III. Signal Similarity Measures

In many situations, we need a quantitative measure of the similarity of two signals. For example, suppose x(t) is the signal some system should ideally produce, y(t) is the signal the system actually produces. Then, as a measure of how well the system has performed, we need a quantitative measure of how similar y(t) is to x(t). As another example, suppose r(t) is a measured signal that is either the "desired" signal $s_1(t)$ plus some measurement noise, or the "desired" signal $s_2(t)$ plus some measurement noise, and suppose a system must be built that decides which of the two desired signals the measured signal r(t) contains. Such a system needs a signal similarity measure in order to compare r(t) to $s_1(t)$ and r(t) to $s_2(t)$.

In summary, signal similarity measures are needed for quantitative performance measures for the systems we design and as an integral piece of certain systems. In the following we introduce and discuss the two most important signal similarity measures.

A. Difference Energy, Mean-Squared Difference and Mean-Squared Error

The difference energy between signals x(t) and y(t) is simply the energy of the difference signal x(t)-y(t). For continuous-time signals, the difference energy over the time interval (t_1,t_2) is

$$E(x-y) = \int_{t_1}^{t_2} (x(t)-y(t))^2 \, dt \ .$$

Similarly, for discrete-time signals x[n] and y[n], the difference energy over the time interval $[n_1,n_2]$ is

$$E(x-y) = \sum_{n_1}^{n_2} (x[n]-y[n])^2 \ .$$

A closely related signal similarity measure is the *mean-squared difference* (MSD) between signals x(t) and y(t), which is simply the mean-squared value of the difference signal x(t)-y(t). For continuous-time signals, the MSD over the time interval (t_1,t_2) is

$$MSD(x,y) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} (x(t)-y(t))^2 \, dt = \frac{1}{t_2-t_1} E(x-y) \ .$$

Similarly, for discrete-time signals, the MSD over the time interval $[n_1,n_2]$ is

$$MSD(x,y) = \frac{1}{n_2-n_1+1} \sum_{n_1}^{n_2} (x[n]-y[n])^2 = \frac{1}{n_2-n_1+1} E(x-y) \ .$$

When one of the signals, say x(t), is considered to be the "desired" signal and the other, say y(t), is considered to be an approximation to it, then the difference signal x(t)-y(t) is considered to be an *error signal*, and the mean-squared difference is called the *mean-squared error* and abbreviated MSE(x,y) or simply MSE. MSE is considered a measure of the quality of y(t) as an approximation to x(t), with small MSE indicating good quality.

In many situations, the significance of a particular value of MSE generally depends on the size or strength of the signal x(t). For example, an MSE value of 10 is considered *large* if the squared signal values of the desired signal are mostly smaller than 10, and is considered small if the squared values of the desired signal are much larger than 10. For such reasons, it is common to use *signal-to-noise ratio* as a measure of signal quality, which is defined by

$$SNR(x,y) = \frac{\sigma^2(x)}{MSE} \ ,$$

where $\sigma^2(x)$, which is t\he variance of x(t), is used as the measure of signal size. Large signal-to-noise ratio indicates good quality.

B. Signal Correlation

Another measure of the similarity of signals x(t) and y(t) is their *correlation*, which is defined

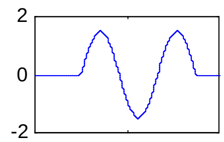
$$C(x,y) = \int_{t_1}^{t_2} x(t)\, y(t)\, dt \ ,$$

where (t_1,t_2) is the time interval of interest. Similarly, the correlation between two discrete-time signals x[n] and y[n] is defined as

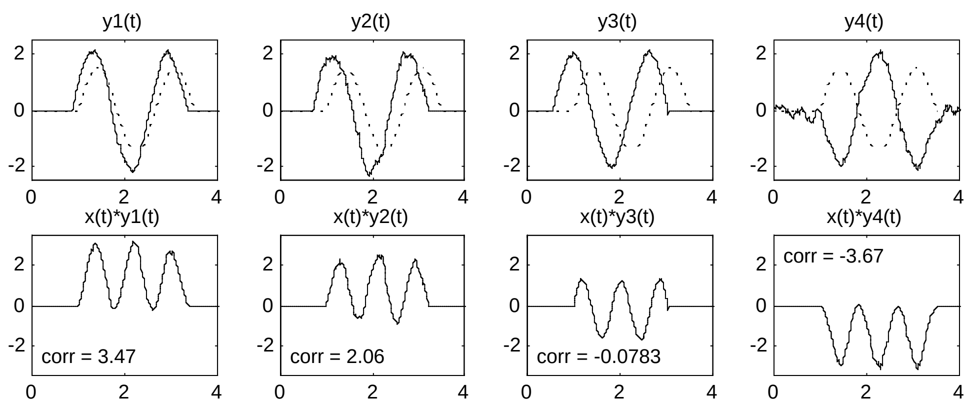
$$C(x,y) = \sum_{n_1}^{n_2} x[n]\, y[n] \ ,$$

where $[n_1,n_2]$ is the time interval of interest. The discussion to follow focuses on continuous-time signals. But everything applies equally to discrete-time signals.

To get a feeling for why correlation is a good measure of signal similarity examine consider the signal x(t) shown below



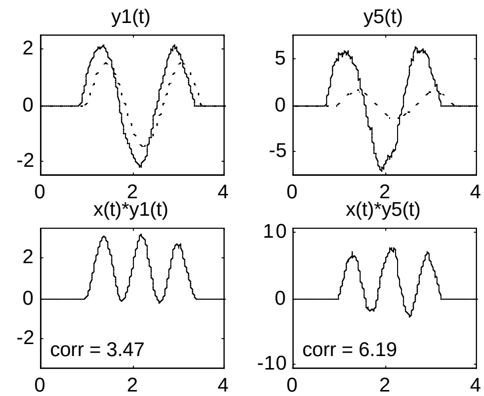
and consider the similarity of each of the signals below, $y_1(t)$, $y_2(t)$, $y_3(t)$, $y_4(t)$, to x(t).



As a reference, x(t) is shown with a dotted line in each of the above plots. Also shown below each signal is a plot of the product of x(t) with the signal. The correlation between x(t) and the given signal, which is the area under this plot, is also marked on the plot. Intuitively, we see that x(t) is more like $y_1(t)$ than the other signals, and this is reflected in $C(x,y_1)$ being larger than the other correlations. What is happening is that $y_1(t)$ tends to be positive where x(t) is positive and negative where x(t) is negative. Thus, the product $x(t)\,y_1(t)$ is mostly positive, and the correlation $C(x,y_1)$ is large. The signal $y_2(t)$ has the same sign as x(t) less often. Thus $x(t)\,y_2(t)$ has negative area cancelling some of the positive area, leading to a smaller value of correlation. This is taken to the extreme in $x(t)\,y_3(t)$, for which the positive area is nearly completely cancelled by the negative area, causing $C(x,y_3)$ to be zero. The fourth signal, $y_4(t)$, almost always has the oppposite sign of x(t), causing $x(t)\,y_4(t)$ to be almost entirely negative, leading to $C(x,y_3)$ being very negative.

These examples show that C(x,y) tends to be large when y(t) follows the same trends as x(t) -- positive at times t that x(t) is positive, negative at times t that x(t) is negative. This explains why the everyday word "correlation" is taken as the name for the similarity measure C(x,y). We say that x(t) and y(t) are positively or negatively correlated, according to whether C(x,y) is positive or negative. When C(x,y) = 0, we say the signals are *uncorrelated*, meaning that they are very different in the sense that the positivity of one at time t gives no clues as to the positivity of the other.

As a next set of examples consider correlating $x(t)$ with $y_1(t)$, shown above, and also with $y_5(t) = 3\,y_2(t)$.



W observe that even though $x(t)$ is intuitively more similar to $y_1(t)$ than to $y_5(t)$, the correlation $C(x,y_1)$ is smaller than the correlation $C(x,y_5)$. What is happening is that correlation is being heavily influenced by the fact that $y_5(t)$ is considerably larger signal than $y_1(t)$, i.e. it has much larger energy. In many situations, it is important to prevent correlation from being influenced by signal size. In such cases, it is customary to use *normalized correlation* as defined by

$$C_N(x,y) \;=\; \frac{C(x,y)}{\sqrt{E(x)}\sqrt{E(y)}} \;=\; \frac{1}{\sqrt{E(x)}\sqrt{E(y)}} \int_{t_1}^{t_2} x(t)\,y(t)\;dt$$

as the signal similarity measure. Here, we have divided $C(x,y)$ by the square root of the energies of both signals. The following lists the values of $C(x,y)$ and $C_N(x,y)$

	$y_1(t)$	$y_2(t)$	$y_3(t)$	$y_4(t)$	$y_5(t)$
E(y)	5.42	5.44	4.95	5.42	49.0
C(x,y)	3.47	2.06	-0.08	-3.67	6.19
$C_N(x,y)$	0.89	0.53	-0.02	-0.94	0.53

We see now that $C_N(x,y_5) = C_N(x,y_2)$, i.e. that normalized correlation is not affected by the size of the $y_5(t)$.

If, as suggested by the example above, normalized correlation is not affected by the sizes of the signals, then there ought to be some largest value that it can have. The following inequality, called the *Cauchy-Schwarz inequality*, shows that the normalized correlation can never be larger than one, nor less than negative one.

$$\sqrt{E(x)}\,\sqrt{E(y)} \;\le\; C(x,y) \le \sqrt{E(x)}\,\sqrt{E(y)}$$

Equivalently,

$$-1 \;\le\; C_N(x,y) \;\le\; 1\;.$$

The proof of this inequality is beyond the scope of the course[11].

