EKOS Linear 1 Pass Focus

Introduction

The concept of the Linear 1 Pass focus algorithm is to provide a mechanism to produce an accurate focus curve in a single pass, then fit a curve to the datapoints, calculate the point of focus and then move the focuser to that point.

The following curves can be chosen:

- Hyperbola.
- Parabola.
- Quadratic. The original EKOS method which fits a quadratic (parabolic) curve.

As to which curve works best, that is upto the user to determine. In general, a hyperbola will work better on a wider range of focus movement. For narrower focus movement ranges where there is insufficient travel for the asymptotes of the hyperbola to be reached, the curve will likely be better approximated by a parabola.

The GNU Science Library (GSL) is used to fit the curve to the datapoints. The generic Levenberg–Marquardt solver is used. See the section on Levenberg–Marquardt for more details.

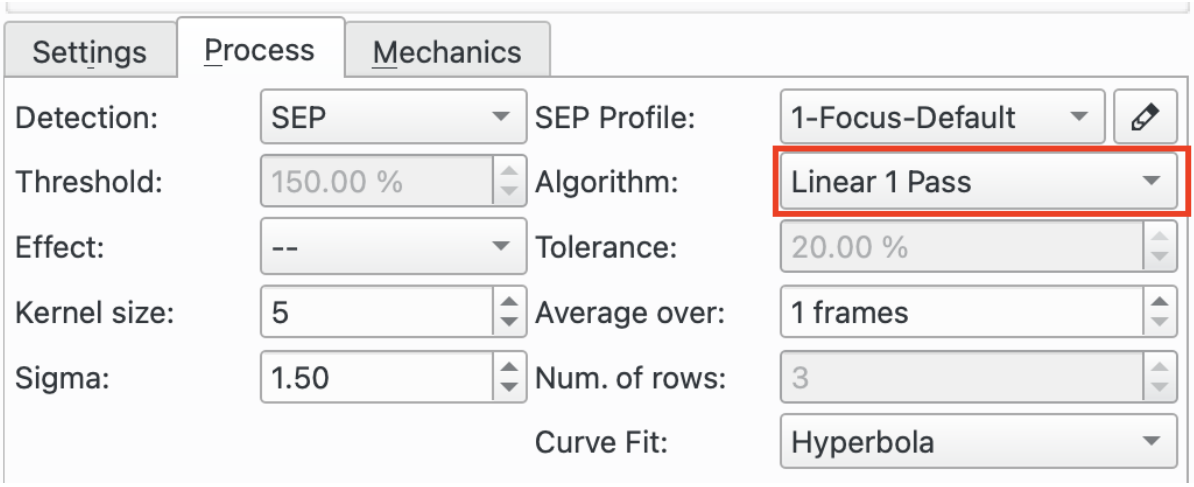
The degree to which the curve fits the measured datapoints is estimated by R2. R2 ranges from 0 (the curve does not fit the data at all) upto 1 (which means all datapoints lie exactly on the curve. See the section on Coefficient of Determination, R2 for more details.

There is an option to set a minimum value of R2 that is acceptable. If the fitted curve fails to meet the minimum, the focus process will be re-run.

There is also the option to weight each datapoint during the curve fitting process, or give each datapoint equal weight (the current EKOS method). The weighting is based on the standard deviation of HFR star values captured during the focus process. The idea is to give more weight to datapoints with less variation in HFR and less weight to datapoints with more variable HFRs.

