Algorithms for use in Fringe Fitting.

I M Stewart

March 16, 2011

Abstract

Some algorithms for use in fringe fitting are described. Particular attention is paid to factors which constrain or limit the use of each algorithm.

CAVEATS!

- This report is *not finished.* I know for a fact that it contains mistakes. Therefore please don't dig into it line by line. It is only intended at this stage as a working version to show the kinds of directions in which I am thinking.
- IMO software without documentation is about as much use as documentation without software. Thus I have not spent 100% of my time so far on either. Thus they both may be expected to be in approximately the same state of incompleteness.
- I have not yet read all the past documentation on fringe fitting and so it is quite probable that I am reinventing some wheels here. However, (i) I will eventually read this stuff, so the wheel designs will converge; and (ii) I think there is virtue in trying to solve problems independently before consulting the wisdom of the ancients.

1 Introduction

The purpose of this report is to discuss some algorithms which it is proposed to employ in fringe fitting, in the process of calibrating radio interferometry measurements. Section 2 presents generic aspects of the theory; in section 3, the use of the 2-dimensional Fourier transform for fringe fitting is discussed; a linear-least-squares algorithm is presented in section 4.

2 Theory

2.1 Fringe visibilities

The generic expression for visibilities for a single baseline as a function of frequency ν and time t is

$$V(\nu, t) = A(\nu, t) \exp(i\phi[\nu, t]), \tag{1}$$

where A and ϕ are amplitude and phase respectively. As is well known in the interferometry community, measurements of source phases are invariably corrupted by random offsets introduced by the troposphere, the ionosphere and by the instrumentation itself. Such offsets must be estimated and removed before high-quality data can be obtained.

In the present report I am concerned solely with phase offsets due to the troposphere and the ionosphere. The former gives rise to a phase offset which may be characterized as

$$\phi_{\rm trop}(\nu, t) = 2\pi\nu\,\tau(t)$$

where τ, known as the delay, represents the time difference between the detection of the signal at the two antennas. Phase offsets due to the ionosphere are (to good approximation) inversely proportional to frequency (****citation, eg Kraus?):

$$\phi_{\rm iono}(\nu, t) = 2\pi K(t)/\nu$$

where K is proportional to the difference in ionospheric total electron content as seen by the two elements of the baseline. K/ν^2 and τ may be thought of as dispersive and non-dispersive delays respectively. Both K and τ are random functions, due to the natural variation in the properties of the atmosphere with time.

If the source contributions are divided out, and instrumental phase distortions removed, the amplitude in equation 1 becomes unity, and the phases reflect only the atmospheric delays. The complete expression for fringe visibilities is therefore

$$V(\nu, t) = \exp(i\phi[\nu, t])$$

where

$$\phi(\nu, t) = \phi_0 + 2\pi(\nu\,\tau[t] + K[t]/\nu). \tag{2}$$

The ϕ_0 term has been included for generality but may (I assume) usually be expected to be insignificant.

2.2 Parameterized delays

Since τ and K are non-trivial functions of time it is desirable to parameterize them in some way, for example as a weighted sum of basis functions $T_i(t)$:

$$\tau(t) \text{ or } K(t) \sim \sum_j^N \theta_j T_j(t).$$

The parameterization employed in the present report is the Taylor series, for which

$$\theta_j = \frac{1}{j!}\left.\frac{d^j\tau}{dt^j}\right|_0$$

and

$$T_j(t) = (t - t_0)^j$$

in respect of some reference time t_0. Keeping only terms to order 1, the expression 2 becomes approximated by

$$\phi(\nu, t) \sim \phi_0 + 2\pi(\nu\,[\tau\{t_0\} + \{t - t_0\}\dot{\tau}\{t_0\}] + [K\{t_0\} + \{t - t_0\}\dot{K}\{t_0\}]/\nu). \tag{3}$$

2.3 Real data

In practice what we have is not the continuous visibility function $V(\nu, t)$ but a rectangular grid of evenly-spaced samples $V_{j,k}$ of it, spanning finite intervals $\Delta\nu$ and Δt. For convenience I define additional quantities M, N for the numbers of samples respectively in the ν and t axes, and the 'pixel widths' $\delta\nu = \Delta\nu/M$ and $\delta t = \Delta t/N$.

