Lecture 18

– Matched filtering
– Localization
Kolmogorov-Smirnov like test for time-frequency Fourier spectrogram analysis in LISA Pathfinder

Luigi Ferraioli · Michele Armano · Heather Audley · Giuseppe Congedo · Ingo Diepholz · Ferran Gibert · Martin Hewitson · Mauro Hueller · Nikolaos Karnesis · Natalia Korsakova · Miquel Nofrarias · Eric Plagnol · Stefano Vitale

Received: 15 April 2013 / Accepted: 1 December 2014 / Published online: 19 December 2014
© Springer Science+Business Media Dordrecht 2014

Abstract A statistical procedure for the analysis of time-frequency noise maps is presented and applied to LISA Pathfinder mission synthetic data. The procedure is based on the Kolmogorov-Smirnov like test that is applied to the analysis of time-frequency noise maps produced with the spectrogram technique. The influence of the finite size windowing on the statistic of the test is calculated with a Monte Carlo simulation for 4 different windows type. Such calculation demonstrate that the test statistic is modified by the correlations introduced in the spectrum by the finite size of the window and by the correlations between different time bins originated by overlapping between windowed segments. The application of the test procedure to LISA Pathfinder data demonstrates the test capability of detecting non-stationary features
Performance analysis of GPU-accelerated filter-based source finding for HI spectral line image data

Stefan Westerlund · Christopher Harris

Abstract Searching for sources of electromagnetic emission in spectral-line radio astronomy interferometric data is a computationally intensive process. Parallel programming techniques and High Performance Computing hardware may be used to improve the computational performance of a source finding program. However, it is desirable to further reduce the processing time of source finding in order to decrease the computational resources required for the task. GPU acceleration is a method that may achieve significant increases in performance for some source finding algorithms, particularly for filtering image data. This work considers the application of GPU acceleration to the task of source finding and the techniques used to achieve the best performance, such as memory management. We also examine the changes in performance, where the algorithms that were GPU accelerated achieved a speedup of around 3.2 times the 12 core per node CPU-only performance, while the program as a whole experienced a speedup of 2.0 times.

Keywords Source finding · Radio astronomy · High performance computing · Parallel processing · GPU computing
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 > \( m \sigma \) 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: \( t_0 \) = 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.
Matched Filtering

- Template too wide
- Template too narrow
- Template matched
Timing Error from Radiometer Noise

Rms TOA error from template fitting with additive noise:

$$\Delta t_{S/N} = \frac{\int \int dt dt' \rho(t - t')U'(t)U'(t')}{\text{SNR} \int dt [U'(t)]^2} \frac{1}{\sqrt{2\pi \ln 2}} \frac{\Delta \rho}{W \sqrt{N}} = \frac{W_{\text{eff}}}{\text{SNR}} \left( \frac{\Delta}{W_{\text{eff}}} \right)^{1/2}$$

Gaussian shaped pulse:

$$\Delta t_{S/N} = \frac{W}{(2\pi \ln 2)^{1/4} \text{SNR}_1 \sqrt{N}} \left( \frac{\Delta}{W} \right)^{1/2}$$

$$\Delta t_{S/N} = 0.69 \mu s W_{\text{ms}} N_6^{-1/2} \text{SNR}_1^{-1} (\Delta/W)^{1/2}$$

Interstellar pulse broadening, when large, increases $\Delta t_{S/N}$ in two ways:

- SNR decreases by a factor $W / [W^2 + \tau_d^2]^{1/2}$
- $W$ increases to $[W^2 + \tau_d^2]^{1/2}$

→ Large errors for high DM pulsars and low-frequency observations

Low-DM pulsars: DISS (and RISS) will modulate SNR

$N_6 = N / 10^6$
Timing Error from Pulse-Phase Jitter

\[ U(\phi) \propto \int d\phi' f_\phi(\phi')a(\phi - \phi') \]

\[ \Delta t_J = N_i^{-1/2} (1 + m_I^2)^{1/2} P(\phi^2)^{1/2} \]

\[ = N_i^{-1/2} (1 + m_I^2)^{1/2} P \left[ \int d\phi \phi^2 f_\phi(\phi) \right]^{1/2} \]

- \( f_\phi \) = PDF of phase variation
- \( a(\phi) \) = individual pulse shape
- \( N_i \) = number of independent pulses summed
- \( m_i \) = intensity modulation index \( \approx 1 \)
- \( f_J \) = fraction jitter parameter = \( \phi_{\text{rms}} / W \approx 1 \)

