Correlation

Correlation can be used to the measure of how similar two signals, x and y, are to each other. Mathematically, it can be defined as follows:

When dealing with signals that are zero valued up to sample n=0, and zero values for all samples greater than N-1, we can rewrite the equation as:

We will run through a basic example to illustrate the process of correlating signals.  
Take the signals x, y and z below, with their respective plots:

|  |  |
| --- | --- |
| x = [ 2 1 -1 3 ]  y = [ 3 2 0 3 ]  z = [ -1 3 2 1 ] |  |

We then use the equation above to calculate the correlations:  
Corr­­­­­x,y = x[0]y[0] + x[1]y[1] + x[2]y[2] + x[3]y[3] = (2)(3) + (1)(2) + (-1)(0) + (3)(3) = 6 + 2 + 0 + 9 = 17  
Similarly, we can find Corr­­­­­x,z = 2 and Corr­­­­­y,z = 6

The results can be interpreted as the larger the result, the more similar signals are. As can be seen from the graph, the signal x is more like y than it is z. Correlation is a method to tell how similar signals are automatically, without needing to plot.   
However, this basic form of correlation runs into issues when the signals we are interpreting have different energy levels.

Let’s see what happens if we change one of the values in z to be much larger:

|  |  |
| --- | --- |
| x = [ 2, 1, -1, 3 ]  y = [ 3, 2, 0, 3 ]  z = [ -1, 30, 2, 1 ] |  |

We can see that the signal z is now much less similar than the others, but the correlation yields the following results:  
Corr­­­­­x,y = 17, Corr­­­­­x,z = 29 and Corr­­­­­y,z = 60  
This is not what is expected of a correlation result. A larger value should imply more similarity. For more accurate results we need a way to scale the correlation function by a factor related to the energy levels we are looking at. Thus, we can use the normalised correlation function:

The numerator of the expression above is the correlation function as above, however the lower determines the energy in the signals and scales the result. Taking the second example above we can determine the following:

These results are in keeping with what you would expect from correlation, where a higher value implies a more similar signal. Normalised correlation will always return values from -1 to 1.   
You would get a value of 1 if you compared a signal with itself, and a value of -1 if you compared a signal with and inverted version of itself, i.e. x[n]\*-1

There are practical applications to using both approaches, and it depends on what is trying to be achieved.

Cross Correlation   
Now that we have an understanding of what correlation is, we can look at practical uses within signal processing. Cross correlation is the measurement of the similarities of two signals at different lag positions. It is defined for discrete functions f and g as:

where denotes the complex conjugate of . It is much easier to consider it in practical terms. Let us look at the cross correlation of the two signals below, x and y.

|  |  |
| --- | --- |
| x = [ 2, 1, -1, 3 ]  y = [ 2, 0, 3, 1 ] |  |

We show signals x and y at lag position 0. That is to say sample 0 of the signal x, x[0], is vertically aligned with sample 0 of the signal y, y[0]. To calculate the correlation when there is no time lag, we use the same formula from earlier. We can say the correlation measure at a lag of 0 is, (2)(2) + (1)(0) + (-1)(3) + (3)(1) = 4.

We can now shift the sequence y one sample to the right, we see the signals at a lag position 1, y[0] is now aligned to x[1], and so on. We now take the correlation of the vertically aligned samples to find the correlation measure at lag position 1: (1)(2) + (-1)(0) + (3)(3) = 11

|  |
| --- |
|  |

We continue to shift to the right until we have calculated all of the overlapping lag positions, as we complete the correlation sequence:

|  |  |
| --- | --- |
|  |  |

Just as we shift to the right to find the positive lag, we should also shift to the left to find the similarity measure at the negative lag values, to build up the full correlation sequence:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Now we have the correlation measurements for these signals at all lag positions, giving us the following results:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lag Position | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| Correlation measure | 2 | 7 | 2 | 4 | 11 | -2 | 6 |

Remember that correlation is a measurement of similarity, so what these results tell us is that signals x and y were most similar at lag position 1.

We use this result later to determine a time delay between two signals which can be used to measure the distance between points.

What is Autocorrelation?   
Autocorrelation is simply the cross correlation of a signal with itself. When dealing with one dimensional, real sequences, autocorrelation will have a peak at a lag of zero, and its size will be the signal energy.

Taking the example of the signal x above, cross correlating the signal with itself yields the following correlation sequence:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lag Position | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| Correlation measure | 6 | 1 | -2 | 15 | -2 | 1 | 6 |

Autocorrelation has some interesting properties for use in digital signal processing. The autocorrelation of a periodic function is itself periodic, with the same period. This can be used to determine frequencies or pitches of musical tones. In a noisy waveform, autocorrelation can be used to reduce the effects of that noise.