Notice that if $y(t)$ is simply an amplitude scaling of $x(t)$, as in $y(t) = a\,x(t)$ for all t, where $a > 0$, then

[11] One may find a version of the Cauchy-Schwarz inequality in most linear algebra textbooks.

$$E(y) = E(ax) = \int_{t_1}^{t_2} (a\,x(t))^2\, dt = a^2 \int_{t_1}^{t_2} x^2(t)\, dt = a^2\, E(x)$$

$$C(x,y) = \int_{t_1}^{t_2} x(t)\; a\,x(t)\, dt = a \int_{t_1}^{t_2} x^2(t)\, dt = a\, E(x)$$

$$\sqrt{E(x)}\sqrt{E(y)} = \sqrt{E(x)}\sqrt{a^2 E(s)} = a\, E(x)\ .$$

Thus in the case we see that

$$C(x,y) = \sqrt{E(x)}\sqrt{E(y)}$$

or equivalently

$$C_N(x,y) = 1\ ,$$

i.e. the Cauchy-Schwarz relation holds with equality. In fact, this is the only way to obtain equality. That is, it can be shown that

$$C(x,y) = \sqrt{E(x)}\sqrt{E(y)},\ \text{ or equivalently, }\ C_N(x,y) = 1,$$

when and only when $x(t)$ and $y(t)$ are the same except for a positive multiplicative scaling, i.e. when and only when

$$y(t) = a\,x(t)\ \text{ for some }\ a > 0\ \text{ and all }\ t\ .$$

Similarly, it can be shown that the only way for $C(x,y)$ to equal $-\sqrt{E(x)}\sqrt{E(y)}$, or equivalently for $C_N(x,y)$ to equal -1, is when and only when $x(t)$ and $y(t)$ are the same except for a negative multiplicative scaling, i.e. when and only

$$y(t) = a\,x(t)\ \text{ for some }\ a < 0\ \text{ and all }\ t\ .$$

A corollary to the Cauchy-Schwarz inequality is the fact that the correlation of a signal with itself equals the signals energy, i.e.

$$C(x,x) = E(x) \quad \text{for any signal } x.$$

The relation between correlation and mean-squared difference energy: The relation between mean-squared difference and signal correlation is

$$E(x-y) = E(x) - 2\, C(x,y) + E(y)\ .$$

Thus, for example, a large positive correlation $C(x,y)$ implies a small difference energy $E(x-y)$. This relation is demonstrated below.

$$E(x-y) = \int_{t_1}^{t_2} \big(x(t)-y(t)\big)^2\, dt = \int_{t_1}^{t_2} \big(x^2(t) - 2\,x(t)\,y(t) + y(t)^2\big)\, dt$$

$$= \int_{t_1}^{t_2} x^2(t) - 2\int_{t_1}^{t_2} x(t)\,y(t)\, dt + \int_{t_1}^{t_2} y^2(t)\, dt$$

$$= E(x) - 2\, C(x,y) + E(y)$$

Since difference energy and correlation are closely related, the choice of which to use is a matter of taste, of convenience, or dependent upon other factors. For example, correlation $C(x,y)$ tends to preferred over difference energy in situations where one signal, say x, is much larger than the other, y, is small. In this case $E(x-y) \cong E(x)$, which indicates that $E(x-y)$ depends very weakly on the smaller signal. Thus, it is very sensitive to noise and computational roundoff errors. In contrast, $C(x,y)$ is always greatly influenced by y. For example, when y is much smaller than x, doubling y causes $C(x,y)$ to double, but has little effect on $E(x-y)$. Thus correlation is less sensitive to noise and roundoff errors.

The uses of correlation in EECS 206

Correlation will be used in a couple of the lab assignments as a method for detecting, classifying or recognizing signals. It will also be seen later that one of the principal analysis techniques that we study (Fourier analysis) and the principal kind of systems we study (linear time-invariant filters) are based on correlation. That Fourier analysis is based on correlation relates to the discussion below about "signal components".

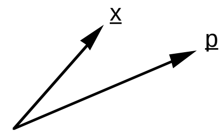
Signal components[12]

The question addressed in this subsection is: What does it mean for one signal to be a component of another? Specifically, suppose we are given signals $x(t)$ and $p(t)$ (or $x[n]$ and $p[n]$ in the discrete-time case).

- Is there a component of $x(t)$ that is like $p(t)$? (or of $x[n]$ that is like $p[n]$?)
- If so, how much $p(t)$ is in $x(t)$? (or $p[n]$ in $x[n]$?)
- How to define "how much of ___ is in ___ "?

For example, is there a component of $x(t)$ that is like $p(t) = \cos(3t)$?

Vector geometry: Such questions are similar to the following traditional questions in vector geometry: Suppose $\underline{x} = (x_1,...,x_N)$ and $\underline{p} = (p_1,...,p_N)$ are N-tuple vectors, illustrated below.

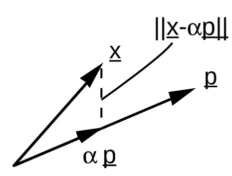


- Is there a component of $\underline{x}$ that is like $\underline{p}$?
- How much of $\underline{p}$ is in $\underline{x}$?

The conventional approach to answering these questions in vector geometry is to find the value α such that $\alpha\underline{p}$ is as close to $\underline{x}$ as possible, i.e. such that $||\underline{x}-\alpha\underline{p}||$ is as small as possible, where $||\underline{u}-\underline{v}||$ denotes the Euclidean distance between $\underline{u}$ and $\underline{v}$, as defined by

$$||\underline{u}-\underline{v}|| = \sqrt{\sum_{i=1}^{N} (u_i-v_i)^2}$$

For example, $\alpha\underline{p}$ for one choice of α is illustrated below.



Actually, it is a bit easier to find the value of α that minimizes $||\underline{x}-\alpha\underline{y}||^2$, because this avoids the square root. To find the proper α, let's equate to zero the derivative of $||\underline{x}-\alpha\underline{y}||^2$ with respect to α, and solve for α. First let's rewrite $||\underline{x}-\alpha\underline{y}||^2$:

$$||\underline{x}-\alpha p||^2 = \sum_{i=1}^{N} (x_i-\alpha p_i)^2 = \sum_{i=1}^{N} x_i^2 - 2\alpha\sum_{i=1}^{N} x_i p_i + \alpha^2\sum_{i=1}^{N} p_i^2$$

$$= ||\underline{x}||^2 - 2\alpha(\underline{x} \circ \underline{p}) + \alpha^2 ||\underline{p}||^2$$

---

[12]This section should be skipped or skimmed on first reading. It becomes suggested reading, but not required, when Fourier analsysi is introduced, as in Chapter 3 of our text.

where $||\underline{x}||$ and $||\underline{p}||$ are the lengths of $\underline{x}$ and $\underline{p}$, respectively, and $(\underline{x} \circ \underline{p})$ is the dot product defined by

$$(\underline{x} \circ \underline{p}) = \sum_{i=1}^{N} x_i \, p_i$$

Now differentiating and equating to zero gives

$$0 = \frac{d}{d\alpha} ||\underline{x}-\alpha p||^2 = \frac{d}{d\alpha}\Big(||\underline{x}||^2 - 2\,\alpha\,(\underline{x} \circ \underline{p}) + \alpha^2\, ||\underline{p}||^2 \Big)$$

$$= -2\,(\underline{x} \circ \underline{p}) + 2\alpha||\underline{p}||^2 ,$$

which yields

$$\alpha = \frac{(\underline{x} \circ \underline{p})}{||\underline{p}||^2}$$

We conclude that the component of $\underline{x}$ that is like $\underline{p}$ is $\frac{(\underline{x} \circ \underline{p})}{||\underline{p}||^2}\,\underline{p}$.

Fact: $\alpha = \frac{(\underline{x} \circ \underline{p})}{||\underline{p}||^2}$ is the unique value of α that makes the residual vector $(\underline{x}-\alpha\underline{p})$ and $\underline{p}$ orthogonal, where $\underline{u}$ and $\underline{v}$ are said to be orthogonal if $\underline{u} \circ \underline{v} = 0$.

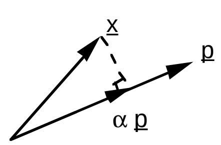
Proof: The dot product of $(\underline{x}-\alpha\underline{p})$ and $\underline{p}$ is

$$(\underline{x}-\alpha\underline{p}) \circ \underline{p} = (\underline{x} \circ \underline{p}) - \alpha(\underline{p} \circ \underline{p}) \quad \text{by the linearity of the dot product}$$

$$= (\underline{x} \circ \underline{p}) - \alpha\, ||\underline{p}||^2$$

which is zero when and only when $\alpha = \frac{(\underline{x} \circ \underline{p})}{||\underline{p}||^2}$, i.e. when and only when $(\underline{x}-\alpha\underline{p})$ and $\underline{p}$ are orthogonal.

With this fact in mind, we see that the component of $\underline{x}$ that is like $\underline{p}$ is the vector in the direction of $\underline{p}$ obtained by projecting $\underline{x}$ onto the direction of $\underline{p}$ as illustrated below.



Back to signals: Let us now return to the original questions for signals: Suppose we are given signals x(t) and p(t).

- Is there a component of x(t) that is like p(t)?
- If so, how much p(t) is in x(t)?
- How to define "how much of ___ is in ___ "?

Our approach will be to find the value α such that the difference energy $E(x(t)-\alpha\, p(t))$ is as small as possible. We will then say that "α p(t) is the component of x(t) that is like p(t)" and "α is the amount of p(t) that is in x(t)". The same approach applies to discrete-time signals.

The idea is that the question we are asking is just like the question for vectors, and we can use the same approach. The only difference is that instead of Euclidean distance as a measure of similarity we use difference energy. Indeed, for discrete-time signals the question is *exactly the same*, because the signals are vectors and difference energy is Euclidean distance squared. Thus in the discrete-time case, we can simply use the answers to the vector question. In doing so, we recognize that what is called "dot product" in the "vector domain", is just what we have called "correlation". Morever, it is easy to check that with "correlation" replacing "dot product", "energy" replacing "length squared", and "uncorrelated" replacing "orthogonal", the answer we found to the vector question applies to continuous-time signals as well as to discrete-time signals. Therefore, we immediately obtain the following:

- The value of α that minimizes the difference energy $E(x(t) - \alpha\, p(t))$ is
$$\alpha = \frac{c(x,p)}{E(p)} \ .$$
- The amount $p(t)$ that is in $x(t)$ is $\dfrac{c(x,p)}{E(p)}$.
- The component of $x(t)$ that is like $p(t)$ is $\dfrac{c(x,p)}{E(p)}\, p(t)$. (++)
- $\alpha = \dfrac{c(x,p)}{E(p)}$ is the unique value that makes the difference signal $(x(t)\text{-}\alpha p(t))$ and $p(t)$ uncorrelated.
- These answers apply to discrete-time signals as well, with $p[n]$ replacing $p(t)$.
- These answers apply to complex-valued signals, in discrete or continuous time. (Correlation between complex-valued signals is discussed below.)

