Algorithms for use in Fringe Fitting.

I M Stewart   S Bourke

21 May 2012

Abstract

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

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 1-dimensional Fourier transform for fringe fitting is discussed, whereas the extension to 2 dimensions is discussed in section 4; a linear-least-squares algorithm is presented in section 5.

2   Theory

2.1   Fringe visibilities

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

$$V(\nu, t) = |G|(\nu, t) \exp(i\phi[\nu, t]) \times |A|_{\mathrm{src}}(\nu, t) \exp(i\phi_{\mathrm{src}}[\nu, t]), \tag{1}$$

where $G = |G| \exp(i\phi)$ represents the complex gain of the instrument for this baseline and $|A|_{\mathrm{src}}$ and ϕ_{src} are respectively the amplitude and phase of the source. In the present report I am going to ignore any variation with the source component across the data range and thus assume that the source contribution can be approximated by a single complex scalar value A_{src}. The justification for this is that it is usual in self-calibration to divide out a source model before attempting to estimate G. The remaining source contribution is in theory therefore simply unity. In practice, although it seems reasonable to assume that this remainder will be slow-varying and perhaps close to unity in amplitude, it seems prudent to retain some latitude for a source residual phase and amplitude.

It will be convenient to merge the remaining source phase offset into the instrumental phase variation $\phi(\nu, t)$ and to combine the remaining source amplitude $|A|_{\mathrm{src}}$ with the amplitude gain $|G|$ into a single (real-valued) product S. With these modifications, equation 1 becomes

$$V(\nu, t) = S \exp(i\phi[\nu, t]). \tag{2}$$

In the present report I am concerned solely with phase offsets due to two different causes, which may be labelled respectively as non-dispersive and dispersive delay times between the reception of signals from the two receivers at the correlator. Non-dispersive delays, primarily due to either clock delays or path length differences within the troposphere, give rise to a phase offset which is proportional to frequency:

$$\phi_{\mathrm{trop}}(\nu, t) = 2\pi\nu T(t).$$

1

Here T is the delay time. Dispersive delays are caused primarily by the ionosphere and produce phase offsets which are (to good approximation) inversely proportional to frequency:

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

where K is proportional to the difference in ionospheric total electron content as seen by the two antennas. K/ν^2 has units of time and may be thought of as the dispersive delay. Both K and T are random functions of time, due to the natural variation in the properties of the atmosphere with time. The complete expression for instrumental phase ϕ is therefore

$$\phi(\nu, t) = \phi_0(t) + 2\pi\left(\nu T[t] + K[t]/\nu\right). \tag{3}$$

The ϕ_0 term has been included for generality. If different frequency standards are used for different antennas, its time variation may be significant.

2.2 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 $\sigma_{j,k}$.

Note that it is NOT assumed that the standard deviation is the same from sample to sample. This is very generally untrue and is the main reason applying weights to the data.

2.3 Inputs and outputs

The input to the fitting routine comprises the entire set of visibilities (I assume the source model has already been divided out), optionally also information from previous fits or other prior knowledge. The set of visibilities is assumed to be divided up into chunks covering a certain time and frequency range, as outlined in the preceding subsection. The entire observation is thus divided up into a 3-dimensional array of such chunks. The observation duration and bandwidth supply two of the dimensions: a third dimension runs over either baseline or antenna. For each baseline/antenna, each chunk in the corresponding 2D array comprises a regular array of complex visibilities. I'll usually assume that we also have a matching array of data weights, each value of which can usually be expected to be inversely proportional to the variance (square of the standard deviation) of the data in the matching pixel. The fitting algorithm should take one of these chunks at a time and fit a phase model to it.

The initial output from the fitting routine should therefore be a 3D array of phase-model parameter values, 1 set of values per chunk. Additional matching arrays of information about the fit quality and precision may also be returned. It may be useful to post-process the parameter values in the frequency direction, for example to perform further fitting so as to transform Taylor components into fundamental delay values; and/or in the time direction, to smooth out measurement noise (some of the noise will be 'real', i.e. produced by variations in the atmosphere, not just measurement error). See section 2.6 for this.

2.4 Model fitting generalities

Fringe fitting or finding involves fitting a model of the phase variation due to delays to the data. This common situation presents several issues to be settled.

1. Selection of an appropriate model. Several different phase models are discussed in section 2.5. Any phase model will unavoidably diverge from the real phases given a large enough field of data. This argues for reducing the amount of data fitted, in contrast to point 2.

2. Finding a global minimum of the fitting function in a field rich in local minima. Additional local minima are often unavoidable when the model to be fitted is not linear in all parameters. The global maximum deepens with more data - this argues for increasing the amount of data fitted, in contrast to point 1.

3. Convergence criteria for the fit. For this it is good to keep in mind that precision in phase of several degrees is usually good enough for calibration of interferometer data.