Gaussian shaped pulse:

\[ \Delta t_J = \frac{f_J W_i (1 + m_I^2)^{1/2}}{2(2N_i \ln 2)^{1/2}} \]

\[ N_6 = N_i / 10^6 \]

\[ \Delta t_J = 0.28\mu s W_{i,\text{ms}} N_6^{-1/2} \left( \frac{f_J}{1/3} \right) \left( \frac{1 + m_I^2}{2} \right)^{1/2} \]
Transdet: a matched-filter based algorithm for transit detection – application to simulated COROT light curves

P. Bordé\footnote{Michelson Postdoctoral Fellow. Current address: Michelson Science Center, Caltech, Pasadena, CA 91125, USA.}
Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

F. Fressin
Gemini, Observatoire de la Côte d’Azur, 06304 Nice, France

M. Ollivier and A. Léger
IAS, Université Paris-Sud, 91405 Orsay, France

D. Rouan
LESIA, Observatoire de Paris, 92195 Meudon, France

Abstract. We present a matched-filter based algorithm for transit detection and its application to simulated COROT light curves. This algorithm stems from the work by Bordé, Rouan & Léger (2003, A&A 405, 1137). We describe the different steps we intend to take to discriminate between planets and stellar companions using the three photometric bands provided by COROT. These steps include the search for secondary transits, the search for ellipsoidal variability, and the study of transit chromaticity. We also discuss the performance of this approach in the context of blind tests organized inside the COROT exoplanet consortium.

1. Introduction

COROT (Baglin 2003), scheduled for launch by the end of 2006, will be the first space mission with the capability to detect extrasolar planets with sizes down to a couple of Earth radii (Rouan et al. 2000). It will observe a total of 60,000 stars in 5 runs of 150 days each. In a previous paper, we worked out a theoretical estimate of the planet detection capability of COROT (Bordé et al. 2003). In this paper, we first describe our transit detection algorithm, dubbed Transdet, then we report its performance on simulated light curves (LCs), and finally, we discuss the tools we are developing to discriminate between genuine exoplanets and astrophysical false positives (binary stars).

2. Key ideas about Transdet

At low signal-to-noise (SNR), a transit signal can be approximated by a rectangular signal oscillating between a high level H (outside the transit) and a low level L (during the transit). The amplitude of this signal depends on the ratio of the planet surface to that of the star: \( H - L = \varepsilon H \), where \( \varepsilon \equiv (R_p/R_*)^2 \). The time spent at level L, or transit duration \( \tau \), is much shorter than the transit period \( T \). We note \( t_0 \) the date of the beginning of the first transit falling inside the observation window (150 days for COROT).

In order to detect the transit signal in a noisy LC, we adopt the matched-filter approach that consists in correlating the LC with rectangular signals of the type described above. We call test-signals these rectangular signals. Each of them is characterized by a parameter triplet \((\hat{T}, \hat{\tau}, \hat{t}_0)\). The correlation reaches a peak when \((\hat{T}, \hat{\tau}, \hat{t}_0) = (T, \tau, t_0)\). Practically, LCs must be first high-pass filtered to remove irrelevant low-frequency variations (stellar fluctuations) that would bias the correlation, and we take the negative of the LCs so that transits would produce positive correlation peaks.

For this method to work, the three-dimensional space \((\hat{T}, \hat{\tau}, \hat{t}_0)\) must be explored with a sufficient resolution in the range accessible to COROT (1 \( \leq \hat{T} \leq 75 \) days, 1 \( \leq \hat{\tau} \leq 10 \) hrs, 0 \( \leq \hat{t}_0 \leq \hat{T} \)). The choice of this resolution results from a trade-off between computing time and detection efficiency. Typically, we use steps \((\delta \hat{T}, \delta \hat{\tau}, \delta \hat{t}_0)\) small enough so that the maximum correlation signal would be no less than 75\% of its theoretical maximum value. This requirement translates into the computation of \(14 \times 10^6\) correlations per LC, a task completed in 90 s with a fully IDL-coded algorithm on a Pentium M at 1.6 Ghz. However, preliminary tests indicate that an optimized IDL–C hybrid code could cut down the computing time by a factor 7–8, which would make possible a higher detection efficiency.