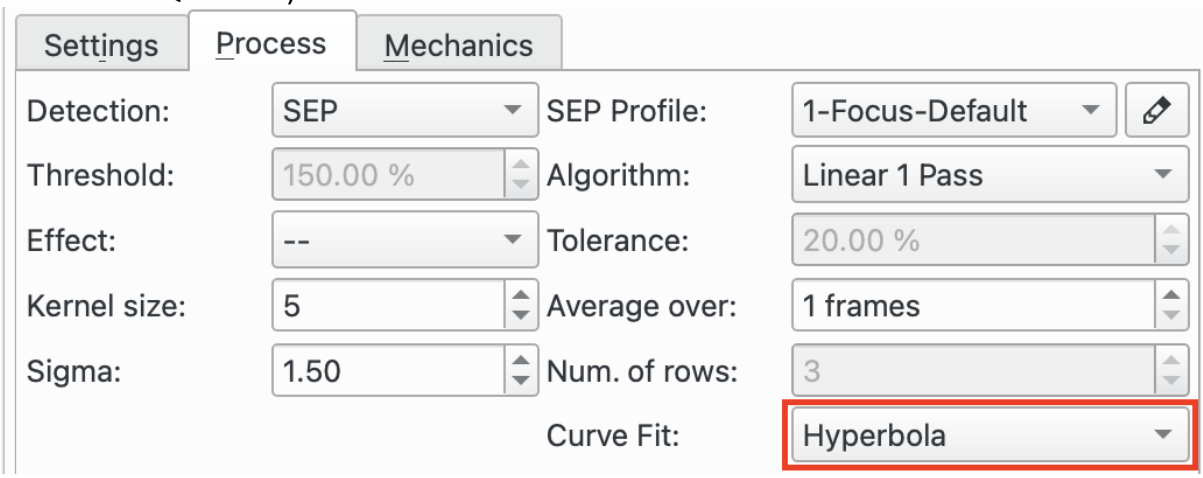
The Focus Run

The Linear 1 Pass algorithm is selected as Algorithm on the Process tab on the Focus screen:

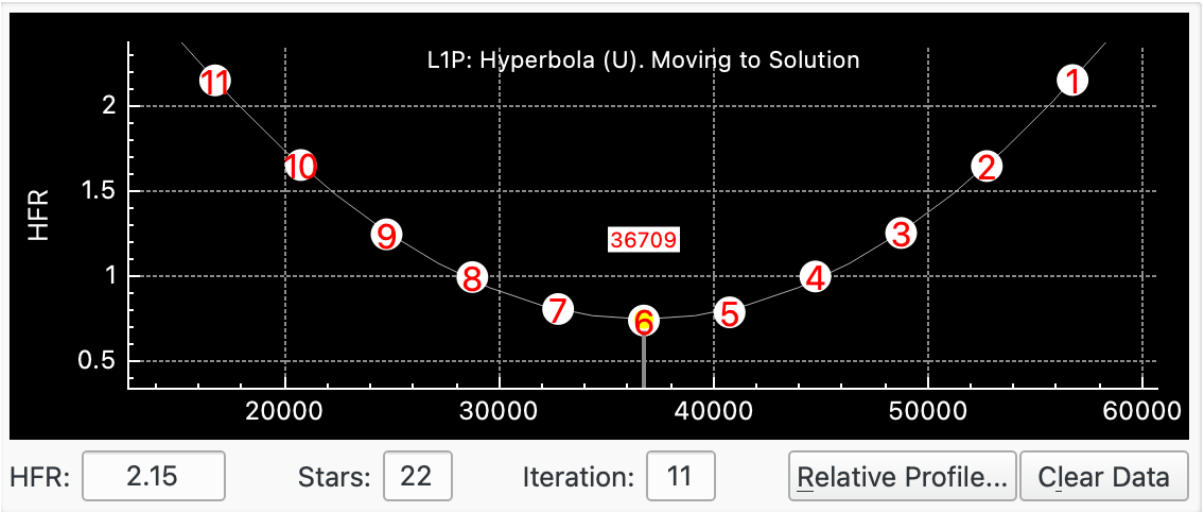


When the Auto Focus button is pressed, the Linear 1 Pass algorithm starts in the same way as the Linear algorithm.

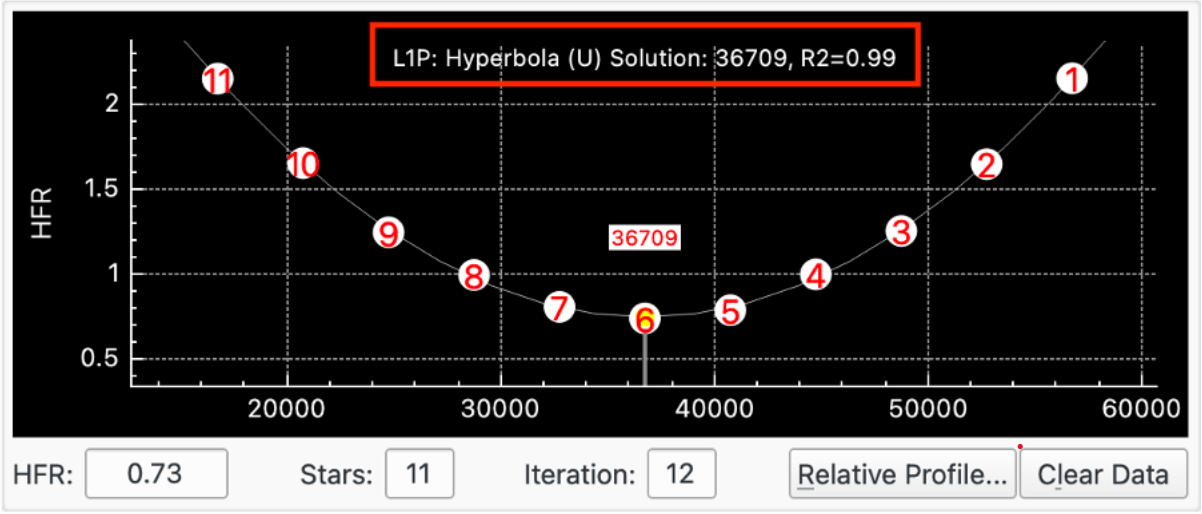
1. The algorithm steps out from the current position by “Initial Step Size” * “Out Step Multiple”. This is done in a backlash independent way (see the section on Backlash for more details). It then takes a frame and measures the HFR of the stars in it.
2. It then steps in by Initial Step Size, takes a frame and calculates the star HFR.
3. When it has 4 or more datapoints it tries to fit a curve. The type of curve (Hyperbola, Parabola or Quadratic) is chosen here:



