Lecture 26
- Localization/Matched Filtering (continued)
- Prewhitening

Lectures next week:
- Bases, principal component analysis, wavelets, Bayesian blocks.

Reading
- Article on Bayesian Blocks (Scargle)
- Article on wavelets
**ROC Curves:**

A plot of $P_d$ vs. $P_{fa}$ is called a *receiver operating characteristics* curve, named after radar detection schemes.

![ROC plot](image)

Figure 5: ROC plot for a pulse with SNR 5 (peak to rms noise) and width of 10.3 samples. Matched filtering is used.
Localization Using Matched Filtering

This handout describes localization of an object in a parameter space. For simplicity we consider localization of a pulse in time. The same formalism applies to localization of a spectral feature in frequency or to an image feature in a 2D image. The results can be extrapolated to a space of arbitrary dimensionality.

I. First consider finding the amplitude of a pulse when the shape and location are known. Let the data be

\[ I(t) = aA(t) + n(t), \]

where \( a \) = the unknown amplitude and \( n(t) \) is zero mean noise. The known pulse shape is \( A(t) \). Let the model be \( \hat{I}(t) = \hat{a}A(t) \).

Define the cost function to be the integrated squared error,

\[ Q = \int dt \left[ I(t) - \hat{I}(t) \right]^2. \]
Taking a derivative, we can solve for the estimate of the amplitude, \( \hat{a} \):

\[
\partial_{\hat{a}} Q = -2 \int dt \left[ I(t) - \hat{I}(t) \right] \partial_{\hat{a}} \hat{I}(t) = 0
\]

\[
\Rightarrow \int dt \left[ I(t) - \hat{I}(t) \right] A(t) = 0
\]

\[
\Rightarrow \int dt \hat{I}(t) A(t) = \int dt I(t) A(t)
\]

\[
\Rightarrow \hat{a} \int dt A^2(t) = \int dt I(t) A(t)
\]

\[
\Rightarrow \hat{a} = \frac{\int dt I(t) A(t)}{\int dt A^2(t)}.
\]

Note that:

a. The model is linear in the sole parameter, \( \hat{a} \)

b. The numerator is the zero lag of the crosscorrelation function (CCF) of \( I(t) \) and \( A(t) \).

c. The denominator is the zero lag of the autocorrelation function (ACF) of \( A(t) \).
II. Now consider the case where we don’t know the location of the pulse in time (the time of arrival, TOA) and that it is the TOA we wish to estimate. We still know the pulse shape, \textit{a priori}.

Let the data, model and cost function be

\[ I(t) = aA(t - t_0) + n(t) \]

\[ \hat{I}(t) = \hat{a}A(t - \hat{t}_0). \]

\[ Q = \int dt \left[ I(t) - \hat{I}(t) \right]^2. \]

Note that the model is \textbf{linear} in \( \hat{a} \) but is \textbf{nonlinear} in \( \hat{t}_0 \).
Minimizing $Q$ with respect to $\hat{a}$, we have

\[
\partial_\hat{a} Q = -2 \int dt \left[ I(t) - \hat{I}(t) \right] \partial_\hat{a} \hat{I}(t) = 0
\]

\[
\Rightarrow \int dt \left[ I(t) - \hat{I}(t) \right] A(t - \hat{t}_0) = 0
\]

\[
\Rightarrow \int dt \hat{I}(t) A(t - \hat{t}_0) = \int dt I(t) A(t - \hat{t}_0)
\]

\[
\Rightarrow \hat{a} \int dt A^2(t - \hat{t}_0) = \int dt I(t) A(t - \hat{t}_0)
\]

\[
\Rightarrow \hat{a} = \frac{\int dt I(t) A(t - \hat{t}_0)}{\int dt A^2(t - \hat{t}_0)}.
\]

This last equation has the same form as in I. except that the estimate for the arrival time $\hat{t}_0$ is involved.
Now, minimizing $Q$ with respect to $\hat{t}_0$, we have

$$\partial_{\hat{t}_0} Q = -2 \int dt \left[ I(t) - \hat{I}(t) \right] \partial_{\hat{t}_0} \hat{I}(t) = 0$$

$$\Rightarrow -\hat{a} \int dt \hat{I}(t) \hat{a} A'(t - \hat{t}_0) = -\hat{a} \int dt I(t) A'(t - \hat{t}_0)$$

$$\Rightarrow \hat{a} \int dt A(t - \hat{t}_0) A'(t - \hat{t}_0) = \int dt I(t) A'(t - \hat{t}_0). \quad (5)$$

**Grid Search:** One approach to finding the arrival time is to search over a 2D grid of $\hat{a}, \hat{t}_0$ to find the values that satisfy equations 4 and 5. This approach is inefficient. Instead, one can search over a 1D space for the single nonlinear parameter, $\hat{t}_0$, and then solve for $\hat{a}$ using either equation 4 or 5.