4. Precision of the fitted parameters. At a minimum, a covariance matrix is required.

5. How good is the model? I prefer a Bayesian Odds Ratio to decide between the model fit and a fit using a null model, simply because it is very general, but reduces to a likelihood ratio in the absence of priors. However, simple p-value tests are often perfectly adequate (see section 3.4 for example).

2.5 Phase models

There are two related issues to consider when choosing a phase model: (i) how large the fractional bandwidth is, and (ii) how large the time window is.

2.5.1 Fractional bandwidth

The behaviour of phase with respect to frequency is quite regular, see for example equation 3; but if the fractional bandwidth is small, equation 3 is not a good way to parameterize the phase, because the covariance between the parameters ϕ_0, T and K becomes large. About the best that can be done in this case is to fit something like a polynomial of order not greater than 2 - e.g. a Taylor expansion of equation 3 about some reference frequency ν_{ref}. In this case the parameters are the derivatives of ϕ with respect to frequency, evaluated at ν_{ref}. The fundamental delay parameters ϕ_0, T and K can then be evaluated (although with uncertainty values which may become very large at low fractional bandwidth) from the following expressions:

$$\phi_0 = \phi_{\mathrm{ref}} - \nu_{\mathrm{ref}}\phi'_{\mathrm{ref}} + 0.5\nu_{\mathrm{ref}}^2\phi''_{\mathrm{ref}}$$

$$T = \frac{1}{2\pi}\left(\phi'_{\mathrm{ref}} + 0.5\nu_{\mathrm{ref}}\phi''_{\mathrm{ref}}\right)$$

$$K = \frac{\nu_{\mathrm{ref}}^3}{4\pi}\phi''_{\mathrm{ref}}.$$

2.5.2 Size of time window

The variation of the phase over time is in principle not regular, since part of the delay difference at the antennas is due to different paths through the atmosphere; and because the atmosphere is a natural system, one expects its properties to vary with time in a random fashion, with the noise spectrum of the variation probably following some sort of inverse power law. Many natural systems exhibit such behaviour and the refractive index for radio waves in the atmosphere is found to be no exception. The red-noise behaviour of the phase signal means that, if the time window is small enough, the phase behaviour may be well approximated by a low-order expansion in an appropriate set of basis functions.

Three treatments are suggested here:

1. Solve only in the frequency domain for each time sample. In this case the resulting time series in ϕ_0, T and K (or any other frequency parameterisation) will probably be dominated by measurement noise and could benefit from post-processing to smooth this out (see section 2.6). This is likely to be possible only for very strong sources, since the amount of data in the fit is minimal.

2. Low-order Taylor expansions about some reference time of the separate variations of ϕ_0, T and K.

3. Low-order joint time-and-frequency Taylor expansion of phase variation.

2.6 Post-processing

3 The 1D Fourier method

The purpose of this section is to investigate the circumstances in which useful values of the delay quantities can be obtained from a one-dimensional Fourier transform of data from a single time sample (or perhaps of several time samples, averaged down to one).

3.1 Theory

Supposing we take the basic expression for fringe visibilities in equation 2, discard the time dependence, and expand the phase in a Taylor series about some reference frequency ν_{ref}, keeping just the linear terms:

$$\phi(\nu) \sim \phi_{\text{ref}} + (\nu - \nu_{\text{ref}})\, \dot{\phi}_{\text{ref}}. \tag{4}$$

(Here I have switched to a dot notation for the derivative since I want to use primes for a different purpose.) We then multiply by a (real-valued, normalized) weight function $w(\nu)$ which is zero outside the bandwidth $[\nu_{\text{lo}}, \nu_{\text{hi}}]$. The Fourier transform of the result is

$$\mathfrak{V}(\tau) = S \ \exp(i[\phi_{\text{ref}} - \dot{\phi}_{\text{ref}}\nu_{\text{ref}}])\, \delta(\dot{\phi}_{\text{ref}} - 2\pi\tau) \star \epsilon(\tau) \tag{5}$$

where τ is the lag-time coordinate; S is the source residual amplitude; the star represents convolution; and ϵ, which I call the envelope function, is given by the Fourier transform of the normalized weight function:

$$\epsilon(\tau) = \frac{1}{W} \int_{-\infty}^{\infty} d\nu \, w(\nu) \, \exp(-2\pi i\nu\tau) \tag{6}$$

where

$$W = \int_{-\infty}^{\infty} d\nu \, w(\nu).$$

In the simple case that w equals $1/(\nu_{\text{hi}} - \nu_{\text{lo}})$ between ν_{lo} and ν_{hi}, ϵ is given by

$$\epsilon(\tau) = \exp(-\pi i\tau[\nu_{\text{hi}} + \nu_{\text{lo}}]) \, \text{sinc}(\tau[\nu_{\text{hi}} - \nu_{\text{lo}}]). \tag{7}$$

Here I have used the Bracewell definition for sinc:

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

The final step is to calculate the 'power spectrum' q of $\mathfrak{V}$. The phase term disappears and the result is

$$q(\tau) = \mathfrak{V}\mathfrak{V}^* = S^2 \delta(\dot{\phi}_{\text{ref}} - 2\pi\tau) \star \epsilon\epsilon^*(\tau). \tag{8}$$

Note that since $w(\nu)$ is real-valued, ϵ is Hermitian, and thus $\epsilon\epsilon^*$ is not only real-valued but also in this instance symmetrical. In practical terms this means that the maximum value of $\epsilon\epsilon^*$ occurs at lag time $\tau = 0$, and that $\epsilon\epsilon^*(0) = \epsilon^2(0) = \epsilon_0^2$.