4. The process continues until it passes the minimum of the curve. There is some logic to ensure it copes with small random fluctuations in the datapoints when looking for the minimum. Once past the minimum is counts for Out Step Multiple datapoints, before finishing the focus run.
5. When the auto focus run has enough datapoints is calculates the curve fit for the last time, again computing the minimum position which is displayed on the V-Curve, as shown below:



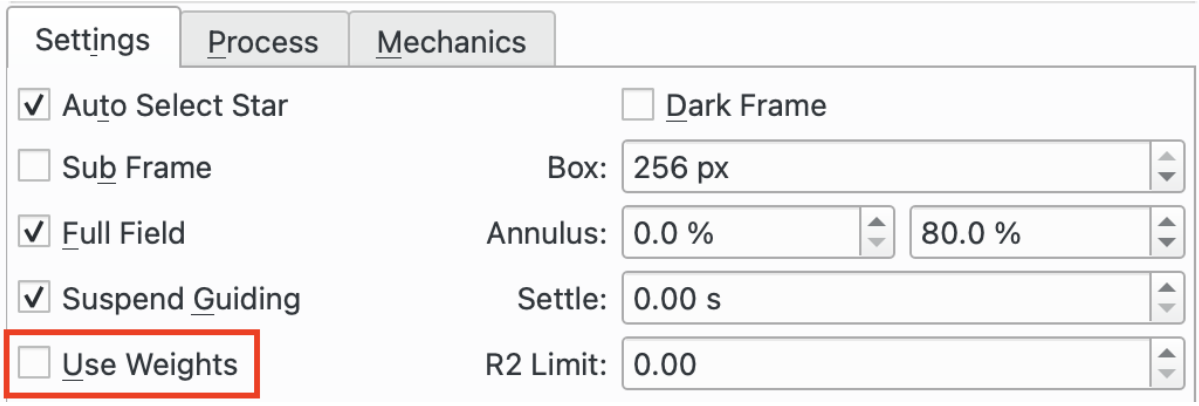
6. Up until this point, the Linear 1 Pass algorithm is very similar to the Linear algorithm. The next step in Linear 1 Pass is different though. The focuser is moved directly to the focus minimum in a backlash independent way (see the Backlash section). The R2 value of the curve fit is calculated, and the final V-Curve displayed. The title summarises details of the solution including the R2 value.