**Linearization + Iteration:** Another method is to find solutions for $\hat{a}$ and $\hat{t}_0$, we can linearize the equations in $\hat{t}_0 - t_0$ by using Taylor-series expansions for $A(t - \hat{t}_0)$ and $A'(t - \hat{t}_0)$.

Let $\hat{t}_0 = t_0 + \delta \hat{t}_0$. Then, to first order in $\delta \hat{t}_0$:

$$A(t - \hat{t}_0) \approx A(t - t_0) - A'(t - t_0) \delta \hat{t}_0$$

$$A'(t - \hat{t}_0) \approx A'(t - t_0) - A''(t - t_0) \delta \hat{t}_0$$

$$A^2(t - \hat{t}_0) \approx A^2(t - t_0) - 2A'(t - t_0)A(t - t_0) \delta \hat{t}_0.$$
Now equations (4) and (5) become
\[
\dot{a} \int dt \left[ A^2(t - t_0) - 2\delta \dot{t}_0A(t - t_0)A'(t - t_0) \right] = \\
\int dt \left[ A(t - t_0) - \delta A'(t - t_0) \right] \\
\dot{a} \int dt \left[ A(t - t_0)A'(t - t_0) - \delta \dot{t}_0A(t - t_0)A''(t - t_0) - \delta \dot{t}_0A^2(t - t_0) \right] = \\
\int dt \left[ A'(t - t_0) - \delta \dot{t}_0A''(t - t_0) \right].
\]

Consider the integral
\[
\int dt \ A(t - t_0)A'(t - t_0).
\]

The integrand may be written as
\[
A(t - t_0)A'(t - t_0) = \frac{1}{2} \frac{d}{dt} A^2(t - t_0)
\]
and so the integral equals
\[
\frac{1}{2} A^2(t - t_0) \bigg|_{t_1}^{t_2} \rightarrow 0
\]
in the limit of (e.g.) \( t_{1,2} = \mp T/2 \) with \( T \gg \) pulse width.
We then have
\[
\dot{a} \int dt \, A^2(t - t_0) = \int dt \, I(t)[A(t - t_0) - \delta t_0 A'(t - t_0)]
\]
\[-\delta t_0 \dot{a} \int dt \, [A(t - t_0)A''(t - t_0) + A^2(t - t_0)] = \int dt \, I(t)[A'(t - t_0) - \delta t_0 A''(t - t_0)].
\]

Solving for \( \dot{a} \) in both cases we have
\[
\dot{a} = \frac{\int dt \, [I(t)A(t - t_0) - \delta t_0 I(t)A'(t - t_0)]}{\int dt \, A^2(t - t_0)} \tag{6}
\]
\[
\dot{a} = \frac{\int dt \, I(t) \left[ -A'(t - t_0) + \delta t_0 A''(t - t_0) \right]}{\delta t_0 \int dt \, \left[ A(t - t_0)A''(t - t_0) + A^2(t - t_0) \right]} \tag{7}
\]
Using the notation

\[ i_0 \equiv \int dt I(t)A(t - t_0) \]
\[ i_1 \equiv \int dt I(t)A'(t - t_0) \]
\[ i_2 \equiv \int dt I(t)A^2(t - t_0) \]
\[ i_3 \equiv \int dt I(t)A''(t - t_0) \]
\[ i_4 \equiv \int dt I(t) \left[ A(t - t_0)A''(t - t_0) + A'^2(t - t_0) \right] . \]  

we have

\[ \dot{\delta} = \frac{i_0 - \delta \hat{t}_0 i_1}{i_2} \]
\[ \dot{\delta} = \frac{-i_1 + \delta \hat{t}_0 i_3}{\delta \hat{t}_0 i_4} . \]

Solving for \( \delta \hat{t}_0 \) (to first order) we have

\[ \delta \hat{t}_0 = \frac{-i_1 i_2}{i_0 i_4 + i_2 i_3} . \]
Iterative Solution for $\hat{t}_0$

This equation can be solved iteratively for $\delta \hat{t}_0$:

0. choose a starting value for $\hat{t}_0$.

1. calculate $\delta \hat{t}_0$ using the linearized equations.

2. is $\delta \hat{t}_0 = 0$?

3a. if yes, stop.

3b. if no, update $\hat{t}_0 \rightarrow \hat{t}_0 + \delta \hat{t}_0$ and go back to step 1.

For the best fit value for $\hat{t}_0$, the change is zero, $\delta \hat{t}_0 = 0$ (top of the hill) and $\hat{a}$ can be calculated using one of the equations 6 or 7.

