Proposal to add UV-plane model fitting to Difmap.

Martin Shepherd (mcs@phobos.caltech.edu)

1 Introduction

It is common practice in VLBI to parameterize maps by fitting simple models directly to the source UV data. These models may be used as plausible starting models for self-calibration, or to augment the CLEAN/self-calibration mapping loop by removing the kinds of extended structure that CLEAN has difficulty describing.

An existing program that has been used extensively in VLBI to fit models to UV data, is the modelfit program in the "Caltech VLBI Package" [2]. This program is capable of fitting models composed of aggregates of up to 15 simple components, each component being chosen from a list that includes delta-functions, elliptical gaussians and elliptical tapered-disks. The free parameters of the model are adjusted to minimize the value of χ^2 formed between the UV-plane representation of the model and the measured amplitudes and closure-phases. The use of closure phases rather than raw phase makes modelfit immune to station-based phase mis-calibrations. Modelfit does have some faults. Its two least-squares minimization algorithms are not optimal and for the best fits and fast convergence it relies on the coincidental match between the data and the arbitrary iteration step sizes used. The use of closure-phases also requires large amounts of memory and CPU time, amounts that grow as the cube of the number of stations in the data, times the number of integrations.

While Difmap still used the same data format as the rest of the Caltech VLBI package, no real need for a new model fitting algorithm was deemed necessary in Difmap. But now that Difmap uses FITS as its data format, observations that make use of the multiple dimensions that FITS allows but the Caltech merge format does not, can not be model fitted by the modelfit program. It is with this in mind that a new model fitting algorithm to be incorporated in difmap is to be written.

2 Requirements

Given that a new model fitting implementation is required, it would be short sighted to simply re-produce the algorithms used in the modelfit program without examining alternative techniques. One such technique which purportedly incorporates the strengths of both of the algorithms used in the modelfit program, and which has the extra advantage that step sizes are automatically determined, is the Levenberg-Marquardt method. This is described in [3] and will be adopted for the new implementation.

The Levenberg-Marquardt technique requires equations for the partial derivatives of the UV-plane model representation with respect to all free parameters, rather than determining them numerically as the modelfit gradient search algorithm does. The next section derives these equations.

Unlike the modelfit program, the new algorithm will be designed to fit models directly to the measured real and imaginary parts of the complex visibilities. The advantage gained is mainly one of speed, memory requirements and simplicity. In addition, the errors are more likely to be Gaussian distributed than say those of the measured amplitudes. The main disadvantage is that the program will not be immune to station based phase mis-calibration. However fringe-fitting, self-calibration and phase-referencing techniques have improved to the extent that this should not be a significant problem.

It has been decided that the new algorithm will fit component positions in a Cartesian rather than a polar coordinate system, although initial guesses will still be presented in the established polar form. This removes the problem of what to do if the model fitting algorithm produces negative radii. The similar problem of what to do if the sizes or aspect ratios of components go out of bounds remains un-answered.

A useful feature that does not exist in the modelfit program, is the estimation of the uncertainties in the fitted parameters. This is not as straight forward as might be expected, since the free parameters can turn out to be correlated. The full generality of this problem will be met in the new algorithm by optionally printing out the covariance matrix, thus leaving the onus on the user to interpret the results.

3 Partial derivatives of UV plane models.

Adopting the Levenberg-Marquardt algorithm requires a knowledge of the partial derivatives of the functions being fitted, calculated separately with respect to each free parameter. This section will serve to derive such equations.

A single model component of flux S, placed at the center of the image plane can be represented in the UV plane by a function of form:

$$F(u, v) = Sf(u, v) \tag{1}$$

Where (u, v) is a position in the UV plane. From the *shift theorem* of Fourier transforms we know that the same component whose centroid has been shifted to a position (x, y) in the image plane appears in the UV plane as:

$$F(u, v) = Sf(u, v)e^{2\pi i(ux+vy)} \tag{2}$$

By the *addition theorem* of Fourier transforms we also know that an aggregate model $M(u, v)$ can be formed from a sum of n individual components as:

$$M(u, v) = \sum_{i=1}^{n} S_i f_i(u, v)e^{2\pi i(ux_i+vy_i)} \tag{3}$$

The partial derivative of $M(u, v)$ with respect to a single free-parameter β of the k'th component is then given by:

$$\frac{\partial}{\partial \beta_k} M(u, v) = \frac{\partial}{\partial \beta_k} \sum_{i=1}^{n} S_i f_i(u, v)e^{2\pi i(ux_i+vy_i)} \tag{4}$$

$$\frac{\partial}{\partial \beta_k} M(u, v) = \frac{\partial}{\partial \beta_k} [S_k f_k(u, v)e^{2\pi i(ux_k+vy_k)}] \tag{5}$$

Given that for all model component types, $f(u, v)$ is independent of flux S and shifted position (x, y) we can determine the following partial derivatives for these parameters:

$$\frac{\partial}{\partial x_k} M(u, v) = 2\pi i u F_k(u, v) \tag{6}$$

$$\frac{\partial}{\partial y_k} M(u,v) = 2\pi i v F_k(u,v) \tag{7}$$

$$\frac{\partial}{\partial S_k} M(u,v) = \frac{1}{S_k} F_k(u,v) \tag{8}$$

All the supported model types can be represented by an azimuthally symmetric function, which is stretched into an elliptical shape along a specific direction. In the following equations the elliptical aspect will be parameterized by the major-axis half-extent, a, the direction of the major axis with respect to North, by the angle ϕ measured North through East, and by the minor-axis half-extent, b, or the alternative axial ratio $\gamma = b/a$.