The overall strategy then is two-fold: firstly to find the location τ_{max} of the maximum in the power spectrum $q(\tau)$, which from equation 8 gives us $\dot{\phi}_{\text{ref}}$ and $S^2\epsilon_0$; then to deduce ϕ_{ref} from the phase of $\mathfrak{V}$ at τ_{max}.

The first part of this, finding the peak in the power spectrum, is reasonably straightforward, even when we move in section 3.2 to consider the situation of real data, which necessitates a discrete Fourier transform. Practical implementation of the second part however could be problematic, because inspection of the example envelope function given in equation 7 shows that $\epsilon(\tau)$ is not likely to be a well-behaved function if the fractional bandwidth is low, because in this case the slowly-varying, sinc-like envelope is going to be modulated by a rapid phase spiral. With noisy, discrete real data it is an advantage to make use of several samples of ϵ around the maximum in the power spectrum *actually no - because they are correlated. The advantage is illusory. However we still may be as much as 1/2 pixel away from the maximum so reducing the phase winding is still good.*. Rapid phase winding can render this impractical.

The root of the problem is the first derivative of ϵ with respect to lag time τ. A rapid phase wind equates to a large value for this derivative. From the Fourier-transform definition of ϵ in equation 6, this derivative is given by

$$\frac{d\epsilon}{d\tau} = \frac{1}{W} \int_{-\infty}^{\infty} d\nu\, w(\nu) \frac{d}{d\tau} \exp(-2\pi i\nu\tau)$$
$$= \frac{-2\pi i}{W} \int d\nu\, \nu\, w(\nu) \exp(-2\pi i\nu\tau)$$

which at $\tau = 0$ is proportional to the first moment of w:

$$\left. \frac{d\epsilon}{d\tau} \right|_0 \propto \bar{\nu} = \frac{1}{W} \int d\nu\, \nu\, w(\nu).$$

Hence the effect of phase winding can be minimized if we make a change of variable $\nu' = \nu - \bar{\nu}$ in equation 4 before performing the Fourier transform, giving:

$$\phi(\nu') \sim \phi_{\text{ref}} + (\nu' + \bar{\nu} - \nu_{\text{ref}})\, \dot{\phi}_{\text{ref}}$$

with transform

$$\mathfrak{V}'(\tau) = S\, \exp(i[\phi_{\text{ref}} - \dot{\phi}_{\text{ref}}\{\nu_{\text{ref}} - \bar{\nu}\}])\, \delta(\dot{\phi}_{\text{ref}} - 2\pi\tau) \star \epsilon'(\tau) \tag{9}$$

for

$$\epsilon'(\tau) = \frac{1}{W} \int_{-\infty}^{\infty} d\nu'\, w(\nu' + \bar{\nu}) \exp(-2\pi i\nu'\tau).$$

Since this translation in the frequency domain affects only phases in the lag domain, the power-spectrum envelope $\epsilon\epsilon^*$ is unaffected.

Note that the secondary goal of determining ϕ_{ref} is made easier (i.e. likely to have a smaller uncertainty) if we choose ν_{ref} to be close to $\bar{\nu}$.

3.2 Locating the peak and estimating the parameters

As mentioned already in section 2.2, what we have is not a continuous function $V(\nu)$, but a vector 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 vector V_j but this vector embedded in a larger vector which is elsewhere filled with zeros. The greater the degree of 'zero-padding', the more finely spaced are the resulting discrete samples of the envelope function ϵ.

As mentioned in the preceding section, in order to reduce the phase spiral modulation of the envelope function in delay space, we want to perform a change of variable to place the first moment of the weight function w at the origin before transforming. In terms of the discrete transform this means performing a cyclic shift, after zero padding, to relocate the first moment as close as possible to the zero pixel. Since we are limited now to discrete shifts we will not in general be able to do so exactly, but a maximum half-pixel discrepancy will leave only a mild residual phase gradient which we can expect to always be much more slowly varying than the amplitude of ϵ. The first moment of w is of course in the discrete case calculated not via an integral but via a sum.

