C3. Matched Filters

Objectives

- Define the inner product between two vectors
- Define the correlation between two signals
- Detect the presence of a given pulse by the matched filter

1. Introduction

A very important problem in signal processing is the determining how two signals compare with each other. To give an example, you can program most cell phones to obey voice commands: how does it do? The cellphone compares the signal from the microphone to a dictionary of possible phonemes and decide which one is the “closest” one.
Another example we will be addressing is the detection of a radar or sonar return. In this case a pulse is transmitted and, if there is a target, it bounces back. Since there is distortion and noise, what we receive is not identical to what we transmit and we need sort of define a measure of similarity, so we can decide whether the received signal is the oulse we expect.

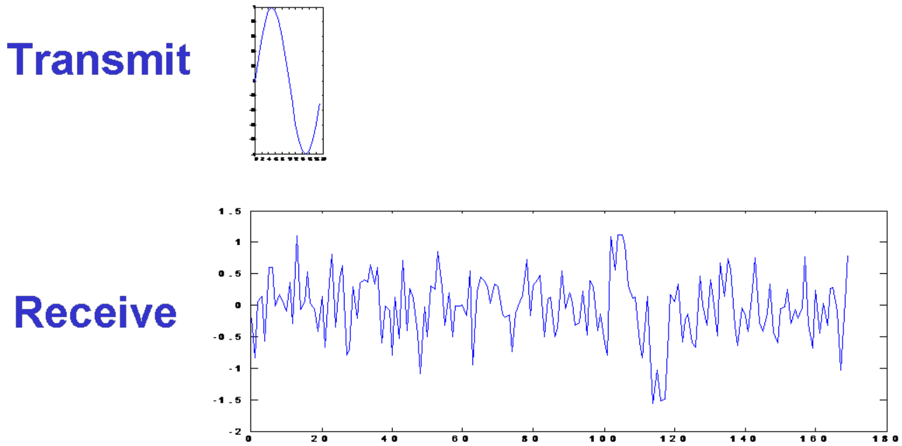
The main subject of this chapter is the correlation between two signals. It has a very specific definition, presented below, and it is a very effective way of deciding whether two signals are “correlated” to each other or not. For example in the case of the radar return, the pulse coming back from the target is “correlated” to the pulse we transmit. Based on this we will be introducing the concepts of the matched filter, which is at the basis of detection mechanisms for both radar and sonar systems.

2. Correlated and Uncorrelated Signals

VIDEO: Introduction (10:53)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_1_introduction.mp4

In this section we address the problem of determining whether or not two different signals are correlated with each other. In particular, given two signals $x[n], y[n]$, with $n = 0,..., N-1$ of the same length N we want to determine a criterion for which we can decide whether or not they are from the same source.

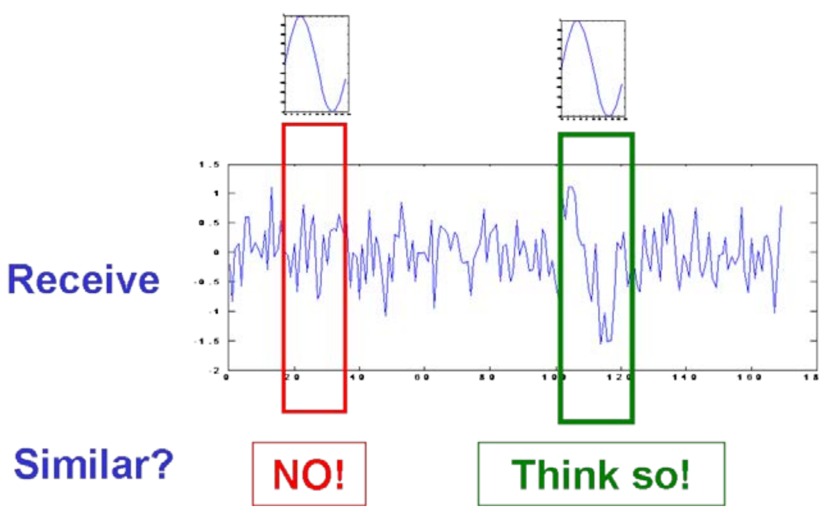
Typical example is in *radar* or *sonar* applications. As it is well known, we determine the distance of a target by transmitting a pulse (say $x[n]$) and detecting the return. The time interval between transmission and reception gives the information on the distance of the target. The figure below illustrates the problem.



Transmitted Pulse (top) and received waveform (bottom)

The goal of the radar or sonar system is to detect the transmitted pulse, which in this case is one period of a sinusoid, from the received signal shown at the bottom of the figure. In this case we can eyeball where the return is, and it is obvious that, what is received, is not an exact replica of the transmitted pulse: there is noise and distortion to be taken into account.

The way to proceed is as shown in the figure below. We compare the transmitted pulse with every segment of the recived signal and, based on some decision criterion, we decide whteher or not there is any “correlation” between what is received and what has been transmitted.



Detection of the transmitted pulse by comparing the received signal with the transmitted pulse.

From this example it is clear that we need a “measure” of how close (or how “correlated”) two signals are. Although the term “correlation” at this point is used very loosely, as an everyday term, we will see that it is a well defined mathematical term to assess how “correlated” two signals are.

