Appendix E

Optimal Filtering

In this appendix I review the formalism for the Fourier domain optimum filter and present the pileup and correlation optimum filters which will help achieve better thresholds in the HV detectors. I also include sections on Wiener filtering, which is the optimal filter in the absence of a template, and on the time-domain version of the optimal filter, which is usually called a matched filter, and includes the correlation matrix in place of the PSD. In this note, the measured signal is $v(f)$ in the frequency domain or $v(t)$ in the time domain, the template is $s(f)$ or $s(t)$, and the noise PSD is $J(f)$.

E.1 Basic Optimal Filter

For the chi-squared function

$$\chi^2 = \int_{-\infty}^{\infty} \frac{|v(f) - As(f)|^2}{J(f)} df \tag{E.1}$$

we find the goodness of fit by minimizing χ^2 with respect to A, as

$$0 = \frac{d\chi^2}{dA} = \frac{d}{dA} \int_{-\infty}^{\infty} \frac{v^*(f)v(f) - 2As^*(f)v(f) + A^2 s^*(f)s(f)}{J(f)} df \tag{E.2}$$

$$0 = 2\int_{-\infty}^{\infty} \frac{-s^*(f)v(f) + As^*(f)s(f)}{J(f)} df \tag{E.3}$$

$$\int_{-\infty}^{\infty} \frac{s^*(f)v(f)}{J(f)} df = A\int_{-\infty}^{\infty} \frac{s^*(f)s(f)}{J(f)} df \tag{E.4}$$

$$A = \frac{\int_{-\infty}^{\infty} \frac{s^*(f)v(f)}{J(f)} df}{\int_{-\infty}^{\infty} \frac{|s(f)|^2}{J(f)} df} \tag{E.5}$$

This suggests that the optimum filter for this signal has the form

$$\phi(f) = \frac{s^*(f)}{J(f)} \tag{E.6}$$

so that we can write the optimal estimate as

$$A = \frac{\int_{-\infty}^{\infty} \phi(f) v(f) df}{\int_{-\infty}^{\infty} \phi(f) s(f) df} \tag{E.7}$$

or simplifying further, we can renormalize the filter as

$$\phi'(f) = \frac{\frac{s^*(f)}{J(f)}}{\int_{-\infty}^{\infty} \phi(f) s(f) df} \tag{E.8}$$

to give the resulting simple estimator

$$A = \int_{-\infty}^{\infty} \phi'(f) v(f) df \tag{E.9}$$

This is a nice and clean derivation which only captures the simplest case, where there is a single, well-known template, one signal per trace, and the start of the pulse is well-known. This is also a derivation for the continuous limit, whereas real data is sampled discretely. In this note I will delve into more complex optimal filters which each relax one of these assumptions, and provide formulae for discrete optimum filters for the different cases along with reduced chi-square formulae to make their implementation simple and straightforward.

E.2 Optimum Filter with Time Offset

Here we make the extension that the signal can slide in time, i.e. that

$$s(f) \rightarrow s(f, t_0) = e^{-i\omega t_0} s(f) \tag{E.10}$$

which has the goodness of fit

$$\chi^2 = \int_{-\infty}^{\infty} \frac{|v(f) - Ae^{-i\omega t_0} s(f)|^2}{J(f)} df \tag{E.11}$$

Here we can go through the same procedure to get an optimal estimate of A, now as a function of t_0:

$$0 = \frac{d\chi^2}{dA}(t_0) = \frac{d}{dA}\int_{-\infty}^{\infty} \frac{|v(f) - Ae^{-i\omega t_0}s(f)|^2}{J(f)}df \tag{E.12}$$

$$0 = \frac{d}{dA}\int_{-\infty}^{\infty} \frac{v^*(f)v(f) - 2Ae^{i\omega t_0}s^*(f)v(f) + A^2 s^*(f)s(f)}{J(f)}df \tag{E.13}$$

$$\int_{-\infty}^{\infty} \frac{e^{i\omega t_0}s^*(f)v(f)}{J(f)}df = \int_{-\infty}^{\infty} \frac{As^*(f)s(f)}{J(f)}df \tag{E.14}$$

