

# Proper motions estimation

Robert Estalella

2009 November

## 1 Calculation of proper motions when there are measurements at more than two epochs

Let us assume that there are  $n$  measurements of the position of a feature (star, knot, maser . . .) in the plane of the sky,

$$(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1}),$$

taken respectively at times

$$t_0, t_1, \dots, t_{n-1}.$$

The error in the measurements are

$$\epsilon(x_i), \epsilon(y_i), \quad i = 0, \dots, n-1.$$

For simplicity, since the proper motions in  $x$  and  $y$  are estimated independently, let us call  $z$  the coordinate we want to determine its proper motion, being either  $z = x$  or  $z = y$ .

Normally, the “absolute” positions  $z_i$  are meaningless (or are not measured), and we measure only the displacements with respect to a reference position, which we take, for instance, as that of epoch  $t_0$ . Let us call the displacements

$$\Delta z_i = z_i - z_0, \quad i = 1, \dots, n-1,$$

corresponding to time intervals

$$\Delta t_i = t_i - t_0, \quad i = 1, \dots, n-1.$$

The error associated with each  $\Delta z_i$  is now

$$\epsilon(\Delta z_i) = \sqrt{\epsilon^2(z_i) + \epsilon^2(z_0)}.$$

Two different approaches are possible:

### 1.1 First approach

The best determination of the proper motion velocity  $m'$  is the value that best approximates

$$\Delta z_i \simeq m' \Delta t_i, \quad i = 1, \dots, n-1.$$

The value of  $m'$  is that who minimizes the objective function

$$Q' = \sum_{i=1}^{n-1} \frac{(\Delta z_i - m' \Delta t_i)^2}{\epsilon^2(\Delta z_i)}.$$

This is equivalent to fit a straight line that passes through the origin  $(0, 0)$ , to a set of points  $(\Delta z_i, \Delta t_i)$ , with errors  $\epsilon(\Delta z_i)$  in the ordinate values. The value of  $m'$  can be easily calculated and is given by

$$m' = \frac{\langle \Delta t \Delta z \rangle}{\langle \Delta^2 t \rangle}.$$

The averages  $\langle \dots \rangle$  are calculated over  $n - 1$  points, using as weights  $w_i = 1/\epsilon^2(\Delta z_i)$ ,  $i = 1, \dots, n - 1$ .

$$\langle \Delta t \Delta z \rangle = \frac{\sum_{i=1}^{n-1} \Delta t_i \Delta z_i / \epsilon^2(\Delta z_i)}{\sum_{i=1}^{n-1} 1 / \epsilon^2(\Delta z_i)},$$

and

$$\langle \Delta^2 t \rangle = \frac{\sum_{i=1}^{n-1} \Delta^2 t_i / \epsilon^2(\Delta z_i)}{\sum_{i=1}^{n-1} 1 / \epsilon^2(\Delta z_i)}.$$

## 1.2 Second approach

The displacements  $\Delta z_i = z_i - z_0$  are only a change in the origin of coordinates. Thus, the best determination of the proper motion velocity  $m$  is the value that best approximates

$$\Delta z_i \simeq m \Delta t_i + p, \quad i = 0, \dots, n - 1,$$

where  $m$  and  $p$  are the slope and intercept of the regression line that minimizes the objective function

$$Q = \sum_{i=0}^{n-1} \frac{(\Delta z_i - m \Delta t_i - p)^2}{\epsilon^2(z_i)}.$$

Note that in the last expression the sum includes the point  $\Delta t_0 = 0$ ,  $\Delta z_0 = 0$ , with error  $\epsilon(z_0)$ . The regression line is no longer forced to pass through the origin  $(0, 0)$ . The errors are no longer  $\epsilon(\Delta z_i)$ , but  $\epsilon(z_i)$ .

The values of  $m$  and  $p$  are given by the standard formula of the regression line,

$$m = \frac{\langle \Delta t \Delta z \rangle - \langle \Delta t \rangle \langle \Delta z \rangle}{\langle \Delta^2 t \rangle - \langle \Delta t \rangle^2},$$

and

$$p = \langle \Delta z \rangle - m \langle \Delta t \rangle,$$

where the averages are taken over  $n$  points, using as weights  $w_i = 1/\epsilon^2(z_i)$ ,  $i = 0, \dots, n - 1$ .

### 1.3 Comparison of the approaches

The two methods give, in general, different results. Which one is the best?