VIDEO: Inner Product (9:55)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_01_innerProduct.mp4

The arguments we will be developing are based on the concept of “Inner Product” between two vectors. In particular

Definition. Given two discrete time signals $x[n], y[n]$ with $n = 0,\ldots, N-1$, we define the Inner Product between the two signals as

$$r_{xy} = \sum_{n=0}^{N-1} x[n] y^*[n]$$

For completeness, we include the complex conjugate (“*”) since the signals can be complex. In most of what we are going to cover in this course, the signals are real. Before going into why this important, let see a couple of properties of the inner product.

Properties of Inner Product. The inner product between two signals (or two vectors) defined above is such that:

a. $r_{xx} = \sum_{n=0}^{N-1} x[n] x^*[n] = \sum_{n=0}^{N-1} |x[n]|^2 \geq 0$

b. $|r_{xy}|^2 \leq r_{xx} r_{yy}$

with the equality $|r_{xy}|^2 = r_{xx} r_{yy}$ valid if and only if $y[n] = C\, x[n]$ for some constant C.

These two properties mean the following:

a. r_{xx} is just the energy of the signal by definition and it is never negative;

b. the square of the inner product itself is upperbounded by $r_{xx} r_{yy}$, the product of the energies of the two signals, and it is maximum when the two signals are the same, apart from a scaling factor.

These properties come from a geometric interpretation of signals and vectors, for which a signal is a vector in an N dimensional space and the inner product is what the reader might have seen in geometry as the “inner product” (also called “internal product” or “dot product”) between two vectors. The geometric definition is the product of the lengths of the vectors (just like $\sqrt{r_{xx}}$ and $\sqrt{r_{yy}}$) times the “cos” of the angle in between.

VIDEO: Correlation Coefficient (13 :35)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_01_correlation.mp4

By this definition of inner product we define the "correlation coefficient" as

$$\rho_{xy} = \frac{|r_{xy}|}{\sqrt{r_{xx} r_{yy}}}$$

with the following properties:

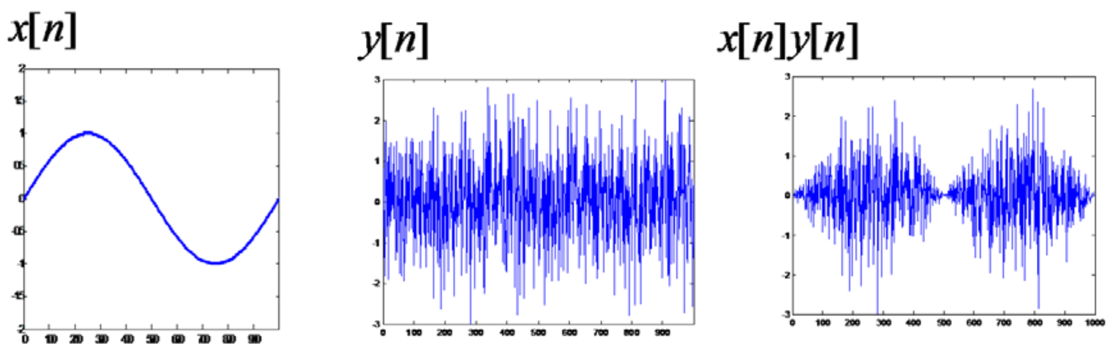
<u>Properties of Correlation Coefficient.</u> Let two signals $x[n], y[n]$ of the same length N be "zero mean". Then the term ρ_{xy} is such that

$$0 \leq \rho_{xy} \leq 1$$

with $\rho_{xy} \approx 1$ if the two signals are similar to each other, and $\rho_{xy} \approx 0$ if they are not.

From this we can see that given two signals we have a measure to determine how closely related two signals are.

Example. Consider again the radar return we discussed before. Then consider the inner product of the signal we transmitted and a part of the received signal with only noise. Referring to the figure below, we see the transmitted pulse $x[n]$, a part of the received signal containing only noise (no return) $y[n]$ and the product $x[n]y[n]$ all with the same length $n = 0,..., N-1$. This corresponds to the part of received signal in the leftmost window in the previous figure.



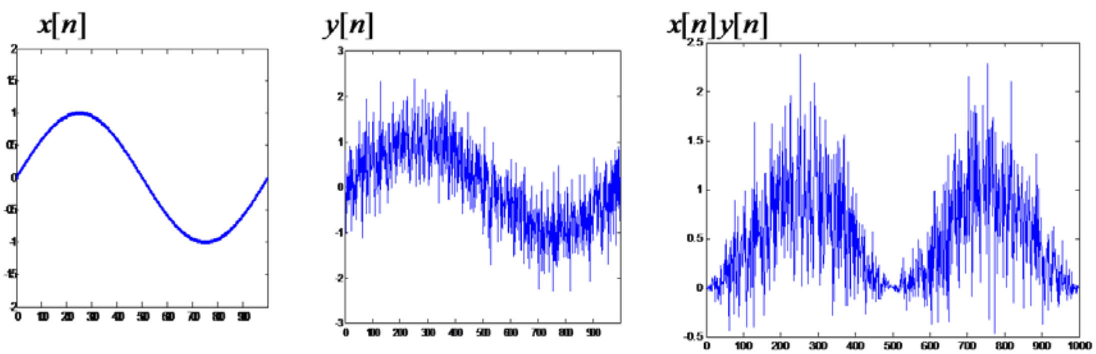
Transmitted Pulse, Noise only and the product of these two signals

Notice that $x[n]y[n]$ has positive and negative values and its sum is fairly small. In fact we can easily compute the inner product and correlation coefficients as

$$r_{xy} = 2.27$$
$$r_{xx} = 500$$
$$r_{yy} = 982$$
$$\rho_{xy} = 0.003$$

A value $\rho_{xy} = 0.003$ close to “zero” shows that there is no “correlation” between the to signals. In other words the reurn pulse is not there.

