8. Cross-Correlation

Cross-correlation

The cross-correlation of two real continuous functions, ϕ_{xy} is defined by

$$\phi_{xy}(t) = \int_{-\infty}^{\infty} x(\tau - t)\, y(\tau)\, d\tau \tag{8-1}$$

If we compare it to convolution

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(t-\tau)\, y(\tau)\, d\tau \tag{8-2}$$

we can see that the only difference is that for the cross correlation, one of the two functions is not reversed. Thus,

$$\phi_{xy}(t) = x(-t) * y(t) \tag{8-3}$$

In the frequency domain we can write the Fourier transform of x(-t) as

$$FT\left[x(-t)\right] = \int_{-\infty}^{\infty} x(-t)\exp(-i2\pi ft)\, dt \tag{8-4}$$

Substituting $t' = -t$ yields

$$FT\left[x(-t)\right] = \int_{\infty}^{-\infty} -x(t')\exp(i2\pi ft')\, dt' = \int_{-\infty}^{\infty} x(t')\exp(i2\pi ft')\, dt' = X^*(f) \tag{8-5}$$

Time reversal is the same as taking the complex conjugate in the frequency domain.
We can thus write

$$\Phi_{xy} = FT\left[\phi_{xy}(t)\right] = X^*(f)\, Y(f) \tag{8-6}$$

Unlike convolution, cross-correlation is not commutative but we can write

$$\phi_{xy}(t) = \phi_{yx}(-t) \tag{8-7}$$

You can show this by letting $\tau' = \tau - t$

In the discrete domain, the correlation of two real time series

$x_i,\ i = 0, 1, \ldots, M-1$

and

$y_j,\ j = 0, 1, \ldots, N-1$

is by analogy to equation (8-1) given by

$$\phi_{xy,k} = \sum_{j=\max(0,k)}^{\min(M-1+k,\,N-1)} x_{j-k}\, y_j, \quad k = -(M+1), \ldots, 0, \ldots, (N-1) \tag{8-8}$$

In Matlab cross-correlations are computed with the function `xcorr` which works in the frequency domain. Note that to obtain the discrete version of ϕ_{xy} as defined by equation (8-8), one reverses the arguments (i.e., one calls `phixy = xcorr(y,x)`). `xcorr` also pads the end of the shorter input with zeros so that they are the same length. Since Matlab cannot have zero or negative indexes the cross correlation sample with zero lag is the central element in the output vector. An alternate way of doing the cross correlation without padding with zeros is using the `conv` command (`phixy = conv(y,x(end:-1:1))`)

Autocorrelation

Autocorrelation is the result of cross-correlating a function with itself. Equation (8-1) becomes

$$\phi_{xx}(t) = \int_{-\infty}^{\infty} x(\tau - t)x(\tau)d\tau \tag{8-9}$$

and equation (8-4) becomes

$$\Phi_{xx} = FT\left[\phi_{xx}(t)\right] = X^*(f)X(f) = \left|X(f)\right|^2 \tag{8-10}$$

From the expression of the Energy used in the derivation of Parseval's theorem (equation 7-9) we can see that energy is given by the zero lag autocorrelation

$$\phi_{xx}(0) = \int_{-\infty}^{\infty} x(\tau)^2 d\tau = \int_{-\infty}^{\infty} \left|X(f)\right|^2 df = E \tag{8-11}$$

Normalized Cross-Correlation

In seismology we often use correlation to search for similar signals that are repeated in a time series – this is known as matched filtering. Because the correlation of two high amplitude signals will tend to give big numbers, one cannot determine the similarity of two signals just by comparing the amplitude of their cross correlation.

The normalized correlation for two time series can be defined as

$$\overline{\phi}_{xy}(t) = \frac{\phi_{xy}(t)}{\sqrt{\phi_{xx}(0)\phi_{yy}(0)}} \tag{8-12}$$

the normalized quantity $\overline{\phi}_{xy}(t)$ will vary between -1 and 1. A value of $\overline{\phi}_{xy}(t)$=1 indicates that at the alignment t, the two time series have the exact same shape (the amplitudes may be different) while a value $\overline{\phi}_{xy}(t)$=-1 indicates that they have the same shape except that they have the opposite signs. A value of $\overline{\phi}_{xy}(t) = 0$ shows that they are completely uncorrelated. In practice when one applies this normalization to real discrete signals, one will find that a correlation coefficient greater than about 0.7 or 0.8 indicates a pretty good match.