Chosing a Curve

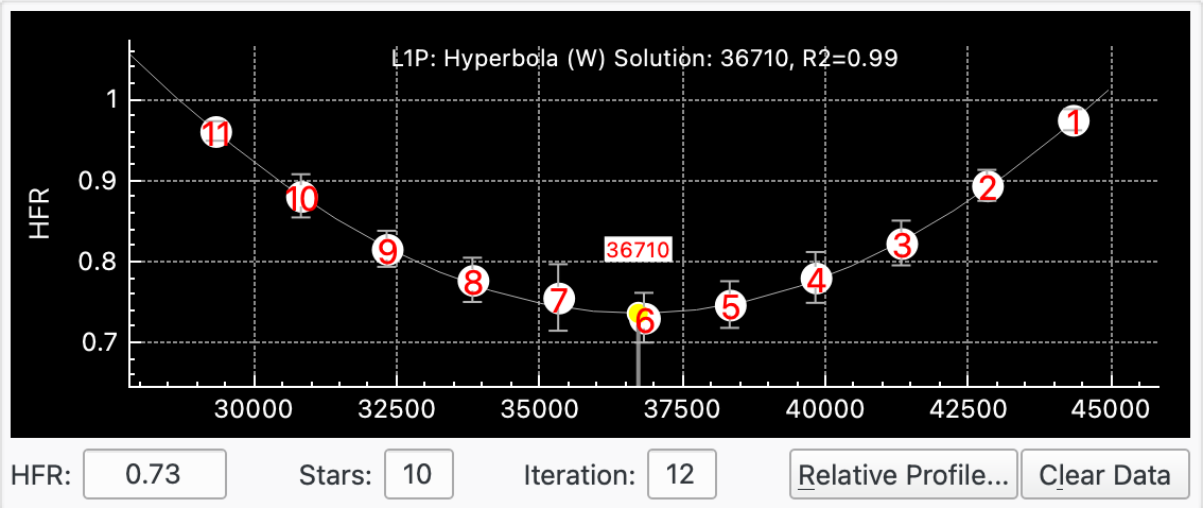
The Linear 1 Pass algorithm supports the existing EKOS methodology called Quadratic which fits a parabola as well as the new Levenberg–Marquardt solver for Hyperbola and Parabola. As to which works best the user should experiment with their equipment.

There is also the option to weight the datapoints used by the LM solver. The option is found on the Settings tab of the Focus window:



When selected the standard deviation of the star HFR values calculated is used to weight each datapoint in the LM solver. The idea here is to give less weight to points with higher standard deviations and more weight to those with smaller values. The standard deviation of each point is shown on the V-Curve as error-bars. The larger the error bar the less certainty there is in the HFR of the datapoint.

In the below example, point 1, 2 and 11 have smaller error bars, corresponding to a smaller standard deviation, so the curve passes close to the centre of those points. Point 7, however, has a larger error bar, corresponding to a larger standard deviation, so the LM solver has fitted a curve that has been allowed to go a bit further away from the centre of that point.

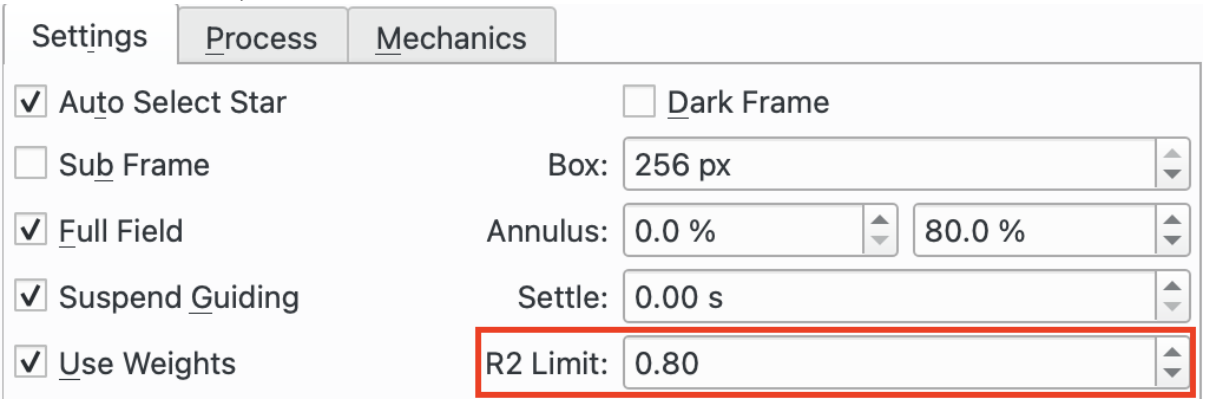


Coefficient of Determination, R2

R2 is calculated in order to give a measure of how well the fitted curve matches the datapoints. More information is available here: https://en.wikipedia.org/wiki/Coefficient_of_determination. In essence, R2 gives a value between 0 and 1, with 1 meaning a perfect fit where all datapoints sit on the curve, and 0 meaning that there is no correlation between the datapoints and the curve.

The user should experiment with their equipment to see what values they can obtain, but as a guide, a value above, say 0.8 would be a good fit.

There is an option to set an “R2 Limit” in the Settings tab of the Focus window that is compared to the calculated R2 after the auto focus run has completed. If the limit value has not been achieved, then the auto focus is rerun:



Setting an R2 Limit could be useful for unattended observation if the focus run produces a bad result for a 1-off reason. Obviously if the reason is not transitory then rerunning will not improve anything.