Example. Now take the case in which there is a return. The figure below shows again the transmitted pulse $x[n]$ and a segment $y[n]$ of the received signal where there is a return, together with their product $x[n]y[n]$.



Transmitted Pulse, Received Pulse and the product of these two signals

Since $x[n]$ and $y[n]$ tend to be the same (or at least similar), the product $x[n]y[n]$ tends to be mostly positive. Therefore the inner product (ie the sum of $x[n]y[n]$) will not be small and the correlation coefficient is computed as

$$r_{xy} = 494$$
$$r_{xx} = 500$$
$$r_{yy} = 754$$
$$\rho = 0.8$$

and it is fearly close to “one”. This shows that the signal is there.

3. Inner Product in Matlab

VIDEO: Inner Product in Matlab (4:48)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_02_innerProductMatlab.mp4

As we have seen several times, in Matlab a finite length signal is a vector of samples. Although in the time domain we tend to start the first index at “zero”, indicating “time zero" as a starting point, in Matlab the index cannot be negative and it starts at "one". For this reason we make a distinction in notation between a discrete time signal $x[n]$, n=0,…,N-1, and the elements $x(n), n = 1,.., N$ of a vector in Matlab, as

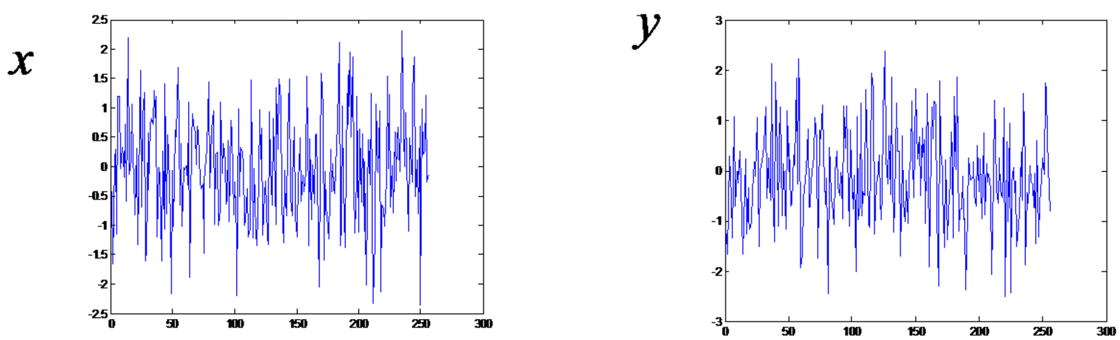
$$x = [x(1),..., x(N)]$$

Given two signals as <u>row vectors</u> x and y of the same length, the inner product is computed as

```
rxy = x*y'
```

where the "prime" stands for "complex conjugate and transpose".
See a couple of examples.

Example. Consider the two signals shown below, each one stored as a row vector.



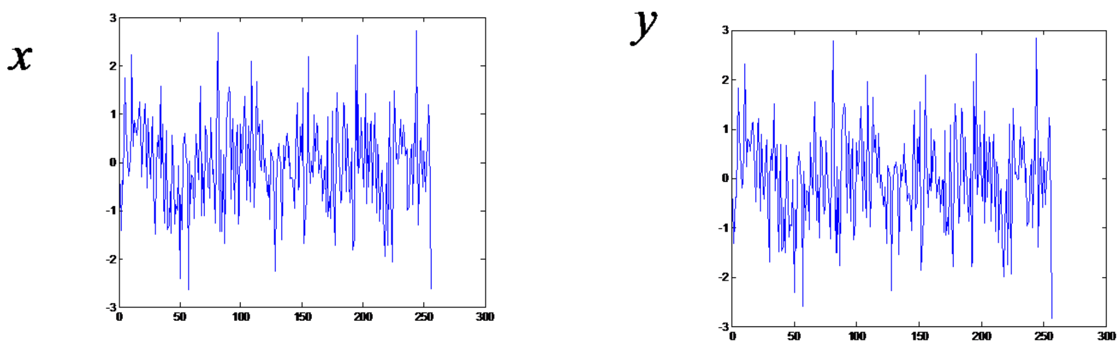
Two signals for the example

Then we can compute

```
rxy  = x*y';
rxx  = x*x';
ryy  = y*y';
rho  = abs(rxy)/sqrt(rxx*ryy)
```

and this yields rxy=-19.7, rxx=218.8, ryy=241.9 and the correlation coefficient rho=0.0856. This shows that the two signals are not correlated with each other.

