15.6: Cauchy-Schwarz Inequality

Introduction

Any treatment of linear algebra as relates to signal processing would not be complete without a discussion of the Cauchy-Schwarz inequality, a relation that enables a wide array of signal processing applications related to pattern matching through a method called the matched filter. Recall that in standard Euclidean space, the angle $\theta$ between two vectors $x, y$ is given by

$$\cos(\theta) = \frac{x \cdot y}{||x|| ||y||}.$$ 

Since $\cos(\theta) \leq 1$, it follows that

$$||x \cdot y||^2 \leq ||x||^2 ||y||^2.$$ 

Furthermore, equality holds if and only if $\cos(\theta) = 0$, implying that

$$||x \cdot y||^2 = ||x||^2 ||y||^2$$

if and only if $y = ax$ for some real $a$. This relation can be extended to all inner product spaces over a real or complex field and is known as the Cauchy-Schwarz inequality, which is of great importance to the study of signals.

The Cauchy-Schwarz Inequality
Statement of the Cauchy-Schwarz Inequality

The general statement of the Cauchy-Schwarz inequality mirrors the intuition for standard Euclidean space. Let $V$ be an inner product space over the field of complex numbers $\mathbb{C}$ with inner product $\langle \cdot, \cdot \rangle$. For every pair of vectors $x, y \in V$, the inequality

$$||x, y||^2 \leq ||x||^2 ||y||^2$$

holds. Furthermore, the equality

$$||x, y||^2 = ||x||^2 ||y||^2$$

holds if and only if $y = ax$ for some $a \in \mathbb{C}$. That is, equality holds if and only if $xx$ and $yy$ are linearly dependent.

Proof of the Cauchy-Schwarz Inequality

Let $V$ be a vector space over the real or complex field $\mathbb{F}$, and let $x, y \in V$. In order to prove the Cauchy-Schwarz inequality, it will first be proven that $||x, y||^2 = \langle x, y \rangle$ for all $x, y \in V$. It will then be shown that $||x, y||^2 \leq ||x||^2 ||y||^2$ if $y \neq ax$ for all $a \in \mathbb{F}$.

Consider the case in which $y = ax$ for some $a \in \mathbb{F}$. From the properties of inner products, it is clear that

$$||x, y|| = ||x, ax|| = ||x|| ||a||^2.$$

Hence, it follows that

$$||x, y||^2 = ||x||^2 ||y||^2 = ||x, ax||^2 = ||a||^2 ||x, x||^2.$$

Similarly, it is clear that

$$||x, y||^2 = ||x, ax||^2 = ||x, x||^2.$$

Thus, it is proven that $||x, y||^2 = ||x||^2 ||y||^2$ if $x = ay = ax$ for some $a \in \mathbb{F}$.

Next, consider the case in which $y \neq ax$ for all $a \in \mathbb{F}$, which implies that $y \neq 0$ so $y, y \neq 0$. Thus, it follows by the properties of inner products that, for all $a \in \mathbb{F}$, $||x-ay, x-ay|| > 0$. This can be expanded using the properties of inner products to the expression

$$||x-ay, x-ay||^2 = ||x, x||^2 - 2||x, a||^2 ||y||^2 - ||a||^2 ||y||^2.$$

Choosing $a = x, y \neq a = x, y \neq y$, we get

$$||x-ay, x-ay||^2 = ||x, x||^2 - 2||x, y||^2 - ||y||^2.$$

Thus, it follows that

$$||x-ay, x-ay||^2 = ||x||^2 ||y||^2$$

for all $a \in \mathbb{F}$.

Updated: Fri, 03 Jul 2020 15:26:28 GMT
Hence, it follows that $x, x ? ? x, y ? ? y, x, y ? > 0$. Consequently, $x, x ? ? y, y ? ? y, y > 0$. Thus, it can be concluded that $||x, y||^2 < |x, y|^2$ if $y \neq ax$ for all $a \in \mathbb{F}$.

Therefore, the inequality

$$ ||x, y||^2 \leq |x, y|^2 $$

holds for all $x, y \in \mathbb{V}$, and equality

$$ ||x, y||^2 = |x, y|^2 $$

holds if and only if $y = ax$ for some $a \in \mathbb{F}$.

### Additional Mathematical Implications

Consider the maximization of $||x||x||, y||y||$ where the norm $|||\cdot|||\cdot|$ is induced by the inner product. By the Cauchy-Schwarz inequality, we know that $|||x||x||, y||y||^2 \leq 1$ and that

$$ |||x||x||, y||y|| = 1 $$

if and only if $y = ax$, for some $a \in \mathbb{C}$. Hence, $|||x||x||, y||y||^2 \leq 1$ attains a maximum where $y = ax$. This result will be particularly useful in developing the matched filter detector techniques.

