Parametric Gradient of the Chi-Squared Statistic

In *Forward Fitting* (or *MEM* or *PIXONS*) one often needs the gradient of the χ^2 statistic with respect to the model parameters. The χ^2 statistic for photon count rates observed ($\mathcal{C}_{obs}$) and modeled ($\mathcal{C}_{mod}$) is:

$$\chi^2 = \frac{1}{N} \sum_j \frac{[\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}]^2}{\sigma_j^2}, \tag{1}$$

where $\sigma_j^2 = \mathcal{C}_{mod}$ in the case of pure Poisson statistics.
An alternative form of the χ^2 statistic is for visibilities observed (V_{obs}) and modeled (V_{mod}) is:

$$\chi^2 = \frac{1}{2N} \sum_j \frac{|V_{mod}^{(j)} - V_{obs}^{(j)}|^2}{\sigma_j^2} \tag{2}$$

where the sum is over the index j of time bins, and the variance σ_j^2 can either be independent of the model parameters (as in radio astronomy) or can be proportional to the mean count rate ($\frac{1}{2}|V_{obs}|$).

1 Visibilities

For the case of visibilities, the gradient of χ^2 is:

$$\frac{\partial \chi^2}{\partial a_k} = \frac{1}{2N} \frac{\partial}{\partial a_k} \sum_j \frac{|V_{mod}^{(j)} - V_{obs}^{(j)}|^2}{\sigma_j^2} \tag{3}$$

It might be thought that evaluation of this expression is only possible numerically, but it is straightforward in many cases to write an analytic expression for (3). This will speed up the search through parameter space, and eliminates the insidious round-off errors commonly found in numerical differencing schemes.

To illustrate this, we consider the particular case:

1. σ_j^2 is not a function of the parameters a_k,
2. V_{mod} derives from a sum of Gaussian sources.

Both of these conditions can be relaxed, and similar analytic expressions can be found where σ_j^2 derives from Poisson statistics, and for more general basis functions than Gaussians. For the Orthogonal Gaussian model, the visibility is given by a sum over terms like:

$$V^{(j)}_{mod} = a_0 e^{-a_1 u_j^2 - a_2 v_j^2} e^{i a_3 u_j + i a_4 v_j}, \tag{4}$$

for sources of width $(\sqrt{a_1}, \sqrt{a_2})/\pi$, location $(a_3, a_4)/2\pi$, and flux a_0. For multiple Gaussians, the a_k parameters would take on an additional subscript n, and $V^{(j)}_{mod}$ would be a sum over n, but for simplicity we ignore that subscript in the following. The derivatives of $V^{(j)}_{mod}$ with respect to a_k are simple:

$$\frac{\partial}{\partial a_k} V^{(j)}_{mod} = f_k \; V^{(j)}_{mod} \tag{5}$$

where the vector $f = [1/a_0, -u_j^2, -v_j^2, iu_j, iv_j]$. Note that f has 3 real elements and 2 imaginary elements. This leads to simplifications in the derivatives below. Recalling our assumption that σ_j^2 is independent of a_k, the gradient of the numerator in (3) is:

$$\frac{\partial}{\partial a_k} |V^{(j)}_{mod} - V^{(j)}_{obs}|^2 = (V^{(j)*}_{mod} - V^{(j)*}_{obs}) f_k V^{(j)}_{mod} + (V^{(j)}_{mod} - V^{(j)}_{obs}) f_k^* V^{(j)*}_{mod} \tag{6}$$

The above expression is real because the 3rd and 4th terms on the RHS are the complex conjugates of the 1st and 2nd terms. If f_k is real, the RHS becomes:

$$RHS = f_k [2|V^{(j)}_{mod}|^2 - (V^{(j)*}_{obs} V^{(j)}_{mod} + V^{(j)}_{obs} V^{(j)*}_{mod})] \tag{7}$$

and if f_k is pure imaginary,

$$RHS = f_k \; [V^{(j)}_{obs} V^{(j)*}_{mod} - V^{(j)*}_{obs} V^{(j)}_{mod}] \tag{8}$$