Example. Take two other signals shown in the figure below.



Two signals for the example

Same computations as for the previous example yield rxy=230.9, rxx=229.6, ryy=234.3 and the correlation coefficient rho=0.995 close to "one", so the two signals are strongly correlated.

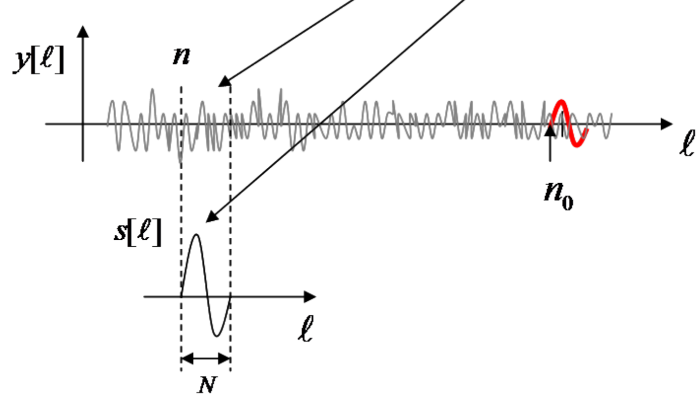
4. The Matched Filter.

[VIDEO: Radar/Sonar Application (4:15)](http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_05_radarSonar.mp4)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_05_radarSonar.mp4

In this section we want to extend what we understand about the inner product and the correlation coefficient to the design of a filter to detect the time of arrival of a pulse. This filter will be called a "matched filter" since it is matched to the particular pulse we try to detect.

Standard application to radar or sonar is shown in the figure below where again we show the received signal with additive noise and the transmitted pulse (one period of a sinusoid).

$$r_{ys}[n] = \sum_{\ell=0}^{N-1} y[n+\ell]s^*[\ell]$$



Correlation between received signal and transmitted pulse

At every time n define the following

$$r_{ys}[n] = \sum_{\ell=0}^{N-1} y[n+\ell]s^*[\ell]$$

It is easy to see that $r_{ys}[n]$ defined in this way is the inner product between the transmitted pulse $s[\ell], \ell = 0,..., N-1$ and the window of received data $y[n+\ell], \ell = 0,..., N-1$, starting at sample *n*. Following what we have seen in the previous section, when $n \neq n_0$ we correlate the pulse with just noise, and the result is small. When $n = n_0$ we correlate the transmitted pulse with the received pulse, and the output is large.

[VIDEO: Matched Filter (18:37)](http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_06_matchedFilter.mp4)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_06_matchedFilter.mp4

A close look at the expression above shows that it can be written as a "convolution" and therefore implemented as a digital filter. In fact write it as

$$\begin{aligned} r_{ys}[n] &= \sum_{n=0}^{N-1} y[n+\ell]s^*[\ell] \\ &= s^*[N-1]y[n+N-1] + ... + s^*[1]y[n+1] + s^*[0]y[n] \end{aligned}$$

and compare it with the response of a digital FIR filter seen in the previous chapters, as

$$\hat{r}[n] = h[0]y[n] + ... + h[1]y[n-1] + h[N-1]y[n-N+1]$$

Comparing the last two equations, we can see that the impulse response

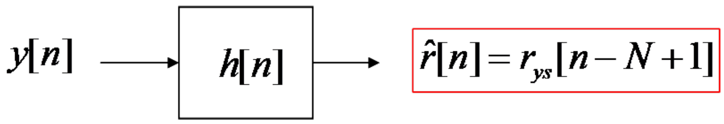
$$h[n] = s^*[N-1-n], \; n = 0,..., N-1$$

yields an FIR filter whose output is given by

$$\hat{r}[n] = r_{ys}[n-N+1]$$

that is to say $r_{ys}[n]$ with a time delay of $N-1$. The impulse response of this filter is just the transmitted pulse, reversed in time and cojugated, since $h[0] = s^*[N-1]$, $h[1] = s^*[N-2]$, ..., $h[N-1] = s^*[0]$.

This is shown in the figure below.



$$h[n] = s^*[N-1-n], \quad n = 0,\dots,N-1$$

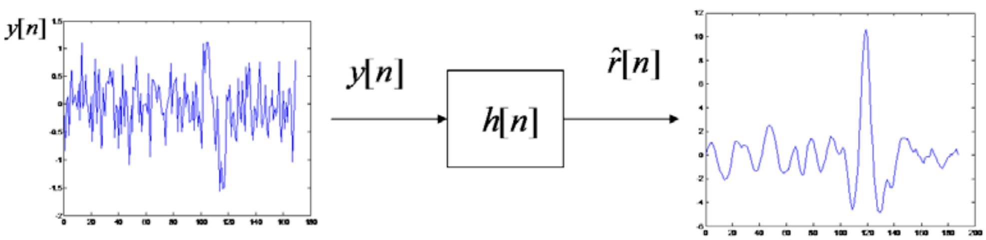
Crosscorrelation as the output of an FIR Filter

If the received signal is like

$$y[n] = As[n-n_0] + w[n]$$

with A indicating attenuation and $w[n]$ additive noise, the output of the matched filter yields a peak at time $n = n_0 + N - 1$, that is to say with a time delay depending on the length of the transmitted pulse.