### Matched Filter Detector

#### Background Concepts

A great many applications in signal processing, image processing, and beyond involve determining the presence and location of a target signal within some other signal. A radar system, for example, searches for copies of a transmitted radar pulse in order to determine the presence of and distance to reflective objects such as building or aircraft. A communication system searches for copies of waveforms representing digital 0s and 1s in order to receive a message.

As has already been shown, the expression $||x||x||, y||y||^2$ attains its upper bound, which is 1, when $y = ax$ for some scalar $a$ in a real or complex field. The lower bound, which is 0, is attained when $x$ and $y$ are orthogonal. In informal intuition, this means that the expression is maximized when the vectors $x$ and $y$ have the same shape or pattern and minimized when $x$ and $y$ are very different. A pair of vectors with similar but unequal shapes or patterns will produce relatively large values of the expression less than 1, and a pair of vectors with very different but not orthogonal shapes or patterns will produce relatively small values of the expression greater than 0. Thus, the above expression carries with it a notion of the degree to which two signals are "alike", the magnitude of the normalized correlation between the signals in the case of the standard inner products.

This concept can be extremely useful. For instance consider a situation in which we wish to determine which signal, if any, from a set $XX$ of signals most resembles a particular signal $yy$. In order to accomplish this, we might evaluate the
above expression for every signal \(x \in \mathbb{R} \), choosing the one that results in maxima provided that those maxima are above some threshold of “likeness”. This is the idea behind the matched filter detector, which compares a set of signals against a target signal using the above expression in order to determine which among them are most like the target signal. For a detailed treatment of the applications of the matched filter detector see the liked module.

**Signal Comparison**

The simplest variant of the matched filter detector scheme would be to find the member signal in a set \(X\) of signals that most closely matches a target signal \(y\). Thus, for every \(x \in \mathbb{R}\) we wish to evaluate

\[
m(x, y) = ||x|| ||y|| = \langle x, y \rangle
\]

in order to compare every member of \(X\) to the target signal \(y\). Since the member of \(X\) which most closely matches the target signal \(y\) is desired, ultimately we wish to evaluate

\[
x_{m} = \arg\max_{x \in \mathbb{R}} ||x|| ||y||
\]

Note that the target signal does not technically need to be normalized to produce a maximum, but gives the desirable property that \(m(x, y)\) is bounded to \([0, 1]\).

The element \(x_{m} \in \mathbb{R}\) that produces the maximum value of \(m(x, y)\) is not necessarily unique, so there may be more than one matching signal in \(X\). Additionally, the signal \(x_{m} \in \mathbb{R}\) producing the maximum value of \(m(x, y)\) may not produce a very large value of \(m(x, y)\) and thus not be very much like the target signal \(y\). Hence, another matched filter scheme might identify the argument producing the maximum but only above a certain threshold, returning no matching signals in \(X\) if the maximum is below the threshold. There also may be a signal \(x \in \mathbb{R}\) that produces a large value of \(m(x, y)\) and thus has a high degree of “likeness” to \(y\) but does not produce the maximum value of \(m(x, y)\). Thus, yet another matched filter scheme might identify all signals in \(X\) producing local maxima that are above a certain threshold.

For example, consider the target signal given in [Figure](#) and the set of two signals given in [Figure](#). By inspection, it is clear that the signal \(g_2\) is most like the target signal \(f\). However, to make that conclusion mathematically, we use the matched filter detector with the \(L_2\) inner product. If we were to actually make the necessary computations, we would first normalize each signal and then compute the necessary inner products in order to compare the signals in \(X\) with the target signal \(f\). We would notice that the absolute value of the inner product for \(g_2\) when normalized is greater than the absolute value of the inner product of \(g_1\) with \(f\) when normalized, mathematically stated as

\[
g_2 = \arg\max_{x \in \mathbb{R}} ||x|| ||f|| = \langle g_2, f \rangle
\]

Template Signal

![Figure](https://eng.libretexts.org/Bookshelves/Electrical_Engineering/Signal_Processing_and_Modeling/Book%3A_Signals_and_Syst...)
We wish to find a match for this target signal in the set of signals below.

**Candidate Signals**

We wish to find a match for the above target signal in this set of signals.

**Pattern Detection**

A somewhat more involved matched filter detector scheme would involve attempting to match a target time limited signal \( y = f \) to a set of time shifted and windowed versions of a single signal \( X = \{ wStg \mid t \in \mathbb{R} \} \) indexed by \( \mathbb{R} \).