Comments: Engineers have long recognized the connnections between signals and vectors. As a result, basic ideas from geometry, and more generally from linear algebra, are commonly used in signals and systems analysis. One of the most beneficial transferences is the idea that we can draw geometric pictures that represent signals and their relationships, such as those on the previous pages. For example, uncorrelated signals are drawn at right angles to one another. It often happens that a geometric picture will help one to understand some complex signal situation. It is also true that studying linear algebra will lead to increased understanding of signals and systems. For example, you might wish to learn as much as possible about linear algebra in Math 216 and to take Math 419 as an elective.

III. Basic Signal Processing Tasks

In this section, we describe three broad and nearly ubiquitous tasks that require the processing of signals. That is, there is need to develop systems that perform these tasks. Much of the remainder of the course will be devoted to developing techniques to design and improve such systems.

The first two tasks have a similar flavor. In each, the signal to be processed contains a component that interests us and a component that does not. That is, the signal $r(t)$ to be processed can be modeled as

$$r(t) = s(t) + n(t) ,$$

where $s(t)$ is the component that interests us and $n(t)$ is the component that does not. For example, the component that interests us might be the signal produced by someone speaking into a microphone, and the component that does not might be the signal produced by background noise. In the first task, called *signal recovery* or *noise reduction*, the goal is to recover the signal component $s(t)$ that interests us. For example, we might wish to recover the speech signal without the background noise. In the second task, called *signal detection* or *signal classification* or *signal recognition*, we wish to make a decision about the signal component that interests us. For example, we might wish to decide the identity of the speaker or what the speaker has said. These two tasks will be introduced in the next two subsections.

In each of the tasks, the noise $n(t)$ is not a known signal. If it were known, we could simply subtract it from $r(t)$, and there would be no need for a signal recovery or signal detection system. We also assume that the desired signal $s(t)$, or some aspect of it, is not known. If $s(t)$ were entirely known, we could dispense with $r(t)$, and simply display the signal $s(t)$. On the other hand, there must be something we do know about $s(t)$ and $n(t)$, such as their signal value or signal shape characteristics. Indeed, there must be something we know that is different for $s(t)$ than for $n(t)$. Otherwise, we will have no way to separate one from the other. For example, much of the course will be devoted to developing systems that work when $s(t)$ and $n(t)$ have spectra that differ in known ways, e.g. one contains only low frequencies and the other contains only high frequencies.

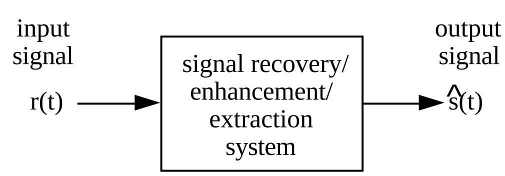
The third task to be discussed is signal digitization. Nowadays, when signals such as audio or pictures or video must be processed, stored or transmitted, it is generally done in digital fashion, i.e. the data is converted to binary. This is done because excellent digital techniques have been found, and because the bits so produced can be processed, transmitted and stored rapidly and reliably.

A. Signal Recovery/Extraction/Enhancement

Suppose we are given a signal $r(t)$ with two components,

$$r(t) = s(t) + n(t) ,$$

and our task is to design a system, such as illustrated below, which processes $r(t)$ in order to produce $s(t)$, or more precisely, an approximation $\hat{s}(t)$ to $s(t)$.

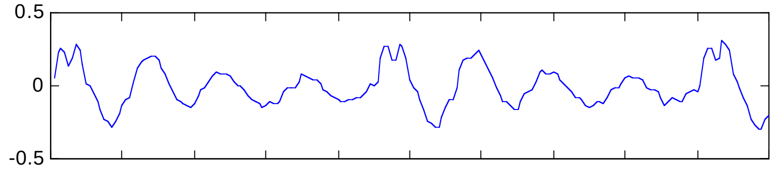


We consider $r(t)$ to be the *original* or *measured* or *received* signal, $s(t)$ to be the *desired signal*, and $n(t)$ to be *noise*. It is sometimes called *signal recovery*, because the system is recovering the signal $s(t)$ from the noise corrupted signal $r(t)$. It is also called *noise reduction* or *noise suppression*, because it attempts to do precisely this.

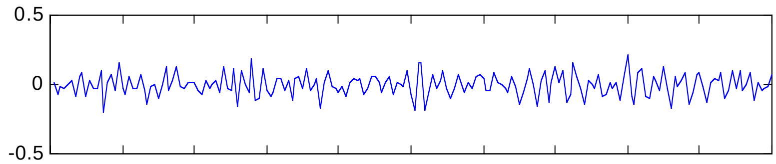
Examples of signals requiring recovery/extraction/enhancement include:

- An audio signal, especially when it is particularly faint, or when the microphone is part of a hearing aid, or when there is much background noise, such as in an automobile or helicopter or crowded cocktail party.
- A photograph or movie or video taken in faint light
- A signal being played back on an analog tape player (video or audio). Magnetic tapes introduce significant amounts of noise due to the granularity of the magnetic media.
- An AM or FM radio signal, or an analog TV signal, as it emerges from the receiving antenna. There is always lots of background noise, much of it due to other radio signals.
- A digital communication signal as it emerges from the receiving wire, antenna or other sensor. This signal must be extracted from background noise and from all other communication signals on the same medium.

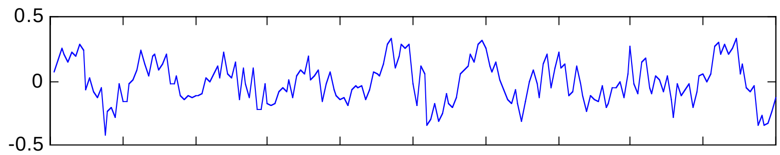
Linear Filters: There are many possible approaches to signal recovery. In this course, we focus mostly on *linear filtering*, which is the most common approach. Let us introduce it with an example. Suppose s(t) is an audio signal, for example the one shown below.



Suppose the measured signal is r(t) = s(t) + n(t), where n(t) looks like the signal below.



Then r(t) is

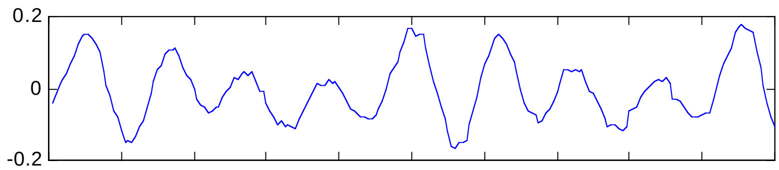


Since the noise signal fluctuates more rapidly than the audio signal[13], a natural approach to reducing the noise is to use a *running-average filter*. That is, we design a system that replaces r(t) by an average of r(t) over an interval up to time t. Specifically, it replaces r(t) with the average over the of r(t) over the time interval (t-T,t), where T is chosen small enough that the audio signal s(t) changes little in the interval and large enough that the noise signal fluctuates a great deal in the interval and, consequently, averages to a small value. In other words, the running-average filter produces the output signal

[13]This is the signal-shape charactertistic that differentiates the s(t) from n(t) in this example.

$$\hat{s}(t) = \frac{1}{T} \int_{t-T}^{t} r(t') \, dt' \ .$$

When such a filter is applied to r(t), it has the effect of smoothing the signal r(t). In our example, it produces the signal shown below, which sounds much more like s(t) than does r(t). Notice that the filtering has not only reduced the noise, but it has also modified the desired signal somewhat.



While the running average filter is fairly common, there are many other linear filters. As a precursor to introducing the full variety of possible linear filters, let us note that by applying the change of variables t" = t'-t to the above integral, we may rewrite the running average filter as producing

$$\hat{s}(t) = \frac{1}{T} \int_{-T}^{0} r(t+t'') \, dt'' \ ,$$

which in turn may be rewritten as

$$\hat{s}(t) = \int_{-\infty}^{\infty} r(t+t'') \, w(t'') \, dt'' \ .$$

where

$$w(t'') = \begin{cases} \frac{1}{T}, & -T \le t'' \le 0 \\ 0, & \text{else} \end{cases} \ .$$

Other linear filters are obtained by replacing the function w(t"), which we call a *weighting function*, by something else. That is, the output is produced by a running average, except that the average is with respect to a weighting function w(t"). We obtain different linear filters by making different choices of w(t"). For example, if we choose

$$w(t'') = \begin{cases} e^{3t''}, & t'' \le 0 \\ 0, & t'' > 0 \end{cases}$$

Then

$$\hat{s}(t) = \int_{-\infty}^{0} r(t+t'') \, e^{3t''} \, dt''$$

In this case, we see that $\hat{s}(t)$ is the average of all past values of r(t). However, in computing the average, past values are multiplied by exponentially decreasing weights.

By careful choice of the weighting function w(t"), one can develop filters that do a better job of extracting a signal from noise than the running average filter. Quite a different sort of weighting function is needed to perform the complex task of extracting a single radio signal from all those at other frequencies. As the course progresses, we will develop better and better techniques for designing filters for recovering signals or suppressing noise.

Actually, in this course, we will focus primarily on discrete-time linear filters for filtering discrete-time signals. (Chapters 5-8 of our text.) Specifically, a discrete-time filter performs the analogous operation

$$\hat{s}[n] \;=\; \sum_{k=-\infty}^{\infty} r[n+k]\, w[k] \;,$$

where the $w[k]$'s are a sequence of weights that distinguish one linear filter from another. For example, if $w[k] = 1/M$, $k = -M+1,...,0$, then we obtain a discrete-time running average filter, which produces

$$\hat{s}[n] \;=\; \frac{1}{M} \sum_{k=n-M+1}^{n} r[k] \;.$$

Performance Measure: As engineers, wherever possible we wish to quantify the goodness of the systems that we build. In this course, for the signal recovery task, we will use mean-squared error (MSE) as our measure of goodness. Specifically, if the signal $s(t)$ has support interval (t_1,t_2), then

$$\text{MSE} \;=\; \frac{1}{t_2-t_1} \int_{t_1}^{t_2} (s(t)-\hat{s}(t))^2 \, dt$$

Our goal, then, is to design a system that makes MSE as small as possible.

