THE FAILURE OF THE GRADIENT

It is obvious that when one of the c_k vanishes, $\mathbf{grad}(\chi^2)$ becomes independent of f_k, the corresponding model count rate. In any iteration that uses the gradient, the map pixels will be unconstrained by those observed count rates that vanish. But zero count rates are significant, so this is a serious (and heretofore ignored) problem for MEM at low count rates. If the likelihood function is used instead of χ^2, the term $(c_k/f_k)^2$ becomes (c_k/f_k), so the problem remains.

A possible fix may be to add a small scalar ϵ to all the observed count rates, which would restores the ability of the gradient to respond to changes in the complete model count rate f_k derived from the iterated model maps. (This has an undesirable consequence: the c_k are not all integers, and the likelihood function is not defined.) Whether this "fix" works remains to be seen.

Maimum Entropy Difficulties for Small and Moderate Flares

One of the most important parts of the Maximum Entropy algorithm is the computation of the grad χ^2 map. This is the gradient of the "goodness-of-fit" function χ^2 (or the equivalent likelihood function–Cash 1979, Sivia, 1996) with respect to the map pixels. In MEM, the way the gradient map appears is in the maximization of the following function (Sakao, 1993, Gary and Hurford, 199x):

$$S' = -\sum_{i=1}^{N} x_i ln(x_i) \; - \; \lambda\chi^2$$

where the first term on the RHS is the entropy (S) , and the second term multiplied by λ is the χ^2 statistic. For Poisson-distributed variables, the latter is given in terms of the observed count rates c_i and the model count rates f_i by:

$$\chi^2 = \sum_{j=1}^{M} (f_j - c_j)^2 / f_j$$

In order to find the maximum of S', we must take its derivatives with respect to the map pixels x_i. The gradient of χ^2 produces what is called the gradient map:

$$\frac{\partial \chi^2}{\partial x_i} = \sum_k \frac{\partial \chi^2}{\partial f_k} \frac{\partial f_k}{\partial x_i}.$$

The last derivatives on the right are the modulation constants M_{ik}, determined by the geometry of the imaging elements. The derivatives preceding them are:

$$\frac{\partial \chi^2}{\partial f_k} = 1 \; - \; (c_k / f_k)^2.$$

So the gradient map is given by the matrix product:

$$\mathbf{grad}(\chi^2) = \mathbf{M}(\mathbf{I} - \mathbf{v})$$

where $\mathbf{M}$ is the modulation pattern matrix, $\mathbf{I}$ is the identity matrix, and $\mathbf{v}$ is a vector whose elements are c_k^2 / f_k^2.