$$A = \frac{\int_{-\infty}^{\infty} \frac{e^{i\omega t_0}s^*(f)v(f)}{J(f)}}{\int_{-\infty}^{\infty} \frac{s^*(f)s(f)}{J(f)}df} \tag{E.15}$$

which looks similar to the original form except for the addition of the time offset. Writing this with the normalized optimum filter makes it a bit more explicit what we're looking at:

$$A(t_0) = \int_{-\infty}^{\infty} e^{i\omega t_0}\phi'(f)v(f)df \tag{E.16}$$

This is just the inverse fourier transform of the filtered trace! So adding a time offset will only marginally increase the complexity of our optimal filter algorithm. We can get the amplitude as a function of time, and then look for the best chi-square in time space. With this amplitude function, we can re-write the chi-square as

$$\begin{aligned}\chi^2(t_0) &= \int_{-\infty}^{\infty} \frac{|v(f)|^2}{J(f)}df - 2A(t_0)\int_{-\infty}^{\infty} e^{i\omega t_0}\phi(f)v(f)df + A^2(t_0)\int_{-\infty}^{\infty}\phi(f)s(f)df \\ &= \int_{-\infty}^{\infty} \frac{|v(f)|^2}{J(f)}df - A^2(t_0)\int_{-\infty}^{\infty}\phi(f)s(f)df\end{aligned}$$

and we find that the optimal amplitude with optimal offset is the condition that maximizes the filtered amplitude function $A(t_0)$. Practically speaking, we can compute the optimal amplitude with delay by computing $A(t_0)$, computing the resulting χ^2 as shown, and finding the amplitude corresponding to the minimum chi-square or the maximum amplitude. Re-writing this in more convenient form allows the process to be more explicit:

$$\begin{aligned}
\phi(f) &= \frac{s^*(f)}{J(f)} \\
N_\phi &= \int_{-\infty}^{\infty} \phi(f)s(f)df \\
\phi'(f) &= \frac{\phi(f)}{N_\phi} \\
A(t_0) &= \int_{-\infty}^{\infty} e^{i\omega t_0}\phi'(f)v(f)df \\
\chi_0^2 &= \int_{-\infty}^{\infty} \frac{|v(f)|^2}{J(f)}df \\
\chi^2(t_0) &= \chi_0^2 - A^2(t_0)N
\end{aligned}$$

Using this fourier transform trick makes the implementation of the OF with delay a trivial extension of the nominal optimum filter.

E.3 Two-Pulse Optimum Filter

Suppose now that instead of one event per trace, we have two. We are still in the small signal limit of our device (in principle) so there should be no different between the templates for each pulse, but now we have two amplitudes and two offsets. In other words,

$$As(f,t_0) \rightarrow A_1 s(f,t_1) + A_2 s(f,t_2) \tag{E.17}$$

which has the goodness of fit

$$\chi^2 = \int_{-\infty}^{\infty} \frac{|v(f) - A_1 e^{-i\omega t_1}s(f) - A_2 e^{-i\omega t_2}s(f)|^2}{J(f)}df \tag{E.18}$$

Let's expand the numerator to see what happens:

$$|v(f) - A_1 e^{-i\omega t_1}s(f) - A_2 e^{-i\omega t_2}s(f)|^2 \tag{E.19}$$

$$= (v^*(f) - A_1 e^{i\omega t_1}s^*(f) - A_2 e^{i\omega t_2}s^*(f))(v(f) - A_1 e^{-i\omega t_1}s(f) - A_2 e^{-i\omega t_2}s(f)) \tag{E.20}$$

$$= v^*(f)v(f) - 2v(f)(A_1 e^{i\omega t_1}s^*(f) + A_2 e^{i\omega t_2}s^*(f)) + |A_1 e^{-i\omega t_1}s(f) + A_2 e^{-i\omega t_2}s(f)|^2 \tag{E.21}$$

$$= |v(f)|^2 - 2v(f)(A_1 e^{i\omega t_1} + A_2 e^{i\omega t_2})s^*(f) + \left[A_1^2 + A_2^2 + 2A_1A_2\cos(\omega(t_1 - t_2))\right]|s(f)|^2 \tag{E.22}$$