One be aware that MSE is sensitive to scale and to time shifts. For example, suppose the signal recovery system has completely eliminated the noise, but has scaled and delayed the somewhat, for example, suppose it prodcues $\hat{s}(t) = 1.2\, s(t-.1)$. Then, even though the system has done well, the measured MSE may be large. In such cases, we may wish to allow $\hat{s}(t)$ to be scaled and time-shifted before measuring MSE.

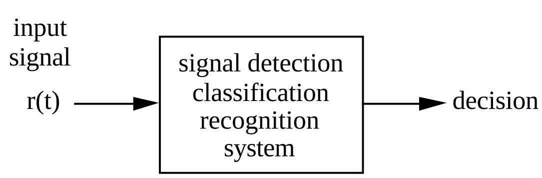
Other Signal Recovery Tasks: There are other situations where the desired signal and noise are not simply added. Rather $r(t)$ depends on the desired signal $s(t)$ in some more complicated way. For example, in AM radio transmission the audio signal we wish to recover is the envelope of the transmitted signal (minus a constant), and it is desired to recover this audio signal from the transmitted signal plus noise. In tomographic imaging (e.g. X-ray, MRI, PET, etc.), the desired signal is a two or three-dimensional image, which must be extracted from a complex set of measurements. The same is true of synthetic aperture radar. These are advanced topics that will not be pursued in this course or in these notes.

B. Signal Detection/Classification/Recognition

Suppose we are given a signal $r(t)$ with two components,

$$r(t) \;=\; s(t) + n(t) \;,$$

and our task is to design a system, such as illustrated below, which processes $r(t)$ and produces a decision about $s(t)$.



There are three closely related versions of this, introduced below along with examples.

1. **Signal/No Signal?** In this case, $s(t) = 0$ or $s(t) = v(t)$, where $v(t)$ is some known or partially known desired signal. From $r(t)$, the system must decide which of these two possibilities has occurred. This is considered to be a *detection* or *recognition* task because the goal is to *detect* or *recognize* whether or not $u(t)$ has occurred. Some specific examples are given below.

   • Radar: Decide if the signal $r(t)$ from the receive antenna contains a reflected pulse at time t_o. The same issues apply to sonar.

• Dollar bill changer: Decide if the signal r(t) obtained by optically scanning a bill is due to a genuine dollar bill.

• Fingerprint recognition: Decide if the signal r(t) obtained by optically scanning a fingerprint contains the fingerprint of John Smith. Similar tasks include recognition from retinal scans or voice prints.

• Heart monitoring. Decide if an ekg signal r(t) contains a characteristic indicating a heart defect.

2. **Which Signal?** Here, $s(t) = v_1(t)$ or $v_2(t)$ or ... or $v_M(t)$, where M is some finite integer and the $v_i(t)$ are known signals. From r(t) decide which of the $v_i(t)$'s is contained in r(t). This is considered to be a *classification* or *recognition* task because the goal is to *classify* r(t) according to which $v_i(t)$ has occurred, or equivalently to *recognize* which $v_i(t)$ has occurred. Some specific examples are give below.

   • Digital communication receiver: Decide if the received signal r(t) contains the signal representing "zero" or the signal representing "one". That is, the system must decide if the transmitter sent "zero" or "one". In some systems, the transmitter has more than two signals that it might send, and so the receiver must make a multivalued decision.

   • Optical character recognition: Decide if a character printed on paper is a or b or c or ... . This is especially challenging when the characters are handwritten.

   • Spoken word recognition: Decide what spoken word is present in the signal r(t) recorded by a microphone.

   • The "signal/no signal" task may be considered to be a special case of the "which signal task".

3. **Signal? And if So Which Signal?** This is a combination of the two previous subtasks. Suppose s(t) equals 0 or $v_1(t)$ or $v_2(t)$ or ... or $v_M(t)$. From r(t) decide whether or not $s(t) = 0$, and if not, decide which of the $v_i(t)$'s is contained in r(t). Examples:

   • Digital communication receiver: Some digital communication systems operate *asynchronously* in the sense that the receiver does not know when the bits will be transmitted. In this case, the receiver must decide if a bit is present, and if so, is it a zero or a one.

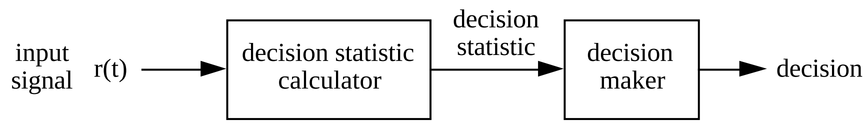
   • Personal identification system: Decide if a thumb has been placed on the electronic thumbpad, and if so, whose thumb.

   • Touch-tone telephone decoder: Decide if the signal from a telephone contains a key press, and if so, which key has been pressed.

   • Spoken word recognition: Decide is a word has been spoken and if so, what word.

For brevity, we will use the term *detection* as a broad term encompassing all of the above.

Detection Systems: As illustrated below, a detection system ordinarily has two subsystems: the first processes the received signal in order to produce a number (or several numbers) from which a decision can be made. The second makes the decision based on the number (or numbers) produced by the first. The number or numbers produced by the first system are called *decision statistics* or *feature values*, and the first subsystem is called a *decision statistic calculator* or a *feature calculator*. The second subsystem is called the *decision maker* or *decision device*. We will discuss two general types of detection systems, corresponding to two types of decision statistic generators -- *energy detectors* and *correlating detectors*.



Quality/Performance Measures: For detection systems, the most commonly used measure of performance is the *error frequency*, which as its name suggests, is simply the frequency with which its decisions are incorrect. We let the symbol f_e denote the error frequency. The typical goal is to design the detection system to minimize f_e.

In some situations, certain types of errors are more significant than others. For example, from the point of view of the owner of a dollar bill recognizer, classifying a counterfeit bill as valid is a more significant error than classifying a genuine dollar bill as invalid. In such cases, one will want to keep track of the frequency of the different types of errors. And one may choose to minimize the total frequency of errors subject to constraints on the frequencies of certain specific types of errors. For example, the owner of a dollar bill recognizer might insist that detector make as few errors as possible, subject to the constraint that it classifiy counterfeit bills as valid no more than one time in a million.

Energy Detectors for Deciding Signal/No Signal: For the "signal/no signal" task, the detector must decide whether r(t) contains signal AND noise, i.e. r(t) = v(t) + n(t), or just noise, i.e. r(t) = n(t). Since it is natural to expect that r(t) will have larger energy in the former case than in the latter, it is natural to choose the energy E(r) of r(t) as the decision statistic. (One would normally measure the energy of r(t) over the support interval of v(t).) The decision maker would then decide that v(t) is present if the energy is sufficiently large, and would decide that v(t) is not present otherwise. To make such a decision, one needs to specify a *threshold*, denoted τ, and the decision rule becomes

$$v(t) \text{ is present if } E(r) \geq \tau, \text{ and } v(t) \text{ is not present if } E(r) < \tau .$$

How to choose the threshold? The first thing to note that is that the noise signal n(t) is usually random. That is, it is not known in advance, and it is different every time we measure it. In particular, the energy of the noise will vary from decision to decision. However, based on past experience, it is usually possible to estimate the average value of the noise energy, which we denote $\overline{E}(n)$. Then we can say that when v(t) is not present, the signal r(t) = n(t) has a random energy value, with average $\overline{E}(n)$. On the other hand, when the signal v(t) is present, the energy of r(t), though still random tends to be larger. Specifically, it ordinarily has average energy equal[14] to $E(v) + \overline{E}(n)$. In summary, when the signal v(t) is present, the average energy of r(t) is $E(v) + \overline{E}(n)$, and when v(t) is not present, the average energy of r(t) is $\overline{E}(n)$. It is natural then to choose a threshold that lies half way between these two average energy values. That is, we choose

$$\tau = \frac{1}{2}\left(E(v) + \overline{E}(n)\right) + \frac{1}{2}\,\overline{E}(n) = \frac{1}{2}\,E(v) + \overline{E}(n) .$$

Energy detectors can also be used for the "which signal" task, provided the signals $v_1(t)$, $v_2(t)$, ... , $v_M(t)$ have sufficiently different energies -- so different that the differences will not be obscured by the noise. In this case, the typical decision maker strategy is to compare E(r) to the average energies $E(v_1)+\overline{E}(n)$, $E(v_2)+\overline{E}(n)$, ..., $E(v_M)+\overline{E}(n)$ that one expects if the various $v_i(t)$'s were present. The decision maker then decides in favor of the signal $v_i(t)$ such that $E(v_i)+\overline{E}(n)$ is closest to E(r).

Correlating Detectors for the "Which Signal Task": For the "which signal" task, an alternate and usually more effective method of detection (than energy detection) is to directly compare r(t) to each of the signals $v_1(t)$, $v_2(t)$, ..., $v_M(t)$.

[14] This is because v(t) and n(t) are usually uncorrelated.