The windowing function is given by

\[
w(t) = u(t-t_1) - u(t-t_2)\]

where \([t_1, t_2] \) is the interval to which \( f \) is time limited. This scheme could be used to find portions of \( g \) that have the same shape as \( f \). If the absolute value of the inner product of the normalized versions of \( f \) and \( wStg \) is large, which is the absolute value of the normalized correlation for standard inner products, then \( g \) has a high degree of “likeness” to \( f \) on the interval to which \( f \) is time limited but left shifted by \( t \). Of course, if \( f \) is not time limited, it means that the entire signal has a high degree of “likeness” to \( f \) left shifted by \( t \).

Thus, in order to determine the most likely locations of a signal with the same shape as the target signal \( f \) in a signal \( g \) we wish to compute

\[
 tm = \arg \max_{t \in \mathbb{R}} \left| \frac{1}{||f||} \| wStg \| \int_{-\infty}^{\infty} f(\tau) g(\tau-t) \, d\tau \right|,
\]

where \( f \) and \( w \) are time limited to the same interval.

Making the substitution \( h(t) = g(-t) \) we have

\[
 tm = \arg \max_{t \in \mathbb{R}} \left| \frac{1}{||f||} \| wStg \| \int_{-\infty}^{\infty} f(\tau) h(\tau-t) \, d\tau \right|,
\]

\[
 tm = \arg \max_{t \in \mathbb{R}} \left| \frac{1}{||f||} \| wStg \| \int_{-\infty}^{\infty} f(\tau) h(\tau-t) \, d\tau \right|.
\]
Noting that this expression contains a convolution operation

\[ t_0 = \arg \max_{t \in \mathbb{R}} | | f(t) \|^2 | | w(t) \|^2 \],

where \( hh \) is the conjugate of the time reversed version of \( gg \) defined by \( hh(t) = g(-t) \).

In the special case in which the target signal \( f \) is not time limited, \( hh \) has unit value on the entire real line. Thus, the norm can be evaluated as \( | | w(t) \|^2 = | | w(t) \|^2 = | | g(t) \|^2 = | | h(t) \|. \) Therefore, the function reduces to

\[ t_0 = \arg \max_{t \in \mathbb{R}} | | f(t) \|^2 | | h(t) \|^2 \] where \( h(t) = g(-t). \) The function

\[ f \ast g = (f \ast h)(t) \] is known as the normalized cross-correlation of \( f \) and \( g \).

Hence, this matched filter scheme can be implemented as a convolution. Therefore, it may be expedient to implement it in the frequency domain. Similar results hold for the \( L_2(\mathbb{R}) \), \( L_2(\mathbb{Z}) \), and \( L_2(\mathbb{Z}[a,b]) \) spaces. It is especially useful to implement the \( L_2(\mathbb{Z}[a,b]) \) cases in the frequency domain as the power of the Fast Fourier Transform algorithm can be leveraged to quickly perform the computations in a computer program. In the \( L_2(\mathbb{R}) \) and \( L_2(\mathbb{Z}[a,b]) \) cases, care must be taken to zero pad the signal if wrap-around effects are not desired. Similar results also hold for spaces on higher dimensional intervals with the same inner products.

Of course, there is not necessarily exactly one instance of a target signal in a given signal. There could be one instance, more than one instance, or no instance of a target signal. Therefore, it is often more practical to identify all shifts corresponding to local maxima that are above a certain threshold.

The signal in Figure contains an instance of the template signal seen in Figure beginning at time \( t = s_1 \) as shown by the plot in Figure. Therefore,

\[ s_1 = \arg \max_{t \in \mathbb{R}} | | f(t) \|^2 | | w(t) \|^2 \].

Pattern Signal

This function shows the pattern we are looking for in the signal below, which occurs at time \( t = s_1 \).

Longer Signal

This signal contains an instance of the above signal starting at time \( t = s_1 \).

Absolute Value of Output
This signal shows a sketch of the absolute value of the matched filter output for the interval shown. Note that this was just an "eyeball approximation" sketch. Observe the pronounced peak at time $t=s_1t=s_1$.

**Cauchy-Schwarz Inequality Video Lecture**

**Proof of the Cauchy-Schwarz Inequality**

Video lecture on the proof of the Cauchy-Schwarz inequality from Khan Academy. Only part of the theorem is proven.

**Cauchy-Schwarz Inequality Summary**

As can be seen, the Cauchy-Schwarz inequality is a property of inner product spaces over real or complex fields that is of particular importance to the study of signals. Specifically, the implication that the absolute value of an inner product is maximized over normal vectors when the two arguments are linearly dependent is key to the justification of the matched filter detector. Thus, it enables the use of matched filters for such pattern matching applications as image detection, communications demodulation, and radar signal analysis.