Pushing the denominator through gives us

$$\frac{|v(f)|^2}{J(f)} - 2\phi(f)v(f)(A_1 e^{i\omega t_1} + A_2 e^{i\omega t_2}) + \left[A_1^2 + A_2^2 + 2A_1A_2\cos(\omega(t_1 - t_2))\right]\phi(f)s(f) \tag{E.23}$$

and minimizing with respect to A_i gives

$$0 = \frac{d\chi^2}{dA_i} = \frac{d}{dA_i}\int_{-\infty}^{\infty} -2\phi(f)v(f)(A_1 e^{i\omega t_1} + A_2 e^{i\omega t_2})df \tag{E.24}$$

$$+\frac{d}{dA_i}\int_{-\infty}^{\infty}\left[A_1^2 + A_2^2 + 2A_1A_2\cos(\omega(t_1 - t_2))\right]\phi(f)s(f)df \tag{E.25}$$

$$= \int_{-\infty}^{\infty} -2\phi(f)v(f)e^{i\omega t_i}df + \left[2A_i + 2A_j\cos(\omega(t_1 - t_2))\right]\phi(f)s(f)df \tag{E.26}$$

$$A_i\int_{-\infty}^{\infty}\phi(f)s(f)df = \int_{-\infty}^{\infty}\phi(f)v(f)e^{i\omega t_i}df - \int_{-\infty}^{\infty}A_j\cos(\omega(t_1 - t_2))\phi(f)s(f)df \tag{E.27}$$

$$A_i = \int_{-\infty}^{\infty}\phi'(f)v(f)e^{i\omega t_i}df - A_j\int_{-\infty}^{\infty}\cos(\omega\delta t)\phi'(f)s(f)df \tag{E.28}$$

$$A_i = A_{0,i}(t_i) - A_j(t_j)\gamma(\delta t) \tag{E.29}$$

where here we've switched to the parameterization where $t_1 = t_0$ and $t_2 = t_0 + \delta t$, and introduced the new variable

$$\gamma(\delta t) = \int_{-\infty}^{\infty}\cos(\omega\delta t)\phi'(f)s(f)df \tag{E.30}$$

For the case of $\delta t = 0$, this trivially tells us that the amplitudes should add. For a given t_0 and δt, we thus have two equations with two unknowns, so we can eliminate one amplitude to find the optimal amplitude as

$$A_i = A_{i,0}(t_i) - (A_{j,0}(t_j) - A_i(t_i)\gamma(\delta t))\gamma(\delta t) \tag{E.31}$$

$$A_i = A_{i,0}(t_i) - A_{j,0}(t_j)\gamma(\delta t) + A_i(t_i)\gamma^2(\delta t) \tag{E.32}$$

$$A_i(1 - \gamma^2(\delta t)) = A_{i,0}(t_i) - A_{j,0}(t_j)\gamma(\delta t) \tag{E.33}$$

$$A_i(t_i, t_j) = \frac{A_{i,0}(t_i) - A_{j,0}(t_j)\gamma(\delta t)}{1 - \gamma^2(\delta t)} \tag{E.34}$$

As a sanity check, this formula tells us that for $\delta t = 0$, the amplitude evaluates to 0, and the other amplitude thus becomes the original value for 0 delay. Putting this in more concrete terms, let's use this formula to estimate A_2. We thus get

$$A_2(t_1, \delta t) = \frac{A_2(t_1 + \delta t) - A_1(t_1)\gamma(\delta t)}{1 - \gamma^2(\delta t)} \tag{E.35}$$

$$A_1(t_1, \delta t) = A_1(t_1) - A_2(t_1, \delta t)\gamma(\delta t) \tag{E.36}$$

$$= \frac{(1 - \gamma^2(\delta t))A_1(t_1) - A_2(t_1 + \delta t)\gamma(\delta t) + A_1(t_1)\gamma^2(\delta t)}{1 - \gamma^2(\delta t)} \tag{E.37}$$

$$= \frac{A_1(t_1) - A_2(t_1 + \delta t)\gamma(\delta t)}{1 - \gamma^2(\delta t)} \tag{E.38}$$