Real data also comes with added noise values $n_{j,k}$. This noise is necessarily complex-valued. I make the following assumptions about it:

- There is no correlation between the noise in different samples.
- The real and imaginary parts of $n_{j,k}$ are uncorrelated for any (j, k).
- The probability distributions of all noise values are Gaussian.
- The real and imaginary parts have the same standard deviation $\sigma_{j,k}$.

2.4 Values returned by the algorithms

Algorithms used in fringe fitting will be expected to return the value of t_0 employed, as well as estimates of the following quantities:

- ϕ_0
- $\tau(t_0)$
- $\dot{\tau}(t_0)$
- $K(t_0)$
- $\dot{K}(t_0)$

3 The 2D Fourier method

The purpose of this section is to investigate the circumstances in which a two-dimensional Fourier transform of a section of delay-fringe visibilities for a given baseline returns useful values of the quantities listed in section 2.4.

3.1 Centred, dimensionless coordinates

It is convenient here to make the substitutions $x = (\nu - \nu_{\mathrm{mid}})/\Delta\nu$ and $y = (t - t_{\mathrm{mid}})/\Delta t$, where ν_{mid} and t_{mid} are respectively the mean frequency and time of the array of visibility samples. With this substitution, equation 3 becomes converted to the following centred, dimensionless form:

$$\phi(x, y) = \phi_{\mathrm{mid}} + 2\pi(G_x x + G_y y + G_{x,y} xy + \phi_K[x, y]) \tag{4}$$

where

$$\phi_{\mathrm{mid}} = \phi_0 + 2\pi\nu_{\mathrm{mid}}\tau_{\mathrm{mid}}$$

$$G_x = \Delta\nu\,\tau_{\mathrm{mid}}$$

$$G_y = \nu_{\mathrm{mid}}\Delta t\,\dot{\tau}_{\mathrm{mid}}$$

$$G_{x,y} = \Delta\nu\Delta t\,\dot{\tau}_{\mathrm{mid}}$$

and

$$\phi_K(x,y) = \frac{K_{\rm mid} + y\Delta t\,\dot{K}_{\rm mid}}{x\Delta\nu + \nu_{\rm mid}}.$$

For the sake of a simpler expression, t_0 has here been set equal to $t_{\rm mid}$; the notation $\tau_{\rm mid}$ for $\tau(t_{\rm mid})$ etc has also been employed.

3.2 The ideal, 'planar' world

Suppose now, instead of equation 4, the fringe phases were given by the following 'planar' expression:

$$\phi'(x,y) = \phi_{\rm mid} + 2\pi(G'_x x + G'_y y).$$

A 2-dimensional Fourier transform F of a 'planar' visibility function

$$V'(x,y) = \exp(i\phi'[x,y]) \text{ for } x,y \in [-0.5,0.5],\ = 0 \text{ else},$$

would result in

$$F(\omega,\zeta) = z \times \delta(\omega - G'_x, \zeta - G'_y) \star E(\omega,\zeta)$$

where z is a simple phase term given by

$$z = \exp(i\phi_{\rm mid})$$

and E is the 'envelope' function

$$E(\omega,\zeta) = \mathrm{sinc}(\omega)\,\mathrm{sinc}(\zeta).$$

The star represents convolution. ***** *Factor of 2 error?*

An algorithm for determining $\phi_{\rm mid}$, G'_x and G'_y suggests itself as follows:

1. Perform the 2D FT.
2. Create a 'power spectral image' $q = FF^*$.
3. Search for the position $(\omega_{\max}, \zeta_{\max})$ of the maximum of q. This position gives (G'_x, G'_y).
4. Obtain z from $F(\omega_{\max}, \zeta_{\max})$ and thus $\phi_{\rm mid}$.