For the purpose of model fitting we will choose to allow a, γ and ϕ to be free parameters. The alternative choice of a, b and ϕ is less useful, because there is then no way for the user to constrain components to have circular aspects, as required in some specialized problems. To fit these parameters requires equations of the partial derivatives of components with respect to them, in addition to those above for x, y, and S.

To give the components the desired rotated elliptical aspect, the value of the function at (x, y) in the image plane is assigned the value of the circularly symmetric function at radius:

$$R_{xy} = \sqrt{(x\cos\phi - y\sin\phi)^2 \frac{1}{\gamma^2} + (x\sin\phi + y\cos\phi)^2} \tag{9}$$

The Fourier transform of a circularly symmetric real function in the image plane is another circularly symmetric function in the UV plane. As in the image plane, the value of the rotated and stretched function at (u, v) in the UV plane is that of the circularly symmetric function measured at UV radius:

$$R_{uv} = \sqrt{(u\cos\phi - v\sin\phi)^2\gamma^2 + (u\sin\phi + v\cos\phi)^2} \tag{10}$$

Since each of the functions of interest multiplies this radius by πa, the following parameter will also be defined:

$$\Gamma = \pi a\sqrt{(u\cos\phi - v\sin\phi)^2\gamma^2 + (u\sin\phi + v\cos\phi)^2} \tag{11}$$

Partial derivatives of individual components that depend on Γ, with respect to one of a, γ, or ϕ (denoted as β) are then given by:

$$\frac{\partial}{\partial \beta_k} M(u,v) = \frac{\partial F_k(u,v)}{\partial \Gamma} \times \frac{\partial \Gamma}{\partial \beta_k} \tag{12}$$

If the latter part is evaluated for each of the desired free parameters, this expands to:

$$\frac{\partial}{\partial \phi_k} M(u,v) = \frac{\partial F_k(u,v)}{\partial \Gamma} \times \frac{2\pi^2 a_k^2}{\Gamma}(1-\gamma_k^2)(u\cos\phi_k - v\sin\phi_k)(u\sin\phi_k + v\cos\phi_k) \tag{13}$$

$$\frac{\partial}{\partial a_k} M(u,v) = \frac{\partial F_k(u,v)}{\partial \Gamma} \times \frac{\Gamma}{a_k} \tag{14}$$

$$\frac{\partial}{\partial \gamma_k} M(u,v) = \frac{\partial F_k(u,v)}{\partial \Gamma} \times \frac{\pi^2 a_k^2 \gamma_k}{\Gamma}(u\cos\phi_k - v\sin\phi_k)^2 \tag{15}$$

Further, since we know that neither the flux nor the position of the components depends on Γ, we can simplify the unknown $\frac{\partial F(u,v)}{\partial \Gamma}$ term to:

$$\frac{\partial F(u,v)}{\partial \Gamma} = S e^{2\pi i(ux+vy)} \frac{\partial f(u,v)}{\partial \Gamma} \tag{16}$$

and thus only $\frac{\partial f(u,v)}{\partial \Gamma}$ need be calculated individually for the different component types. This will be done in the following sections.

3.1 Gaussian components

The equation for the value of a gaussian component at position (u, v) in the UV plane is given in [2] as:

$$f(u,v) = e^{-\frac{\Gamma^2}{4 \ln 2}} \tag{17}$$

This has a partial derivative *wrt* Γ of:

$$\frac{\partial}{\partial \Gamma} f(u,v) = -\frac{2\Gamma}{4 \ln 2} f(u,v) \tag{18}$$

3.2 Uniformly Bright Disk components

The equation for the value of a uniformly bright disk component at position (u, v) in the UV plane is given in [2] as:

$$f(u,v) = 2 \frac{J_1(\Gamma)}{\Gamma} \tag{19}$$

Using a standard derivative from [1], the partial derivative of $f(u, v)$ *wrt* Γ is:

$$\frac{\partial}{\partial \Gamma} f(u,v) = -2 \frac{J_2(\Gamma)}{\Gamma} \tag{20}$$

3.3 Optically Thin Sphere

The equation for the value of a optically thin sphere component at position (u, v) in the UV plane is given in [2] as:

$$f(u,v) = \frac{3}{\Gamma^3} (\sin \Gamma - \Gamma \cos \Gamma) \tag{21}$$

The partial derivative of this function *wrt* Γ is:

$$\frac{\partial}{\partial \Gamma} f(u,v) = \frac{9 \cos \Gamma}{\Gamma^3} - \frac{9 \sin \Gamma}{\Gamma^4} + \frac{3 \sin \Gamma}{\Gamma^2} \tag{22}$$