Finishing the problem, we can rearrange the terms in the chi-square to make them look similar to the single-template with time shift:

$$\chi^2 = \chi_0^2 - \int_{-\infty}^{\infty} \left[2\phi(f)v(f)(A_1 e^{i\omega t_1} + A_2 e^{i\omega t_2}) - \left[A_1^2 + A_2^2 + 2A_1A_2\cos(\omega(t_1 - t_2))\right]\phi(f)s(f)\right] df \tag{E.39}$$

$$= \chi_0^2 - 2A_1 \int_{-\infty}^{\infty} \left[\phi(f)v(f)e^{i\omega t_1} - A_2\cos(\omega(t_1 - t_2))\phi(f)s(f)\right] df + A_1^2 \int_{-\infty}^{\infty} \phi(f)s(f)df \tag{E.40}$$

$$- 2A_2 \int_{-\infty}^{\infty} \left[\phi(f)v(f)e^{i\omega t_2} - A_1\cos(\omega(t_1 - t_2))\phi(f)s(f)\right] df + A_2^2 \int_{-\infty}^{\infty} \phi(f)s(f)df \tag{E.41}$$

$$- 2A_1A_2 \int_{-\infty}^{\infty} \cos(\omega(t_1 - t_2))\phi(f)s(f)df \tag{E.42}$$

$$= \chi_0^2 - \left[A_1^2 + A_2^2 + 2A_1A_2\gamma(\delta t)\right] N_\phi \tag{E.43}$$

We can make this algorithm more efficient, practically speaking, by substituting the cosine transform with an inverse fourier transform and taking the real part. This can be see here:

$$\gamma(\delta t) = \mathrm{Re}\left[\int_{-\infty}^{\infty} e^{\omega\delta t}\phi'(f)s(f)df\right] \tag{E.44}$$

$$= \mathrm{Re}\left[\int_{-\infty}^{\infty} \left(\cos(\omega\delta t) + i\sin(\omega\delta t)\right)\phi'(f)s(f)df\right] \tag{E.45}$$

$$= \mathrm{Re}\left[\int_{-\infty}^{\infty} \cos(\omega\delta t)\phi'(f)s(f)df + i\int_{-\infty}^{\infty} \sin(\omega\delta t)\phi'(f)s(f)df\right] \tag{E.46}$$

$$= \int_{-\infty}^{\infty} \cos(\omega\delta t)\phi'(f)s(f)df \tag{E.47}$$

Here we see that we could also use the direct transform, but we want the normalization to be consistent with that of the amplitudes, so we choose to use the inverse transform for both. Note that this works because the product $\phi(f)'s(f)$ is real.

So to summarize as in the previous section, the calculation procedes as follows. We first calculate the one-dimensional quantities:

$$\phi(f) = \frac{s^*(f)}{J(f)}$$
$$N_\phi = \int_{-\infty}^{\infty} \phi(f)s(f)df$$
$$\phi'(f) = \frac{\phi(f)}{N_\phi}$$
$$A_0(t_i) = \int_{-\infty}^{\infty} e^{i\omega t_i}\phi'(f)v(f)df$$
$$\gamma(\delta t) = \mathrm{Re}\left[\int_{-\infty}^{\infty} e^{i\omega\delta t}\phi'(f)s(f)df\right]$$

We then project these into the full two-dimensional space and compute the chi-square as follows:

$$A_2(t_1,\delta t) = \frac{A_0(t_1+\delta t) - A_0(t_1)\gamma(\delta t)}{1-\gamma^2(\delta t)}$$
$$A_1(t_1,\delta t) = A_0(t_1) - A_2(t_1,\delta t)\gamma(\delta t)$$
$$\chi_0^2 = \int_{-\infty}^{\infty} \frac{|v(f)|^2}{J(f)}df$$
$$\chi^2(t_1,\delta t) = \chi_0^2 - (A_1^2(t_1,\delta t) + A_2^2(t_1,\delta t) + 2A_1(t_1,\delta t)A_2(t_1,\delta t)\gamma(\delta t))N$$