Accordingly, we need a measure of similarity, and we will choose correlation. Specifically, the correlation between two continuous-time signals x(t) and y(t) is defined to be

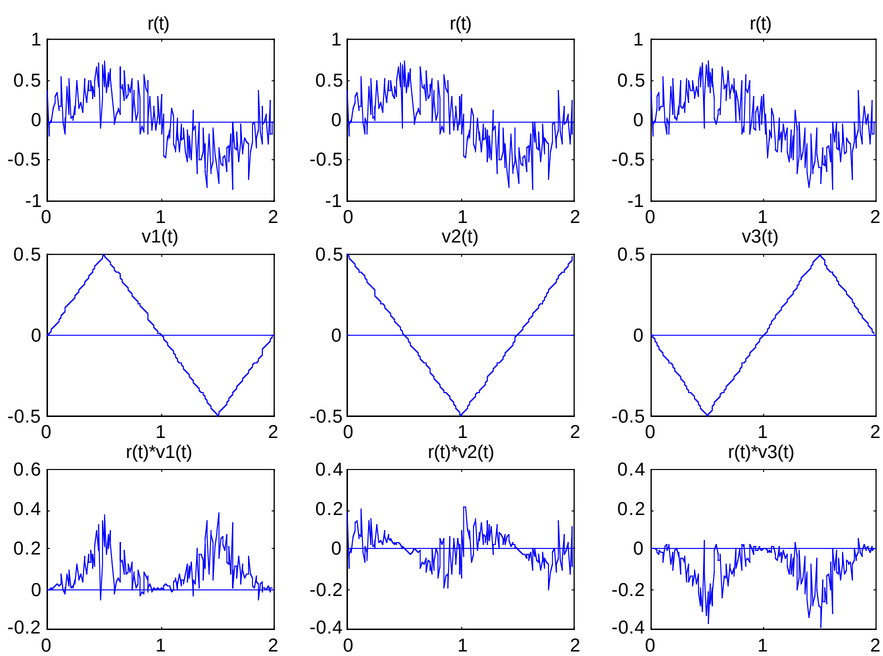
$$C(x,y) = \int_{t_1}^{t_2} x(t)\, y(t)\, dt \ ,$$

where (t_1,t_2) is the time interval of interest. Similarly, the correlation between two discrete-time signals x[n] and y[n] is defined to be

$$C(x,y) = \sum_{n_1}^{n_2} x[n]\, y[n] \ .$$

For brevity, we will continue the discussion presuming continuous-time signals. To see why correlation is a good measure of similarity to use in detection, consider the signal pairs shown below, in which a signal r(t) is compared to the three possibilities $v_1(t)$, $v_2(t)$ and $v_3(t)$. To aid the comparisons, r(t) is plotted above each signal. One can see that r(t) and $v_1(t)$ are similar in that, roughly speaking, where one is positive, the other is as well; where one is negative the other is as well. Moreover, r(t) roughly follows the shape of $v_1(t)$. On the other hand, the signals r(t) and $v_2(t)$ are rather dissimilar. Where $v_2(t)$ is positive, r(t) is sometimes negative; where $v_2(t)$ is increasing, r(t) is sometimes decreasing. Finally, r(t) and $v_3(t)$ are very dissimilar. Indeed, r(t) is very much like the negative of $v_3(t)$. If one were to make a decision about which of the three signals $v_1(t)$, $v_2(t)$, $v_3(t)$ was contained in r(t) based on visually comparing r(t) to the these signals, one would clearly choose $v_1(t)$. And indeed this is correct, because r(t) was generated by adding noise to $v_1(t)$.

Let us now consider how the same decision could be based on correlation. To do so, let's examine the value of correlation for each pair of signals. The product of each pair of signals is shown below the pair. Correlation is the integral of the product, i.e. the area under the plot of the product signal. For the first pair, the product is almost entirely positive, and the correlation is large. For the second pair, the product is approximately half negative and half positive, and the correlation is small because the positive and negative areas of the product tend to cancel each other. Finally, for the third pair, the product is mostly negative, and the correlation gives a large negative value.



$C(r,v_1) = 16.7$ $\quad$ $C(r,v_2) = 0.82$ $\quad$ $C(r,v_3) = -16.7$

If a detection system had to decide from the three correlation values which of the three signals $v_1(t)$, $v_2(t)$, $v_3(t)$ was contained in $r(t)$, clearly it should choose the one corresponding to the largest correlation, namely, $v_1(t)$.

Though correlation would work well in the example above, consider what would have happened if, for example, $v_2(t)$ were 100 times larger. In this case, it is easy to see that the correlation $C(r,v_2) = 82$, rather than 0.82. Thus even though v_2 has a very different shape than $r(t)$, a decision based solely on the size of the correlation would make the wrong decision. We can remedy this potential shortcoming by normalizing correlation. That is, it is better to make a decision based on normalized correlation, which is defined by

$$C_N(x,y) = \frac{C(x,y)}{\sqrt{E(x)}\sqrt{E(y)}} = \frac{1}{\sqrt{E(x)}\sqrt{E(y)}} \int_{t_1}^{t_2} x(t)\,y(t)\,dt$$

where $E(x)$ and $E(y)$ are the energies over the interval (t_1,t_2) of x and y, respectively. If the energies of the $v_i(t)$'s are the same, then signal $v_i(t)$ that has the largest correlation $C(r,v_i)$ also has the largest normalized correlation $C_N(r,v_i)$. However, when the $v_i(t)$'s have different energies, the normalized correlation accounts properly for such and permits the decision to be properly based.

Having discussed correlation, we can now completely describe a typical correlating detector. Suppose we must decide which of the signals $v_1(t), v_2(t), ..., v_M(t)$ is contained in $r(t)$. The decision statistic calculator computes and outputs $C_N(r,v_1)$, $C_N(r,v_2)$, ..., $C_N(r,v_M)$. The decision maker makes finds the largest of these, and outputs the corresponding decision.

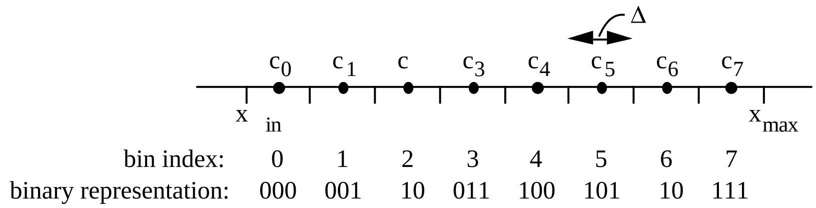
Comparison of Energy and Correlating Detectors: There are some situations where energy detectors cannot be used and some where correlating detectors cannot be used. For example, energy detectors cannot be used for the "which signal" problem when the signals have the same energy, which is often the case in digital communications. On the other hand, correlating detectors cannot be used when the precise shape of the signals is not known. For example, in Marconi's original transatlantic radio transmission, the transmitted signal was generated by a spark, with no known signal shape. Clearly, a correlating detector was out of the question!

In situations where both energy and correlating detectors can be used, it is usually found that the latter performs significantly better than the former, i.e. it makes fewer errors.

C. Signal Digitization for Data Storage and Transmission

In today's world where signal processing is increasingly done by general or special purpose computers, it is necessary to convert signals into digital form. Moreover signal storage and transmission are increasingly done in digital fashion. Again, this necessitates conversion to digital form. Such conversion involves two steps: (1) sampling, and (2) representing each sample as a binary number. Both of these steps generally involve losses, i.e. changes to the signal. Sampling is the topic of Chapter 4 and will be extensively discussed there. Converting to bits will be the subject of one of our lab assignments. However, let us describe here the most elementary method of converting samples to bits, called *uniform scalar quantization*.

With uniform scalar quantization, if we wish to represent a sample value $x[n]$ with b bits, then as illustrated below for the case that $b = 3$, we divide the range of sample values, (x_{min},x_{max}) into 2^b nonoverlapping bins of width $\Delta = (x_{max}-x_{min})/2^b$. These bins are indexed from left to right by the integers $0, 1, 2, ..., 2^b-1$, and each of these integers is represented as a b-bit binary number. For example, if $b = 3$, then $5 \Leftrightarrow 101$. Let $x_i = x_{min} + \Delta/2 + i\Delta$ denote the center of the ith bin. Now, if the sample $x[n]$ to be quantized lies in the ith bin, then we represent it by the binary representation of i, and we consider $x[n]$ to have been *quantized* to the value c_i. Note that when using this binary number in a processing task, we consider it to represent the value c_i, and must act accordingly. Actually, if the processing is done in a general purpose computer, we might convert i to binary using one of the standard conventions that are convenient for doing arithmetic, such as "two's complement".



A system that does both sampling and uniform scalar quantization is called an *analog-to-digital converter*.

There are more sophisticated methods for converting samples to bits that produce many fewer bits. These are generally called *data compression* methods. Examples include JPEG image compression, MP3 audio compression, and CELP speech compression, which is the system used in digital cellular telephones, digital answering machines, and the like. A simplified version of a JPEG like image compression system is included in one of the lab assignments. Generally speaking, data compression is done in order to reduce the amount of memory needed to store a signal or the amount of time needed to transmit a signal. When the signal actually needs to be processed or played, the compressed representation must ordinarily be changed back into a representation like the one produced by a uniform scalar quantizer. This is called *decompression*.

Concluding Remarks

Having discussed several basic signal processing tasks, it should be mentioned that from now on, we will not focus on them in future lectures. Instead we will focus on developing tools and techniques that enable systems to perform these tasks well. In particular, we will discuss sampling (Chapter 4 of our text), spectra (Chapter 3 and handouts) and linear filters (Chapters 5-8). Although these signal processing tasks will not be the focus of the lectures, from time to time we will discuss how the techniques being developed in lecture apply to them. On the other hand, these basic signal processing tasks will be the focus of a number of the lab assignments in this course.

Appendix A: Complex-Valued Signals

Complex-valued signals will be introduced in Chapter 2 as a way to simplify certain calculations involving sinusoidal signals. This appendix briefly summarizes the properties, statistics of complex signals and elementary operations on them. It should be read after complex signals are introduced in lecture.

Definition: A complex-valued signal is simply a signal whose values at each time are complex. As such, it has a real part and an imaginary part, a magnitude and a phase. For example, if

$$z(t) = x(t) + j\,y(t) = r(t)\,e^{j\phi(t)},$$

then $x(t)$ is the real part, $y(t)$ is the imaginary part, $r(t)$ is the amplitude and $\phi(t)$ is the angle or phase.

Signal Characteristics and Statistics:

The following table shows the definitions of the signal characteristics mentioned previously for real-valued signals, with the exception of signal value distribution, which is not easily summarized in table form.