In reality, the envelope function may be more complicated than such a simple product of symmetrical, real-valued functions. If we use instead of the simple 'planar' expression the more complicated (but still approximate) expression for phase given in equation 4, we have

$$E(\omega,\zeta) = \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} dx\ dy\ \exp(-i[\omega x + \zeta y])\ \exp(i2\pi[G_{x,y}xy + \phi_K\{x,y\}]). \qquad (5)$$

Under certain circumstances, the shape of E may become so distorted that its maximum value will no longer be anywhere near (G_x, G_y). It is clearly important to explore such circumstances. This is done for two different approximations in sections 3.6 and 3.5.

Note however that there is no intention at present to induce the Fourier algorithm to produce any 'non-planar parameters', via measuring the distortion of the peak or by any other method. This seems to go against the 'quick and dirty' intention of this algorithm.

3.3 Real data

As mentioned already in section 2.3, what we have is not a continuous function $V(\nu, t)$, but a grid of samples of V, with added noise. The closest we can come to a 'pure' Fourier transform is the discrete Fourier transform. This can be made to approximate the 'pure' transform arbitrarily closely, however, if we transform not the raw array $V_{j,k}$ but this array embedded in a larger array which is elsewhere filled with zeros (****citation - Bracewell?).

Another modification that is useful to make before venturing on the transform is to cyclic-shift the zero-padded array such that the pixel nearest to the central frequency and time values $\nu_{\rm mid}$ and $t_{\rm mid}$ is shifted to location (0,0). If this is not done, the envelope function E will be modulated by a phase spiral. This phase will not appear in the power spectral image, thus it will have no effect on locating the peak, but it has the potential to disrupt the determination of $\phi_{\rm mid}$.

If either of the dimensions M and N of the original unpadded array are even, it will be impossible to centre the non-zero values about the (0,0) pixel. Some phase slope in E will thus be present. Provided however that $M, N \gg 1$, the effect is unlikely to be severe.

3.3.1 Masking and weighting

A more problematical case may occur if some of the pixels of $V_{j,k}$ are flagged as bad, or otherwise masked from consideration. The simplest way to implement such masking in the present algorithm is to set the appropriate values of V to zero before performing the transform. An alternative procedure would be simply to multiply all $V_{j,k}$ by their statistical weights (usually inversely proportional to the variance at that pixel) before transforming (which is a good thing to do anyway, since it reduces the noise power to minimum - ****citation??), and express the mask via a zero weight value. Either procedure will however in general modify the position, phase and amplitude of E. A way to compensate for this might be to also transform an array of the weights - analogous to the formation of the dirty beam in imaging - and use the result to precalculate the offsets.

3.3.2 Noise power

If the transform is performed on pure noise then the average value $\eta =< q_{\rm noise} >$ of the power spectrum image is

$$\eta = \frac{2}{(MN)^2} \sum_j^M \sum_k^N \left(\sigma_{j,k}^2\right).$$

In the case that the standard deviations all have the same value σ, this becomes

$$\eta = \frac{2\sigma^2}{MN}. \tag{6}$$

***** *Include effects of weights as in section 3.3.1.* ***** *Inconsistent notation, y vs q.*

The probability distribution $p(y)$ of such values is given by a suitably scaled χ^2 distribution with 1 degree of freedom:

$$p(y) = \frac{e^{-y/2\eta}}{\sqrt{y\eta}}.$$

The frequency f with which y will exceed a certain value $y_{\max}$ in the set of visibility samples is given by $M \times N$ times the corresponding survival function (a.k.a. complementary cumulative distribution):

$$f(y > y_{\max}) = MN\mathrm{erfc}\left(\sqrt{\frac{y_{\max}}{2\eta}}\right).$$

If the noise variance is even across the chunk of visibilities, equation 6 gives

$$f(y > y_{\max}) = MN\mathrm{erfc}\left(\sqrt{\frac{MNy_{\max}}{4\sigma^2}}\right).$$

Since $x\ \mathrm{erfc}(\sqrt{x})$ decreases (after passing a maximum at about $x = 0.67$) asymptotically as $\sqrt{x}\exp(-x)$, it is clear that one should always use the maximum number of visibility samples consistent with the constraints laid out in sections 3.5 and 3.6.

