Lecture 19

- Localization: cross correlation and least-squares approaches
- Prewhitening
- Source finding in surveys
Matched Filtering in Astronomy

- Point source detection in surveys
- Asteroid detection
- Gravitational wave detection
- Galaxy detection
- Galaxy cluster finding
- Match filtering approach for signal acquisition in radio-pulsar navigation
- Large scale structure in the universe
- Cluster detection in databases
- Radio images
- Precision localization (time, frequency, space)
  - Pulsar timing
  - Radial velocities and exoplanets
  - Astrometry
Matched Filtering and Detection

- MF gives the optimal S/N and, if required, the best point estimates for localization.
- **Localization** tasks (timing, Doppler shifts) are less concerned with the actual amplitude. Different cases include:
  - Template known: find translation parameter via least-squares or by cross correlation, paying attention to any needed interpolation. The highest precision is far easier in the Fourier domain.
  - Template not known: can use a family of templates (a template bank ordered by one or more parameters) and find maximum S/N over the set.
- **Detection** is more dependent on S/N than on localization precision.
  - Detection criteria:
    - S/N > mσ where m corresponds to some false-alarm probability if the underlying PDF is known or can be assumed to be Gaussian.
    - Want $N_{\text{trials}}P_{\text{fa}} < 1$ in an analysis with $N_{\text{trials}}$ statistical trials.
  - Issue: If the template is not known, a large template bank increases $N_{\text{trials}}$ significantly.
  - E.g. detection of chirped waveforms in LIGO data (NS-NS, NS-BH binaries): templates ~known from GR, but depend on unknown masses; the occurrence time is also not known so the search parameters include: $t_0 = \text{event time}$ and the chirp mass; also the amplitude.
  - One can calculate the probability of detection $P_{\text{det}}$ for different detection criteria used on the same data set. Similarly for the null case (no signal) one can calculate the false-alarm probability for each detection scheme. A plot of $P_{\text{det}}$ vs $P_{\text{fa}}$ is an ROC curve.
Localization in Space, Time, or Frequency

Matched Filtering

Matched filtering is an optimal method for detecting a signal of known shape in the presence of additive noise. It also plays a role in estimation of signal parameters.

We will use a time-domain signal as a prototype. Consider the model where $A$ is deterministic and known and $n$ is additive, zero-mean WSS noise with arbitrary spectrum:

$$x(t) = a_0A(t) + n(t).$$

We want a filter that whose output maximizes the signal-to-noise ratio of the output.

Matched filtering is different from Wiener filtering, which yields an estimate for a signal that has a minimum least-squares error from the true signal.

Signals in other domains can be treated identically:

- Frequency domain: $x(\nu) = a_0A(\nu) + n(\nu)$.
- Image domain: $I(\theta) = a_0A(\theta) + n(\theta)$.
- Spatial domain: $I(x) = a_0A(x) + n(x)$.
- Arbitrary domain: $U(y) = a_0A(y) + n(y)$ where (e.g.) $y = (x, \nu, t)$. 


Figure 2: Matched filtering of a narrower Gaussian pulse. One realization of the template and pulse are shown while ten realizations of the CCF are shown. The SNR of the pulse is 5 (peak to rms noise). The pulse width (HW @ 1/e) is 10.3 samples.

**Detection:** find $CCF_{\text{max}}$ and $\text{rms}(CCF)$ and apply S/N threshold

$\rightarrow$ Detection probability and false-alarm probability

**Localization:** find location of $CCF$ maximum

General result: $\text{rms localization} = \text{width of } CCF / (\text{S/N of } CCF) $
**ROC Curves:**

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

![ROC Curve](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.
Arbitrary WSS Noise: Generally, the solution is gotten by Fourier transforming Equation 3. Let the noise spectrum be $R_n(\tau) \iff S_n(f)$ and define the FTs of $A$ and $h$ as $\tilde{A}$ and $\tilde{h}$. Using 3 (repeated here),

$$\int dt' R_n(t - t') h(t') = \frac{a_0 \sigma^2_y}{\langle y \rangle} A(t),$$

we have

$$\text{FT} \left\{ \int dt' R_n(t - t') h(t') \right\} = S_n(f) \tilde{h}(f),$$

where

$$\sigma_y^2 = \iint dt dt' R_n(t - t') h(t) h(t') = \int df S_n(f) |\tilde{h}(f)|^2$$

and

$$\langle y \rangle = a_0 \int dt A(t) h(t) = a_0 \int df \tilde{A}(f) \tilde{h}^*(f).$$

so the expression for $h$ becomes

$$\tilde{h}(f) = \frac{\tilde{A}(f) a_0 \sigma^2_y}{S_n(f) \langle y \rangle} = \frac{\tilde{A}(f)}{S_n(f)} \frac{\int df S_n(f) |\tilde{h}(f)|^2}{\int df \tilde{A}(f) \tilde{h}^*(f)}.$$

and the solution is

$$\tilde{h}(f) = \text{constant} \times \frac{\tilde{A}(f)}{S_n(f)}.$$  

This filter “whitens” the data, in effect, to yield an optimal test statistic.
Comments on the general matched filtering solution

• For white noise, $S_n(f) = \text{constant}$ and we get our previous result. More generally, the matched filter favors frequencies where the ratio of signal to noise is larger.