	Continuous-time signal $z(t)$ $z(t) = x(t) + j\,y(t)$	**Discrete-time signal $z[n]$** $z[n] = x[n] + j\,y[n]$
support interval	$[t_1,t_2]$	$\{n_1,n_1+1,...,n_2\}$
duration	t_2-t_1	n_2-n_1+1
mean value:	$M(z) = \frac{1}{t_2-t_1}\int_{t_1}^{t_2} z(t)\,dt$ $= M(x) + j\,M(y)$	$M(z) = \frac{1}{n_2-n_1+1}\sum_{n=n_1}^{n_2} z[n]$ $= M(x) + j\,M(y)$
magnitude:	$\lvert z(t)\rvert = \sqrt{x^2(t)+y^2(t)}$	$\lvert z[n]\rvert = \sqrt{x^2[n]+y^2[n]}$
squared value, aka instantaneous power:	$\lvert z(t)\rvert^2 = x^2(t)+y^2(t)$	$\lvert z[n]\rvert^2 = x^2[n]+y^2[n]$
mean-squared value, aka average power:	$MS(z) = \frac{1}{t_2-t_1}\int_{t_1}^{t_2} \lvert z(t)\rvert^2\,dt$ $= MS(x) + MS(y)$	$MS(z) = \frac{1}{n_2-n_1+1}\sum_{n=n_1}^{n_2} \lvert z[n]\rvert^2$ $= MS(x) + MS(y)$
RMS value:	$RMS(z) = \sqrt{MS(z)}$	$RMS(z) = \sqrt{MS(z)}$
energy:	$E(z) = \int_{t_1}^{t_2} \lvert z(t)\rvert^2\,dt$ $= E(x) + E(y)$	$E(x) = \sum_{n=n_1}^{n_2} \lvert z[n]\rvert^2$ $= E(x) + E(y)$

Periodicity of complex continuous-time signals:

A complex continuous-time signal $z(t)$ is said to be *periodic with period* T if $z(t+T) = z(t)$ for all values of t. This is equivalent to saying that both $x(t)$ and $y(t)$ are periodic with period T.

1. A continuous-time signal $z(t)$ with period T is also periodic with period nT for any positive integer n.
2. The *fundamental period* T_o is the smallest period. The reciprocal of T_o is called the *fundamental frequency* f_o of the signal. That is, $f_o = 1/T_o$.
3. $z(t)$ is periodic with period T if and only if T is an integer multiple of T_o.

4. If signals $z(t)$ and $z'(t)$ are both periodic with period T, then the sum of these two signals, $w(t) = z(t) + z'(t)$ is also periodic with period T. This same property holds when three or more signals are summed.
5. The sum of two signals with fundamental period T_o is periodic with period T_o, but its fundamental period might be less than T_o.
6. The sum of two signals with differing fundamental periods, T_1 and T_2, will be periodic when and only when the ratio of their fundamental periods equals the ratio of two integers. The fundamental period of the sum is the least common multiple of T_1 and T_2. The fundamental frequency of the sum is the greatest common divisor of the fundamemental frequencies of the two sinusoids.

Periodicity of complex discrete-time signals:

A complex discrete-time signal $z[n]$ is said to be *periodic with period* N if $z[n+N] = z[n]$ for all integers n. This is equivalent to saying that both $x[n]$ and $y[n]$ are periodic with period N.

1. A discrete-time signal with period N is also periodic with period mN for any positive integer m.
2. The *fundamental period*, denoted N_o, is the smallest period. The reciprocal of N_o is called the *fundamental frequency* f_o of the signal. That is, $f_o = 1/N_o$.
3. $z[n]$ is periodic with period N if and only if N is an integer multiple of N_o.
4. If signals $z[n]$ and $z'[n]$ are both periodic with period N, then the sum of these two signals, $w[n] = z[n] + z'[n]$ is also periodic with period N. This same property holds when three or more signals are summed.
5. The sum of two signals with fundamental period N_o is periodic with period N_o, but its fundamental period might be less than N_o.
6. The sum of two signals with differing fundamental periods, N_1 and N_2, is periodic with fundamental period equal to the least common multiple of N_1 and N_2 and fundamental frequency equal to the greatest common divisor of their fundamental frequencies f_1 and f_2. Note that unlike continuous-time case, the ratio of the fundamental periods of discrete-time periodic signals is always the ratio of two integers. Therefore, the sum is always periodic.

Elementary Operations On One Complex Signal

These are illustrated for continuous-time signals, but apply equally to discrete-time signals.

Adding a constant: $z'(t) = z(t) + c$, where c is a real or complex number.

Amplitude scaling: $z'(t) = c\, z(t)$, where c is a real or complex number.

This has the effect of scaling both the average and the mean-squared values. Specifically, $M(z') = c\, M(z)$ and $MS(z') = |c|^2 MS(z)$.

Time shifting: If $z(t)$ is a signal and T is some number, then the signal

$$z'(t) = z(t-T) = x(t-T) + j\, y(t-T)$$

is a *time-shifted* version of $x(t)$.

Time reflection/reversal: The time reflected or time reversed version of a signal $z(t)$ is

$$z'(t) = z(-t).$$

Time scaling: The operation of *time-scaling* a signal $x(t)$ produces a signal

$$z'(t) = z(ct)$$

where c is some positive real-valued constant.

Combinations of the above operations: In the future we will frequently encounter signals obtained by combining several of the operations introduced above, for example,

$$z'(t) = 3\ z(-2(t-1)) \ .$$

Elementary Operations on Two or More Complex Signals

These are illustrated for continuous-time signals, but apply equally to discrete-time signals.

Summing: $w(t) = z(t) + \hat{z}(t)$.

Linear combining: $w(t) = c_1\, z_1(t) + c_2\, z_2(t) + c_3\, z_3(t)$, where c_1,c_2,c_3 are real or complex numbers.

Multiplying: $w(t) = z(t)\, z(t)$.

Concatenating: *Concatenation* is the process of appending one signal to the end of another.

Correlation

The correlation between continuous-time complex signals $z(t)$ and $\hat{z}(t)$ is

$$C(z,\hat{z}) = \int_{t_1}^{t_2} z(t)\, \hat{z}^*(t)\, dt \ ,$$

where (t_1,t_2) is the time interval of interest. Similarly, the correlation between discrete-time complex signals $z[n]$ and $\hat{z}[n]$ is defined to be

$$C(z,\hat{z}) = \sum_{n_1}^{n_2} z[n]\, \hat{z}^*[n] \ .$$

Why the complex conjugate? The reason is that this enables the relation $E(z) = C(z,z)$ to continue to be valid. Specifically,

$$C(z,z) = \int_{t_1}^{t_2} z(t)\, z^*(t)\, dt = E(z) \ .$$

Unfortunately, correlation for complex-valued signals is not symmetric, i.e. $C(z,\hat{z}) \neq C(\hat{z},z)$. However, there is a close relation between $C(z,\hat{z})$ and $C(\hat{z},z)$, namely,

$$C(\hat{z},z) = C^*(z,\hat{z}) \ .$$

This is because

$$C(\hat{z},z) = \int_{t_1}^{t_2} \hat{z}(t)\, z^*(t)\, dt = \left(\int_{t_1}^{t_2} z(t)\, \hat{z}^*(t)\, dt \right)^* = C^*(z,\hat{z}) \ .$$

The normalized correlation between signals z and $\hat{z}$ is

$$C_N(z,\hat{z}) = \frac{C(z,\hat{z})}{\sqrt{E(z)}\sqrt{E(\hat{z})}} \ .$$

The Cauchy-Schwarz Inequality continues to hold for complex signals. That is,

$$|C_N(z,\hat{z})| \leq 1,$$

with equality if and only if one signal is an amplitude scaling of the complex conjugate of the other; i.e. $y(t) = c\, x(t)$ for some real or complex constant c.

Appendix B: Trigonometric Identities and Facts About Complex Exponentials

Trigonometric Identities

We will not use these much, but nevertheless it is nice to have a table. The first five comprise Table 2.2 on p. 14 of *DSP First*.

1. $\sin^2\theta + \cos^2\theta = 1$
2. $\cos 2\theta = \cos^2\theta - \sin^2\theta$
3. $\sin 2\theta = 2\sin\theta\cos\theta$
4. $\sin(\alpha\pm\beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$
5. $\cos(\alpha\pm\beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$
6. $\sin\alpha\sin\beta = \frac{1}{2}\cos(\alpha-\beta) - \frac{1}{2}\cos(\alpha+\beta)$
7. $\cos\alpha\cos\beta = \frac{1}{2}\cos(\alpha-\beta) + \frac{1}{2}\cos(\alpha+\beta)$
8. $\sin\alpha\cos\beta = \frac{1}{2}\sin(\alpha+\beta) + \frac{1}{2}\sin(\alpha-\beta)$
9. $\cos\alpha\sin\beta = \frac{1}{2}\sin(\alpha+\beta) - \frac{1}{2}\sin(\alpha-\beta)$
10. $\sin\alpha + \sin\beta = 2\sin\frac{1}{2}(\alpha+\beta)\cos\frac{1}{2}(\alpha-\beta)$
11. $\sin\alpha - \sin\beta = 2\cos\frac{1}{2}(\alpha+\beta)\sin\frac{1}{2}(\alpha-\beta)$
12. $\cos\alpha + \cos\beta = 2\cos\frac{1}{2}(\alpha+\beta)\cos\frac{1}{2}(\alpha-\beta)$
13. $\cos\alpha - \cos\beta = -2\sin\frac{1}{2}(\alpha+\beta)\sin\frac{1}{2}(\alpha-\beta)$
14. $\sin^2\theta = \frac{1}{2}(1-\cos 2\theta)$
15. $\cos^2\theta = \frac{1}{2}(1+\cos 2\theta)$
16. $\sin\theta = \cos(\theta-\frac{\pi}{2})$
17. $\cos\theta = \sin(\theta+\frac{\pi}{2})$

Useful Facts About Complex Exponentials

1. $e^{j\theta} = \cos\theta + j\sin\theta$ (Euler's formula)

2. $\cos\theta = \frac{1}{2}\left(e^{j\theta}+e^{-j\theta}\right)$ (Inverse Euler formula)

3. $\sin\theta = \frac{1}{2j}\left(e^{j\theta}-e^{-j\theta}\right)$ (Another Inverse Euler formula)

4. $1 = e^{j2\pi} = e^{j2\pi n}$ for any integer n

5. $-1 = e^{j\pi} = e^{-j\pi}$

6. $(-1)^n = e^{j\pi n}$

7. $j = e^{j\pi/2}$

8. $-j = \frac{1}{j} = e^{-j\pi/2}$

Problems

Elementary Signal Characteristics

1. State the defining formula for:

   (a) The support interval of a continuous-time signal signal $x(t)$.

   (b) The duration of a continuous-time signal $x(t)$.

   (c) The mean value of the continuous-time signal $x(t)$ over the time interval $[t_1,t_2]$.