3.2.1 Delay

After zero-padding, cyclic-shifting and Fourier-transforming, the centre of $\delta \star \epsilon\epsilon^*$ is estimated from the power spectrum as follows. Suppose we isolate a stretch of 3 pixels of the power spectrum centred on our initial coarse guess $j_{\max}$ at the pixel with the maximum value. Let us label these 3 values $q_{j_{\max}+l}$ where l runs from -1 to 1, and for convenience also let us work for the moment in continuous l-pixel coordinates x, without yet worrying about the proper scaling to time-lag values. We will assume that the q values near the peak may be approximated by a parabolic bell, viz:

$$q_{j_{\max}+l} \sim A_{\mathrm{PS}}^2 \left(1 + \frac{c}{2A_{\mathrm{PS}}^2}[l - x_{\max}]^2\right), \tag{10}$$

with the curvature c (the second derivative of q with respect to x) naturally expected to be < 0. The expectation value of the amplitude A_{PS}^2 will not equal S^2 as might be expected from equation 8, but rather S^2 plus the average noise power, which is calculated in section 3.2.3.

For the simple top-hat weighting example already mentioned in the previous section, the discrete form of ϵ' is given by

$$\epsilon'_k = \mathrm{sinc}(k/z),$$

where z is the zero-padding ratio. The power-spectrum envelope $\epsilon\epsilon^*$ is the square of this. In the limit of large z, then, the curvature c of the power-spectrum envelope will tend to the value

$$c \to A_{\mathrm{PS}}^2 \frac{2}{3} \frac{\pi^2}{z^2}. \tag{11}$$

Note that the noise correlation function of this example, discussed in section 3.3, remains a sinc function of unity amplitude, therefore with a central curvature equal to $c/2$.

The coefficients are obtained by parabolic fit as follows:

$$c = q_{j_{\max}-1} + q_{j_{\max}+1} - 2q_{j_{\max}}, \tag{12}$$

$$x_{\max} = \frac{q_{j_{\max}-1} - q_{j_{\max}+1}}{2c}, \tag{13}$$

then A_{PS}^2 is found trivially by reversing equation 10 with $q = q_{j_{\max}}$. The position of the peak in lag units, $\tau_{\max}$, will then be given by

$$\tau_{\max} = \frac{j_{\max} + x_{\max}}{z\,\Delta\nu},$$

where $\Delta\nu$ is the bandwidth $\nu_{\mathrm{hi}} - \nu_{\mathrm{lo}}$.

Note that, although centroiding can proceed perfectly well without any prior knowledge in the case that the noise power is negligible, 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.

3.2.2 Phase

After estimating the position of the peak and thus deriving an estimate for the delay T, the second task here is to estimate the phase at the peak and thus derive ϕ_{ref}. This is done by fitting the same parabolic bell formula separately to the real and imaginary parts $\mathbf{d}_{\mathrm{R}}$ and $\mathbf{d}_{\mathrm{I}}$ of the lag data $\mathfrak{V}$. The fitting is however now performed slightly differently. Firstly, there is no point in fitting $x_{\max}$: we already have the best value we can get for this. Secondly, we expect that the value of the relative curvature c/A_{PS}^2 will now be about half what it was when fitting to the squared envelope distribution. Thus the shape of the function can be assumed to be known and thus left fixed: all we need to do is fit its amplitude. In mathematical terms, the expected value at the jth pixel is

$$\langle d_j \rangle = A m_{j-j_{\max}}$$

where the fixed shape function is given by

$$m_l = \left(1 + \frac{c}{4A_{\mathrm{PS}}^2}[l - x_{\max}]^2\right).$$

A_{PS}^2 here is the amplitude fitted to the power spectrum peak. Given the real and imaginary amplitudes A_{R} and A_{I}, the derivation of phase is then straightforward.

For fitting the amplitude we could perform a standard least-squares using as many d_j values as we like. Normally we would expect that using more values would reduce the uncertainty in phase. However here the extra degrees of freedom are illusory. The only reason the peak spreads over several channels in the lag spectrum is because we have zero-padded the visibilities before transforming. This of course cannot increase the signal-to-noise ratio. The resulting noise values end up highly correlated and at the end of the day there is nothing to be gained by using several values. The amplitude is thus just as well obtained from the central value:

$$A = \frac{d_{j_{\max}}}{m_0}. \tag{14}$$

3.2.3 Noise in the estimated power spectrum

Finally, a word about the behaviour of noisy input under the weighting, padding and transforming scheme here proposed. All real data includes a noise component and its behaviour will affect both the uncertainties in the parameters here estimated and the reliability of the peak detection.

For a visibility spectrum of M channels, the real and imaginary values of which in any single jth channel comprise Gaussian-distributed random values which are independent from one another (and from the values in any other channel) but which have the same standard deviation σ_j, it is not difficult to show that the noise values in the weighted, Fourier-transformed data all have the same standard deviation σ given by

$$\sigma = \frac{\sqrt{\sum_{j=1}^{M} w_j^2 \sigma_j^2}}{\sum_{j=1}^{M} w_j}. \tag{15}$$

Zero-padding will not change this relation, it just introduces correlations between the output noise values. These random values are still Gaussian-distributed; for the power spectrum however, the distribution $p(q)$ of random PS values is a simple exponential function:

$$p(q) = \frac{1}{2\sigma^2} \exp\left(\frac{-q}{2\sigma^2}\right) \text{ for } q > 0, = 0 \text{ else.} \tag{16}$$

The mean of this distribution, which is equal to the average noise power, is $2\sigma^2$.