• Equivalently the filter reduces the noise in the output $y$ at frequencies where the noise spectrum is large and increases it where it is small.

• The amplitude of the signal $a_0$ does not appear in the solution. It can be considered to be part of the proportionality constant (normalization). Also the actual noise variance cancels in the solution for $\tilde{h}(f)$.

  Thus the amplitudes of both the signal and the noise factor out of the solution.

• The matched filter $\tilde{h}(f)$ has an inverse Fourier transform $h(t)$ that is the optimal smoothing function. It maximizes the S/N of the output.

  – The OSF depends on both the signal and the noise.

  – For red noise with a steep power spectrum, the low frequencies are de-weighted in the filtering compared to the high frequencies.

  – Usage of Fourier methods is still susceptible to leakage effects so prewhitening methods need to be used.
When the Template is Unknown

Matched filtering is based on knowing the shape of the signal in the appropriate domain. What if you don’t know the shape?

1. Guess!
   (a) While the MF is optimal, a detection experiment may not depend critically on the detailed shape.
   (b) Maybe you know the shape but not the size (width, etc.). In that case, use a family with different shape parameters.
   (c) You may not even know the actual shape. In that case, use a family of default shapes with different shape parameters.

2. Perhaps you are searching for a wide variety of shapes (e.g. stars, galaxies, nebulae, filaments, etc.) in images. One goal might be that you are trying to detect and classify everything you see in an image or data cube.
   (a) You will have to use multiple families of shapes with a range of parameters for each.
   (b) What are the consequences of seeking a wider variety of shapes? (Hint: number of statistical trials.)
   (c) Think about how you would go about setting up a detection pipeline for this kind of survey.
Additive Noise in MF Applications

• Generally the matched filter depends on both the signal $A$ and the noise $n$

• Specifically, the signal $A$ is assumed to be deterministic while the noise is considered WSS with a known spectrum $S_n$.
  – In some applications, $A$ is stochastic but has a well-defined ensemble average $\langle A \rangle$
  – $S_n$ is the *ensemble average* spectrum of the noise

• What if we do not know $S_n$ a priori?

• How do we proceed?
Noise Spectrum in Matched Filtering Contexts

• Guess the shape of the noise spectrum
• Argue that the CLT applies:
  – Time series: measurement errors often Gaussian
  – Images:
    • Photon counts in detector >> 1: Poisson → Gaussian
    • Images from visibility functions via van Cittert-Zernike theorem: FT → sum of large N random variables
• Other cases:
  – May need to bootstrap from measured data (use the same data to estimate $S_n$ (tricky … chicken and egg issues)
  – Ancillary data → estimate for $S_n$
    • E.g. Confusion noise from astrophysical sources
      – Power-law amplitude distribution
      – Spectrum = shape of resolution function (dirty beam)
  – Coupling of temporal phenomena with image formation
    • Calibration errors that confound self-calibration algorithms
For circular apertures first null is at $1.22 \lambda/a$
Source Confusion

Figure 6. Confusion profile. This profile plot shows the 3 GHz confusion amplitude in an 8 arcsec FWHM beam, truncated at μmJy beam\(^{-1}\).
Active Galactic Nuclei

Inner Structure of an Active Galaxy
- Shock
- Relativistic Jet
- Supermassive Black Hole
- Accretion Disk
- Opaque Torus (Inner Regions)

FR Class I source: radio galaxy 3C31

FR Class II source: quasar 3C175
Cygnus A

Radio image
VLA, NM
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, *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 **linear** in \( \hat{a} \) but is **nonlinear** in \( \hat{t}_0 \).
Minimizing $Q$ with respect to $\hat{a}$, we have

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

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

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

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

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

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

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

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

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

**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
\[
\hat{a} \int dt \left[ A^2(t - t_0) - 2\delta\hat{t}_0 A(t - t_0) A'(t - t_0) \right] = \\
\int dt I(t) \left[ A(t - t_0) - \delta\hat{t}_0 A'(t - t_0) \right] 
\]
\[
\hat{a} \int dt \left[ A(t - t_0) A'(t - t_0) - \delta\hat{t}_0 A(t - t_0) A''(t - t_0) - \delta\hat{t}_0 A^2(t - t_0) \right] = \\
\int dt I(t) \left[ A'(t - t_0) - \delta\hat{t}_0 A''(t - t_0) \right].
\]