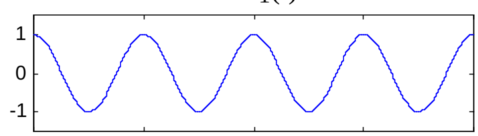
   (d) The average power of the continuous-time signal $x(t)$ over the time interval $[t_1,t_2]$.

   (e) The energy of the continuous-time signal $x(t)$ over the time interval $[t_1,t_2]$.

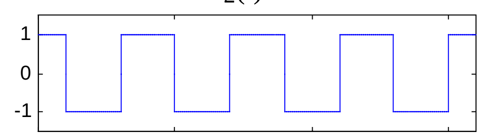
2. The continuous-time signal $x(t) = 3\cos(2\pi 100t)$ is sampled with sampling interval $T_s = 0.005$ msec, creating the discrete-time signal $x[n]$. Find a simple formula for $x[n]$ that does not include a cosine or any other trigonometric function.

3. Consider the two continuous-time signals shown below:

$x_1(t)$

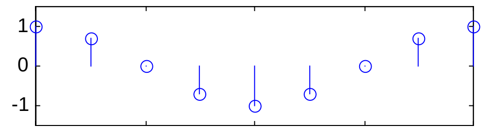


$x_2(t)$

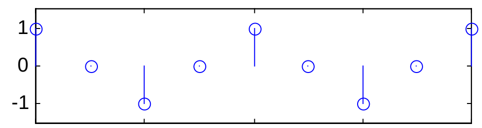


   From which of the above signals could each of the following discrete-time signals be obtained by sampling. Find the sampling interval in each case. If more than one sampling interval is possible, find the smallest among those that are possible.

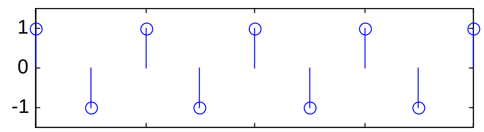
   (a) $y_1[n]$



   (b) $y_2[n]$

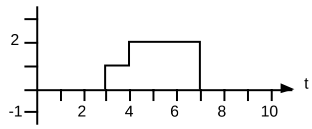


   (c) $y_3[n]$

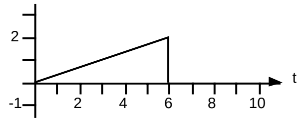


4. Find the support interval, the mean value, the mean-squared value, and the energy of each the following signals, with the last three values computed over the support interval of the signal.

   (a) $x_1(t)$



(b) $x_2(t)$



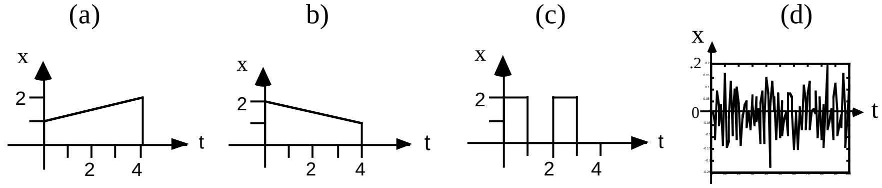
5. Derive the relationship between the mean-squared value, the variance and the mean value:

$$MS(x) = \sigma^2(x) + M^2(x)\ (x)$$

6. Match each signal below with its signal value distribution.

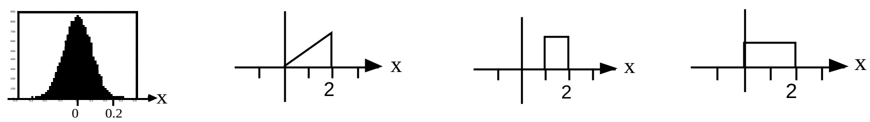
Signals:

(a) b) (c) (d)

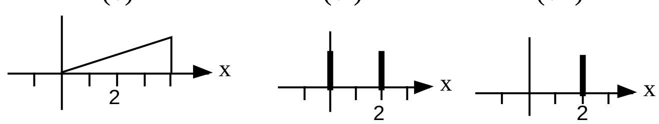


Signal value distributions:

(i) (ii) (iii) (iv)

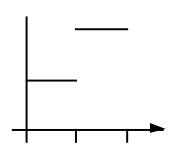


(v) (vi) (vii)

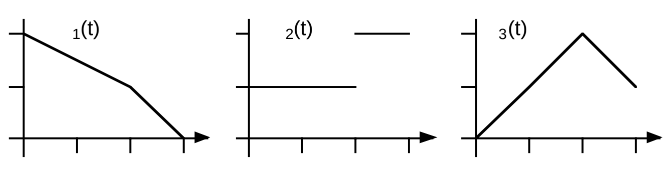


(Tip: You can work this problem from both ends. For each signal, you can look at the range of signal values, and see what you can deduce about which values occur more frequently than others. Also, look at each signal value distribution and see what you can deduce about the signal from which it came.)

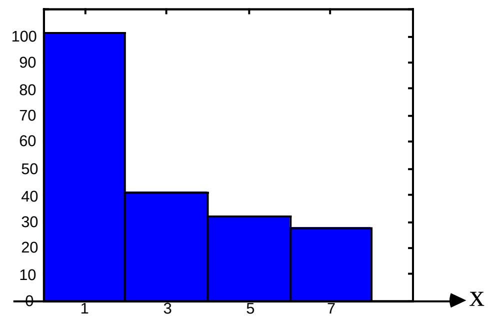
7. The function below



is the signal value distribution of which of the following signals



8. A discrete-time signal $x[n]$ has the histogram shown below.



(a) Find, approximately, the mean value of $x[n]$.

(b) Find, approximately, the mean-squared value of $x[n]$.

Periodicity

9. (a) State the condition defining the periodicity of a signal $x(t)$.

(b) State the definition of the fundamental period of a periodic signal $x(t)$.

10. Which of the signals shown below are periodic? For those that are periodic, find their fundamental period.

(a) $x_a(t) = 3 \sin(2t)$

(b) $x_b(t) = 4 \sin(e^t)$

(c) $x_c(t) = \cos(t^2+2t+1)$

(d) $x_4(t) = 4(-1)^{\text{floor}(t/3)}$, where floor(z) = largest integer $\leq z$

11. Let $s(t) = A \cos(\omega t+\phi)$.

(a) Show that $s(t)$ is periodic with fundamental period of $2\pi/\omega$.

(b) Find the mean value of $s(t)$ over one period.

(c) Show that the average power of this signal over one period is $A^2/2$.

12. Show that if $x(t)$ and $y(t)$ are periodic with period T, and a and b are arbitrary numbers, then $z(t) = a\,x(t) + b\,y(t)$ is also periodic with period T.

13. (a) Show that if $x(t)$ is periodic with period T and a is a positive number, then $y(t) = x(at)$ is periodic with period T/a.

(b) Repeat Part (a) with the word "period" replaced by "fundamental period".

14. Which of the signals below are periodic? For those that are periodic, find their fundamental period.

(a) $x_a(t) = \cos(2t) + \sin(3t)$

(b) $x_b(t) = \cos(2\pi t) + \sin(6\pi t)$

(c) $x_c(t) = \cos(2t) + \sin(6\pi t)$

15. Let $x(t) = 3 \cos(2t)$. Is $y(t) = 4\,x(2t-3)$ periodic? If so, find its fundamental period.

Envelope

16. Find a formula for the envelope of the signal $x(t) = \sin(100t) \sin(3t)$.

Elementary Operations on One Signal

17. Let $y(t) = x(t) + c$. Let (t_1,t_2) be the time interval of interest.

(a) Derive a formula for the mean $M(y)$ of $y(t)$ in terms of c and the mean $M(x)$ of x. (Hint: Start by writing the defining formula for what you need to find, namely, for $M(y)$.)

(b) Derive a formula for the mean-squared value $MS(y)$ of $y(t)$ in terms of c, the mean-squared value $MS(x)$ of x, and the mean value $M(x)$. (Hint: Start by writing the defining formula for what you need to find, namely, for $MS(y)$.)

18. Let $y(t) = c\,x(t)$. Let (t_1,t_2) be the time interval of interest.

(a) Derive a formula for the average $M(y)$ of $y(t)$ in terms of c and the mean $M(x)$ of x.

(b) Derive a formula for the mean-squared value $MS(y)$ of $y(t)$ in terms of c, the mean-squared value $MS(x)$ of x, and the mean value $M(x)$.

19. Let $y(t) = a\,x(t) + b$. Let (t_1,t_2) be the time interval of interest.

(a) Derive a formula for the mean $M(y)$ of $y(t)$ in terms of a, b, and the mean $M(x)$ of x.

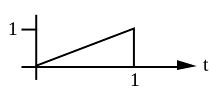
(b) Derive a formula for the mean-squared value $MS(y)$ of $y(t)$ in terms of a, b, the mean-squared value $MS(x)$ of x, and the mean value $M(x)$.

20. Let $y(t) = x(at)$, where $x(t)$ is a signal with support interval (t_1,t_2).

(a) Find the support interval of $y(t)$.

(a) Derive a formula for the mean $M(y)$ of $y(t)$, over its support interval, in terms of a and the mean $M(x)$ of x.

(b) Derive a formula for the mean-squared value $MS(y)$ of $y(t)$, over its support interval, in terms of a and the mean-squared value $MS(x)$ of x.

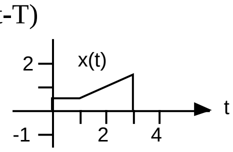
21. Let $x(t)$ = 

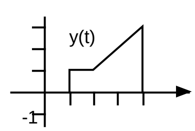
. Plot the following signals:

(a) $y_1(t) = -2\,x(3t-2)$

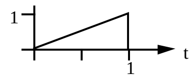
(b) $y_2(t) = 3\,x(-2t+6)$

22. Let $x(t)$ and $y(t)$ be as shown below. Find numbers a and T such that $y(t) = a\,x(t-T)$

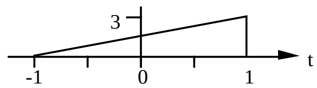


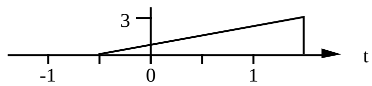


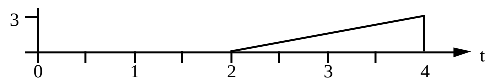
(No systematic procedure has been developed to solve this problem. Use your creativity.)