3.3 Uncertainties

3.3.1 Delay

The delay $\tau_{\max}$ is proportional to the position $x_{\max}$ of the maximum in the power spectrum. Thus I'll concentrate on the uncertainty in $x_{\max}$.

One can via standard propagation of uncertainty formulae derive an approximate[1] formula for this uncertainty, but there are two slightly fiddly aspects to this. The first is that the zero-padding mentioned in section 3.2 introduces a correlation between the error values of separate data channels. This has to be taken into account by including the (sometimes neglected) cross-correlation terms in the propagation-of-uncertainty formula. The second aspect arises because the position parameter is determined by fitting to the power spectrum, for which the distribution of random values is no longer Gaussian. However, with care, both difficulties are perfectly addressable and a working expression can be obtained.

Starting from equations 12 and 13, one derives the full correct linear approximation for σ_x, the uncertainty in $x_{\max}$, to be

$$c^2 \sigma_x^2 = \sigma_0^2 (6x_{\max}^2 + 0.5) - 8\sigma_1^2 x_{\max}^2 + 2\sigma_2^2 (x_{\max}^2 - 0.25). \tag{17}$$

---

[1]It is a linear approximation which is technically only valid in the limit of small σ.

Here σ_0^2, σ_1^2 and σ_2^2 are respectively the values on the central, the first outer (on either side) and the second outer diagonals of the noise correlation matrix. Note that this relies on the assumption that the variance σ_0^2 of the noise in the power spectrum is not a function of the delay coordinate x - an assumption we will question in a moment.

For heavily correlated data, i.e. if the zero padding factor z is greater than about 4, which will probably be usual, σ_1^2 and σ_2^2 are given by

$$\sigma_1^2 = \sigma_0^2 Y(1)$$

and

$$\sigma_2^2 = \sigma_0^2 Y(2),$$

where $Y(x)$ is the Fourier transform of the power spectral density of the noise (Wiener-Kinchin theorem); and the PSD of the noise is just the square of the weight function w. For the simple top-hat example w described in section 3.1, w^2 is of course the same shape as w, and therefore

$$Y(x) = \text{sinc}(x/z) \sim \left(1 - \frac{1}{6}\left[\frac{\pi x}{z}\right]^2\right) \sim (1 + 0.25cx^2).$$

In this case all the terms in equation 17 involving $x_{\max}$ cancel one another to leave

$$\sigma_x^2 = \frac{\sigma_0^2}{|c|}.$$

We now come to the second 'fiddly aspect' mentioned at the beginning of the present section, namely the proper value of σ_0. There is a trap here and that is to consider the distribution of random values in a power spectrum (PS) from real and imaginary inputs R and I which comprise purely Gaussian noise, with no added signal. In this case one deduces that the distribution $p(q)$ of random PS values is exponential as given in equation 16. This distribution has both mean and standard deviation equal to $2\sigma^2$. However this is not the situation of present interest, because our data do include a signal, and in the lag domain we are sitting right on top of it. If we arbitrarily rotate the phase of the peak to zero (which doesn't affect the statistical properties of the noise), we see that the PS data value q at the top of the peak (assuming the peak has height unity) is given by

$$q = R^2 + I^2 = (1 + n_R)^2 + n_I^2$$

where n_R and n_I are respectively the noise in the real and imaginary parts respectively - both of which have standard deviation σ. Clearly if $\sigma \ll 1$ the uncertainty σ_0 in q is going to be approximately equal to 2σ rather than $2\sigma^2$.

A more exact treatment is not difficult. The variance of q is formally equal to $E[q^2] - E[q]^2$ where $E[]$ represents expectation value. The nth moment $E[q^n]$ is given by:

$$E[q^n] = \frac{1}{2\pi\sigma^2} \int dR \int dI \, (R^2 + I^2)^n \exp\left[\frac{-(R - A_{PS})^2 - I^2}{2\sigma^2}\right],$$

which integral is always separable into products of moments of 1-dimensional Gaussian distributions. Thus $E[q] = A_{PS}^2 + 2\sigma^2$ and $E[q^2] = A_{PS}^4 + 8A_{PS}^2\sigma^2 + 8\sigma^4$ and from that, it is easy to arrive at the expression for the variance of the power spectrum on the peak:

$$\sigma_0^2 = 4\sigma^2(A_{PS}^2 + \sigma^2).$$

This derivation is only exact for the centre of the peak, whereas our hat-fitting technique relies also on values at either side of it. Technically this means that our earlier assumption that σ_0 was independent of $x \propto \tau$ is incorrect. However it ought to be a good approximation provided that the (power-envelope) curvature $|c|$ is less than about 0.2, which translates (see equation 11) to a zero padding factor $z >\sim 4$.