The convenient aspect of this algorithm is that it allows us to easily restrict our search windows in either t_1 or δt without any additional tricks, we simply limit the domain of the two-dimensional projection. A final note: we have to take care to treat the $\delta t = 0$ case independently; mathematically this will reduce correctly, but we have to show that by evaluating a limit in which the numerator and denominator both go to 0. Taking the derivatives, we can show that for $\delta t = 0$, this solution reduces to the single template solution:

$$\lim_{\delta t\to 0} A_2(t_1,\delta t) = \lim_{\delta t\to 0}\frac{A_0(t_1+\delta t) - A_0(t_1)\gamma(\delta t)}{1-\gamma^2(\delta t)} = \lim_{\delta t\to 0}\frac{A_0'(t_1+\delta t) - A_0'(t_1)\gamma(\delta t) - A_0(t_1)\gamma'(\delta t)}{1-2\gamma(\delta t)\gamma'(\delta t)} = \frac{0}{1-0} = 0 \tag{E.48}$$

$$\to A_1(t_1,0) = A_0(t_1) - A_2(t_1,0) = A_0(t_1), \;\; \chi^2(t_1,\delta t) = \chi_0^2 - A_1^2(t_1)N \tag{E.49}$$

which is just the single-template optimum filter with time-offset.

E.4 Joint Channel Optimum Filter

Moving back to the case of a single event in a trace, let's consider a joint fit of all channels with any number time-shifted templates. Let's add a channel index i into our notation, such that the signal from channel i is $v_i(f)$ and the template at $t_0 = 0$ in channel i for a template γ is $s_{i\gamma}(f)$. In this way, we allow for n channels and m templates, where each template can take a separate form in each channel but has one unified amplitude across all channels, allowing for correlated signals between channels.

We can construct the least-squares statistic as usual:

$$\chi^2 = \sum_i \int_{-\infty}^{\infty} \frac{|v_i(f) - \sum_\gamma A_\gamma e^{-i\omega t_0} s_{i\gamma}(f)|^2}{J_i(f)} df \tag{E.50}$$

Here we sum templates for a given channel, and sum over least-squares for all channels to get a joint goodness of fit estimator. We can expand the numerator to get

$$|v_i(f) - \sum_\gamma A_\gamma e^{-i\omega t_0} s_{i\gamma}(f)|^2 \tag{E.51}$$

$$= (v_i^*(f) - \sum_\gamma A_\gamma e^{i\omega t_0} s_{i\gamma}^*(f))(v_i(f) - \sum_\beta A_\beta e^{-i\omega t_0} s_{i\beta}(f)) \tag{E.52}$$

$$= v_i^*(f) v_i(f) - 2\sum_\gamma A_\gamma e^{i\omega t_0} s_{i\gamma}^*(f) v_i(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta s_{i\gamma}^* s_{i\beta}(f) \tag{E.53}$$

$$= |v_i(f)|^2 - 2\sum_\gamma A_\gamma e^{i\omega t_0} s_{i\gamma}^*(f) v_i(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta s_{i\gamma}^* s_{i\beta}(f) \tag{E.54}$$

pushing the denominator through we get the integrand

$$\frac{|v_i(f)|^2}{J_i(f)} - 2\sum_\gamma A_\gamma e^{i\omega t_0} \phi_{i\gamma}(f) v_i(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta \phi_{i\gamma} s_{i\beta}(f) \tag{E.55}$$

where we now have the optimum filter $\phi_{i\gamma}$ for channel i, template γ

$$\phi_{i\gamma} = \frac{s_{i\gamma}^*(f)}{J_i(f)} \tag{E.56}$$

If we minimize the fit with respect to a given amplitude A_α, we find

$$\frac{d}{dA_\alpha}\chi^2 = 0 = \frac{d}{dA_\alpha} \sum_i \int_{-\infty}^{\infty} \left[\frac{|v_i(f)|^2}{J_i(f)} - 2\sum_\gamma A_\gamma e^{i\omega t_0} \phi_{i\gamma}(f) v_i(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta \phi_{i\gamma} s_{i\beta}(f) \right] df \tag{E.57}$$

$$= \sum_i \int_{-\infty}^{\infty} \left[-2e^{i\omega t_0} \phi_{i\alpha}(f) v_i(f) + 2\sum_\beta A_\beta \phi_{i\alpha} s_{i\beta}(f) \right] df \tag{E.58}$$

$$\sum_i \int_{-\infty}^{\infty} e^{i\omega t_0} \phi_{i\alpha}(f) v_i(f) df = \sum_\beta A_\beta \sum_i \int_{-\infty}^{\infty} \phi_{i\alpha} s_{i\beta}(f) df \tag{E.59}$$

