Newton–Raphson Fitting

Kaloyan Penev

December 13, 2017

1 Definitions

V Combined flux variance of all sources and observations (the quantity we are trying to minimize).

σ_i A vector of the sub–pixel sensitivities (the independent variables, $i = 1 \dots N$).

b^{st} The background for each observation of each source. The s index identifies the star, t identifies the observation (frame).

$\bar{b}^s$ $\equiv \sum_t \xi_{st} b^{st}$.

w^{st}_{pi} A set of matrices which can be used to compute the fluxe each source has for each observation given a subpixel map. The flux of the s-th star in the t-th frame is calculated as

$$f^{st} \equiv \sum_p \frac{1}{\sum_i w^{st}_{pi} s_i} - b^{st}$$

q_s A vector of the relative weights of the various images. Should satisfy: $\sum_s q_s = 1$.

ξ_{st} A matrix of weights. Each row gives the relative weights of the different observations of a star. Each row should satisfy: $\sum_t \xi_{st} = 1$.

Λ^{st}_k $\equiv \sum_q \frac{w^{st}_{ql}}{\left(\sum_i w^{st}_{qi} s_i\right)^2}$

$\bar{\Lambda}^s_k$ $\equiv \sum_t \xi_{st} \Lambda^{st}_k$

Ω^{st}_{kl} $\equiv \sum_p \frac{2 w^{st}_{pk} w^{st}_{pl}}{\left(\sum_i w^{st}_{pi} s_i\right)^3}$

2 The Variance and its Derivatives

In general the combined flux variance can be written as:

$$V \equiv \sum_s q_s \left\{ \sum_t \xi_{st} \left[\sum_p \frac{1}{\sum_i w_{pi}^{st} s_i} - b^{st} \right]^2 - \left[\sum_t \xi_{st} \left(\sum_p \frac{1}{\sum_i w_{pi}^{st} s_i} - b^{st} \right) \right]^2 \right\}$$

Differentiating:

$$\frac{\partial V}{\partial s_k} = 2 \sum_s q_s \sum_t \left\{ \xi_{st} \left[\sum_p \frac{w_{pk}^{st}}{\left(\sum_i w_{pi}^{st} s_i\right)^2} \right] \left[\sum_{t'} \xi_{st'} \left(\sum_q \frac{1}{\sum_i w_{qi}^{st'} s_i} - b^{st'} \right) - \sum_q \frac{1}{\sum_i w_{qi}^{st} s_i} + b^{st} \right] \right\}$$

Differentiating a second time:

$$\begin{aligned}
\frac{\partial^2 V}{\partial s_k \partial s_l} &= -2 \sum_{s,t} q_s \xi_{st} \left\{ \left[\sum_p \frac{2 w_{pk}^{st} w_{pl}^{st}}{\left(\sum_i w_{pi}^{st} s_i\right)^3} \right] \left[\sum_{t'q} \frac{\xi_{st'}}{\sum_i w_{qi}^{st'} s_i} - \sum_q \frac{1}{\sum_i w_{qi}^{st} s_i} + b^{st} - \bar{b}^s \right] + \right. \\
&\quad \left. \left[\sum_p \frac{w_{pk}^{st}}{\left(\sum_i w_{pi}^{st} s_i\right)^2} \right] \left[\sum_{t'q} \frac{\xi_{st'} w_{ql}^{st'}}{\left(\sum_i w_{qi}^{st'} s_i\right)^2} - \sum_q \frac{w_{ql}^{st}}{\left(\sum_i w_{qi}^{st} s_i\right)^2} \right] \right\} \\
&= -2 \sum_s q_s \left\{ \sum_t \xi_{st} \left[\Omega_{kl}^{st} \left(\bar{f}^s - f^{st} \right) - \Lambda_k^{st} \Lambda_l^{st} \right] + \bar{\Lambda}_k^s \bar{\Lambda}_l^s \right\}
\end{aligned}$$

3 Non-Degenerate Uniformly Distributed Variables

The subpixel sensitivities (s_i) that appear in the expression for the variance and its derivatives are not ideal, since they must average to 1, otherwise the overall pixel sensitivity is modified and that is separately taken care of by flat fielding.

One can define variables x_i ($i = 1 \ldots N-1$) which satisfy this automatically, and further if one wants the s_i values to be uniformly distributed over the volume allowed, x_i will each individually be uniformly distributed:

$$s_k = \prod_{i=0}^{k-1} (1 - x_i)^{\frac{1}{N-i}} \left[1 - (1 - x_k)^{\frac{1}{N-k}} \right]$$

In order to transform the derivatives of V with respect to s_i to derivatives with respect to x_i one needs to multiply by the following matrix:

$$J_{lk} \equiv \frac{\partial s_k}{\partial x_l} = \begin{cases} 0 & l > k \\ \frac{\prod_{i=0}^{k} (1-x_i)^{\frac{1}{N-i}}}{(N-l)(1-x_l)} & l = k \\ \frac{s_k}{(N-l)(1-x_l)} & l < k \end{cases}$$

The Newton–Raphson iterations then follow:

$$\Delta x = \gamma(\partial^2 V(x))^{-1}\, \partial V(x)$$

Where Δx is the change in the vector x, $(\partial^2 V(x))$ is a matrix where the k, l-th entry is $\frac{\partial^2 V}{\partial x_k \partial x_l} = J \cdot (\partial^2 V(s)) \cdot J^T$ evaluated at the given sensitivities, similarly $(\partial V(x)) = J \cdot (\partial V(s))$,

On each step γ starts out 1 and is halved until the updated value results in a smaller S than the current.