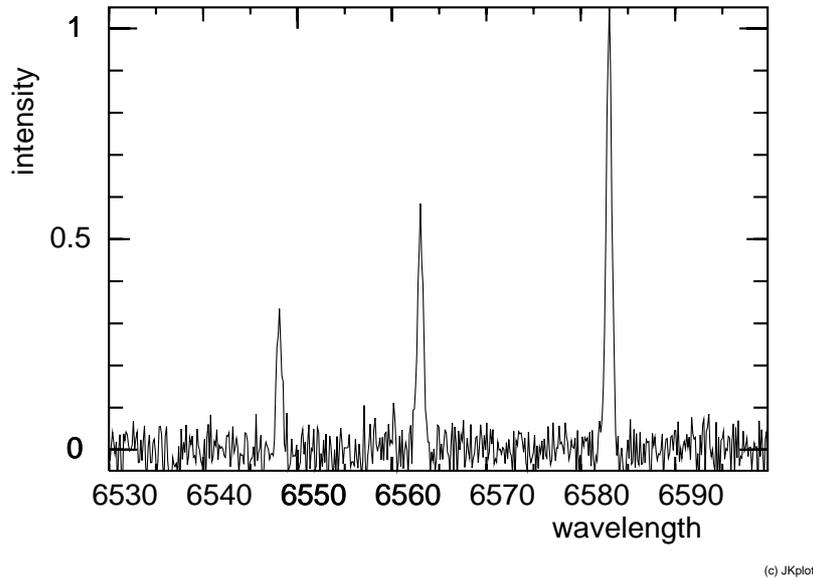
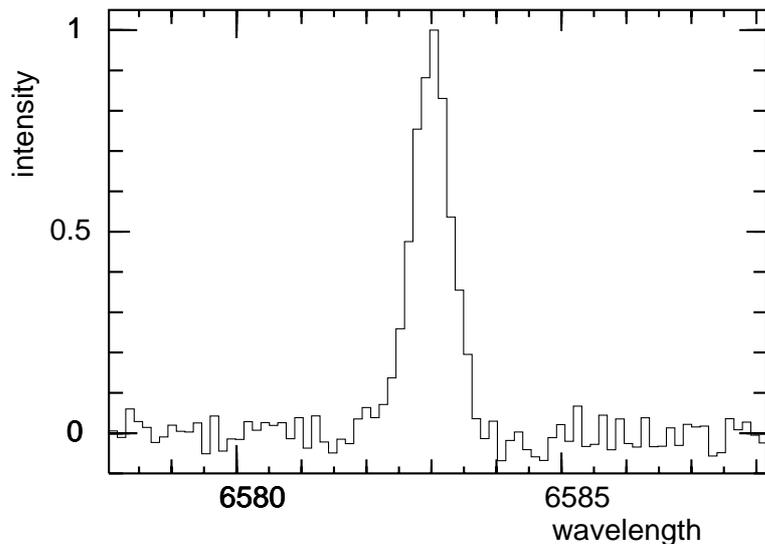


Figure 1: The spectrum of a (simulated) ionized nebula, showing the emission lines of $\text{H}\alpha$ 6563 Å, and $[\text{N II}]$ 6583 and 6548 Å



(c) JKplot

Figure 2: The line profile of the stronger [N II] line showing the binned structure of the data in form of a staircase curve

However, this is not all: The spectrograph has only a finite spectral resolution $R = \lambda/\Delta\lambda$ which smears out the true spectrum. One may well assume that the function with which the true spectrum is convolved is a Gaussian of width $\Delta\lambda$, so that the spectrum which is registered by the CCD camera is really:

$$F_{\text{CCD input}}(\lambda) = \int_{-\infty}^{\infty} F_{\text{true}}(\lambda)\phi(\lambda + x)dx \quad (17)$$

In principle there are further modifications of the spectrum due to the instrument and the Earth atmosphere, but as they change slowly with wavelength, the data can be corrected for these effects more easily. So we shall assume that all these corrections have already been done properly.

When one computes the line profiles predicted by the model, we have to apply all these same effects to get the predicted flux values for each pixel. This will mean that we have to convolve the theoretical line profile with the spectrograph's resolution. For Gaussian line profiles, we can make use of a nice property of the Gaussians: that the convolution of them (with widths σ_1 and σ_2) gives just another Gaussian with width $\sqrt{\sigma_1^2 + \sigma_2^2}$. This saves evaluating this integral numerically.

5 Noise and Error Distribution

For the formulation of the likelihood, we need to specify the distribution function of the error amplitude. Above, we used a Gaussian distribution, which is appropriate for white noise.

If there is not a single noise source, we might use a different distribution, such as a sum of two Gaussians of different widths, or a Lorentz function, or whatever else. Do make experiments!