8

3.3.2 Phase

To calculate the uncertainty in phase we first need the uncertainty in the amplitude A of the parabolic bell fitted separately to the real and imaginary parts of the lag data. From equation 14

$$\sigma_A = \frac{\sigma}{m_0}$$

in either case. Given

$$\phi = \arctan(A_I/A_R)$$

where A_R and A_I are respectively the real and imaginary amplitudes of the peak in the complex-valued lag data, standard propagation of uncertainty gives the following approximation for the uncertainty in ϕ:

$$\sigma_\phi \sim \frac{\sigma}{m_0 A_{PS}}$$

where

$$A_{PS}^2 = A_I^2 + A_R^2.$$

It is probably quite an acceptable approximation to replace m_0 by unity.

3.3.3 Estimating the noise power

As shown in equation 15, if we know both the weights and the uncertainties of the original data, we can calculate the standard deviation σ of the Fourier-transformed values. However the original uncertainties may not be well known. It turns out that it is fairly easy to estimate σ from the transformed data. The total power in the power spectrum will be the sum of the noise power, which from equation 16 is just $2\sigma^2$ multiplied by the number of channels M, plus the area under $\epsilon\epsilon^*$. From Parseval's theorem, the latter is just equal to

$$\int \epsilon\epsilon^*(\tau) = \int w^2(\nu).$$

In terms of the discrete Fourier transform, including also for generality the measured amplitude A_{PS}^2 of the peak in the power spectrum, the mean noise power $2\sigma^2$ is estimated from the following:

$$2\sigma^2 = \frac{1}{M}\sum_{j=1}^{M} q_j - \frac{\sum_{j=1}^{M} w_j^2}{\left(\sum_{j=1}^{M} w_j\right)^2}(A_{PS}^2 - 2\sigma^2).$$

The first term on the RHS gives the mean total power; the second gives the mean power under the peak. (Note the correction of the measured peak height A_{PS}^2 to remove the bias due to the mean noise power itself.) This reduces to the formula

$$2\sigma^2 = \frac{\frac{1}{M}\left(\sum_{j=1}^{M} w_j\right)^2 \sum_{j=1}^{M} q_j - A_{PS}^2 \sum_{j=1}^{M} w_j^2}{\left(\sum_{j=1}^{M} w_j\right)^2 - \sum_{j=1}^{M} w_j^2}.$$

Note however that since this involves a subtraction, it will only return reliable values if the ratio between the noise power and the power under the peak is larger than about unity. Normally one would expect this to be the case, since decoherence effects encourage the user to look for fringes in as small a section of data as possible, therefore in a regime in which the signal-to-noise ratio (SNR) is as low as can be managed. A number of channels which is only just large enough to provide an acceptable value of SNR will however still provide a significant value for the integrated noise power.

3.4 Reliability

Noisy data will always have a maximum value somewhere. Our algorithm commences with finding this maximum. We assume that this corresponds to the location of the peak in lag space which represents the instrumental and tropospheric delay. There is a certain probability however that the peak we have found is simply a random fluctuation of the noise values. To estimate this probability we return to the exponential distribution $p(q)$ of power spectrum values given in equation 16. The probability that a q value greater than the peak height A_{PS}^2 is simply $\exp(-A_{\mathrm{PS}}^2/2\sigma^2)$. The expected frequency of such occurrences in each spectrum of M values is M times that probability.

3.5 Decoherence due to dispersive (ionospheric) delay

As can be seen from equation 3, any significant amount of ionospheric contribution to the variation in phase across the band is going to manifest itself as a departure of the phase relation from linearity. The Fourier result given in equation 5 will then no longer be exact: the peak in the lag spectrum will be broadened and reduced in height. To estimate the degree of this height reduction, I start by extending the Taylor expansion of $\phi(\nu)$ about ν_{ref} to include the second-order term. This is equivalent to multiplying the previous first-order expression for the visibilities by

$$\exp\left(0.5i\,\ddot{\phi}[\nu - \nu_{\mathrm{ref}}]^2\right).$$

If we lump this together with the weight function, and translate the coordinates again by $\bar{\nu}$, then we end up with equation 9 as before but with a new envelope function

$$\epsilon_{\mathrm{disp}}(\tau) = \int_{-\infty}^{\infty} d\nu'\, w(\nu' + \bar{\nu})\, \exp\left(0.5i\,\ddot{\phi}[\nu' + \bar{\nu} - \nu_{\mathrm{ref}}]^2 - 2\pi i\nu'\tau\right).$$

Obviously the precise form of this is going to depend upon w, but to make the estimate we require, let us take the simple top-hat example in which

$$w(\nu + \bar{\nu}) = \frac{1}{\Delta\nu} \text{ for } \nu \in [-\Delta\nu/2, \Delta\nu/2], = 0 \text{ else.}$$

Let us also for the sake of simplicity choose $\nu_{\mathrm{ref}} = \bar{\nu}$. The height of the peak is given by evaluating the integral at $\tau = 0$, viz:

$$\epsilon_{\mathrm{disp}}(0) = \frac{1}{\Delta\nu} \int_{-\Delta\nu/2}^{\Delta\nu/2} d\nu\, \exp\left(i\frac{\ddot{\phi}}{2}\nu^2\right).$$

(Here I've also suppressed the now pointless prime on the variable of integration.) This could be evaluated by Fresnel integrals but a simple series expansion of the integrand suffices to provide the result

$$\epsilon_{\mathrm{disp}}(0) \sim 1 - \frac{(\ddot{\phi})^2 \Delta\nu^4}{640}.$$

(Actually, we should add $2\sigma^2$ to this since this is the expected average of the power-spectrum noise.)