This looks odd until we realize that it is just a system of equations with one equation for each of the m templates. We can thus turn it into the matrix operation

$$\mathbf{V} = \mathbf{N}\mathbf{A} \rightarrow \mathbf{A} = \mathbf{N}^{-1}\mathbf{V} \tag{E.60}$$

where

$$V_\alpha = \sum_i \int_{-\infty}^{\infty} e^{i\omega t_0} \phi_{i\alpha}(f) v_i(f) df \tag{E.61}$$

$$N_{\alpha\beta} = \sum_i \int_{-\infty}^{\infty} \phi_{i\alpha} s_{i\beta}(f) df \tag{E.62}$$

As we'll see later, these are just the normal equations for any linear model, but we've left this problem slightly nonlinear by introducing the time offset, so we'll still need to evaluate the chi-square at each allowable time offset. In this notation the chi-square becomes

$$\chi^2 = \chi_0^2 - \mathbf{A}^{\mathbf{T}}\mathbf{V} = \chi_0^2 - \mathbf{A}^{\mathbf{T}}\mathbf{N}\mathbf{A}$$

which again, looking forward to the next appendix, just resembles the normal equations chi-square, but we've made the time-shift implicit in the amplitudes and data vector.

This is functionally a generalization of the original time-shifted optimum filter because it reduces to that filter in the limit that $S_{\alpha\beta} = \delta_{\alpha\beta} S_{\alpha\alpha}$. This is either the case for orthonormal templates, which we don't expect in our detectors, or for channels which are completely uncorrelated in their signal, which is also not a correct assumption. For these reasons we should expect this OF to perform better than fitting the channels independently. It's also interesting to note that adding arbitrary linear dimensionality to our model space is not nearly as involved as adding one additional non-linear element to the model, as can be seen by comparing this and the pileup OF sections.

E.5 Joint Channel Correlated Optimum Filter

The final generalization of this technique which can accommodate any type of detector where the channels some set of known templates is to include the effect of noise covariance between channels. As in the previous section, we have a model made of M templates for N channels. The least-squares statistic takes on the modified form

$$\chi^2 = \sum_i \sum_j \int_{-\infty}^{\infty} \left(v_i(f) - \sum_\gamma A_\gamma e^{-i\omega t_0} s_{i\gamma}(f) \right)^* \Sigma_{ij}^{-1}(f) \left(v_j(f) - \sum_\beta A_\beta e^{-i\omega t_0} s_{j\beta}(f) \right) df \tag{E.63}$$

This of course reduces to the form of the previous section in the limit that $\Sigma_{ij}^{-1} = \delta_{ij} J_i^{-1}(f)$, so we can see that now we're going to modify the chi-square for the same model if we allow for noise correlation between channels. Here the matrix $\Sigma_{ij}(f)$ is the cross-power spectral density of the channels, where the diagonals are the auto-power spectral densities (what we have been calling simply the PSDs) of each channel.

We should note that this is a slightly different form than how the generalized linear least-squares model is written, as to write this as a matrix product we would have to express the covariance matrix in a block-diagonal form for each frequency which would add dimensionality to the problem which has been removed by use of the Fourier basis. As in the previous section, the integrals will end up in our matrix definitions, and the result will resemble the linear least-squares case.

Let's expand out the terms as usual in our least-squares statistic. We have the model-independent term

$$\chi_0^2 = \sum_i \sum_j \int_{-\infty}^{\infty} v_i^*(f)\Sigma(f)_{ij}^{-1} v_j(f) df \tag{E.64}$$

and the model-dependent term

$$\sum_i \sum_j \int_{-\infty}^{\infty} \left[-2 \sum_\gamma A_\gamma e^{i\omega t_0} s_{i\gamma}^*(f) \Sigma_{ij}^{-1}(f) v_j(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta s_{i\gamma}^*(f) \Sigma_{ij}^{-1}(f) s_{j\beta}(f) \right] df \tag{E.65}$$

$$= \sum_i \sum_j \int_{-\infty}^{\infty} \left[-2 \sum_\gamma A_\gamma e^{i\omega t_0} \phi_{ij\gamma}(f) v_j(f) + \sum_\gamma \sum_\beta A_\gamma A_\beta \phi_{ij\gamma} s_{j\beta}(f) \right] df \tag{E.66}$$