Correlation Function Approach

The iterative solution for $\hat{t}_0$ is similar to the following procedure that uses a crosscorrelation approach more directly:

1. cross correlate the template $A(t)$ with $I(t)$ to get a CCF.

2. find the lag of peak correlation as an estimate for the arrival time, $\hat{t}_0 = \tau_{max}$.

3. calculate $\hat{a}$ if needed.

Subtleties of the Cross Correlation Method

The CCF is calculated using sampled data and therefore is itself a discrete quantity. Often one wants greater precision on the arrival time than is given by the sample interval. I.e. we want a
floating point number for $\hat{t}_0$, not an integer index. Therefore we want to calculate the peak of the CCF by interpolating near its peak. The interpolation should be done properly by using the appropriate interpolation formula for sampled data (using the sinc function). Using parabolic interpolation yields excessive errors for the arrival time.

In practice, the proper interpolation is effectively done in the frequency domain by calculating the phase shift of the Fourier transform of the CCF, which is the product of the Fourier transform of the template and the Fourier transform of the data.

**Arrival Time Errors**

Here we wish to localize the occurrence of a function $A(t)$. We will consider this to be a pulse whose arrival time $t_0$ we want to estimate, along with its expected error.

Let the signal be

$$I(t) = a_0 A(t - t_0) + n(t),$$

where $a_0$ is the amplitude and $n(t)$ is zero mean noise.

We will find the time of arrival (TOA) by cross correlating the presumed known pulse shape $A(t)$ with the signal:

$$C_{AI}(\tau) = \int dt I(t) A(t - \tau).$$

First, assume that the signal has been coarsely shifted so that the template is already aligned with the signal and that template is centered on $t = 0$. This way we can assume that the arrival time estimate is a small correction to the coarse estimate.
AN OPTIMIZED ALGORITHM
FOR DETERMINING PULSAR ARRIVAL TIMES

J. H. Taylor
Joseph Henry Laboratories and Physics Department
Princeton University

Berkeley Workshop on Pulsar Timing, June 7-9, 1990

Binary and millisecond pulsars are thought to be relatively old objects, recycled
into active status by the accretion of mass and angular momentum from an evolving
companion star. They have relatively short periods and small period derivatives, and
they make particularly attractive targets for precision timing experiments — partly
because they are even better time-keepers than ordinary pulsars, and partly because
of the special characteristics associated with their complicated evolutionary histo-
ries. My collaborators and I have recently been using accurate measurements of pulse
times of arrival (TOA's) for a wide range of applications including milli-arcsecond
astrometry, tests of fundamental physical laws, cosmology, and timekeeping metro-
logy. Accumulated experience has only strengthened our motivation for achieving the
highest possible accuracies in pulsar timing measurements. In this paper, I describe
an algorithm used at Princeton for the optimum extraction of information from pulsar
timing data.

To a reasonably good approximation, the integrated profiles for a given pulsar at a
particular observing frequency are all the same except for a DC bias, a multiplicative
scale factor, a shift of time origin, and additive random noise. In other words, an
observed profile \( p(t) \) for a given pulsar is related to the corresponding standard profile,
\( s(t) \), by an equation of the form

\[
p(t) = a + b s(t - \tau) + g(t),
\]

where \( a, b, \) and \( \tau \) are constants and \( g(t) \) is a random variable representing radiometer
and background noise. To measure TOA's we want to determine \( \tau \) as accurately as
possible in the presence of a given amount of noise. This might be done by using a
"least squares" or "maximum likelihood" method to fit a scaled, shifted version of
the standard profile to the observed profile, obtaining optimized estimates of the bias,
scale factor, and time offset. The parameter uncertainties would also be calculated;
their magnitudes would depend on the pulse shape and the signal-to-noise ratio.

In practice, observed profiles are digitized representations of the detected signals,
with a well defined sampling interval and known amount of instrumental smoothing.
Let's assume that \( p(t) \) and \( s(t) \) have been recorded at \( N \) equally spaced intervals
covering the full period, say \( t_j = j \Delta t, j = 0, 1, \ldots, N - 1 \). Good experiment design
requires that the detected signals be low-pass filtered at a cutoff frequency \( f_c < \)
Notice that the sample interval, $\Delta t = P/N$, appears nowhere in Eqs. (3-12). Our formalism places no limit on timing accuracy expressed as a fraction of the sampling interval. In contrast, experience has shown that the time-domain methods widely in use do not readily produce arrival-time accuracies much smaller than about $0.1\Delta t$ (see, for example, Rawley 1986).