Note that the Fourier algorithm knows nothing of the value of $\ddot{\phi}$; this decoherence factor must therefore be calculated external to the algorithm as part of further processing.

4 The 2D Fourier method

As mentioned in section 2.1, the variation of the delay quantities with time is fundamentally a random quantity. However, the $1/f$ behaviour of typical natural processes means that the amplitude of variations becomes smaller over shorter time intervals. Over a short enough time interval, we may expect that a 2-dimensional Taylor expansion of phase, keeping only linear terms, will be an acceptable approximation to the true phase behaviour. Such a linear variation of phase with both radio frequency *and* time can be analysed by a 2-dimensional extension of the Fourier technique discussed in section 3. The linear variation in the frequency direction is proportional to delay as before, whereas the variation with time now is associated with a new quantity, *fringe rate*, measured in frequency units.

To be precise, the phase model to be used here is as follows:

$$\phi(t, \nu) \sim \phi_{\text{ref}} + (t - t_{\text{ref}}) \left.\frac{\partial \phi}{\partial t}\right|_{\text{ref}} + (\nu - \nu_{\text{ref}}) \left.\frac{\partial \phi}{\partial \nu}\right|_{\text{ref}}. \tag{18}$$

Nearly all the results from section 3 apply also here and can be extended trivially to cover the time axis. There is no interaction between the axes.

4.1 Modifications to the 1D procedure

To locate the peak in the 2D data we isolate a 3×3 patch of power spectrum image values centred on our initial coarse guess $(j_{\text{max}}, k_{\text{max}})$ at the pixel with the maximum value. The position of the maximum in real-valued, i.e. continuous, pixel coordinates we label $(x_{\text{max}}, y_{\text{max}})$. Let us label the 3×3 block of values $q_{j_{\text{max}}+l, k_{\text{max}}+m}$ where l and m here both run from -1 to 1. We assume that these power spectrum image values may be approximated by a product of two parabolas, viz:

$$q_{j_{\text{max}}+l, k_{\text{max}}+m} \sim A_{\text{PS}}^2 \left(1 + \frac{c_{\text{X}}}{2A_{\text{PS}}^2}[l - x_{\text{max}}]^2\right)\left(1 + \frac{c_{\text{Y}}}{2A_{\text{PS}}^2}[m - y_{\text{max}}]^2\right). \tag{19}$$

The values c_{X} and c_{Y} are now the second derivatives of q with respect to x and y respectively, but evaluated now at $(x, y) = (x_{\text{max}}, y_{\text{max}})$. They are again expected to be negative-valued.

The location of the peak in two dimensions is obtained by separate 1-dimensional parabolic fits as follows. Let us define

$$Q_{\text{X},j} = \sum_{k=k_{\text{max}}-1}^{k_{\text{max}}+1} q_{j,k}$$

and

$$Q_{\text{Y},k} = \sum_{j=j_{\text{max}}-1}^{j_{\text{max}}+1} q_{j,k}.$$

Then

$$Q_{\text{X},j_{\text{max}}+l} = A_{\text{X}}\left(1 + \frac{c_{\text{X}}}{2A_{\text{PS}}^2}[l - x_{\text{max}}]^2\right)$$

and

$$Q_{\text{Y},k_{\text{max}}+m} = A_{\text{Y}}\left(1 + \frac{c_{\text{Y}}}{2A_{\text{PS}}^2}[m - y_{\text{max}}]^2\right)$$

where

$$A_{\text{X}} = A_{\text{PS}}^2\left(3 + \frac{c_{\text{Y}}}{2A_{\text{PS}}^2}[2 + 3y_{\text{max}}^2]\right) \tag{20}$$

and

$$A_{\text{Y}} = A_{\text{PS}}^2\left(3 + \frac{c_{\text{X}}}{2A_{\text{PS}}^2}[2 + 3x_{\text{max}}^2]\right). \tag{21}$$

We then have

$$x_{\text{max}} = \frac{1}{2}\frac{Q_{\text{X}j_{\text{max}}-1} - Q_{\text{X}j_{\text{max}}+1}}{Q_{\text{X}j_{\text{max}}-1} + Q_{\text{X}j_{\text{max}}+1} - 2Q_{\text{X}j_{\text{max}}}}$$

and

$$y_{\text{max}} = \frac{1}{2}\frac{Q_{\text{Y}j_{\text{max}}-1} - Q_{\text{Y}j_{\text{max}}+1}}{Q_{\text{Y}j_{\text{max}}-1} + Q_{\text{Y}j_{\text{max}}+1} - 2Q_{\text{Y}j_{\text{max}}}}.$$