Example. Assume again that the transmitted pulse $x[n], n = 0,\dots,19$ is one period of a sinusoid and the transmitted signal is shown below.



Received Signal $y[n]$ ***and output of matched filter***

If the maximum at the output of the match filter is (say) at n=120, then the time of arrival of the pulse (time delay between transmitted and received signal) is

$$n_0 = 120 - 19 = 101$$

5. Autocorrelation of a Signal

VIDEO: Autocorrelation (14:03)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_07_autocorrelation.mp4

Now the question is how do we choose a "good pulse", ie a pulse with a sharp maximum with no ambiguities. In the absence of noise, let us see how we can characterize the output of a match filter. Again assume the input to be $y[n] = As[n-n_0]$ with no noise, and the output of the match filter becomes

$$\hat{r}[n] = r_{ys}[n-N+1]$$

with

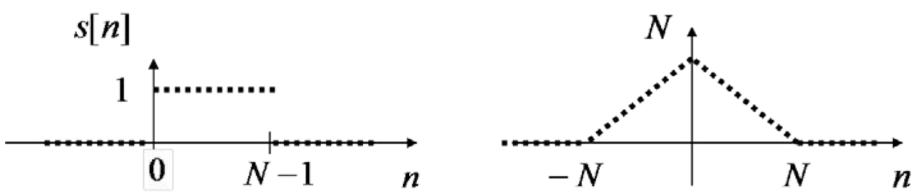
$$r_{ys}[n] = A\sum_{n=0}^{N-1} s[n-n_0+\ell]s^*[\ell]$$
$$= A r_{ss}[n-n_0]$$

The sequence $r_{ss}[n]$ defined as

$$r_{ss}[n] = \sum_{\ell=0}^{N-1} s[n+\ell]s^*[\ell]$$

is called the autocorrelation of the signal and it contains quite a bit of information on the signal characteristics. See an example.

Example. Let $s[n], n = 0,\ldots, N-1$ be a rectangular pulse as shown on the left of the figure below.



Rectangular Pulse (left) and its Autocorrelation (right)

Its autocorrelation $r_{ss}[n]$ is shown n the right, and it can be easily computed by applying the definition. In fact, see a few terms

$$r_{ss}[0] = \sum_{\ell=0}^{N-1} s[\ell]s^*[\ell] = \sum_{\ell=0}^{N-1} \left|s[\ell]\right|^2 = N$$
$$r_{ss}[1] = \sum_{\ell=0}^{N-2} s[1+\ell]s^*[\ell] = \sum_{\ell=0}^{N-2} 1 = N-1$$

and in general, for any k between 1 and N-1 we can write

$$r_{ss}[k] = \sum_{\ell=0}^{N-1-k} s[k+\ell]s^*[\ell] = \sum_{\ell=0}^{N-1-k} 1 = N-k$$
$$r_{ss}[-k] = \sum_{\ell=0}^{N-1-k} s[-k+\ell]s^*[\ell] = \sum_{\ell=k}^{N-1} 1 = N-k$$

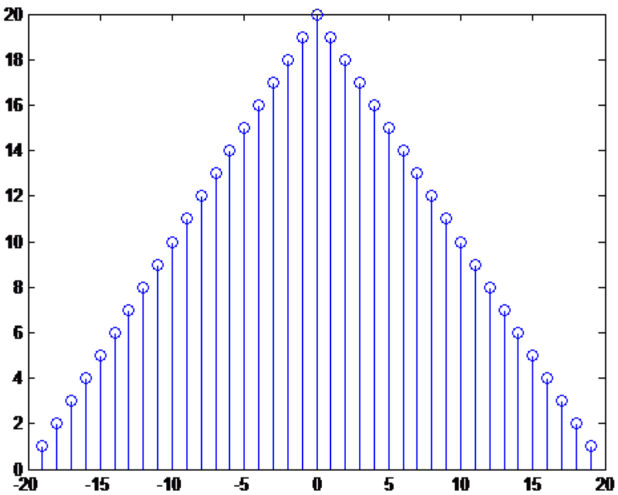
The plot of this data yields a triangle as shown above.

Clearly for more complex signals we can use "xcorr" in Matlab as follows

```
N=20;           % data length
s=ones(1,N);    % rectangular pulse of length N
rss=xcorr(s);   % compute autocorrelation
```

```
n=-N+1:N-1;     % vector of indeces for the plot
stem(n,rss)     % plot the autocorrelation
```

The end result of the plot is shown in the figure below and it is as expected.