3.4 Centroiding

3.4.1 The algorithm

We want to obtain as accurately as possible the position of the peak in the power spectrum image. It would also be nice to obtain an estimate of the amplitude and phase of the delay image at this place. The way this is done here is as follows. Suppose we isolate a 3×3 patch of power spectrum image values centred on our initial coarse guess $(j_{\max}, k_{\max})$ at the pixel with the maximum value. Let us label the values $q_{j,k}$ where j and k here both run from -1 to 1. Let us assume that these power spectrum image values may be approximated by a product of two parabolas, viz:

$$q_{j,k} \sim A^2(1 - a[j - \nu_0]^2)(1 - b[k - t_0]^2).$$

The necessary coefficients are obtained by separate 1-dimensional parabolic fits as follows. Let us first sum together the values in each rown of z, which gives:

$$Q_j = A^2(1 - a[j - \nu_0]^2)(3 - b[2 + 3t_0^2]).$$

ν_0 is given by

$$\nu_0 = \frac{1}{2}\ \frac{Q_1 - Q_{-1}}{2Q_0 - Q_{-1} - Q_1}.$$

A similar procedure after adding columns of q will produce t_0.

**** *Mention obtaining amp, phi0 by fitting half-curvature hats to the real and imag delay images.*

Note that, although centroiding can proceed perfectly well without any prior knowledge in the case that the noise power is negligible (see section 3.4.2), if that is not the case, a prior expectation of the location of the peak can be useful. I have not yet investigated such a Bayesian approach however.

3.4.2 Uncertainties

By a bit of propagation of variables one can show that the uncertainty σ_ν in the 'frequency' position coordinate is given by

$$\sigma_\nu^2 = \frac{3\sigma_q^2}{(Q")^2}\left(6\nu_0^2 + \frac{1}{2}\right)$$

where Q" $\sim Q_{-1} + Q_1 - 2Q_0$. A similar relation holds for the 'time' coordinate. σ_q is the standard deviation of the noise power... which might be tricky to calculate... also considering that these noises are correlated in the FT of a padded image. Hmm. Maybe do some Monte Carlos to get a conversion factor..?

3.5 Non-dispersive delay

If the ionospheric contribution can be neglected, equation 5 for the envelope function becomes

$$E(\omega, \zeta) = \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} dx\ dy\ \exp(-i[\omega x + \zeta y])\ \exp(i2\pi G_{x,y} xy). \tag{7}$$

where $G_{x,y} = \Delta\nu\Delta t\dot{\tau}_{\rm mid}$. In the limit that $G_{x,y} \to 0$, E is just the product of sinc functions mentioned in section 3.2. The magnitude $|E(0,0)|$ of its central value is equal to unity.

The integral in equation 7 can be solved by making use of the result (*****citation of Gradshteyn and Rhyzik)

$$\int \frac{\exp(cx)\ dx}{x} = \ln|x| + \sum_{n=1}^{\infty} \frac{(cx)^n}{n \times n!}.$$

It can thence be shown that the power spectrum peak height is approximately given by

$$y_{\rm max} = E^2(0,0) \sim 1 - \frac{(2\pi\Delta\nu\Delta t\dot{\tau}_{\rm mid})^2}{144}.$$

Distortions in the shape of the peak become relatively severe for $y_{\rm max} < 0.5$. Note however that, because the kernel in equation 7 is symmetrical, there is no offset in either the position or phase of the peak.

3.6 Dispersive delay

**** *Unfinished - need to rework.*

If we now include the influence of the ionosphere then the expression for phase in equation 2.1 becomes

$$\phi(\nu, t) = \phi_0 + 2\pi[\nu\ \tau(t) + K/\nu]$$

where K is proportional to the difference in ionospheric total electron content as seen by the two elements of the relevant baseline.