3. Detection threshold

3.1. Theoretical detection threshold

Practically, we use as our primary detection metric not the correlation itself, but the SNR on the correlation defined as

\[ \text{SNR}_1 \equiv (kr)^{1/2} \varepsilon H/\sigma, \]

where \( \sigma \) is the standard deviation of the noise affecting the high-pass filtered LC, and \( k \) is the number of transits in the observation window.

A theoretical detection threshold can be simply set assuming: (1) a total number of LCs, (2) a false detection rate, and (3) Gaussian white noise (see Bordé 2003, p. 45). Requiring less than 1 false detection, we obtain \( \text{SNR}_1 > 6.5 \) for 1,000 LCs, and \( \text{SNR}_1 > 7.0 \) for 60,000 LCs.

3.2. Empirical detection threshold

In order to train the co-investigators of the COROT exoplanet consortium and compare the performances of existing transit detection algorithms, a first transit detection blind test (BT1) was organized inside the consortium in 2004.
<table>
<thead>
<tr>
<th>Plasma Propagation Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Radio propagation phenomena (many)</td>
</tr>
<tr>
<td>• Dedispersion methods</td>
</tr>
<tr>
<td>• Kolmogorov-like microstructure in the ionized ISM</td>
</tr>
<tr>
<td>• Astronomical impacts:</td>
</tr>
<tr>
<td>• Pulsar surveys</td>
</tr>
<tr>
<td>• Precision pulsar timing to detect gravitational waves</td>
</tr>
<tr>
<td>• Pulsars orbiting Sgr A* (GR tests, dark matter)</td>
</tr>
<tr>
<td>• Scattering in the intergalactic medium</td>
</tr>
</tbody>
</table>
Matched Filtering and Deconvolution of the Interstellar Medium

- Dispersion in the ISM
  - Frequency dependent index of refraction $\rightarrow$ change in the phase of EM waves
  - $\delta$(phase) $\rightarrow$ time delay
  - The phase itself is frequency dependent so the time delay is too

- Dispersion acts as a linear filter

- It can be removed exactly once a single parameter is known (the ‘dispersion measure’ DM)
Nanoshots from the Crab Pulsar

\[ T_{eff} \sim 10^{42} \text{ K} \]

Requirements:

- Requires coherent dedispersion
- \( \sim 0.4 \text{ ns (unresolved)} \)
- \( \sim 2 \text{ MJy at 9.2 GHz} \)

Hankins et al. (2006)
The pulse is actually spread out over time (ms-sec) by interstellar dispersion.

~ 6000 years

Image courtesy of NRAO/AUI and Joeri van Leeuwen (UC Berkeley) / ESO / AURA
Dispersion

\[ t_{DM}(\nu) = \frac{4.15 \times DM}{\nu^2_{\text{GHz}}} \]

High frequencies arrive earlier than low frequencies

DM is unique to each pulsar

DM changes with time due to motion of the pulsar and turbulence in the interstellar medium
Starting Point: Index of Refraction

• Dispersion relation in free space:
  • \( \omega = kc \)
  • phase velocity = group velocity = \( \omega/k = c \)

• Cold plasma (collisions not important):
  – No magnetic field:
    \[
    \omega = kc/n_r(\omega) \quad n_r(\omega) = \left(1 - \omega_p^2/\omega^2\right)^{1/2}
    \]
    \( \omega_p \) = plasma frequency = \( \frac{4\pi n_e e^2}{m} \)
  • phase velocity = \( c/n_r(\omega) > c \) (speed of phase fronts)
  • group velocity = \( d\omega/dk = cn_r(\omega) < c \) (speed of a wave packet)
  – With magnetic field: birefringent (left, right)
    \[
    n_{l,r} \approx 1 - \omega_p^2/2\omega^2 \mp \omega_p^2 \omega B_{||}/2\omega^3
    \]
    – cyclotron frequency
    \[
    \omega B_{||} = \frac{eB \cos \theta}{m_e c}
    \]
Propagation through the interstellar plasma

Refractive indices for cold, magnetized plasma

\[ n_{\ell,r} \sim 1 - \frac{\nu_p^2}{2\nu^2} \mp \frac{\nu_p^2 \nu_B^2}{2\nu^3} \]
\[ \nu \gg \nu_p \sim 2 \text{ kHz} \quad \nu \gg \nu_B \| \sim 3 \text{ Hz} \]

Propagation velocities are frequency dependent:

Phase velocity: \( v_p = \frac{\omega}{k} = \frac{c}{n_{\ell,r}} \)

Group velocity: \( v_g = \frac{\partial \omega}{\partial k} = \frac{\partial}{\partial k} \left( \frac{kc}{n_{\ell,r}} \right) \)

Group delay = \( \Delta (\text{Time of Arrival}) \)

\[
\begin{align*}
t & = t_{\text{DM}} \pm t_{\text{RM}} \\
t_{\text{DM}} & = 4.15 \text{ ms DM } \nu^{-2} \\
t_{\text{RM}} & = 0.18 \text{ ns RM } \nu^{-3}
\end{align*}
\]

Dispersion Measure DM = \[ \int ds \, n_e \] units: pc cm\(^{-3}\)

Rotation Measure RM = 0.81 \[ \int ds \, n_e B_{\parallel} \] units: rad m\(^{-2}\)
Dedispersion

Two methods:

**Coherent:**
- operates on the voltage proportional to the electric field accepted by the antenna, feed and receiver
- computationally intensive because it requires sampling at the rate of the total bandwidth
- “exact”

**Post-detection:**
- operates on intensity $= |\text{voltage}|^2$
- computationally less demanding
- an approximation
Basic data unit = a dynamic spectrum

- Basic data unit: Dynamic spectrum
- 64 to 2024 channels
- 10^6 – 10^8 samples x 64 µs

Fast-dump spectrometers:
- Analog filter banks
- Correlators
- FFT (hardware)
- FFT (software)
- Polyphase filter banks
Post-detection Dedispersion:
Sum intensity over frequency after correcting for dispersion delay

Residual time smearing:
\[ \Delta t = \left[ \Delta t_{\text{DM}}^2 + \left( \frac{1}{\Delta \nu} \right)^2 \right]^{1/2} \]
\[ = \left[ (a \Delta \nu)^2 + (\Delta \nu)^{-2} \right]^{-1/2} \]
\[ \implies \text{minimum smearing time across a channel when} \]
\[ \Delta \nu = \left[ 8.3 \, \mu s \, \text{DM} \nu^{-3} \right]^{-1/2} \]

A consequence of the uncertainty principle for Fourier transforms:
\[ \Delta \nu \Delta t \sim 1 \]
Dispersed Pulse

\[ \Delta t = 8.3 \mu s \ DM \nu^{-3} \Delta \nu \]

Coherently dedispersed pulse

\[ DM = 200 \ pc \ cm^{-3} \]

No Dispersion

Coherent Dispersion
Temporal Coherence

- **Temporal coherence**: describes the relationship of different Fourier components that make up a signal vs. time:

  - **Coherent**: a delta function!
  - **Incoherent**: white noise
Coherent Dedispersion
pioneered by Tim Hankins (1971)

Dispersion delays in the time domain represent a phase perturbation of the electric field in the Fourier domain:

$$\tilde{E}_{\text{measured}}(\omega) = \tilde{E}_{\text{emitted}}(\omega)e^{ik(\omega)z}$$

Coherent dedispersion involves multiplication of Fourier amplitudes by the inverse function,

$$e^{-ik(\omega)z}$$

For the non-uniform ISM, we have

$$k(\omega)z \rightarrow \int dz k(\omega) \propto \omega^2 \text{DM} + \text{constant}$$

which is known to high precision for known pulsars.

The algorithm consists of

$$\text{IFFT}\{\text{FFT} [E_{\text{measured}}(t)] \times e^{-ik(\omega)z}\} \approx E_{\text{emitted}}(t)$$

Application requires very fast sampling to achieve usable bandwidths.
Micropulses coherently dedispersed
(Hankins 1971)
Backend stage of mixing: quadrature baseband mixing

\[ \epsilon(t) = I(t) + iQ(t) = \text{complex baseband signal} \]

voltage at intermediate frequency in heterodyned system:
\[ v(t) = \Re \{ \epsilon(t) e^{2\pi i f_0 t} \} = \text{real} \]
Coherent Dedispersion
pioneered by Tim Hankins (1971)

Coherent dedispersion works by explicit deconvolution:

\[ E_{\text{measured}}(t) = E_{\text{emitted}}(t) \ast \text{FFT}\{e^{ik(\omega)z}\} \]
\[ \Rightarrow E_{\text{emitted}}(t) \approx E_{\text{measured}}(t) \ast \text{FFT}\{e^{-ik(\omega)z}\} \]