If the R2 Limit is not achieved and the focus process is rerun, and again fails to achieve the R2 Limit, then the focus run is marked as successful to avoid the process getting stuck rerunning auto focus forever.

This feature is turned off by setting the R2 Limit to 0.

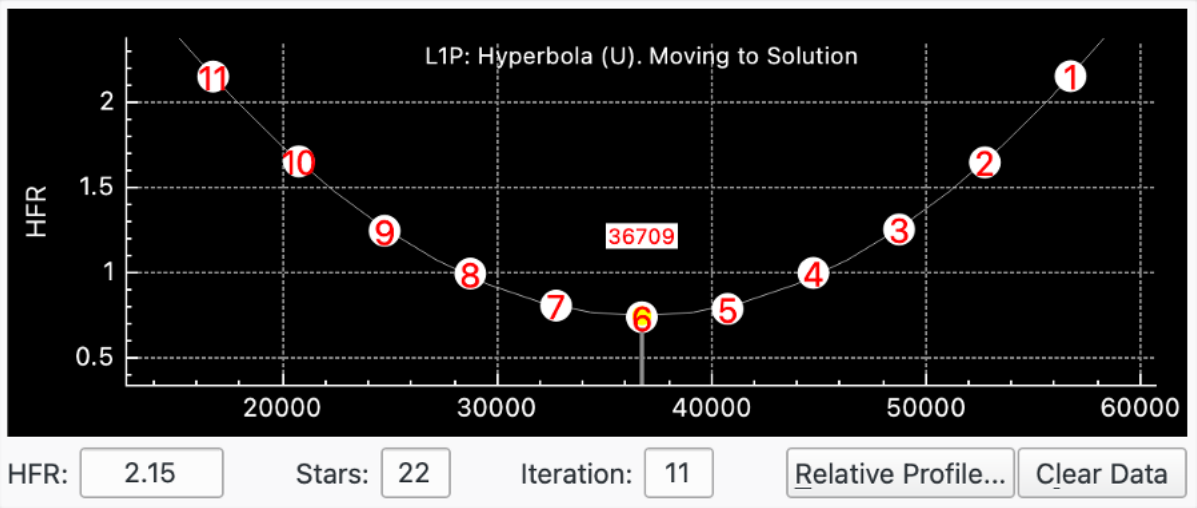
How to Setup for an Auto Focus Run

The exact settings that work best for a given astronomical setup need to be worked out by the user using trial and error, but this section gives some pointers:

1. Start near to focus by manually finding focus.
2. Select Linear 1 Pass and your curve of choice, say Hyperbola. Select Use Weights.
3. Make sure you are finding enough stars.
    a. Start with the SEP star detection method and the 1-Focus-Default profile unless you have reason to use a different setup.
    b. On the Settings tab, use Full Field (rather than Sub Frame) which will use many stars rather than just one. Select Auto Select Star to let the system select the stars to use.
    c. Make sure Annulus is set to use a large amount of frame to make use of as many stars as possible. Note that there may be other factors that prevent you using all of the field such as issues with an un-flat field in the corners of the sensor, but don’t overdo the restriction.
    d. Set the camera settings such as exposure, gain and bining such that you are getting a good number of stars. Its impossible to be prescriptive here but try for between 20 and 100 (obviously if your focal length and target can’t find that many then the process should still work but may be less optimal from a curve fitting perspective).