By analogy with the previous section, we find that

$$V_\alpha = \sum_i \sum_j \int_{-\infty}^{\infty} e^{i\omega t_0} \phi_{ij\alpha}(f) v_j(f) df \tag{E.67}$$

$$N_{\alpha\beta} = \sum_i \sum_j \int_{-\infty}^{\infty} \phi_{ij\alpha}(f) s_{j\beta}(f) df \tag{E.68}$$

and as usual

$$A_\alpha = \sum_\beta N_{\alpha\beta}^{-1} V_\beta \tag{E.69}$$

and

$$\chi^2 = \chi_0^2 - \mathbf{A^T N A} \tag{E.70}$$

As a concluding remark, I'll point out that the logical continuation of the continually more general solutions (which are less and less like filters as we generalize) would be to revert to the time domain and remove the assumption of stationary noise; in this case we would just be dealing with the normal generalized least-squares, which is described in the next section, but the dimensionality of our problem would increase by orders of magnitude. The assumption of stationary, relatively uncorrelated noise allows us to work with PSDs, in the Fourier domain, and reduce the dimensionality of the problem to a tolerable level. That being said, we still need to take the inverse of a potentially large matrix twice in the construction phase of the filter, once to determine the covariance matrix, and once again to determine the filter normalization S^{-1}.

E.6 Optimal Filter Resolution

I've written the above filter algorithms in consistent terminology in order to utilize a general result, discussed in the next appendix, for chi-square fits. The covariance of the parameters of our model A_i is just

$$\Sigma_{ij} = N_{ij}^{-1} \tag{E.71}$$

where for the simplest 1D optimum filter, this is just the inverse of the filter norm, but for the more complex optimum filters it is properly the inverse matrix of the problem's design matrix. Consider, for example, the correlated optimum filter with a single template. This filter has only one variable to fit, the joint channel amplitude A_0, and has the expected resolution

$$\sigma_0^2 = \left[\sum_i \sum_j \int_{-\infty}^{\infty} \phi_{ij}(f) s_j(f) df\right]^{-1} = \left[\sum_i \sum_j \int_{-\infty}^{\infty} s_i^*(f) \Sigma_{ij}^{-1}(f) s_j(f) df\right]^{-1} \tag{E.72}$$

To see that this makes sense, consider the case of uncorrelated noise, with equally weighted templates and noise. We then see that the cross-terms go to 0, and the remaining terms are the resolutions of the 1D optimum filter, thus

$$\sigma_0^2 = \left[\frac{N}{\sigma_{1D}^2}\right]^{-1} = \frac{\sigma_{1D}^2}{N} \tag{E.73}$$

where

$$\sigma_{1D}^2 = \left[\int_{-\infty}^{\infty} s^*(f) J^{-1}(f) s(f) df\right]^{-1} \tag{E.74}$$

This is exactly what we expect for measurements of the same quantity under equal and uncorrelated noise. If the noise is uncorrelated, but some measurements are noisier, this is just the resolution of the weighted sum of observations.

E.7 Wiener Filtering

In this section I describe the more general Wiener filter, in which a new filter is constructed for each trace and the shape of the filter does not depend on the templates used. A good summary and discussion of Wiener filtering can be found in Numerical Recipes, section 13.3.

Suppose we have some Fourier signal $S(f) = R(f)U(f)$, where $U(f)$ is the signal we want to infer from the noise-free measured signal $S(f)$, and an expected noise $N(f)$, but we don't want to apply a template during filtering. If we measure the signal $C(f) = N(f) + S(f)$, then we'd like to construct a filter to estimate $\tilde{U}(f)$ as

$$\tilde{S}(f) = R(f)\tilde{U}(f) = \Phi(f)C(f) \tag{E.75}$$

Here I've tried to explicitly keep the template and measurements separate. If we minimize the difference between $U(f)$ and $\tilde{U}(f)$, we arrive at the maximum likelihood filter estimate

$$\Phi(f) = \frac{|S(f)|^2}{|S(f)|^2 + |N(f)|^2} \tag{E.76}$$

Contrasting this with the single-pulse optimum filter, we see that it discards the frequency information in the signal. So far we still have an unknown signal in this equation, but given that the signal and noise are uncorrelated, we can re-write this as

$$\Phi(f) \approx \frac{|C(f)|^2 - J(f)}{C(f)} = 1 - \frac{J(f)}{|C(f)|^2} \tag{E.77}$$

This filter has the nice property that the bandwidth of the filter increases as the signal becomes stronger, and the filter goes to 0 for the case that there's no signal. This of course relies heavily on the assumption that our noise and signal estimates are fairly precise in the highest signal/noise region.