*** FT is like fitting a plane by least squares. Slice through this at given time is a line $a + b(\nu - \nu_{\rm mid})$ which minimizes

$$S = \int_{\nu_{\rm mid}-\Delta\nu/2}^{\nu_{\rm mid}+\Delta\nu/2} d\nu\ (a + b(\nu - \nu_{\rm mid}) - K/\nu)^2.$$

Change of variable $x = (\nu - \nu_{\rm mid})/\Delta\nu$ and setting $\partial S/\partial a$ and $\partial S/\partial b$ to zero gives

$$a = \frac{K}{\Delta\nu} \ln\left(\frac{\nu_{\rm mid} + \Delta\nu/2}{\nu_{\rm mid} - \Delta\nu/2}\right)$$

and

$$b = \frac{1}{(\Delta\nu)^2}(K - a\nu_{\rm mid}).$$

For simplicity I will in the present section ignore the time axis and thus consider only a 1-dimensional Fourier transform of the frequency axis. The first step in the analysis is to expand the K term in a Taylor series about the central frequency $\nu_{\rm mid}$. Up to order 2 this gives us

$$\phi(\nu) \sim \phi_0 + 2\pi \left(\nu\ \tau + K \left[\frac{1}{\nu_{\rm mid}} - \frac{\{\nu - \nu_{\rm mid}\}}{\nu_{\rm mid}^2} + \frac{\{\nu - \nu_{\rm mid}\}^2}{\nu_{\rm mid}^3} \right] \right).$$

After shifting the frequency coordinate $\nu \rightarrow \nu - \nu_{\rm mid}$, the resulting phase terms of order 0, 1 and 2 in frequency are

$$\phi_0 + 2\pi \left(\nu_{\rm mid}\, \tau + \frac{K}{\nu_{\rm mid}} \right),$$

$$2\pi\nu \left(\tau - \frac{K}{\nu_{\rm mid}^2} \right),$$

$$2\pi\nu^2 \frac{K}{\nu_{\rm mid}^3}.$$

The second of these shows that the position of the peak on the power spectrum will be offset by $2\pi K/\nu_{\rm mid}^2$; the third controls the shape of the peak. This envelope function $E(W)$ will be given by

$$E(W) = \frac{1}{\Delta\nu} \int_{-\Delta\nu/2}^{\Delta\nu/2} d\nu\ \exp(-iW\nu)\ \exp\left(2\pi i \nu^2 \frac{K}{\nu_{\rm mid}^3} \right).$$

This can be evaluated in terms of the Fresnel integrals S and C. For $W = 0$,

$$E(0) = C\left(\frac{\Delta\nu}{2} \sqrt{\frac{2\pi K}{\nu_{\rm mid}^3}} \right) \sim 1 - \frac{\Delta\nu^4 \pi^2 K^2}{40 \nu_{\rm mid}^6}.$$

There is something not quite correct about this analysis, because although the deviation in the height of the peak in the power spectrum should be approximately twice the value above, in fact numerical experiments indicate it is about 10% smaller. A factor of root(pi) on two is consistent with the mathematical picture and gives about the right correction.

The dependence on $\Delta\nu^4$ seems to hold however.

4 The linear algorithm

I refer again to the general exression for fringe visibilities at equation 1. If, over the region of interest, the phase $\phi(\nu, t)$ is small, and the amplitude $|A|$ close to 1, then to good approximation we may write

$$\phi \sim Im(V).$$

Where this holds, we can just fit to the imaginary values of $V_{j,k}$ and ignore the reals altogether. Shortly I will show that ϕ is a linear function of all the parameters on interest. These parameters may therefore be obtained via linear least squares, thus offering a potentially speedy way to obtain the delay and delay rate.

Usually the phase is not 'small' but can on the contrary be expected to make several cycles through 2π over the breadth of the data - which is what necessitates fringe fitting in the first place. The present method can only work if the bulk of the phase variation is subtracted before fitting is attempted. Thus this method absolutely requires a prior guess at the delay and delay rate.

There are various forms which such prior information could take, but it seems to me to be sensible at this stage to consider only two:

- All five parameters listed in section 2.4.
- Just 'planar' parameters, for example those in equation 4: $\phi_{\rm mid}$, G_x and G_y.

In fact at present I have only implemented the 'planar' correction. This is implemented by dividing the input visibilities by

$$\exp(i\phi_{0,\rm est})\exp(2\pi i\nu\,\tau_{\rm est})\exp(2\pi i\nu_0[t-t_0]\,\dot\tau_{\rm est}).$$

The suffix 'est' here represents the estimates. The benefit of this multiplicative form is that only $M+N+2$ trig function evaluations are required instead of $O(M\times N)$ if the full 5-parameter correction is performed.