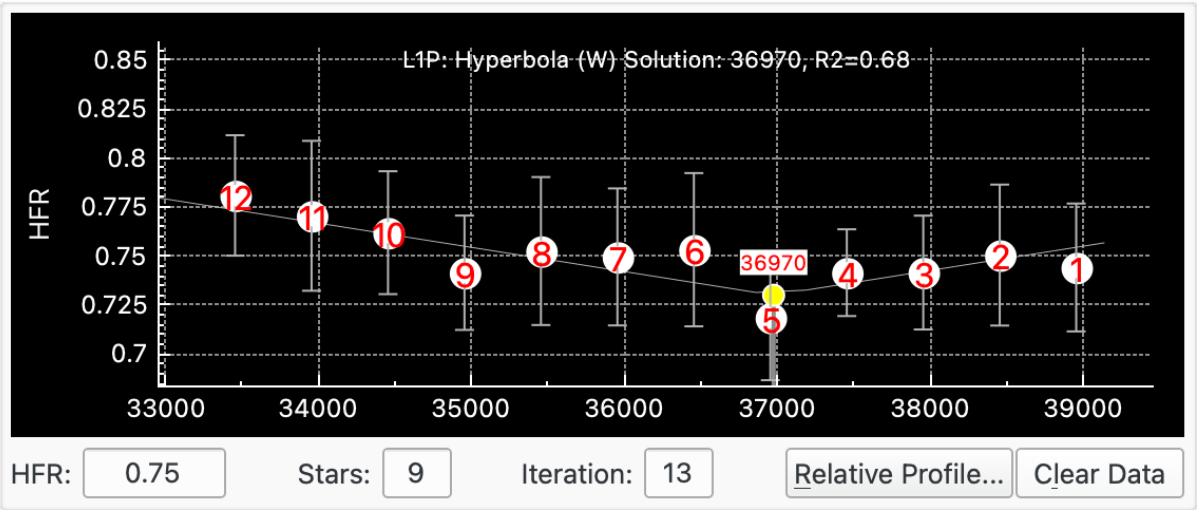
e. Upping the exposure usually finds more stars (but makes the focus process longer). You can also try taking multiple frames at the same point by setting the Average Over field > 1 frame.

4. Setup the Mechanics tab:
   a. Setup Backlash. See the Backlash section for more details but if you don't know the value for your equipment then set to 0.
   b. Setup Max Travel. This should be appropriate for your focuser to prevent overextending it. It needs to be big enough to support the values set in Initial Step Size and Out Step Multiple. It will need a minimum of just over 2 * Initial Step Size * Out Step Multiple. If you can, set it to, say, double this.
   c. Settle. If your focuser causes vibration in the optical train then you need to set this value so that after moving, the system waits for Settle seconds before taking the next frame. Try moving the focuser then taking a few frames at the same position. If the first frame after movement has bigger HFRs than subsequent frames, then you probably need to up the value in Settle.
   d. Out Step Multiple. Start with 4 or 5. This will give you 9-10 or 11-12 datapoints which is a good place to start. You need enough datapoints to form a curve, but the more you have the longer the process will take to complete.
   e. Initial Step Size. The following shows a "good curve". There is significant movement in the HFR axis to clearly demonstrate the curve, in this case max HFR is about 2.2 whilst min is 0.75 which gives a max / min of about 3.



   f. In contract, the next picture shows an Initial Step Size that has been set too low. The HFR varies from about 0.78 to 0.72. Which gives a max / min just over 1. The other clue that this is a poor setup is that the Error Bar range is very large compared to HFR movement which means that the LM Solver is drawing a curve through a lot of noise, which means the results will not be

very accurate.



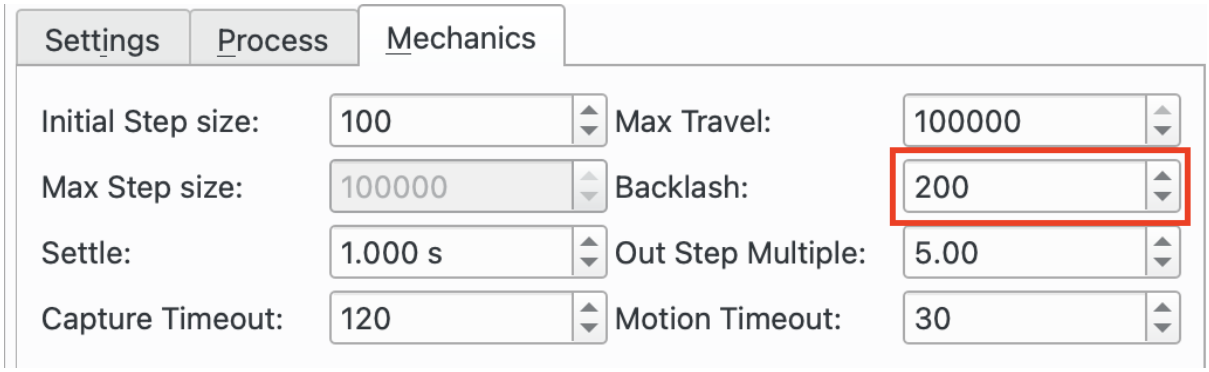
Backlash

All mechanical devices with gears suffer from backlash. In a typical focuser there will be a dead zone where changing direction (e.g. from “in” to “out”) results in movement of the focuser by a few ticks, but no actual movement of the optical train.