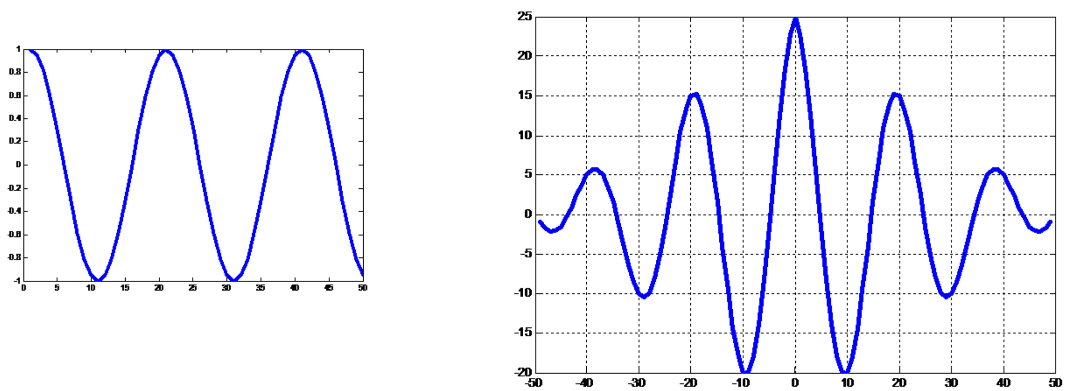


Autocorrelation of rectangular pulse of length N=20 samples

One thing we can see from this plot is that the maximum is not very sharp and, with a bit of noise, it can exhibit some uncertainties.

Let's see some other signals.

Example. Let $s[n]$ be a sinusoidal signal of length $N=50$ as shown on the left of the figure below. Its autocorrelation is shown on te right hand side of the same figure.



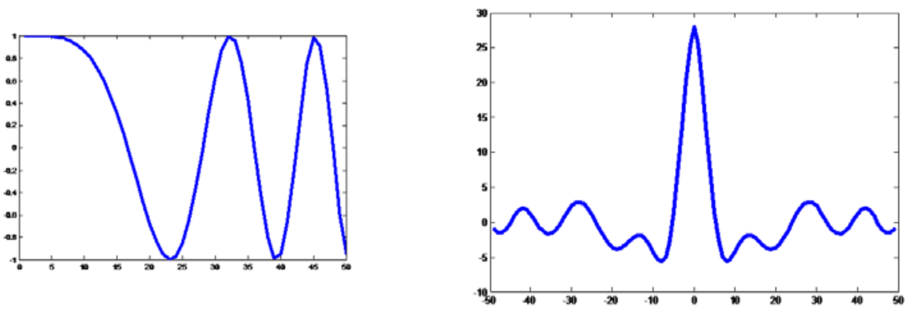
Sinusoid (left) of length N=50 samples and its autocorrelation (right)

Notice that the autocorrelation hs several peaks, negative and positive and still it is not very concentrated around te mximum.

Example. Let the pulse $s[n]$ be a chirp of length $N = 50$ samples. It is generated in Matlab as

```
s=chirp(0:49,0,49,0.1)
```

The signal (left) and its autocorrelation (right) are shown in the figure below. Notice now that the peak is sharp and this is very good choice for target detection.

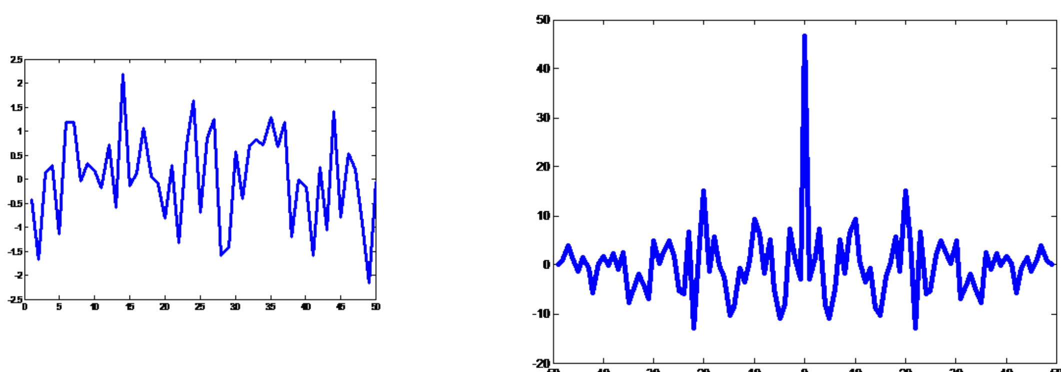


Chirp (left) and its autocorrelation (right)

Example. Last example is a signal $s[n]$ which is generated as pseudonoise as

```
s=randn(1,N);
```

with N again the pulse length in number of samples. The pulse and its autocorrelation are shown in the figure below. Notice that, even for this signal here, the autocorrelation has a sharp peak, just suitable for signal detection.



Pseudonoise (left) and its autocorrelation (right)

6. Detection with Noise

VIDEO: Detection with Noise (24.39)
http://faculty.nps.edu/rcristi/eo3404/c-filters/videos/c3_08_detectionWithNoise.mp4

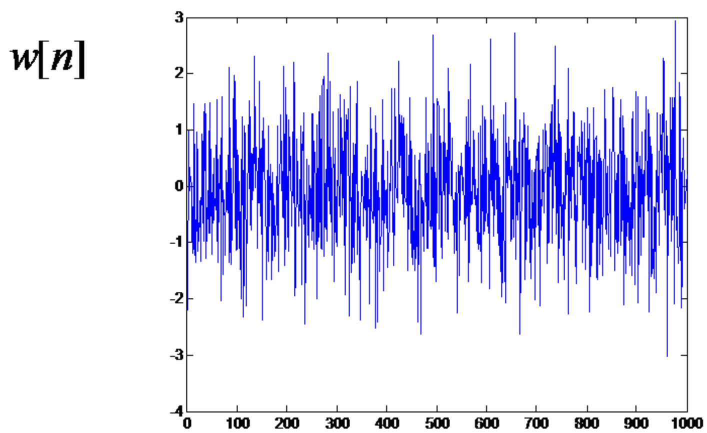
In this last section we see a very important property of the matched filter. When the pulse is detected, this filter yields a peak. But the question now is how is this peak affected by noise.