Figure 1 illustrates the results of fitting a set of 1000 artificially generated pulse profiles, with known time offsets, using both time-domain and frequency-domain algorithms. The frequency-domain uncertainties, represented by filled circles in Figure 1, scale faithfully with the signal-to-noise ratio, reaching values around $0.005\Delta t$ at signal-to-noise ratios of several hundred. On the other hand, the time-domain solutions, in which $\tau$ was estimated by parabolic interpolation of the $\chi^2$ minimum, achieve accuracies no better than about $0.05\Delta t$ even at very high signal-to-noise ratios.

![Graph showing uncertainty in TOA vs. signal to noise ratio]

Figure 1: Average uncertainties obtained for arrival times from simulated observations, plotted as a function of signal-to-noise ratio. Open circles represent TOA's determined by parabolic interpolation in the time domain; filled circles are based on data using the algorithm described in this paper.
1. Least squares in the time domain:

- Sampled profiles, $s_j = s(t_j) = s(j\Delta t)$, $j = 0, \ldots, N - 1$ ($N\Delta t =$ period).

- Find minimum with respect to $\tau$ of

$$\chi^2(a, b, \tau) = \sum_{j=0}^{N-1} \left( \frac{p_j - a - bs_{j-\tau}}{\sigma} \right)^2$$

- Equivalent: find maximum of CCF,

$$c(\tau) = p_j \ast s_j$$

⇒ Problem: profiles are quantized in time; cannot get analytic derivative $\frac{\partial \chi^2}{\partial \tau}$. 
We can expand the signal as
\[ I(t) \approx a_0 A(t) - a_0 t_0 A'(t) + n(t). \]
We also expand the template, but to second order in \( \tau \) because we will be taking a derivative to find the lag of maximum correlation:
\[ A(t - \tau) \approx A(t) - \tau A'(t) + \frac{\tau^2}{2} A''(t) \]
\[ A'(t - \tau) \approx A'(t) - \tau A''(t). \]

Then we can write
\[ C'_{IA}(\tau) = \frac{d}{d\tau} C_{AI}(\tau) = 0 \]
\[ = \int dt \; I(t) A'(t - \tau) \]
\[ \approx C_{IA'}(0) - \tau C_{IA''}(0) \]

An estimator for \( \tau \) is then
\[ \tau = \frac{C_{IA'}(0)}{C_{IA''}(0)} = \frac{a_0 C_{AA'}(0) - a_0 t_0 C_{A'A'}(0) + C_{nA'}(0)}{a_0 C_{AA''} - a_0 t_0 C_{A'A''}(0) + C_{nA''}(0)}. \]

Using previous approximations we encounter the terms \( C_{AA'} \) and \( C_{A'A''}(0) \) that vanish for pulses that are zero at \( \pm \infty \). Also the \( C_{nA''} \) term in the denominator yields a second-order term that can be ignored. Then the TOA estimator becomes
\[ \tau = \frac{-a_0 t_0 C_{A'A'}(0) + C_{nA'}(0)}{a_0 C_{AA''}(0)}. \]
**Noiseless case:**

When there is no noise we have

\[ \tau = -t_0 \frac{C_{A' A'}(0)}{C_{AA''}(0)}. \]

Using a trial function, such as a Gaussian shape, it can be shown that \( \tau = t_0 \), as we would expect!

Even better, the denominator can be integrated by parts to show that \( C_{AA''}(0) = -C_{A' A'}(0) \) for pulses that vanish at \( \pm \infty \), so the equality is general.

**With noise:**

We can now write

\[ \tau = t_0 + \frac{C_{n A'}(0)}{a_0 C_{AA''}(0)}. \]
Then the mean-square TOA error is, using $\delta \tau = \tau - t_0$,

$$\left\langle (\delta \tau)^2 \right\rangle = \frac{C_{nA'}^2(0)}{a_0^2 C_{AA'}^2(0)}$$

$$= \frac{C_{nA'}^2(0)}{a_0^2 C_{A'\alpha}^2(0)}$$

$$= \int \int dt dt' \langle n(t) n(t') \rangle A(t) A(t')$$

$$= \frac{a_0^2 C_{A'\alpha}^2(0)}{a_0^2 C_{A'\alpha}^2(0)}.$$
Now assume *white noise* (for specificity) so that

\[ \langle n(t)n(t') \rangle = \sigma_n^2 w_n \delta(t-t') \]

where \( w_n \) is a short characteristic time scale (such as an inverse bandwidth) to keep the units correct. Then

\[
\left\langle (\delta \tau)^2 \right\rangle = \left( \frac{\sigma_n}{a_0} \right)^2 \frac{w_n}{C_{A'A'}(0)} \int dt A'^2(t)
\]

\[
= \left( \frac{\sigma_n}{a_0} \right)^2 \frac{w_n}{C_{A'A'}(0)} 
\]