Let us return to the simple Gaussian:

$$\psi(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (18)$$

This means that we have to specify the noise amplitude σ . There are two ways: (a) We take a portion of the spectrum where we know that there is no emission, so that all the signal there will be noise, and we simply take the mean and the dispersion of a suitably large number of pixels. The mean will define the zero level of the flux, while the dispersion we can take as σ^2 . (b) Another way is to use σ as another free parameter of our problem, and to determine which pair (v_{rad}, σ) gives the largest value of the likelihood. And one gets the confidence region, and the confidence intervals for either parameter. If you do this, you should in principle get the same value as in the first simple method, within the confidence region.

6 Artificial Data

In order to test your program and the methods it contains, the best way is to analyze artificial data created with known parameters.

To simulate the spectral data, we must do

- compute the theoretical line profile, $F_{\text{mod}}(\lambda, v_{\text{rad}}, \Delta\lambda_D)$
- convolve this profile with the resolution profile of the spectrograph. Note that if both profiles are Gaussians, the result is simply a Gaussian with width $\sigma_{\text{eff}}^2 = \sigma_{\text{line}}^2 + \sigma_{\text{spectrograph}}^2$!
- for every pixel, integrate this theoretical flux over the wavelength interval
- to simulate the detector noise, we add in each bin an error value to the flux value. This is a random number distributed like a Gaussian with width σ (the noise amplitude) and centered at 0. To generate this kind of random numbers, I recommend the transformation method which transforms the numbers generated by a random number generator (see "Numerical Recipes" available at the DEA library).

7 Details on the Real Data

Here are some details about the data that you'll be able to interpret. These are high resolution spectra of a recently discovered planetary nebula. The observations were done by F.Cuisinier and S.Durand with a 1.2m telescope in Brazil. Each CCD image has a 1024 by 1024 pixels. The wavelength scale is 0.13 Å per pixel. Perpendicular to the wavelength axis, there is spatial information. The emission of the nebula extends over several rows of pixels. From one of the images, we have extracted a couple of spectra for each row of pixels close to the center of the nebula. The wavelength region is around the H α and [N II] 6548 and 6583 Å emission lines. The resolution of the spectrograph was 0.25 Å.

The data is in the form of an ASCII file containing essentially a simple table of the flux values for several positions. Each line of the file corresponds to one of the 1024 pixels in wavelength direction, and contains the flux values in arbitrary units. The wavelength values are not given, but the first data line is at 6493.3 Å, and the step is 0.13 Å.

Some Notes:

- For our exercise, we effectively need only a portion of the entire spectrum from 6500 to 6630 Å.
- When you search for an emission-free portion, check whether it is also free from spurious "emissions" caused by cosmic rays that hit the CCD.
- The line fluxes are in some arbitrary units different from what you compute from the model. For simplicity, we shall scale the observed and theoretical to the same value, say that the flux in the line centre should be normalized to 1.0.

8 Types of analysis

In principle, there are a number of free parameters in the interpretation of the spectra: radial velocity, line width, line shape, noise level. What one could do:

- If we make (reasonable) assumptions about three of the four parameters, we can determine the remaining one. Thus we deal with one dimension only, so that we simply plot the curve of the likelihood as a function the parameter. Finding the peak and determining the confidence interval is easy. The obvious example is the derivation of radial velocity for a Gaussian line shape of theoretical width, and using the noise level determined from the dispersion of fluxes in an emission-free portion of the spectrum. This is shown in Fig. 3.
- Two parameters could be simultaneously determined, if we fix the other two. This is now a two-dimensional problem, and one scans both parameters independently. Graphically, we can make a contour plot. The confidence region now is that contour line within which say 90 percent of the total probability are found. The contour lines often are circles or ellipses (as in our problem), but they may have more complicated shapes. What does that tell you? Examples for this kind of analysis are radial velocity and line width, or radial velocity and noise level (Fig. 4).

- In the two-dimensional problem, one may also compute the "marginal" distributions such as

$$P(q) = \int_{p_a}^{p_b} P(p, q) dp \quad (19)$$

which is the distribution projected into the space of parameter q . It gives the probability distribution for the parameter q irrespective of what the best value for p might be. This is of course not the same as the probability $P(p_{\text{fix}}, q)$ for a fixed value p_{fix} ! For instance, we may compute the radial velocity irrespective of the line width. Example is shown in Fig. 5.

- One might well determine three parameters by scanning through all parameter triples, and finding the best triple. However, graphical representation is difficult! A better, and quicker way is to limit oneself to two-dimensional problems, and trying various pairs of parameters.

