The benefits of optimal extraction

A.L.A.

December 22, 2017

1 Our PSF

In classic optimal extraction we consider that the slit is not tilted and its spread in the dispersion direction is smaller than the pixel size. Under these assumptions we can consider that one column of pixels across the order contains exclusively the photons of a single wavelength. We need just to collect the photons of that column into a flux measurement at the given wavelength. This would be the raw case. We shall compare this raw case with optimal extraction in an illustrative case.

Optimal extraction on the other hand is based upon assuming that we know what the PSF shape is along those pixels and that we can take advantage of that information to improve our results. Of course the core of the method is how to retrieve exact or at least sound information on that PSF. But let us suppose that this has been done. We want just to see if that is an improvement.

We will consider the following illustrative case. Our order spans 4 pixels across dispersion. We add one extra pixel on each side, top and bottom, and our 6-pixel PSF is

$$PSF = \frac{1}{7}\begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}$$

We will also consider two cases: the first one, our measurement really corresponds to that PSF multiplied by an unknown flux F; the second one, there has been a drift and the actual PSF during movement has moved 10% towards the top.

2 The brute extraction

In our first observation the pixels result in the following measurement

$$M_1 = \begin{pmatrix} 0 \\ 2 \\ 4 \\ 6 \\ 2 \\ 0 \end{pmatrix}$$

Since our assumption is that all those pixels correspond to a unique known wavelength, we conclude that the flux at that measurement is 14, but since this is a measurement there is an associated error bar which we will assume gaussian (as in many photons noise), so that the variance is 14, and the root mean square $\sqrt{(14)} = 3.74$. So our brute extraction results in the measurement

$$F_1 = 14 \pm 3.74$$

Our second observation suffers from a drift across dispersion of 10%. What we measure is therefore

$$M_2 = \begin{pmatrix} 0.2 \\ 2.2 \\ 4.2 \\ 5.6 \\ 1.8 \\ 0 \end{pmatrix}$$

We notice that $5.6 = 6 - 0.1 * 6 + 0.1 * 2$, that is, 10% of the flux is lost to the upper pixel and we recover 10% of the lower pixel. Our brute extraction is still 14, no change, and it also has the same noise 3.74. Because the amount of pixels was sufficiently large to cover up for drifts, the total amount of photons is still conserved, and nothing changes in spite of the drift.

$$F_2 = 14 \pm 3.74$$

3 The optimal extraction, case 1

Let us take now advantage of our knowledge of the PSF to see if we can improve upon the previous result. Improving means retrieving the same result with more confidence, with smaller error bars. So let us look in detail to those error bars. Actually photon noise affects each individual pixel. In the previous case we just quadratically added those individual pixel noises. But our actual measurement was rather

$$M_1 = \begin{pmatrix} 0 \pm 0\sigma \\ 2 \pm \sqrt{2}\sigma \\ 4 \pm 2\sigma \\ 6 \pm \sqrt{6}\sigma \\ 2 \pm \sqrt{2}\sigma \\ 0 \pm 0\sigma \end{pmatrix}$$

We are not sure any more that the actual values (0,2,4,6,2,0) are the real figures or how much have they been corrupted by noise. But we do know the PSF. So we can write the following equation

$$M_1 = F * PSF$$

or explicitly

$$0 \pm 0\sigma = 0$$

$$2 \pm \sqrt{2}\sigma = \frac{1}{7}F$$

$$4 \pm 2\sigma = \frac{2}{7}F$$

$$6 \pm \sqrt{6}\sigma = \frac{3}{7}F$$

$$2 \pm \sqrt{2}\sigma = \frac{1}{7}F$$

$$0 \pm 0\sigma = 0$$

We want to find an appropriate solution to that system of equations. Of course, with the chosen numbers, we see that the solution is evidently $F = 14$. But let us solve explicitly because in other cases the measurements will not be so nicely set. The presence of σ is what indeed allows us to play the game. It is going to be our χ^2, or our Lagrange multiplier or many other names with the same meaning. We are going to find a single value of F and σ that satisfy those equations with the constraint of σ being as small as possible. Rather than let our usual least-squares fitting algorithms do the job, let us do it by hand. We are bothered by the $\pm$ sign. To make it disappear let us first isolate that term:

$$\pm\sqrt{2}\sigma = \frac{1}{7}F - 2$$

$$\pm 2\sigma = \frac{2}{7}F - 4$$

$$\pm\sqrt{6}\sigma = \frac{3}{7}F - 6$$

$$\pm\sqrt{2}\sigma = \frac{1}{7}F - 2$$

Next we square the equations

$$2\sigma^2 = (\frac{1}{7}F - 2)^2$$

$$4\sigma^2 = (\frac{2}{7}F - 4)^2$$

$$6\sigma^2 = (\frac{3}{7}F - 6)^2$$

$$2\sigma^2 = (\frac{1}{7}F - 2)^2$$

We develop the square and sum all the equations

$$14\sigma^2 = \frac{15}{49}F^2 + 60 - \frac{60}{7}F$$

Since we want a minimum of σ, we derivate the right hand and make it equal to zero

$$\frac{30}{49}F - \frac{60}{7} = 0$$

The solution is $F = 14$, which we introduce in the equation with σ to see what is the error:

$$14\sigma^2 = \frac{2940}{49} + 60 - \frac{840}{7} = 0$$

Our final result is

$$F_{Opt} = 14 \pm 0$$

The zero noise is funny, but it was also improbable that the 6 pixel had exactly the expected flux distribution *as if noise had done nothing to any of them*, so this is exactly what we retrieved: zero noise. One can try (see next case) with slightly different numbers to see that the recovered noise is not zero.

The important message is that through brute extraction our solution was 14 ± 3.74 and now it is 14 ± 0. We are much more (infinitely more) confident on the result. Brute extraction did not brought up the important information that all pixels were proportional to a single flux times the (known) PSF. Optimal extraction allowed us to realise this important information and consequently noise dropped (to zero).

4 Optimal extraction. Case 2

One could play with measurements that were not exactly proportional to the PSF as in Case 1 to see how noise grows in the optimal extraction without actually never reaching the value of the brute extraction. Optimal extraction always improves upon brute extraction or, if there is no signal, provides the same result.

But let us first try with case 2 in which the measurement is different than the PSF, not because noise played tricks with it, but because there was a drift towards the top. As before we want to explicitly write noises in each pixel, so we rewrite

$$M_2 = \begin{pmatrix} 0.2 \pm \sqrt{0.2}\sigma \\ 2.2 \pm \sqrt{2.2}\sigma \\ 4.2 \pm \sqrt{4.2}\sigma \\ 5.6 \pm \sqrt{5.6}\sigma \\ 1.8 \pm \sqrt{1.8}\sigma \\ 0 \pm 0\sigma \end{pmatrix}$$

We equate this measurement to the PSF times a unique flux value. The PSF is unchanged: we have NOT corrected for the drift, we are going to make a mistake here. Let's see what kind of mistake. The explicit equations are:

$$0.2 \pm \sqrt{0.2}\sigma = 0$$

$$2.2 \pm \sqrt{2.2}\sigma = \frac{1}{7}F$$

$$4.2 \pm \sqrt{4.2}\sigma = \frac{2}{7}F$$

$$5.6 \pm \sqrt{5.6}\sigma = \frac{3}{7}F$$

$$1.8 \pm \sqrt{1.8}\sigma = \frac{1}{7}F$$

There is of course one extra equation now, since the top pixel receives light though this is not contemplated by the model. We isolate the σ term, square the equations and add them, to recover the fundamental relation:

$$\frac{15}{49}F^2 - \frac{58.4}{7}F + 57.12 = 14\sigma^2$$

We notice that the right hand has not changed! This is so because all photons are still counted on, even if they are not where they are expected. We look for the minimum by equating the derivative of the left side to zero