The phase residual after subtraction of a planar prior estimate can be written as

$$\Delta\phi(\nu,t)=\phi_0+2\pi(\nu\,[\tau_0+\{t-t_0\}\dot\tau_0]+[K_0+\{t-t_0\}\dot K_0]/\nu)-\phi_{0,\rm est}-2\pi\nu\,\tau_{\rm est}-2\pi\nu_0(t-t_0)\,\dot\tau_{\rm est}.$$

This may be expressed as

$$\Delta\phi(\nu,t)=g(\nu,t)+\sum_{m=1}^{5}q_m h_m(\nu,t).$$

Here g is

$$g(\nu,t)=-2\pi\nu_0(t-t_0)\,\dot\tau_{\rm est}$$

and the five pairs of qs and hs are:

m	q_m	$h_m(\nu,t)$
1	$\phi_0-\phi_{0,\rm est}$	1
2	$2\pi(\tau_0-\tau_{\rm est})$	ν
3	$2\pi\dot\tau_0$	$(t-t_0)\nu$
4	$2\pi K_0$	$1/\nu$
5	$2\pi\dot K_0$	$(t-t_0)/\nu$

The expression for χ^2 is

$$\chi^2=\sum_{j=1}^{M}\sum_{k=1}^{N}\frac{(Im[V_{j,k}]-\Delta\phi_{j,k})^2}{\sigma^2_{j,k}}.$$

Setting the derivatives wrt the qs to 0 in the standard way gives

$$\mathbf{Aq}=\mathbf{b}, \tag{8}$$

where

$$A_{l,m}=\sum_{j=1}^{M}\sum_{k=1}^{N}\frac{h_l(\nu_j,t_k)\,h_m(\nu_j,t_k)}{\sigma^2_{j,k}}$$

and

$$b_l=\sum_{j=1}^{M}\sum_{k=1}^{N}\frac{(Im[V_{j,k}]-g[\nu_j,t_k])\,h_l(\nu_j,t_k)}{\sigma^2_{j,k}}.$$

Equation 8 (equivalent to the normal equations for this problem) is simply inverted to give the coefficients q_m. Further (see e.g. *********Press et al chapter 15.4), $\mathbf{A}^{-1}$ is the covariance matrix of the q_m.

4.1 Notes:

1. It is well known that the normal equations can be ill-conditioned. What that means in practice is that the off-diagonal elements of $\mathbf{A}$ may be relatively large, implying that two of the coefficients are highly correlated. This means that a large increase in one coefficient, matched with a proportionately large decrease in the other, may have only a small effect on χ^2. In these circumstances a singular value decomposition (SVD) is recommended. A measure of the degree of ill-conditioning is the condition number:

$$\kappa = ||\mathbf{A}||\,||\mathbf{A}^{-1}||.$$

   $\mathbf{A}$ is considered ill-conditioned if $\ln(\kappa) >$ number of sig figures of the matrix entries.

2. For this form of model (ie, linear in the coefficients), the Levenberg-Marquardt method gives no advantage because it involves solving exactly the same matrix equation. (I think.*******)

3. The ill-conditioning of the matrix could be relieved by an appropriate shearing transform. However this would require definition of new parameters q'_m which would be related to the old via a matrix equation. The ill-conditioning would be transferred to this new matrix and the upshot would be that each calculation of q_m would require an error-amplifying subtraction between two or more values of q'_m. The ill-conditioning is actually a feature of the way we have chosen to parameterize the delays.

Notice finally that the matrix $\mathbf{A}$ is the same for any same-sized block of visibilities, provided that the frequency range is also unchanged, and that t_0 maintains the same relationship to the start and end times of the block. Therefore $\mathbf{A}$ (also $\mathbf{A}^{-1}$) only needs to be calculated once for the whole set of time samples. This speeds things up a lot. (In fact it is possible that judicious divisions by M and N could make A the same for any chunk of data.*****)

**** *Limits.*