In order to see this we need to first characterize noise and disurbances. In numerous applications a disturbance $w[n]$ is characterized by what is called “white noise”. It sounds exactly like the colorless white noise we experience in real life when, for example, we turn on the radio or the TV and there is no reception: we just hear some “white noise”. It is called white because it contains all the frequencies like the white light.

A plot of a white noise signal, generated in Matlab as

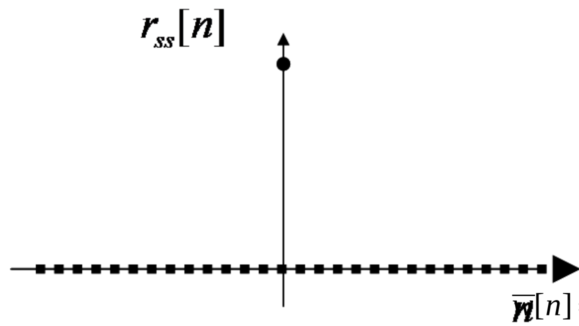
```
w = randn(1,N);
```

is shown in the figure below.



White Noise

As the sequence tends to an infinite length, its autocorrelation $r_{ww}[n]$ tends to become a “delta” function as shown in the figure below.

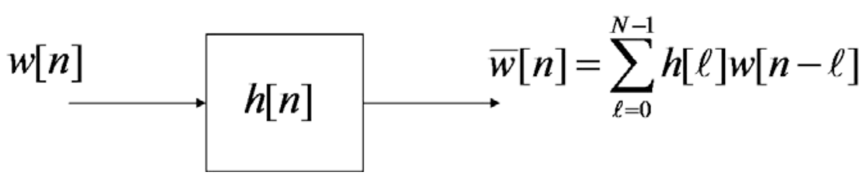


$$\overline{w}[n] = \sum_{\ell=0}^{N-1} h[\ell]w[n-\ell]$$

Autocorrelation of White Noise

What is important now is to understand how a white noise disturbance affects the output of the matched filter. In other words we want to see what is the relationship between the Signal to Noise Ratio of the received signal, at the input of the filter, and the Signal to Noise Ratio after the matched filter. We will see that the SNR itself will be improving by a factor of *N*, the length of the transmitted pulse.

In order to see this, consider the figure below, where we have a filter with impulse response $h[n], n = 0,.., N-1$, of length *N* samples and the input is a white noise signal $w[n]$. The goal is to relate the power of the signal at the output with the power of the signal at the input.



White Noise Input $w[n]$ ***to an FIR filter and the corresponding output*** $\overline{w}[n]$

If we call $\overline{w}[n]$ the output of the filter, it is related to the input by the convolution as

$$\overline{w}[n] = \sum_{\ell=0}^{N-1} h[\ell]w[n-\ell]$$

From this expression we can compute the average Power by taking the average over a long sequence as

$$\begin{aligned}\frac{1}{M}\sum_{n=0}^{M-1}\left|\overline{w}[n]\right|^2 &= \frac{1}{M}\sum_{n=0}^{M-1}\left(\sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1} h[\ell_1]h[\ell_2]w[n-\ell_1]w[n-\ell_2]\right)\\ &= \sum_{\ell_1=0}^{N-1}\sum_{\ell_2=0}^{N-1} h[\ell_1]h[\ell_2]\left(\frac{1}{M}\sum_{n=0}^{M-1} w[n-\ell_1]w[n-\ell_2]\right)\end{aligned}$$

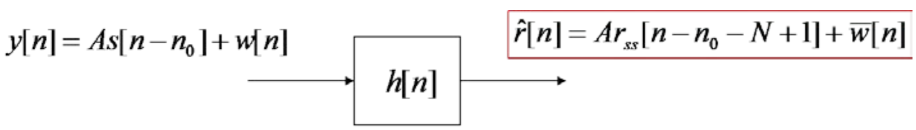
For a length M large, the rightmost sum tends to the ideal correlation of the white noise, as

$$\frac{1}{M}\sum_{n=0}^{M-1} w[n-\ell_1]w[n-\ell_2] \to \begin{cases} 0 & \text{if } \ell_1 \neq \ell_2 \\ \left(\frac{1}{M}\sum_{n=0}^{M-1} |w[n]|^2\right) & \text{if } \ell_1 = \ell_2 \end{cases}$$

As a consequence we can see that as $M \to \infty$,

$$\frac{1}{M}\sum_{n=0}^{M-1} |\overline{w}[n]|^2 \to P_{\overline{W}} = \left(\sum_{n=0}^{N-1} |h[n]|^2\right) P_w$$

Now let us go back to the matched filter which is recalled in the figure below.