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
\[
\frac{1}{2} A^2(t - t_0) \bigg|_{t_1}^{t_2} \to 0
\]
in the limit of (e.g.) \( t_{1,2} = \mp T/2 \) with \( T \gg \) pulse width.
We then have
\[
\hat{a} \int dt \, A^2(t - t_0) = \int dt \, I(t)[A(t - t_0) - \delta \dot{t}_0 A'(t - t_0)]
\]
\[-\delta \dot{t}_0 \hat{a} \int dt \, [A(t - t_0)A''(t - t_0) + A^2(t - t_0)] = \int dt \, I(t)[A'(t - t_0) - \delta \dot{t}_0 A''(t - t_0)].
\]

Solving for \(\hat{a}\) in both cases we have
\[
\hat{a} = \frac{\int dt \, [I(t)A(t - t_0) - \delta \dot{t}_0 I(t)A'(t - t_0)]}{\int dt \, A^2(t - t_0)} \quad (6)
\]
\[
\hat{a} = \frac{\int dt \, I(t) [-A'(t - t_0) + \delta \dot{t}_0 A''(t - t_0)]}{\delta \dot{t}_0 \int dt \, [A(t - t_0)A''(t - t_0) + A^2(t - t_0)]}. \quad (7)
\]
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

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

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 \textit{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.

\section*{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 histories. 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 metrology. 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 \( f_c \leq \frac{1}{2 \pi \Delta t} \).
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 as a function of signal-to-noise ratio.](image)

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, \( s_j = s(t_j) = s(j \Delta t) \),
  \( j = 0, \ldots, N - 1 \) \( (N \Delta t = \text{period}) \).

- 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,
  \[
  c(\tau) = p_j \ast s_j
  \]

\[\Rightarrow\] Problem: profiles are quantized in time; cannot get analytic derivative \( \partial \chi^2 / \partial \tau \).
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

$$\tau = t_0 + \frac{C_{nA'}(0)}{a_0 C_{AA''}(0)}.$$

31
Then the mean-square TOA error is, using $\delta \tau = \tau - t_0$,

$$
\langle (\delta \tau)^2 \rangle = \frac{C_{nA'}^2(0)}{a_0^2 C_{AA'}^2(0)}
= \frac{C_{nA'}^2(0)}{a_0^2 C_{AA'}^2(0)}
= \frac{\int \int dt dt' \langle n(t)n(t') \rangle A(t)A(t')}{a_0^2 C_{AA'}^2(0)}.
$$
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
\[
\langle (\delta\tau)^2 \rangle = \left( \frac{\sigma_n}{a_0} \right)^2 \frac{w_n \int dt A'^2(t)}{C_{A'A'}^2(0)}
\]
\[
= \left( \frac{\sigma_n}{a_0} \right)^2 \frac{w_n}{C_{A'A'}(0)}
\]
We can then write this out as a TOA error
\[
\sigma_\tau = \left( \frac{\sigma_n}{a_0} \right) \left[ \frac{w_n}{\int dt [A'(t)]^2} \right]^{1/2}
\]
\[
= \frac{1}{\text{SNR}} \left[ \frac{w_n}{\int dt [A'(t)]^2} \right]^{1/2}.
\]
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

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

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 esimate $\hat{t}_0 = \ell_{\text{max}}$.

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

\[
\hat{X}(f) = a\hat{A}(f)e^{-2\pi ift_0} + \hat{n}(f).
\]

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

\[
\hat{X}(f) = \hat{X}_r(f) + i\hat{X}_i(f) = |\hat{X}(f)|e^{i\phi(f)}
\]

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

See example.

**With noise:** The phase will have a noise-like error that is nonlinearly related to \( \hat{n}(f) \). 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.
$W = 2 \quad t_0 = 0 \text{ samples} \quad \text{SNR} = 50$

Amplitude

Time

Amplitude

Time (samples)

$|\text{FFT}|$

Frequency

Phase $\phi$

Frequency
$W = 2 \quad t_0 = 1.990000248 \text{ samples} \quad SNR = 100000$
\[ W = 2 \quad t_0 = 1.990000248 \text{ samples} \quad \text{SNR} = 50 \]
$W = 2 \quad t_0 = 8 \text{ samples} \quad \text{SNR} = 10$
$W = 5 \quad t_0 = 8 \text{ samples} \quad \text{SNR} = 50$
**Best approach:** Fit to the complex FFT rather than to the phase.

Use as a model

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

Then define the product

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

and integrate over frequency:

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

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

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,

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

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

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

which can be implemented using the DFT.