Comments and Caveats:

• Software implementation with FFTs to accomplish deconvolution (Hankins 1971)
• Hardware implementations: real-time FIR filters (e.g. Backer et al. 1990s-present)
• Resulting time resolution: \(1 / (\text{total bandwidth})\)
• Requires sampling at Nyquist rate of 2 samples \(\times\) bandwidth
  \[\Rightarrow\text{Computationally demanding}\]
• Actual time resolution often determined by interstellar scattering (multipath)
• Most useful for low-DM pulsars and/or high-frequency observations
Matched Filtering Methods

• What is it generically?
• Formal derivation (two domains)
• Examples
• Issues related to sampling and interpolation
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
A6523

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.) $y = (x, \nu, t)$.  

1
Comments and variations on the theme:

1. In simple cases, all we need to know is the time-domain shape (e.g. a pulse shape) to find the optimal filter for estimating arrival times; the spectral shape for Doppler velocities; the image shape to do astrometry; the object shape to localize spatially.

2. $A(t)$ could be stochastic but if we know its shape (somehow), the matched filter solution is the same as for the deterministic case.

   A more complicated situation might be where we use one kind of tracer to detect sources and use the results as a matched filter to detect the sources in another tracer.

   Example: Use locations of galaxies from a continuum survey as a matched filter for detecting a spectral line in the same galaxies by stacking.

3. Any departure of actual data from the assumed model will have consequences:

   (a) the filter is no longer optimal
   (b) detection probabilities are likely to be smaller
   (c) estimated parameters will have larger errors (random and systematic)

4. Types of departures include:

   (a) $A(t)$ not known.
   (b) $a_0$ not constant, e.g. $a_0(t)$.
   (c) $a_0$ could itself be stochastic due to multiplicative effects such as atmospheric seeing, plasmas, gain variations in instrumentation.
Matched filter:
We want the linear filter $h(t)$ that maximizes the SNR of the output and thus maximizes the detection probability. Usually we would convolve the filter with the data but it is convenient to define $h$ so that we cross correlate it with the data instead. [The equivalent filter that would be convolved with the data is $h(-t)$.

Assume initially that $h(t)$ is aligned with the signal $A(t)$ so that we only need consider the SNR of the output when $h$ and $A$ are aligned to give the maximum correlation. More generally we would have a signal

$$x(t) = a_0 A(t - t_0) + n(t)$$

where the event time $t_0$ is not known and we want to estimate it.

Thus while we would generally consider the full correlation function,

$$y(\tau) = h(t) \ast x(t) = \int dt \, x(t)h(t + \tau)$$

(where ‘*’ means correlation), here we initially consider the special time $\tau = 0$ where, by construction, we say that $y$ is maximized if $t_0 = 0$.

In the following we use the noise statistics

$$\langle n(t) \rangle = 0 \quad \langle n(t)n(t') \rangle = R_n(t - t').$$
The mean and variance of the output correlation $y$ are

$$\langle y \rangle = a_0 \int dt \, A(t)h(t)$$

$$\sigma_y^2 = \iint dt dt' \, \langle n(t)n(t') \rangle h(t)h(t') = \iint dt dt' \, R_n(t-t')h(t)h(t').$$

Define the SNR as

$$S = \frac{\langle y \rangle}{\sigma_y} = \frac{a_0 \int dt \, A(t)h(t)}{\sigma_y}.$$ 

We want the smoothing function $h(t)$ that maximizes $S$ (or $S^2$). We can write the variational problem as the change in $S^2$ due to a change in the filter function, $\delta h$

$$\delta(S^2) = 0 = \delta \left( \frac{\langle y \rangle^2}{\sigma_y^2} \right) = \frac{\sigma_y^2 \delta(\langle y \rangle^2) - \langle y \rangle^2 \delta(\sigma_y^2)}{\sigma_y^4}.$$ 

(2)

This implies

$$2\sigma_y^2 \delta(\langle y \rangle) - \langle y \rangle^2 \delta(\sigma_y^2) = 2\sigma_y^2 \langle y \rangle a_0 \int dt \, A(t) \delta h(t) - \iint dt dt' \, R_n(t-t') \delta [h(t)h(t')]$$

$$= 2\sigma_y^2 \langle y \rangle a_0 \int dt \, A(t) \delta h(t) - 2 \iint dt dt' \, R_n(t-t') \delta h(t)$$

$$= 0.$$
Collecting terms,

\[
\int dt \delta h(t) \left[ \sigma_y^2 a_0 A(t) - \langle y \rangle \right] \int dt' R_n(t - t') h(t') = 0.
\]