In order to estimate the phase via an extension of the technique described in section 3.2.2 we again require the fitted amplitude and curvature(s). To obtain these values is however no longer trivial. The simplest way may be via an iterative procedure as follows:

1. Calculate x_{max} and y_{max} via the procedure already described.

2. Make the initial estimate $A_{\text{X}}/A_{\text{PS}}^2 \sim 3$.

3. Estimate c_X from

$$c_X = \frac{A_{PS}^2}{A_X}(Q_{X,j_{max}-1} + Q_{Xj_{max}+1} - 2Q_{X,j_{max}}),$$

then A_{PS}^2 from

$$A_{PS}^2 = Q_{X,j_{max}}\left(\frac{A_{PS}^2}{A_X}\right) - \frac{c_X x_{max}^2}{2}.$$

4. Estimate A_Y/A_{PS}^2 from equation 21.
5. Estimate c_Y from

$$c_Y = \frac{A_{PS}^2}{A_Y}(Q_{Y,j_{max}-1} + Q_{Yj_{max}+1} - 2Q_{Y,j_{max}}),$$

then A_{PS}^2 from

$$A_{PS}^2 = Q_{Y,k_{max}}\left(\frac{A_{PS}^2}{A_Y}\right) - \frac{c_Y y_{max}^2}{2}.$$

6. Estimate A_X/A_{PS}^2 from equation 20.
7. Go to 3 until converged.

4.2 Decoherence due to fringe rate

In section 3.5 it was described how a significant contribution from ionospheric (dispersive) delay could limit the range over which the 1D Fourier technique could be used. The 2D Fourier technique suffers from an additional decoherence effect in the presence of a non-zero fringe rate. To see this, we start with equation 3, omit the ionospheric term for simplicity, then rewrite it in terms of frequency displacement from a reference frequency ν_{ref}. The result is similar to equation 4, but with both the proportionality between $\partial\phi/\partial\nu$ and non-dispersive delay T, as well as the possible time variation of ϕ_{ref}, now made explicit.

$$\phi(\nu, t) \sim \phi(\nu_{ref}, t) + 2\pi T(t)(\nu - \nu_{ref}).$$

Now if instead of simply expanding the phase as a whole in a first-order Taylor series with respect to time, we do this explicitly with the component functions $\phi_{ref}(t)$ and $T(t)$, we obtain

$$\phi(\nu, t) \sim \phi(\nu_{ref}, t_{ref}) + \left.\frac{\partial\phi}{\partial t}\right|_{\nu_{ref}, t_{ref}} (t - t_{ref}) + 2\pi\left[T(t_{ref})(\nu - \nu_{ref}) + \left.\frac{\partial T}{\partial t}\right|_{\nu_{ref}, t_{ref}} (t - t_{ref})(\nu - \nu_{ref})\right].$$

This can be seen to be identical in form to equation 18, plus an additional term which depends on the product of ν and t. If we absorb the visibility contribution from this additional phase term into the weight function as in section 3.5, and make again all the other approximations in that section, we arrive at the following altered expression for the envelope function:

$$\epsilon_{rate}(0,0) = \frac{1}{\Delta\nu\Delta t} \int_{-\Delta\nu/2}^{\Delta\nu/2} d\nu \int_{-\Delta t/2}^{\Delta t/2} dt \, \exp\left(2\pi i \, \dot{T}\nu t\right),$$

where I have used the following shorthand for the fringe rate:

$$\dot{T} = \left.\frac{\partial T}{\partial t}\right|_{\nu_{ref}, t_{ref}}.$$

The integral can be evaluated by making use of the result, but it is easy to get it via Taylor expansion of the exp)

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

Keeping terms in the series to $n = 3$ we obtain

$$\epsilon_{rate}(0,0) \sim 1 - \frac{(\pi\Delta\nu\Delta t\dot{T})^2}{72}.$$

In the limit of small $|\dot{T}|$ the deviation in $\epsilon\epsilon^*$ will of course be twice as large as that in ϵ.

5 The linear algorithm

I refer again to the general expression 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:

- Just 'planar' parameters, for example those in equation **??**: ϕ_{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

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

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 $O(MN)$ if the full 5-parameter correction is performed.

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

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

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

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

and the five pairs of qs and hs are:

m	q_m	$h_m(\nu, t)$
1	$\phi_0 - \phi_{0,\mathrm{est}}$	1
2	$2\pi(\tau_0 - \tau_{\mathrm{est}})$	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

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

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

$$\mathbf{Aq} = \mathbf{b}, \tag{22}$$

where

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

and

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

Equation 22 (equivalent to the normal equations for this problem) is simply inverted to give the coefficients q_m. Further, $\mathbf{A}^{-1}$ is the covariance matrix of the q_m.

5.1 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 $\mathbf{A}$ the same for any chunk of data.*****)