The Linear 1 Pass algorithm (like to the Linear algorithm) provides backlash compensation in the 2 places during an auto focus run when it moves outwards:

1. The initial outward movement of Initial Step Size * Out Step Multiple at the start of the run.
2. When the inward pass is complete and EKOS has determined the best focus position and moves outward to it.

Linear 1 Pass will extend (by x ticks) the outward movement, and then, as a separate movement it will move inward by x ticks. So it always approaches the desired position in an inward direction.



There are 2 schemes that can be used:

1. Set Backlash to 0. EKOS will use a value of x = 5 * Initial Step Size
2. Set Backlash > 0. EKOS will use a value of x = Backlash. There will probably be instructions with your focuser for determining the value of Backlash. It is not necessary to set an exact value for Linear 1 Pass to work correctly; all that is required is that the value set in Backlash is >= the actual backlash value. For example, if you

measure backlash and get a value around 100, any value >=100 will work equally well. For example, set Backlash = 200.

Levenberg–Marquardt

The Levenburg-Marquardt (LM) algorithm is used to solve non-linear least squares problems. The GNU Science Library provides an implementation of the solver. These resources provide more details:

- https://en.wikipedia.org/wiki/Levenberg–Marquardt_algorithm
- https://www.gnu.org/software/gsl/doc/html/nls.html

The Levenberg-Marquart algorithm is a non-linear least-squares solver and thus suitable for many different equations. The basic is idea is to adjust the equation y = f(x,P) so that the computed y values are as close as possible to the y values of the datapoints provided, so that the resultant curve fits the data as best as it can. P is a set of parameters that are varied by the solver in order to find the best fit. The solver measures how far away the curve is at each data point, squares the result and adds them all up. This is the number to be minimised, lets call is S. The solver is supplied with an initial guess for the parameters, P. It calculates S, makes an adjustment to P and calculates a new S1. Provided S1 < S the solver has moved in the right direction and reduced S. It iterates through this procedure until S1 - S is:

- less than a supplied limit (convergence has been reached), or
- the maximum number of iterations has been reached, or
- the solver encountered an error.

The solver is capable of solving either an unweighted or weighted set of datapoints. In essence, an unweighted set of data gives equal weight to each datapoint when trying to fit a curve. An alternative is to weight each datapoint with a measure that corresponds to how accurate the measurement of the datapoint actually is. Given we are calculating HFRs of stars we can calculate the variance (standard deviation squared, SD^2) in the measurement of HFR and use 1 / SD^2 as weighting. What this means is that if we have a datapoint where the SD in HFR measurements is small we would give this datapoint a relatively high weighting when fitting the curve, and if there was a datapoint with a higher SD in HFR measurements it would receive a lower weighting when trying to fit the curve to that point.

There are several optimisations to LM that speed up convergence for certain types of equations. This code uses the basic LM solver which keeps it generic if more equations are to be added. Of course, if required this could easily be tweaked. An optimisation that has been implemented is that since, in normal operation, the solver is run multiple times with the same or similar parameters, the initial guess for a solver run is set to the solution from the previous run.

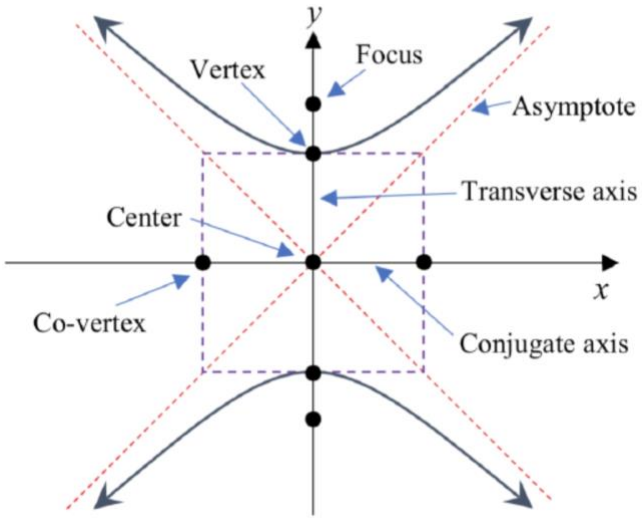
The documents referenced provide the maths of the LM algorithm. What is required is the function to be used f(x) and the derivative of f(x) with respect to the parameters of f(x). This partial derivative forms a matrix called the Jacobian, J(x). LM then finds a minimum. Mathematically it will find a minimum local to the initial guess, but not necessarily the

global minimum but for the equations used this is not a problem as hyperbolas and parabolas have just 1 minimum.