So the equation for the gradient simplifies considerably. We can get a still simpler form by using the amplitude and phase of V_{obs} and by using the location parameter $a_3 \; u_j + a_4 \; v_j$ as a phase:

$$V_{obs} = |V_{obs}| \; e^{i\psi} \tag{9}$$

$$\phi \;=\; a_3 \; u_j + a_4 \; v_j \tag{10}$$

Then we find the gradient to be:

$$\frac{\partial\chi^2}{\partial a_k} = \frac{a_0}{N} \times \tag{11}$$

$$\sum_j \quad (f_k/\sigma_j^2)e^{-a_1u_j^2-a_2v_j^2}[a_0 - |V_{obs}|cos(\psi_j - \phi_j)] \quad k = 0, 1, 2 \tag{12}$$

$$\sum_j \quad (if_k/\sigma_j^2)e^{-a_1u_j^2-a_2v_j^2}[-|V_{obs}|sin(\psi_j - \phi_j)] \quad k = 3, 4 \tag{13}$$

These can be combined in the form:

$$\frac{\partial\chi^2}{\partial a_k} = \frac{a_0}{N}\ \mathbf{Re}\{\sum_j (f_k/\sigma_j^2)e^{-a_1u_j^2-a_2v_j^2}[a_0 - V_{obs}e^{-i(a_3u_j+a_4v_j)}]\} \tag{14}$$

It is important to use these equations starting with a_1 and a_2 very small, because there is a spurious zero to the gradient of χ^2 at $a_1 = \infty$ or $a_2 = \infty$.

2 Count Rates

The model count rate profile can be written as a function of the *phase at map center* Φ:

$$\mathcal{C}_{mod} = a_0\tau(1 + c_1\ cos\Phi_{jm}) \quad + \text{higher harmonics} \tag{15}$$

The phase term Φ_{jm} is the sum of the *phase at map center* Φ_j and the phase offset of a point source at $\Delta x, \Delta y$ relative to map center:

$$\Phi_{jm} = \Phi_j + 2\pi\ (u_j\ \Delta x + v_j\ \Delta y) \tag{16}$$

The variables (u_j, v_j) are the Fourier coordinates on the u, v circle for a given subcollimator. In practice, during an iteration of the reconstruction, Φ_j need be computed only once, while the offset terms change during the iteration. The fundamental amplitude c_1 for an orthogonal Gaussian source is given by:

$$c_1 = e^{-u_j^2a_1-v_j^2a_2} \cdot A_m, \tag{17}$$

where A_m is the modulated amplitude (modamp) obtained from the calibrated event list for a point source at map center.
For the case where the σ_j are independent of a_k, the derivatives of the χ^2 function with respect to the a_k are:

$$\frac{\partial\chi^2}{\partial a_0} = \frac{2}{N}\sum_j \frac{\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}}{\sigma_j^2}\mathcal{C}_{mod}/a_0 \tag{18}$$

$$\frac{\partial\chi^2}{\partial a_1} = \frac{2}{N}\sum_j \frac{\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}}{\sigma_j^2}a_0\tau(-u_j^2)cos\Phi_{jm} \tag{19}$$

$$\frac{\partial\chi^2}{\partial a_2} = \frac{2}{N}\sum_j \frac{\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}}{\sigma_j^2}a_0\tau(-v_j^2)cos\Phi_{jm} \tag{20}$$

$$\frac{\partial\chi^2}{\partial a_3} = \frac{2}{N}\sum_j \frac{\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}}{\sigma_j^2}(-a_0\tau c_1 u_j)sin\Phi_{jm} \tag{21}$$