Let us compare both in a simple case, with only 3 points,  $(t_-, z_-)$ ,  $(t_0 = 0, z_0 = 0)$ , and  $(t_+, z_+)$ , with the same error  $\epsilon$  associated to each point. (In the case of equal errors, the weighted averages are the same as unweighted averages.) The first approach gives

$$m' = \frac{t_- z_- + t_+ z_+}{t_-^2 + t_+^2},$$

while the second approach gives a different expression,

$$m = \frac{3(t_- z_- + t_+ z_+) - (t_- + t_+)(z_- + z_+)}{3(t_-^2 + t_+^2) - (t_- + t_+)^2}.$$

If the first and last point are symmetric in time with respect to  $t_0$ , that is,  $t_+ = -t_- = t$ , and  $t_+ - t_- = 2t$ , both methods give the same result,

$$m' = m = \frac{z_+ - z_-}{2t},$$

result that be easily interpreted geometrically. Let us now consider the non-symmetric case, in the limiting case when  $t_- \rightarrow t_0 = 0$ . The first approach gives

$$m' = \frac{z_+}{t_+},$$

while the second approach gives a different value,

$$m = \frac{z_+ - \frac{1}{2}z_-}{t_+}.$$

Thus, the first approach gives a result that does not depend on the first measurement  $z_-$ , but only depends on  $z_0 = 0$  and  $z_+$ . This does not seem to be a good determination of the proper motion. On the contrary, the second approach gives a reasonable result. The factor  $1/2$  that multiplies  $z_-$  can be seen as the weight that affects  $z_-$  and  $z_0 = 0$ , because both are measured at the same epoch,  $t_- = t_0 = 0$ .

### 1.4 Conclusion

In conclusion, at least in some cases, the second approach appears to work better than the first one.

## 2 Cross-correlation measurements: where to put the errors?

### 2.1 Errors in Cross-correlation

Cross-correlation is the choice method for measuring proper motions of extended objects, like HH objects, or molecular outflow knots. When measuring proper motion through cross-correlation of the images at two different epochs, two sources of error can be taken into account.

**Alignment error,  $\epsilon^{\text{al}}$ .** This is the error in the alignment of the images at the two epochs. It can be measured, for instance, as the rms deviation of the position of the set of reference stars used to align the images.

**Cross-correlation error,  $\epsilon^{\text{cor}}$ .** The cross correlation function can be narrow, with a well-defined peak, or, on the contrary, wide and flat. In addition, the cross-correlation can be very sensitive to the box where it is calculated. A measure of the error in the cross-correlation peak can be the dispersion of results obtained by modifying the box used in  $\pm 1$  or  $\pm 2$  pixels in each side.

Thus, for each measurement of the displacement  $\Delta z_{0i}$  between epochs  $t_0$  and  $t_i$ ,  $i = 1, \dots, n - 1$ , we have associated an error

$$\epsilon_{0i}^2 = (\epsilon_{0i}^{\text{al}})^2 + (\epsilon_{0i}^{\text{cor}})^2.$$

Since we want to fit a straight line to the set of points  $(\Delta t_0 = 0, \Delta z_0 = 0)$ ,  $(\Delta t_i, \Delta z_i)$ ,  $i = 1, \dots, n - 1$ , what are the errors associated with each point  $\Delta z_i$ ?

## 2.2 Errors in the displacements $\Delta z_i$

A reasonable approach is to distribute the error  $\epsilon_{0i}$  evenly between the two epochs,  $t_0$  and  $t_i$ . That is,

$$\epsilon_0^2 = \frac{1}{2}\epsilon_{0i}^2, \quad \epsilon_i^2 = \frac{1}{2}\epsilon_{0i}^2,$$

and the corresponding weights

$$w_0 = \frac{2}{\epsilon_{0i}^2}, \quad w_i = \frac{2}{\epsilon_{0i}^2},$$

Now we have to take into account that the point  $(\Delta t_0 = 0, \Delta z_0 = 0)$  appears for each pair of epochs  $t_0$  and  $t_i$ , in total  $n - 1$  times. So its weight has to be the sum of weights

$$w_0 = \sum_{i=1}^{n-1} \frac{2}{\epsilon_{0i}^2}.$$

## 2.3 Conclusion

Procedure for  $n$  observations at epochs  $t_i$ ,  $i = 0, \dots, n - 1$ :

- For each pair of epochs  $t_0$ ,  $t_i$ , measure the displacement  $\Delta z_i$ , with an associated error  $\epsilon_{0i}$ .
- Calculate the weights  $w_0 = \sum_{i=1}^{n-1} 2/\epsilon_{0i}^2$ ,  $w_i = 2/\epsilon_{0i}^2$ ,  $i = 1, \dots, n - 1$ .
- Fit a straight line to the set of points  $(\Delta t_0 = 0, \Delta z_0 = 0)$ ,  $(\Delta t_i, \Delta z_i)$ , using the weights  $w_0$ ,  $w_i$ ,  $i = 1, \dots, n - 1$ .