3.4 Ring components

The equation for the value of a ring component at position (u, v) in the UV plane is given in [2] as:

$$f(u, v) = J_0(\Gamma) \tag{23}$$

Using a standard derivative from [1], the partial derivative of $f(u, v)$ *wrt* Γ is:

$$\frac{\partial}{\partial \Gamma} f(u, v) = -J_1(\Gamma) \tag{24}$$

4 Alternative parameterizations.

The above partial derivatives were coded up for use in a simple implementation of the Levensburg-Marquardt non-linear least-squares algorithm. It turned out that the parameterization of the elliptical part of the models had significant problems. In particular in equation 13 the partial derivative of the model *wrt* ϕ is zero when $\gamma_k = 1$. This means that the parameter can be changed by an infinite amount without changing the model. This results in zero column and row vectors in the Levensburg-Marquardt Hessian matrix, which in turn results in a singular matrix when inverted.

A better parameterization that is free of this problem was initially suggested by Steve Myers and formalized by Tim Pearson. The essential parts of Tim Pearson's suggestion are included below.

4.1 Reformulation of Model Fitting [Written by Tim Pearson]

I start from Martin's definition (11) of parameter Γ for an elliptical component with major axis a, minor axis $b = \gamma a$, and position angle ϕ:

$$\Gamma = \pi \sqrt{a^2 (u \sin\phi + v \cos\phi)^2 + b^2 (u \cos\phi - v \sin\phi)^2}. \tag{25}$$

Expanding this expression and collecting terms,

$$\Gamma^2/\pi = \cos^2\phi (a^2 v^2 + b^2 u^2) + \sin^2\phi (a^2 u^2 + b^2 v^2) + 2uv \cos\phi \sin\phi (a^2 - b^2). \tag{26}$$

Applying the double-angle formulae:

$$\begin{aligned} \sin 2z &= 2 \sin z \cos z, && (27) \\ \cos 2z &= 2\cos^2 z - 1 = 1 - 2\sin^2 z, && (28) \end{aligned}$$

we obtain

$$\begin{aligned} \Gamma^2/\pi &= (1 + \cos 2\phi)(a^2 v^2 + b^2 u^2)/2 + (1 - \cos 2\phi)(a^2 u^2 + b^2 v^2)/2 + uv \sin 2\phi (a^2 - b^2) && (29) \\ &= (a^2 + b^2)(u^2 + v^2)/2 + \cos 2\phi (a^2 - b^2)(v^2 - u^2)/2 + uv \sin 2\phi (a^2 - b^2) && (30) \\ &= (v^2 - u^2) X + 2uvY + (u^2 + v^2) Z, && (31) \end{aligned}$$

where I have defined

$$\begin{aligned} X &= (a^2 - b^2)\cos 2\phi/2, & (32)\\ Y &= (a^2 - b^2)\sin 2\phi/2, & (33)\\ Z &= (a^2 + b^2)/2. & (34) \end{aligned}$$

A point in the (X, Y, Z) system represents an ellipse. Circles $(a = b)$ are on the line $X = Y = 0$.

4.2 Partial derivatives.

We need the derivatives of

$$\Gamma = \pi\sqrt{(v^2 - u^2)X + 2uvY + (u^2 + v^2)Z} \quad (35)$$

with respect to the model parameters X, Y, Z:

$$\begin{aligned} \frac{\partial\Gamma}{\partial X} &= \frac{\pi^2}{2\Gamma}(v^2 - u^2), & (36)\\ \frac{\partial\Gamma}{\partial Y} &= \frac{\pi^2}{2\Gamma}(2uv), & (37)\\ \frac{\partial\Gamma}{\partial Z} &= \frac{\pi^2}{2\Gamma}(v^2 + u^2). & (38) \end{aligned}$$

4.3 Disadvantages

Use of (X, Y, Z) as variable parameters instead of (a, γ, ϕ) has the disadvantage that axial ratio and position angle cannot be constrained. Fixing Z is somewhat equivalent to fixing major axis. There are some non-physical domains in X, Y, Z space, e.g., $Z < 0$.

4.4 Advantages

X, Y, Z all have the same dimensions so equal increments in each parameter are likely to affect χ^2 by similar amounts. Components can be constrained to be circular by fixing $X = Y = 0$. Note that convolving two elliptical gaussians is equivalent to adding their X, Y, Z parameters—not that this has any relevance to model fitting.

References

[1] Abramowitz, M. and Stegun, I.A (eds) 1972 *Handbook of Mathematical functions* § 9.1.30, 9th printing, Dover publications

[2] Pearson, T.J. 1991 *Introduction to the Caltech VLBI Programs* § 7.2

[3] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1989, *Numerical Recipes*, Cambridge University Press.