We can then write this out as a TOA error

\[
\sigma_\tau = \left( \frac{\sigma_n}{a_0} \right) \left[ \int dt \left[ A'(t) \right]^2 \right]^{1/2}
\]

\[
= \frac{1}{\text{SNR}} \left[ \int dt \left[ A'(t) \right]^2 \right]^{1/2}.
\]

We see that the error scales as the inverse of the *signal to noise ratio* (SNR). The denominator also involves the integral of the squared derivative of the pulse shape, suggesting that *sharper* pulses with larger derivatives will produce smaller arrival time errors.
**Localization: Time vs. Frequency Domains**

For a Nyquist sampled, bandlimited process with bandwidth $B$ the sampling theorem implies that the continuous-time signal can be reconstructed from the sampled data as

$$x(t) = \sum_{-\infty}^{\infty} x_n \text{sinc}(t - n\Delta t)/\Delta t$$

where $\Delta t = 1/B$ and, as usual, $\text{sinc} x \equiv (\sin \pi x)/x$.

Consider a model appropriate for matched filtering,

$$x(t) = aA(t - t_0) + n(t),$$

where $a$ is the amplitude, $A(t)$ is the known template, and $n$ is noise.

Suppose we have sampled versions of the data and template, $x_n, A_n$.

One approach is to calculate the discrete CCF

$$C_{xA}(\ell) = \frac{1}{N} \sum_n x_n A_{n-\ell}$$

and find the maximum to determine an estimate $\hat{t}_0 = \ell_{max}$.

However, we really want the CCF of the *continuous time* quantities, $x(t)$ and $A(t)$,

$$C_{xT}(\tau) = \int dt \ x(t)A(t - \tau).$$
We cannot get the true lag of maximum correlation, $\tau_{\text{max}}$, by interpolating the sampled correlation function unless we interpolate according to the sampling theorem. If we interpolate differently, we will get biased results.

**Another approach: the frequency domain**

Take the FT of the model equation to get

$$\tilde{X}(f) = a\tilde{A}(f)e^{-2\pi ift_0} + \tilde{n}(f).$$

**No noise:** Write the FT in terms of its real and imaginary parts and find the phase:

$$\tilde{X}(f) = \tilde{X}_r(f) + i\tilde{X}_i(f) = |\tilde{X}(f)|e^{i\phi(f)}$$

$$\phi(f) = \tan^{-1}\left[\frac{\tilde{X}_i(f)}{\tilde{X}_r(f)}\right] = \tan^{-1}\left[\frac{-\sin 2\pi ft_0}{\cos 2\pi ft_0}\right] = -2\pi ft_0.$$

See example.

**With noise:** The phase will have a noise-like error that is nonlinearly related to $\tilde{n}(f)$. In the limit of large SNR, the rms phase error will scale as $1/\text{SNR}$.

Working directly with the phase to determine $t_0$, however, is numerically problematic because the phase will wrap for large offsets and for low SNR.
$W = 2 \quad t_0 = 0 \text{ samples} \quad \text{SNR} = 50$
$W = 2 \quad t_0 = 1.99000248$ samples \quad SNR = 100000
$W = 2 \quad t_0 = 1.990000248$ samples \quad SNR = 50
$W = 2 \quad t_0 = 8 \text{ samples} \quad \text{SNR} = 10$

Amplitude vs. Time

Amplitude vs. Time (samples)

$|\text{FFT}|$ vs. Frequency

Phase $\phi$ vs. Frequency
$W = 2 \quad t_0 = 8 \text{ samples} \quad \text{SNR} = 50$
\begin{align*}
W &= 5 & t_0 &= 8 \text{ samples} & \text{SNR} &= 50 \\
\text{Amplitude} & \\
\text{Time} & \\
\text{Amplitude} & \\
\text{Time (samples)} & \\
|\text{FFT}| & \\
\text{Frequency} & \\
\text{Phase } \phi & \\
\text{Frequency} & 
\end{align*}
**Best approach:** Fit to the complex FFT rather than to the phase.

Use as a model

\[
\tilde{M}(f) = \tilde{A}(f)e^{-2\pi if\hat{t}_0}
\]

Then define the product

\[
\tilde{J}(f) = \tilde{X}(f)\tilde{M}^*(f) = a|\tilde{A}(f)|^2e^{-2\pi if(t-\hat{t}_0)} + \tilde{n}(f)\tilde{A}^*(f)e^{2\pi if\hat{t}_0}
\]

and integrate over frequency:

\[
S \equiv \int_{\Delta f} df \tilde{J}(f)
\]

\[
= a \int_{\Delta f} df |\tilde{A}(f)|^2e^{-2\pi if(t_0-\hat{t}_0)} + \int_{\Delta f} df \tilde{n}(f)\tilde{A}^*(f)e^{2\pi if\hat{t}_0}.
\]

This quantity can be maximized vs. \( \hat{t}_0 \). For no noise, we expect \( \hat{t}_0 = t_0 \). The maximum can be found using standard search methods over the nonlinear parameter \( \hat{t}_0 \) (grid search, linearization, etc.).