Say you now have a signal template such that $S(f) = As(f)$ where $s(f)$ is normalized to 1. You would then estimate the signal amplitude as

$$\tilde{A} = \mathrm{Re}\left[\int_{-\infty}^{\infty} \tilde{S}(f)df\right] = \mathrm{Re}\left[\int_{-\infty}^{\infty} \Phi(f)C(f)df\right] \tag{E.78}$$

This is obviously not as clean as the normal optimum filtering formalism, but it does allow you to go further through the process before needing a signal template, and may be a useful trigger quantity to determine whether a pulse exists in data.

E.8 Time-Domain OF: Matched Filtering

Rather than re-doing the full derivation of the matched filter, I want to lay down the basics here so people can appreciate what the trade-offs are between time-domain and frequency-domain optimal filtering. For a full derivation and an extension of the matched filter to include time offset, see Ref [38].

We construct the matched filter to estimate the amplitude of a known signal as

$$\hat{A}(t) = \int_{-\infty}^{\infty} \phi(t' - t)' v(t')dt' \tag{E.79}$$

We already know how this goes from the frequency domain optimal filter when we take the maximum likelihood estimate for the filter. In this case, instead of using the noise PSD, we use the noise auto-correlation matrix

$$R(\tau_1,\tau_2) = \frac{1}{T}\int_0^T (n(t+\tau_1)-\mu)(n(t+\tau_2)-\mu)dt = \left[\frac{1}{T}\int_0^T n(t+\tau_1)n(t+\tau_2)-\mu^2\right] \tag{E.80}$$

which is obviously over-defined; this is a matrix with Toeplitz symmetry and is fully specified by the first row or column (this should help to speed up computation). We then construct the filter

$$\phi(\tau_2) = \int_0^T s(-\tau_1)R^{-1}(\tau_1,\tau_2)d\tau_1 \tag{E.81}$$

which needs to be normalized by the template using

$$N_\phi = \int_0^T\int_0^T s(-\tau_1)R^{-1}(\tau_1,\tau_2)s(\tau_2)d\tau_1 d\tau_2 \tag{E.82}$$

which we know from all linear least-squares problems gives us the filter resolution

$$\sigma^2 = N_\phi^{-1} \tag{E.83}$$

We then get the normalized filter

$$\phi(\tau_2) = \frac{\phi'(\tau_2)}{N} \tag{E.84}$$

which allows us to calculate our expected amplitude.

Let's see what happens for this filter for the special case that the noise is completely uncorrelated and white. We find

$$R(\tau_1,\tau_2) = \delta(\tau_2-\tau_1)\sigma^2_{white} \tag{E.85}$$

which gives us the OF resolution

$$\sigma^2 = N_\phi^{-1} = \left[\sigma^{-2}_{white}\int_0^T\int_0^T s(-\tau_1)\delta(\tau_2-\tau_1)s(\tau_2)d\tau_1 d\tau_2\right]^{-1} \tag{E.86}$$

$$= \sigma^2_{white}\left[\int_0^T s(-\tau_2)s(\tau_2)d\tau_2\right]^{-1} \tag{E.87}$$

So the filter just has some resolution reduced relative to the variance of the trace according to the convolution of of the templates (defined to have unit amplitude). If we have a flat template we just get an averaging reduction

$$\sigma^2 = \frac{\sigma^2_{white}}{N} \tag{E.88}$$

If we have a normalized exponential template, we find the convolution of the exponential gives T/τ^2, so we get the resolution

$$\sigma^2 = \sigma^2_{white} \frac{(\tau/\Delta t)^2}{N} \tag{E.89}$$

where I've implicitly done the discrete math to get the correct dimensionality. In essence, we find that resolution scales linearly with pulse fall-time, and as the inverse square root of the number of samples taken. Thus longer traces and shorter pulses will give us better resolution.