$$\frac{30}{49}F - \frac{58.4}{7} = 0$$

Our solution is therefore $F = 13.62$ (not 14!!) and the associated noise is $\sigma = 0.14$ so our final solution to the problem is

$$F = 13.62 \pm 0.14$$

On the positive side, we notice that optimal extraction is doing a fine job even when there is a mistake in the model: The error bar is much smaller than in the brute case. On the negative side, an error in the model translates into a larger error bar: we go from 0 to 0.14. This is the main effect of the drift upon optimal extraction: with exactly the same amount of photons detected, brute extraction produced the same result, drifted or not. But with the same amount of photons detected but a model error due to the drift optimal extraction reacts with a larger error bar. Rather than producing the same result, optimal extraction responds with a lesser confidence.

We also notice that the measurement in the case of optimal extraction is not 14 but 13.62. Is this not also a problem? Do we have a systematic effect in here that goes beyond the increase in the error bar?

The answer requires a Montecarlo simulation of different measurements possible with the sum flux and same PSF, but different just because of noise. We have therefore taken both measurements M_1 and M_2 and rather than handling the error just by its variance, we have simulated 5000 cases of each one. In each case a gaussian error was added to each one of the pixels with zero mean and variance equal to the pixel signal. The cases have been color coded red and blue, but we omit to say which one corresponds to M_1 and which one to M_2. After mixing randomly the 10000 experiments we have plotted in the top graph 200 cases (roughly half and half of each) with the measured flux and rms bar computed as in our explicit cases above. The shadow area reminds us of the error bar from the brute extraction. The case made above that error bars for optimal extraction are much smaller than for brute extraction is now obvious. But can we tell which cases, red or blue, correspond to M_1 and which ones to M_2?

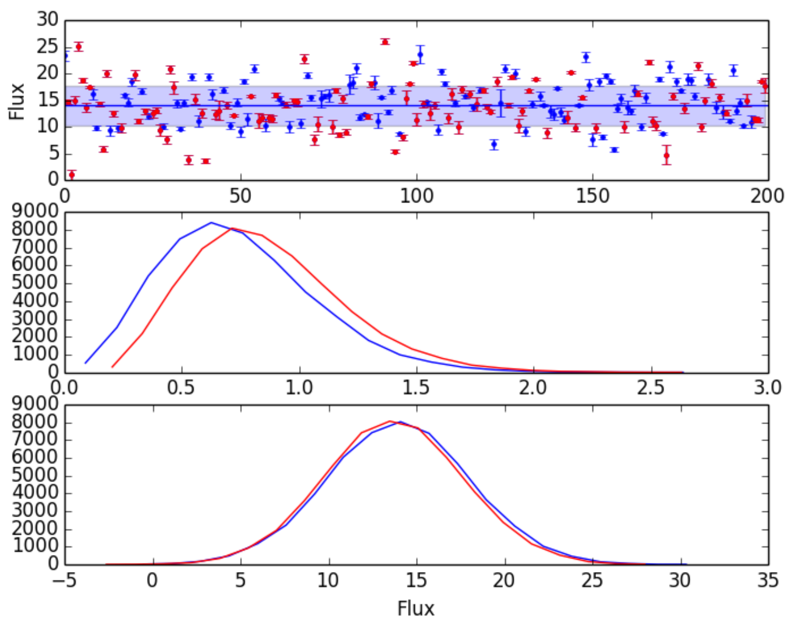

Figure 1:

To help disentangling red and blue cases the central and bottom graphs show corresponding histograms of the error bar and of the measured flux. This last one, the bottom one, shows almost identical histograms. Measured fluxes are almost identical in the red and blue cases. Certainly, the difference cannot be considered as meaningful in view of the variance of the histogram. The case is however more clear for the standard deviations (central plot). Here the red histograms is clearly shifted towards larger error bars. The red dots have systematically larger errors than the blue dots. Can we now tell who is M_1 and who M_2?