Note that the integrand is naturally weighted by the actual signal.

An equivalent approach is to use a different test statistic,

\[
S' = \int_{\Delta f} df \left| \tilde{X}(f) - \tilde{M}(f) \right|^2,
\]

that we would minimize to find \( \hat{t}_0 \).
Note that we have used continuous notation here. With sampled data we can reconstruct the continuous FT as

\[ \hat{X}(f) = \Pi(f/B) \sum_n x_n e^{-2\pi i f n \Delta t}, \]

which can be implemented using the DFT.
AN OPTIMIZED ALGORITHM
FOR DETERMINING PULSAR ARRIVAL TIMES

J. H. Taylor
Joseph Henry Laboratories and Physics Department
Princeton University

Berkeley Workshop on Pulsar Timing, June 7-9, 1990

Binary and millisecond pulsars are thought to be relatively old objects, recycled
into active status by the accretion of mass and angular momentum from an evolving
companion star. They have relatively short periods and small period derivatives, and
they make particularly attractive targets for precision timing experiments — partly
because they are even better time-keepers than ordinary pulsars, and partly because
of the special characteristics associated with their complicated evolutionary histo-
ries. My collaborators and I have recently been using accurate measurements of pulse
times of arrival (TOA's) for a wide range of applications including milli-arcsecond
astrometry, tests of fundamental physical laws, cosmology, and timekeeping metro-
logy. Accumulated experience has only strengthened our motivation for achieving the
highest possible accuracies in pulsar timing measurements. In this paper, I describe
an algorithm used at Princeton for the optimum extraction of information from pulsar
timing data.

To a reasonably good approximation, the integrated profiles for a given pulsar at a
particular observing frequency are all the same except for a DC bias, a multiplicative
scale factor, a shift of time origin, and additive random noise. In other words, an
observed profile \( p(t) \) for a given pulsar is related to the corresponding standard profile,
\( s(t) \), by an equation of the form

\[
p(t) = a + b s(t - \tau) + g(t),
\]

where \( a, b, \) and \( \tau \) are constants and \( g(t) \) is a random variable representing radiometer
and background noise. To measure TOA's we want to determine \( \tau \) as accurately as
possible in the presence of a given amount of noise. This might be done by using a
"least squares" or "maximum likelihood" method to fit a scaled, shifted version of
the standard profile to the observed profile, obtaining optimized estimates of the bias,
scale factor, and time offset. The parameter uncertainties would also be calculated;
their magnitudes would depend on the pulse shape and the signal-to-noise ratio.

In practice, observed profiles are digitized representations of the detected signals,
with a well defined sampling interval and known amount of instrumental smoothing.
Let's assume that \( p(t) \) and \( s(t) \) have been recorded at \( N \) equally spaced intervals
covering the full period, say \( t_j = j\Delta t, j = 0, 1, \ldots, N - 1 \). Good experiment design
requires that the detected signals be low-pass filtered at a cutoff frequency \( f_c \leq \)
Notice that the sample interval, $\Delta t = P/N$, appears nowhere in Eqs. (3-12). Our formalism places no limit on timing accuracy expressed as a fraction of the sampling interval. In contrast, experience has shown that the time-domain methods widely in use do not readily produce arrival-time accuracies much smaller than about $0.1\Delta t$ (see, for example, Rawley 1986).

Figure 1 illustrates the results of fitting a set of 1000 artificially generated pulse profiles, with known time offsets, using both time-domain and frequency-domain algorithms. The frequency-domain uncertainties, represented by filled circles in Figure 1, scale faithfully with the signal-to-noise ratio, reaching values around 0.005$\Delta t$ at signal-to-noise ratios of several hundred. On the other hand, the time-domain solutions, in which $\tau$ was estimated by parabolic interpolation of the $\chi^2$ minimum, achieve accuracies no better than about 0.05$\Delta t$ even at very high signal-to-noise ratios.

![Figure 1: Average uncertainties obtained for arrival times from simulated observations, plotted as a function of signal-to-noise ratio. Open circles represent TOA's determined by parabolic interpolation in the time domain; filled circles are based on data using the algorithm described in this paper.](image)

m4

4
**Prewhitening**

**What is Prewhitening?** Prewhitening is an operation that processes a time series (or some other data sequence) to make it behave statistically like white noise. The ‘pre’ means that whitening precedes some other analysis that likely works better if the additive noise is white.