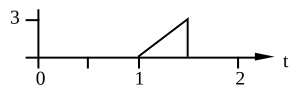
23. Let $x(t)$ = 

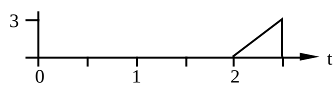
Which of the following shows $y(t) = 3x\left(\frac{t}{2} - 1\right)$?

(a) 

(b) 

(c) 

(d) 

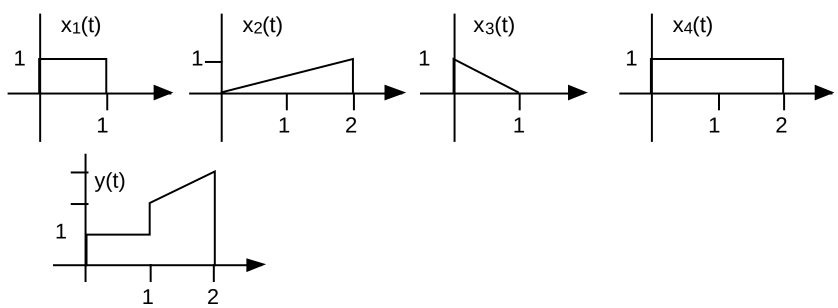
(e) 

Elementary Operations on Two or More Signals

24. For the signals shown below, find numbers a_1, a_2, a_3, a_4 such that

$$y(t) = a_1\, x_1(t) + a_2\, x_2(t) + a_3\, x_3(t) + a_4\, x_4(t)\ .$$

(The a_i's can be positive, negative or zero.)



(Hint: You might think about choosing the a_i's to match some particular feature of $y(t)$ first.)

25. For each of the following, either prove the statement or provide a simple example demonstrating that the statement is false.

(a) The duration of $z(t) = x(t) + y(t)$ equals the sum of the durations of $x(t)$ and $y(t)$.

(b) The energy of $z(t) = x(t) + y(t)$ equals the sum of the energies of $x(t)$ and $y(t)$.

(c) The mean of $z(t) = x(t) + y(t)$ equals the sum of the means of $x(t)$ and $y(t)$.

26. For each of the following collections of signals, determine whether or not the sum is periodic. Find the fundamental period of the periodic ones.

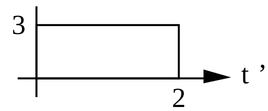
(a) $x(t) = 3\sin(20\pi t)$, $y(t) = 4\cos(40\pi t)$

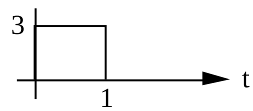
(b) $x(t) = 3\sin(20\pi t)$, $y(t) = 4\cos(21\pi t)$

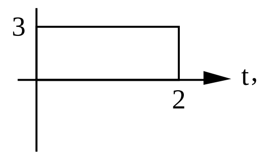
(c) $x(t) = 3\sin(2t)$, $y(t) = 4\cos(\sqrt{2}t)$

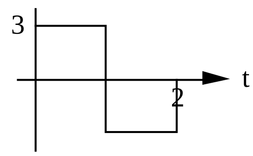
(d) $x(t) = 3\sin(20\pi t)$, $y(t) = 4\cos(40\pi t)$, $z(t) = 3\sin(50\pi t)$

27. For each of the following pairs of signals, find $E(x)$, $E(y)$ and $E(x+y)$ and compare $E(x) + E(y)$ to $E(x+y)$.

(a) x(t) = 

, y(t) = 

(b) x(t) = 

, y(t) = 

28. Let $z(t) = x(t) + y(t)$. Show that $E(z) = E(x) + 2\,C(x,y) + E(y)$.

Mean-Squared Error

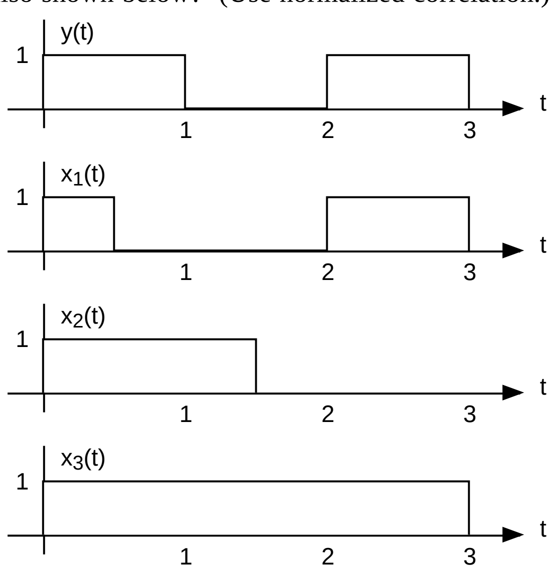
29. Assuming $x(t)$ and $y(t)$ are known signals, find an expression for the value of α that $MSE(x,\alpha y)$ is minimized.

Signal Correlation

30. (a) State the defining formula for the correlation between two continuous-time signals $s(t)$ and $r(t)$, where the time interval of interest is $(3,4)$.

(b) State the defining formula for the correlation between two continuous-time signals $s(t)$ and $r(t)$, where the time interval of interest is $(3,4)$.

31. Which of the signals $x_1(t)$, $x_2(t)$, $x_3(t)$ shown below is most correlated with $y(t)$, also shown below? (Use normalized correlation.)



32. Show that normalized correlation $C_N(x,y)$ between signals $x(t)$ and $y(t)$ equals the unnormalized correlation $C(x',y')$ between the normalized versions of these signals $x'(t) = x(t)/\sqrt{E(x)}$ and $y'(t) = x(t)/\sqrt{E(y)}$, which have energy 1.

33. Show that normalized correlation is not affected by a scaling of either signal, i.e. $C_N(ax,y) = C_N(x,y)$.

34. Show that correlation is linear, i.e. $C(a\,x_1 + b\,x_2, y) = a\,C(x_1,y) + b\,C(x_2,y)$.

35. Show that if $x(t)$ and $y(t)$ are uncorrelated, then energy of their sum equals the sum of their energies, i.e. $E(x+y) = E(x+y)$.

36. Show that a constant signal $x(t) = a$ is uncorrelated with any signal $y(t)$ that has zero mean, i.e. if $M(y) = 0$, then $C(x,y) = 0$.

37. Show that normalized correlation between $x(t) = \cos(\omega t)$ and $y(t) = \cos(\omega t + \phi)$ is $C_N(x,y) = \cos(\phi)$.

38. Suppose $x(t) = 2t$, $0 \le t \le 1$ and $x(t) = 0$, for other t, and suppose $y(t)$ is a signal with energy 3 such that $C_N(x,y) = -1$. Find $y(t)$.

39. Continuous-time signals $x(t)$ and $y(t)$ each have support interval $[0,3]$ and average value 2. Their (unnormalized) correlation is $C(x,y) = 1$. If $z = 3y+2$, find $C(x,z)$.

40. Continuous-time signals $x(t)$ and $y(t)$ have $E(x) = 1$, $E(y) = 2$, and $E(x+y) = 5$, where $E(\,)$ denotes energy. Find the (unnormalized) correlation $C(x,y)$.

41. Continuous-time signals $x(t)$ and $y(t)$ have energies $E(x) = 2$, $E(y) = 2$, and are uncorrelated. Find $E(x-y)$.

42. Let $z(t) = x(t) + y(t)$. Let (t_1,t_2) be a time interval of interest. Show that

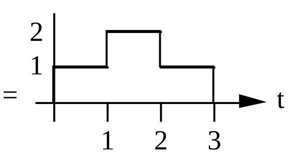
$$E(z) = E(x) + 2\,C(x,y) + E(y),$$

where as usual E denotes energy and C denotes correlation.
(One may conclude from this that $E(z) = E(x) + E(y)$ when and only when x and y are uncorrelated.)

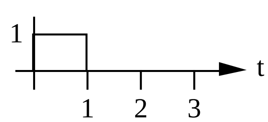
43. Given $E(x)$, $E(y)$ and $E(x+y)$ find $C_N(x,y)$.

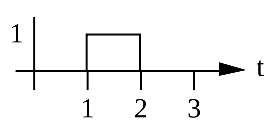
44. Given $E(x)$, $C(x,y)$, what can be said about $E(y)$?

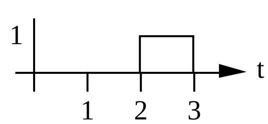
Signal components

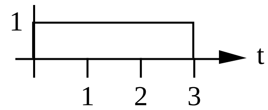
45. Find the component of the signal $x(t) =$ 

that is like:

(a) $p_1(t) =$ 

(a) $p_2(t) =$ 

(a) $p_3(t) =$ 

(a) $p_4(t) =$ 

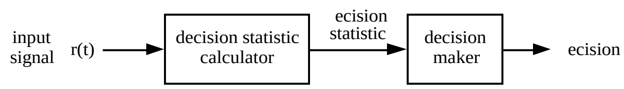
Basic Signal Processing Tasks

46. Consider the running average filter such that when the input signal is $x(t)$, the output signal is

$$y(t) = \frac{1}{2} \int_{t-2}^{t} x(s)\, ds$$

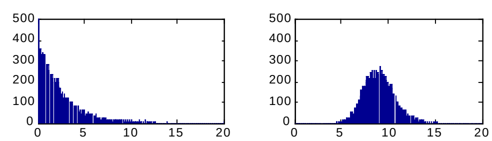
Find an expression for $y(t)$ when $x(t) = \sin(3t)$. Use trigonometric identities to simplify as much as possible.

47. Consider the signal/no signal detection system shown below:



Two histograms for the decision statistic are shown below. Each histogram is based on 10,000 decision statistic measurements.

"no signal present" histogram "signal present" histogram



Assuming we want to minimize the number of decision errors, when the decision statistic is 5, what decision should the decision maker make.

48. A standard audio CD is produced by uniform scalar quantizing the samples of an audio signal and producing 16 bits per sample. The sampling rate is 44,100 samples sec. The quantizer produces 16 bits per sample.

(a) How many quantization bins does the quantizer have?

(b) How many bits per second are produced by the quantizer? (Consider just the bits representing the left channel.