If the solution \( h(t) \) yields a maximum SNR then the integrand factor that multiplies \( \delta h(t) \) is zero for any \( \delta h(t) \), so

\[
\int dt' R_n(t - t') h(t') = \frac{a_0 \sigma_y^2}{\langle y \rangle} A(t) .
\]

(3)

To use this equation we have to know (or assume) the shape of the autocorrelation function for the noise as well as the signal shape \( A(t) \).
White Noise:
This case, the simplest, has $R_n(t - t') = \sigma_n^2 W_n \delta(t - t')$ where $W_n$ has time units to keep the overall units correct. This gives

$$\langle y \rangle = a_0 \int dt \ A(t) h(t)$$

$$\sigma_y^2 = \sigma_n^2 W_n \int dt \ h^2(t)$$

and the maximum S/N is given by the expression,

$$h(t) = A(t) \frac{\int dt \ h^2(t)}{\int dt \ A(t) h(t)}.$$  

While appearing to be an implicit solution for $h(t)$, the integral quantities are just a proportionality constant.

A particular solution is $h(t) = A(t)$, as is anything proportional to $A(t)$. It is now obvious why this approach is called matched filtering.

Interpretation:

- The optimal smoothing function is the shape of the signal when the noise is white.
- When the signal has an unknown time shift $t_0$, the entire cross-correlation function is calculated and its maximum found and assessed as statistically significant or not.
**Demonstration that the MF solution gives the best signal to noise ratio:** The S/N for an arbitrary filter is

$$S = \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt A(t) h(t)}{\left[ \int \int dt dt' \rho_n(t - t') h(t) h(t') \right]^{1/2}}.$$  

Let $S_0$ be the S/N for the MF solution.

Let the filter be $h(t) = A(t) + b(t)$ where $b(t)$ is an arbitrary function but we will consider $b$ and $A$ to be orthogonal functions, i.e. $\int dt A(t)b(t) = 0$. (If they weren’t to begin with we could use the Gram-Schmidt orthogonalization procedure to define them as such.)¹

Evaluate $S$ and calculate the ratio $S/S_0$, which is generally

$$\frac{S}{S_0} = \frac{\int dt A(A + b)}{\left[ \int dt A^2 \int dt (A + b)^2 \right]^{1/2}}$$

$$\rightarrow \text{orthogonal} \quad \frac{\int dt A^2}{\left[ \int dt A^2 \int dt (A^2 + b^2) \right]^{1/2}}$$

$$= \left[ \frac{\int dt A^2}{\int dt (A^2 + b^2)} \right]^{1/2}$$

$$\leq 1$$

¹For two functions $u, v$, we can define two orthogonal functions as $p = u$ and $q = v - u(u, v)/(u, u) = v - u \int dt uv/ \int dt u^2$.  

7
**Low-pass Noise:** Non-white noise with a short correlation time (relative to the width of the signal) can be treated in the same way.

Let $R_n(t - t') = \sigma_n^2 \rho_n(t - t')$ where $\rho_n(0) = 1$ and its width $\ll$ width of $A$.

Returning to our original expression for $S$, we have

$$S = \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt \, A(t)h(t)}{\left[ \int dt dt' \, \rho_n(t - t')h(t)h(t') \right]^{1/2}}$$

$$\approx \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt \, A(t)h(t)}{\left[ \int d\tau \, \rho_n(\tau) \int dt \, h^2(t) \right]^{1/2}}$$

$$\approx \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt \, A(t)h(t)}{\left[ W_n \int dt \, h^2(t) \right]^{1/2}},$$

where the characteristic time scale of the noise is

$$W_n = \int d\tau \, \rho_n(\tau).$$

We can now apply the Cauchy-Schwarz inequality for the numerator

$$\left[ \int dt \, A(t)h(t) \right]^2 \leq \int dt \, A^2(t) \int dt \, h^2(t).$$
which means that

\[ S = \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt A(t) h(t)}{[W_n \int dt h^2(t)]^{1/2}} \]

\[ \leq \left( \frac{a_0}{\sigma_n} \right) \frac{\left[ \int dt A^2(t) \int dt h^2(t) \right]^{1/2}}{[W_n \int dt h^2(t)]^{1/2}} \]

\[ \leq \left( \frac{a_0}{\sigma_n} \right) \frac{\int dt A^2(t)}{\sqrt{W_n}} \]

We have equality when \( h(t) = A(t) \), i.e. when the filter matches the signal.

