Optimising FITACF fitting

Background

The original FITACF utilises the linear least-square fit for estimating velocity (ACF phase), lag 0 power and spectral width (ACF log power) and elevation (XCF phase). The phase and log power fitting and calculation of the respective errors are performed based on the conventional least square expressions for a linear fit (see e.g. Chapter 15.2 in *Numerical Recipes*).

* There is also a parabolic model fit applied to ACF log power (Gaussian ACF decay), but it's results are rarely used, so we will leave it out for the time being.

There are two types of linear fit which are utilised in FITACF:

(a) More general case with an offset, $y = a + bx$, applicable for log-power fit (exponential model) and interferometer phase fit (elevation).

To get the optimum fit, we need to minimise a following sum

$$\chi^2(a,b) = \sum_{i=0}^{N-1} \frac{(y_i - a - b\,x_i)^2}{\sigma_i^2}$$

Here, σ_i represents a standard deviation of the measured parameter, y, at the i^{th} lag. This value can be either measured or estimated theoretically (see next section).

Minimisation of the above sum requires two derivatives to be equal to zero,

$$\frac{\partial \chi^2}{\partial a} \equiv 0\,, \frac{\partial \chi^2}{\partial b} \equiv 0\,.$$

The solutions for a and b can be derived analytically:

$$a = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta}$$

$$b = \frac{S S_{xy} - S_x S_y}{\Delta}\,,$$

where

$$S = \sum_{i=0}^{N-1} \frac{1}{\sigma_i^2},\quad S_x = \sum_{i=0}^{N-1} \frac{x_i}{\sigma_i^2},\quad S_y = \sum_{i=0}^{N-1} \frac{y_i}{\sigma_i^2},$$

$$S_{xx} = \sum_{i=0}^{N-1} \frac{x_i^2}{\sigma_i^2},\quad S_{xy} = \sum_{i=0}^{N-1} \frac{x_i y_i}{\sigma_i^2},\quad \Delta = S S_{xx} - S_x^2\,.$$

The respective statistical errors for the offset and the slope are

$$\sigma_a^2 = \frac{S_{xx}}{\Delta},\quad \sigma_b^2 = \frac{S}{\Delta}.$$

(b) Specific case with no offset (a = 0, applicable for phase in velocity measurements):

$$b = \frac{S_{xy}}{S_{xx}}, \ \sigma_b^2 = \frac{1}{S_{xx}}.$$

Problem formulation

The FITACF implementation of these algorithms, while being rather reliable, contains several non-optimal procedures based on some *ad hoc* assumptions. The major ones are:

1. **During fitting, both ACF power and phase are weighted by inverted power.**
   In other words, $\sigma_i = 1/P_i$, where P_i is an ACF power value at the i^{th} lag.
   However, in reality, both power and phase variances depend on
   (i) magnitude of correlation coefficient
   and
   (ii) number of averages at a given lag.

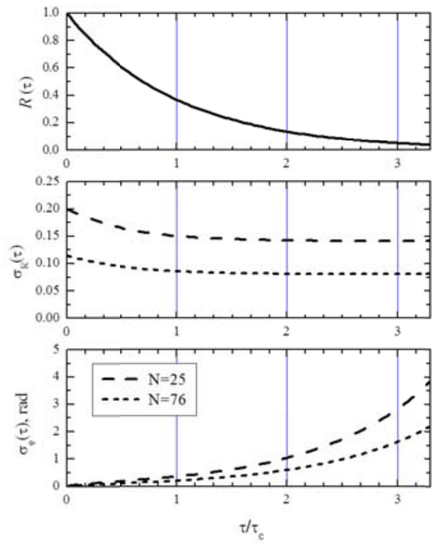
   More importantly, these dependencies are very different (see graphs below) so that we have to apply separate sets of weighting coefficients for power and phase.