These operations can be viewed in either the time domain or the frequency domain:

1. Make the ACF of the time series appear more like a delta function.
2. Make the spectrum appear flat.

Example data sets that may require prewhitening:

1. A well behaved noise process with an additive low frequency (or polynomial) trend added to it.
2. A deterministic signal with an additive red-noise process.

Viewed in the frequency domain, prewhitening means that the *dynamic range* of the measured data is reduced.
Why bother? Recall from our discussions of spectral analysis the issues of leakage and bias. These arise from sidelobes inherent to spectral estimation. We can minimize leakage in two ways: (1) make sidelobes smaller and (2) minimize the power that is prone to leaking into sidelobes. Spectral windows address the former while prewhitening mitigates the latter. Leakage into sidelobes also constitutes bias in spectral estimates. However bias appears in other data analysis procedures. Consider least-squares fitting of a sinusoid to a signal of the form

\[ x(t) = A \cos(\omega t + \phi) + r(t) + n(t), \]

where \( n(t) \) is WSS white noise and \( r(t) \) is red noise with a steep power spectrum. Red noise can strongly bias fitting of a model \( \hat{x}(t) = \hat{A} \cos(\hat{\omega} + \hat{\phi}) \) because its power can leak across the underlying spectrum causing a least-square fit to give highly discrepant values of \( \hat{A}, \hat{\omega}, \) and \( \hat{\phi} \).

Prewhitening of the time series ideally would yield a transformed time series of the form

\[ x'(t) = A' \cos(\omega t + \phi) + n'(t) \]

to which fitting a sinusoidal model will be less biased.
Procedures:

We have already seen one analysis that is related to prewhitening: the matched filter (MF). The MF doesn’t whiten the spectrum of the output but it does weights the frequency components of the measured quantity to maximize the S/N of the signal.

The signal model in this case is \( x(t) = a_0 A(t) + n(t) \). Recall for an arbitrary spectrum \( S_n(f) \) for additive noise that the frequency-domain MF for a signal \( A(t) \) is

\[
\tilde{h}(f) \propto \frac{\tilde{A}(f)}{S_n(f)}.
\]

Taking equality for simplicity, when the filter is applied to the measurements \( x(t) \), we have

\[
\tilde{y}(f) = \tilde{x}(f) \tilde{h}^*(f) \propto \frac{a_0|\tilde{A}(f)|^2}{S_n(f)} + \frac{\tilde{n}(f) \tilde{A}^*(f)}{S_n(f)}.
\]

This means that the ensemble-average spectrum of the filter output is

\[
\langle |\tilde{y}(f)|^2 \rangle = \frac{a_0^2|\tilde{A}(f)|^4}{S_n^2(f)} + \frac{\langle |\tilde{n}(f)|^2 \rangle |\tilde{A}(f)|^2}{S_n^2(f)}
\]

\[
= \frac{a_0^2|\tilde{A}(f)|^4}{S_n^2(f)} + \frac{S_n(f) |\tilde{A}(f)|^2}{S_n^2(f)}
\]

\[
= \frac{a_0^2|\tilde{A}(f)|^4}{S_n^2(f)} + \frac{|\tilde{A}(f)|^2}{S_n(f)}
\]

\[
= \frac{|\tilde{A}(f)|^2}{S_n(f)} \left[ a_0^2|\tilde{A}(f)|^2 + 1 \right].
\]
Signals with trends: A common situation is where a quantity of the form $a_0 A(t) + n(t)$ is superposed with a strong trend, such as a baseline variation. Similar issues arise in measurements of spectra.

Consequences of trends include:

1. **Bias** in estimating parameters of $A(t - t_0)$ or its spectral analog $A(\nu - \nu_0)$.

2. Erroneous estimates of cross correlations between two time series such as

$$x(t) = s_1(t) + n_1(t) \quad \text{and} \quad y(t) = s_2(t) + n_2(t),$$

where $s_{1,2}$ are signals of interest and $n_{1,2}$ are measurement errors. I.e. we may be interested in the correlation

$$C = \frac{1}{N_t} \sum_t s_1(t)s_2(t) \quad \text{or} \quad C = \frac{1}{N_t} \sum_t [s_1(t) - \bar{s}_1][s_2(t) - \bar{s}_2]$$

where $\bar{s}_{1,2} = (1/N_t) \sum_t s_{1,2}(t)$ are the sample means.

If there are trends $p_{1,2}(t)$ added to $x(t)$ and $y(t)$ the correlation $\hat{C}$ of $x$ and $y$ used to estimate $C$ may be dominated completely by the trends and not the signal parts of the measurements.