The Cauchy-Schwartz inequality generalizes the statement that the sum of two vectors has a length that is less than the sum of the magnitudes of the two vectors:

\[ |x + y|^2 \leq (x + y)^2 \]

\[ x^2 + y^2 + 2x \cdot y \leq x^2 + y^2 + 2xy \]

\[ x \cdot y \leq xy \]

This can be written as

\[ \sum_j x_j y_j \leq \left[ \sum_j x_j^2 \sum_j y_j^2 \right]^{1/2} \]

or in integral form

\[ \int dt x(t) y(t) \leq \left[ \int dt x^2(t) \int dt y^2(t) \right]^{1/2} \]
Simplification and Interpretation

For the matched filter case (equality) we can consider $A(t)$ to be dimensionless with $A(0) = 1$. Then the integral

$$\int dt \, A^2(t) \equiv W_A$$

defines the characteristic time scale $W_A$. We can also define the signal to noise of the time series as the signal maximum $a_0$ divided by the rms noise $\sigma_n$,

$$S_t \equiv \left( \frac{a_0}{\sigma_n} \right)$$

and then rewrite the SNR of the correlation function as

$$S = S_t \left( \frac{W_A}{W_n} \right)^{1/2} = S_t \sqrt{N_{\text{eff}}}.$$

- As we have seen in other contexts, the ratio $W_A/W_n$ represents the number of independent noise fluctuations that have been averaged to get the correlation amplitude. We then see that the correlation function enhances the SNR in the time series by a factor $\sqrt{N_{\text{eff}}}$ where $N_{\text{eff}}$ is the effective number of independent fluctuations.

- Note that this simple result holds for WSS noise with a correlation function that is narrow compared to the length of a time series.

- For red noise with a steep power-law spectrum, longer data sets have larger variance because more of the power spectrum is sampled. For these processes, $N_{\text{eff}} \sim 1$. 
**Arbitrary WSS Noise:** Generally, the solution is gotten by Fourier transforming Equation 3. Let the noise spectrum be $R_n(\tau) \leftrightarrow S_n(f)$ and define the FTs of $A$ and $h$ as $\tilde{A}$ and $\tilde{h}$. Using 3 (repeated here),

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

we have

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

where

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

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

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

and the solution is

$$
\tilde{h}(f) = \text{constant} \times \frac{\tilde{A}(f)}{S_n(f)}.
$$
Comments on the general matched filtering solution

• For white noise, \( S_n(f) = \) 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.
Practical Applications in Signal Detection:

Example: we know the shape of the function $A(t)$ and the autocorrelation function of the noise (or its spectrum). The optimal detection scheme is to construct the filter using this information and investigate the cross correlation function Eq. 1.

Output amplitudes can be tested against a threshold $y_T$ that is some multiple of $\sigma_y$. Over an ensemble one can then define the detection probability $P_d$ and the false-alarm probability $P_{fa}$.

By changing the threshold both $P_d$ and $P_{fa}$ will change. Ideally one would like $P_d = 1$ with $P_{fa}$ but in reality there is a tradeoff between the two.

**Periodic Signals:** The matched filter for a periodic signal is simply a periodic train of the pulse shape. When correlated with the measurements, the output is the same thing as “folding” the data with the underlying period.

**Spectra lines:** Change $t \rightarrow \lambda$ or $\nu$ and the results are the same.

**Source detection in images:** Change $A(t) \rightarrow I(\theta)$. This might be the PSF if point sources are of interest or it might be a specific kind of source (e.g. spiral galaxies).

**Multidimensional:** $A(t) \rightarrow A(x)$ where $x$ is $N$-dimensional.
Figure 1: Matched filtering of Gaussian pulse with itself as a template. 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 25.3 samples.
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.
Figure 3: Matched filtering of a narrow Gaussian pulse with a broader template. 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 while the template is 45.3 samples wide.
Figure 4: Matched filtering of Gaussian pulse with a narrower template. The SNR of the pulse is 5 (peak to rms noise). The pulse width (HW @ 1/e) is 10.3 samples while the template is 4.3 samples wide.