   The respective expressions for the power and phase variance can be derived based on *Bendat & Piersol* (1986):

$$\sigma_P(\tau) = \sqrt{\frac{|R(\tau)|^2 + 1}{2N}},$$

$$\sigma_\varphi(\tau) = \sqrt{\frac{|R(\tau)|^{-2} - 1}{2N}}.$$

   Here $|R(\tau)| = P(\tau)/P(0)$ is the correlation coefficient magnitude at a given lag, τ, and N is the number of averages (pulse sequences). The difference between the power and phase variance is illustrated by the following plot obtained for the exponential power decay model:



The top panel shows ACF power (correlation coefficient) *vs* normalised time lag, while the middle and bottom panels represent theoretical variance for power and phase, respectively. The dashed and dotted lines correspond to 25 and 76 averages, respectively.
The most important distinction between the middle and bottom panels is that the power variance slowly decreases with decreasing correlation while at the same time the phase variance quickly grows.

2. **All "bad" lags with excessive cross-range interference (CRI) are removed from fitting.**
The problem here is that, conventionally, all measured data points should be used in fitting, but those with larger variance, regardless of its origin, should be assigned proportionally lower weighting coefficients.
Based on *Bendat & Piersol* (1986), the CRI effects on variance can be estimated using signal power from the interfering ranges. The respective expressions for CRI-affected variance in power (correlation coefficient) and phase are:

$$\sigma_P(\tau_{ij}) = \sqrt{\frac{|R(\tau_{ij})|^2 + \alpha(\tau_{ij})^{-2}}{2N}}$$

$$\sigma_\varphi(\tau_{ij}) = \sqrt{\frac{\alpha(\tau_{ij})^{-2}|R(\tau_{ij})|^{-2} - 1}{2N}}$$

where

$$\alpha(\tau_{ij})^2 = \frac{P_0^2}{(P_0 + P_i)(P_0 + P_j)},$$

and P_0 is a lag 0 power for the analysed range and P_i and P_j are cumulative lag 0 powers from the interfering ranges contributing to CRI in pulses i and j, respectively, which are combined for estimating ACF parameters at a lag τ_{ij}. It is easy to see that for zero CRI levels in both samples $\alpha \equiv 1$.

3. **In estimating fitting errors, the fitted parameter's variance is considered to be the same at all lags of a given ACF and being estimated as a standard deviation of the measured values from the fitted line.**
   The problem here, as one can see from the above expressions, is that the variance values can be very different at different lags of the same ACF, depending on the correlation coefficient and CRI level.

4. **A number of degrees of freedom (DoF) in calculating fitting errors is calculated assuming that statistical variations of the measured parameters at different lags are independent.**
   The problem here is that we actually use only N_{pul} =7 or 8 independent measurements of power and phase (pulses) to calculate N_{lag}=17 or 23 ACF lags. Therefore, the maximum possible DoF number is equal to N_{pul} -n_{par} rather than N_{lag} - n_{par} , where n_{par} is the number of estimated parameters of the fitted model.

Proposed action

The existing software has to be modified in such a way that it performs a following sequence of operations for each range gate:

1. Estimate cumulative CRI levels contributing to each pulse in the sequence and calculate α for each lag.
2. Estimate ACF power variance at each lag based on the measured |R| and α.
3. Fit linear model model with an offset to ACF power using the above variance values for weighting.
4. Estimate ACF phase variance at each lag using fitted |R| and measured α. The use of fitted power is necessary because the analytic expression for the phase variance produces imaginary values if at a given lag its power exceed that at lag 0 (i.e. $|R(\tau)|>1$). This situation is possible because of the statistical variability in the measured ACF power and/or significant contribution from CRI. However, we don't need to worry about the same problem with the power variance because in this case the expression inside the square root is always positive.
5. Fit linear model without (with) an offset to ACF (XCF) phase using the variance values estimated at step 3 for weighting.
6. Estimate the related fitting errors based on the variance values obtained at steps 1 and 3. In doing this, we need to estimate the respective number of DoF using the effective number of available pulses, N_{pul}, rather than that of available ACF lags, N_{lag}.

   * Currently, I have only a heuristic expression for the effective number of DoF, so more literature search has to be done.