$$\frac{\partial\chi^2}{\partial a_4} = \frac{2}{N}\sum_j \frac{\mathcal{C}_{mod}^{(j)} - \mathcal{C}_{obs}^{(j)}}{\sigma_j^2}(-a_0\tau c_1 v_j)sin\Phi_{jm} \tag{22}$$

$$\tag{23}$$

For the pure Poisson statistics case, $\sigma_j^2 = \mathcal{C}_{mod}$, which is a function of the parameters a_k, so χ^2 must be written as:

$$\chi^2 = \frac{1}{N}\sum_j\{\mathcal{C}_{mod}^{(j)} - 2\mathcal{C}_{obs}^{(j)} + [\mathcal{C}_{obs}^{(j)}]^2/\mathcal{C}_{mod}^{(j)}\}; \tag{24}$$

and the χ^2 gradient assumes the form:

$$\frac{\partial\chi^2}{\partial a_0} = \frac{2}{N}\sum_j\{1 - [\mathcal{C}_{obs}^{(j)}/\mathcal{C}_{mod}^{(j)}]^2\}\mathcal{C}_{mod}/a_0 \tag{25}$$

$$\frac{\partial\chi^2}{\partial a_1} = \frac{2}{N}\sum_j\{1 - [\mathcal{C}_{obs}^{(j)}/\mathcal{C}_{mod}^{(j)}]^2\}a_0\tau(-u_j^2)cos\Phi_{jm} \tag{26}$$

$$\frac{\partial\chi^2}{\partial a_2} = \frac{2}{N}\sum_j\{1 - [\mathcal{C}_{obs}^{(j)}/\mathcal{C}_{mod}^{(j)}]^2\}a_0\tau(-v_j^2)cos\Phi_{jm} \tag{27}$$

$$\frac{\partial\chi^2}{\partial a_3} = \frac{2}{N}\sum_j\{1 - [\mathcal{C}_{obs}^{(j)}/\mathcal{C}_{mod}^{(j)}]^2\}(-a_0\tau c_1 u_j)sin\Phi_{jm} \tag{28}$$

$$\frac{\partial\chi^2}{\partial a_4} = \frac{2}{N}\sum_j\{1 - [\mathcal{C}_{obs}^{(j)}/\mathcal{C}_{mod}^{(j)}]^2\}(-a_0\tau c_1 v_j)sin\Phi_{jm} \tag{29}$$

There is a singularity in these equations when $\mathcal{C}_{mod}^{(j)}] = 0$, and any (u, v) points for which this is true must be treated separately.

3 Cash Statistic

In the case of very low count rates, one must use what might be called the *Cash statistic* (Cash, 1979):

$$\chi^2_{Cash} = \frac{2}{N}\sum_{j=1}^{N}[n_j ln(\frac{n_j}{e_j}) - (n_j - e_j)], \tag{30}$$

where the n_j are integers, representing the observed counts per bin, and the e_j are arbitrary real numbers representing the model count rates. (In terms of previous variables, $n_j = \mathcal{C}_{obs}$ and $e_j = \mathcal{C}_{mod}$.) In the limit of large n_j and e_j, χ^2_{Cash} becomes the standard χ^2 statistic, and when $e_j \to n_j$, $\chi^2_{Cash} \to 0$. The gradient of the *Cash statistic* with respect to the model e_k is:

$$\frac{\partial\chi^2_{Cash}}{\partial e_k} = \frac{2}{N}(1 - n_k/e_k) \tag{31}$$

Just as in the Poisson version above (equation 25 - 29), there is a singularity where $e_k = 0$, and such bins must be treated separately. Note that for low count rates, the derivatives in equations 25 - 29 must have the coefficients $\{1 - [\mathcal{C}^{(j)}_{obs}/\mathcal{C}^{(j)}_{mod}]^2\}$ replaced by $2\{1 - [\mathcal{C}^{(j)}_{obs}/\mathcal{C}^{(j)}_{mod}]\}$, and these are equal when n_j is large.