The original algorithm used in Ekos is a quadratic equation, which represents a parabolic curve. In order not to disturb historic code the original solution is called Quadratic and left untouched (as much as possible). It uses a linear least-squares model, not LM. The LM solver has been applied to a hyperbolic and a parabolic curve.

Hyperbola

Lets start with a diagram of a hyperbola, centred on the origin:



The equation is given by:

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1.$$

In our case it will be shifted away from the origin to xy point (c, d) so lets replace x by (x – c) and y by (y – d). So the equation becomes:

$$\frac{(y - d)^2}{b^2} - \frac{(x - c)^2}{a^2} = 1$$

Rearranging gives:

$$y = f(x) = b * \sqrt{\left(1 + \left(\frac{(x - c)}{a}\right)^2\right)} + d$$

This can be re-written as:

$$f(x) = b * \phi + d$$

$$where\ \phi = \sqrt{\left(1 + \left(\frac{(x - c)}{a}\right)^2\right)}$$

So, there are 4 parameters associated with the parabola: a, b, c, d

We need to partially differentiate f(x) with respect to the parameters (a, b, c, d) to get the Jacobian:

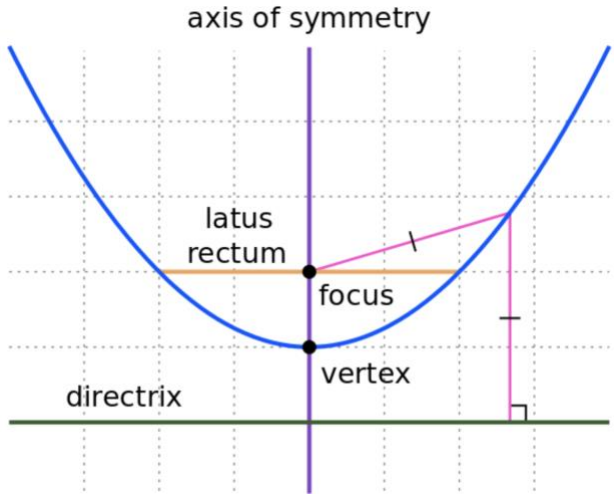
$$
\begin{aligned}
Jacobian\ J &= \left\{\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}, \frac{\partial f}{\partial c}, \frac{\partial f}{\partial d}\right\} \\
\frac{\partial f}{\partial a} &= b * \left(\frac{1}{2}\right) * \left(\frac{1}{\phi}\right) * -2 * \frac{((x - c)^2)}{a^3} \\
&= -b * \frac{((x - c)^2)}{a^3 * \phi} \\
\frac{\partial f}{\partial b} &= \phi \\
\frac{\partial f}{\partial c} &= b * \left(\frac{1}{2}\right) * \left(\frac{1}{\phi}\right) * 2 * \frac{(x - c)}{a^2} * (-1) \\
&= -b * \frac{(x - c)}{a^2 * \phi} \\
\frac{\partial f}{\partial d} &= 1
\end{aligned}
$$

For a valid solution:

- c must be > 0 and within the range of travel of the focuser.
- b must be > 0 for a V shaped curve (b < 0 for an inverted V and b = 0 for a horizontal line).
- b + d > 0. The minimum solution when x = c gives f(x) = b + d which must be > 0.

Parabola

Here is a diagram of a parabola centred on the origin:



The equation is given by:

$$y = b * x^2$$

In our case the parabola is shifted away from the origin to xy point (c, a) so lets replace y with (y – a) and x with (x – c).
The equation now becomes:

$$y = f(x) = a + b((x - c)^2)$$

So, there are 3 parameters associated with the parabola: a, b, c

We need to partially differentiate f(x) with respect to the parameters (a, b, c) to get the Jacobian:

$$Jacobian\ J = \left\{\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}, \frac{\partial f}{\partial c}\right\}$$
$$\frac{\partial f}{\partial a} = 1$$
$$\frac{\partial f}{\partial b} = (x - c)^2$$
$$\frac{\partial f}{\partial c} = -2 * b * (x - c)$$

For a valid solution:

- c must be > 0 and within the range of travel of the focuser.
- b must be > 0 for a V shaped curve (b < 0 for an inverted V and b = 0 for a horizontal line).
- a > 0. The minimum solution when x = c gives f(x) = a which must be > 0.