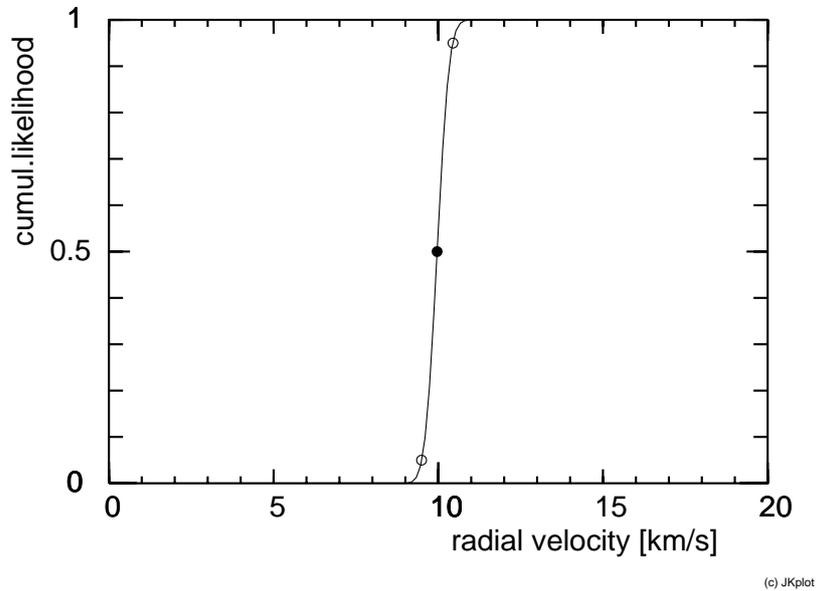
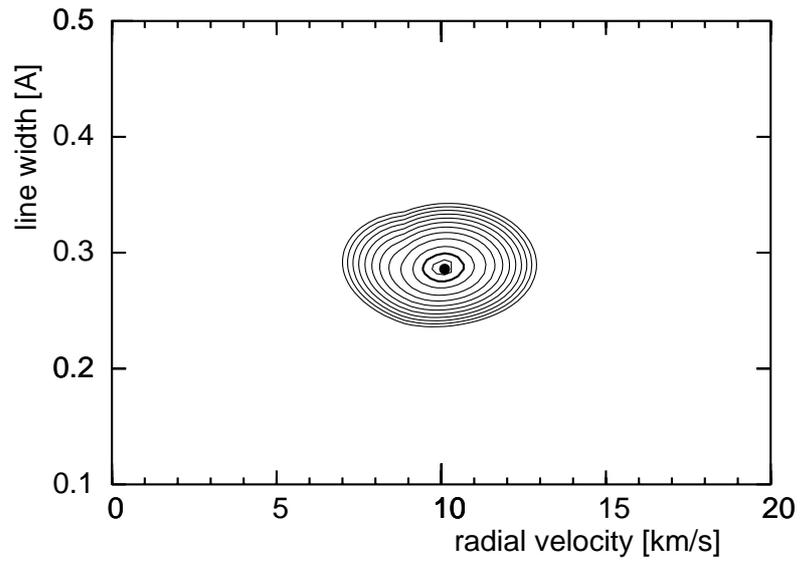


Figure 3: The cumulative likelihood distribution for the radial velocity with median value and the borders of the central 90 percent confidence interval



(c) JKplot

Figure 4: The two-dimensional distribution of the likelihood density shown as contour lines. The thick contour depicts the region which contains 90 percent of the total probability

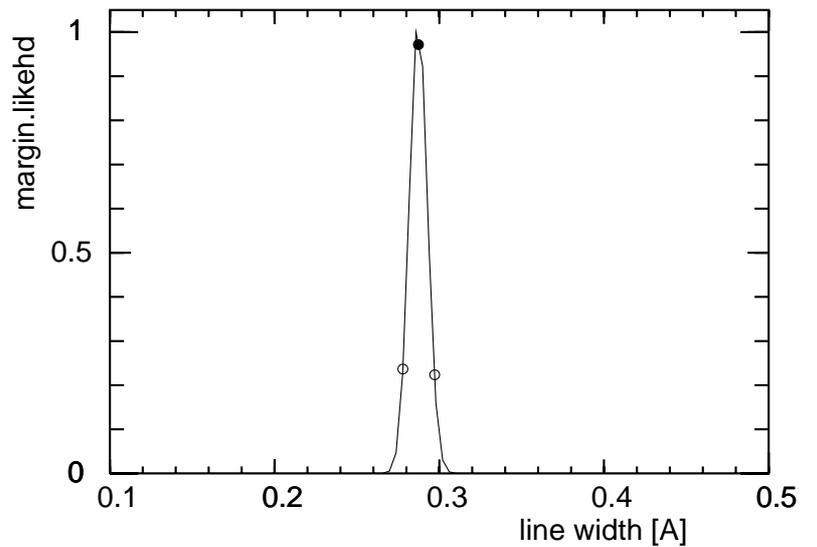


Figure 5: The marginal probability distribution density for the line width, computed from Fig. 4. The median value and confidence borders are indicated

9 Tasks

Write a program that interprets the emission line profiles, verify its proper functioning with artificial data, and analyze the real spectra.

- taking several of the spectra, what can you say about the radial velocity of the nebula?
- does the observed line width agree well with the theoretical expectations?
- what can you say about the validity of our assumption that the line profile is a Gaussian?