$$h[n] = s^*[N-1-n], \quad n = 0,\ldots,N-1$$

Summary of Matched Filter with Additive Noise

At the peak of the output, ie at $n = n_0 + N - 1$, the output of the filter is

$$\hat{r}[n_0 + N - 1] = Ar_{ss}[0] + \overline{w}[n_0 + N - 1]$$

where

$$|Ar_{ss}[0]|^2 = \left(\sum_{n=0}^{N-1} |As[n]|^2\right)\left(\sum_{n=0}^{N-1} |s[n]|^2\right)$$

is the square of the peak of the signal, what we can call the instantaneous power (there is only one sample!) at detection. Also the Power of the noise, from what we have seen before, becomes

$$P_{\overline{W}} = \left(\sum_{n=0}^{N-1} |s[n]|^2\right) P_W$$

Therefore the Signal to Noise Ratio at the peak of the output signal becomes

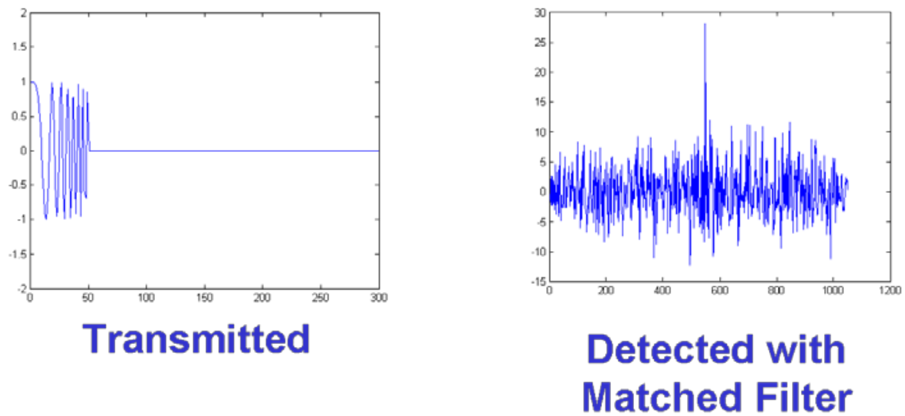
$$SNR_{peak} = \frac{N \times P_S \times \left(\sum_{n=0}^{N-1} |s[n]|^2\right)}{\left(\sum_{n=0}^{N-1} |s[n]|^2\right) \times P_W} = N \times SNR$$

since the pulse at the receiver has a Power

$$P_S = \frac{1}{N}\sum_{n=0}^{N-1} | As[n] |^2$$

The above expression shows the relation between the received SNR and the SNR after the matched filter, when we have a detection. The very good news is that there is a “Processing Gain” *N* which depends on the length of the pulse. Let ‘s see and example.

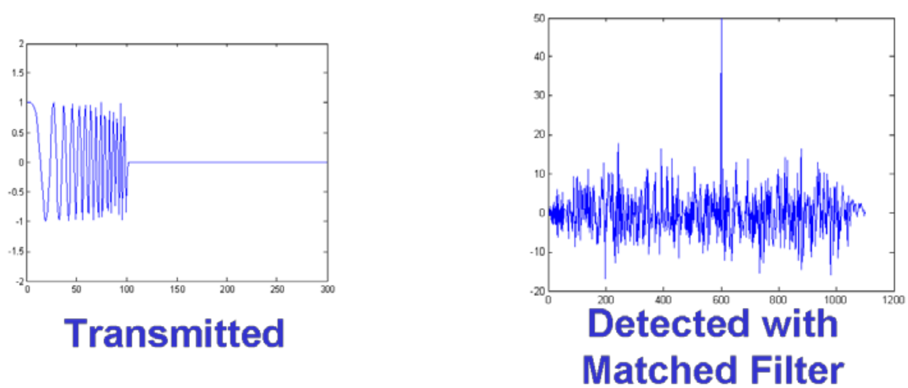
Example. Suppose we transmit the chirp shown, with length *N=50* samples. Let the SNR at the receiver be SNR=0dB.



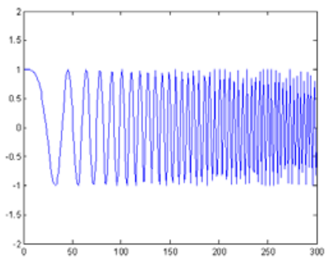
Transmitted and Detected Chirp with length N=50 and received SNR=0dB

The peak of the output shows that we have a detection of the pulse and, at the peak, the SNR is $10\log_{10}(50) = 17dB$

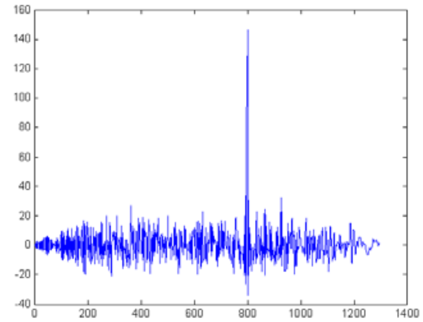
Long chirps of lengths *N=100* and *N=300* are shown below, for the same SNR=0dB at the reception. As expected, the values of the peak get larger as the pulse length *N* in creases.



Transmitted and Detected Chirp with length N=100 and received SNR=0dB



Transmitted



Detected with Matched Filter

Transmitted and Detected Chirp with length N=300 and received SNR=0dB