A fix: Trends can often be modeled as a polynomial of some order that can be fitted to the measurements. The order of the polynomial needs to be chosen ‘wisely.’ For a pulse or spectral line confined to some range of $t$ or $\nu$ this is straightforward. But for a detection problem where the signal location is not known, the situation is very tricky.
**Prewhitening filter:** Consider again \( x(t) = a_0 A(t) + n(t) \) and let’s trivially construct a frequency-domain filter that whitens the measurements.

We want a filter \( h(t) \) that flattens the noise \( n(t) \) in the frequency domain. Let \( y(t) = x(t) \otimes h(t) \) where \( \otimes \) means convolution. All we need is \( \tilde{h}(f) = \sqrt{S_n(f)} \). Then the ensemble spectrum of the output \( \tilde{y}(f) \) is

\[
\langle |\tilde{y}(f)|^2 \rangle = \langle |\tilde{x}(f)|^2 \rangle \langle |\tilde{h}(f)|^2 \rangle = \frac{\langle |\tilde{x}(f)|^2 \rangle}{S_n(f)} = \frac{a_0^2 \langle |\tilde{A}(f)|^2 \rangle}{S_n(f)} + 1
\]

Note how this differs from the result for a matched filter. But the result is that in the mean the spectrum of the additive noise has been flattened.

Prewhitening is important in both detection and estimation applications.
Leakage and Bias
Prewhitening in the least-squares estimation context:

Consider our standard linear model

\[ y = X\theta + n, \]

which has a least-squares solution for the parameter vector

\[ \theta = (X^\dagger C_n^{-1}X)^{-1}X^\dagger C_n^{-1}y, \]

where the covariance matrix of the noise vector \( n \) is

\[ C_n = \langle nn^\dagger \rangle. \]

This is also the maximum likelihood solution in the right circumstances (which are?).

As with any covariance matrix, \( C_n \) is Hermitian and positive, semi-definite. This means that the quadratic form for an arbitrary vector \( z \) satisfies

\[ z^\dagger C_n z \geq 0. \]

Such matrices can always be factored according to the Cholesky decomposition:

\[ C_n = LL^\dagger \]

where \( L \) is a lower-diagonal matrix; e.g.

\[ L = \begin{pmatrix}
    a & 0 & 0 & 0 \\
    b & c & 0 & 0 \\
    d & e & f & 0 \\
    g & h & i & j \\
\end{pmatrix}. \]
Utility: we can transform the model as follows using $\mathbf{L}$:
\[
\mathbf{y} = \mathbf{L}
\mathbf{y}_w \\
\mathbf{X} = \mathbf{L}
\mathbf{X}_w.
\]
Substituting into the solution vector for $\theta$ and using
\[
\mathbf{y}^\dagger = (\mathbf{L}
\mathbf{y}_w)^\dagger = \mathbf{y}_w^\dagger
\mathbf{L}^\dagger, \quad
\mathbf{X}^\dagger = (\mathbf{L}
\mathbf{X}_w)^\dagger = \mathbf{X}_w^\dagger
\mathbf{L}^\dagger, \quad \text{and} \quad
\mathbf{C}_n^{-1} = (\mathbf{LL}^\dagger)^{-1} = \mathbf{L}^\dagger
\mathbf{L}^{-1}
\]
yields
\[
\theta = (\mathbf{X}^\dagger
\mathbf{C}_n^{-1}
\mathbf{X})^{-1}
\mathbf{X}^\dagger
\mathbf{C}_n^{-1}
\mathbf{y}
\]
\[
= (\mathbf{X}_w^\dagger
\mathbf{L}^\dagger
\mathbf{C}_n^{-1}
\mathbf{L}
\mathbf{X}_w)^{-1}
\mathbf{X}_w^\dagger
\mathbf{L}^\dagger
\mathbf{C}_n^{-1}
\mathbf{L}
\mathbf{X}_w
\]
\[
= (\mathbf{X}_w^\dagger
\mathbf{X}_w)^{-1}
\mathbf{X}_w^\dagger
\mathbf{y}.
\]

**So what?** The solution is identical to the least-squares case where the noise covariance matrix is diagonal; i.e. the noise vector $\mathbf{n}_w = \mathbf{L}
\mathbf{n}$ has been transformed to white noise. We have whitened the data.

**When is this useful?** An example is the fitting of a sinusoidal function amid red noise where leakage effects are important just as they are for spectral analysis. A specific example is the fitting of astrometric parameters or periodicities in radial velocity data.

**What’s the catch?** You need to know the covariance matrix of the noise $\mathbf{n}$ to do the Cholesky decomposition. This can be easier said than done!