AutoCorrelation

February 7, 2014

Abstract

Autocorrelation is the cross-correlation of a signal with itself. Informally, it is the similarity between observations as a function of the time lag between them.

1 Definition and physical nature

The autocorrelation is the cross-correlation of a signal with itself. The cross-correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them. It is the "sliding inner product". For continuous functions f and g, the cross-correlation is defined[1] as

$$(f \star g)(t) \equiv \int_{-\infty}^{\infty} \bar{f}(\tau)\, g(t+\tau)\, d\tau, \tag{1}$$

where $\bar{f}$ denotes the complex conjugate of f. Similarly, for discrete functions, the cross-correlation is defined as

$$(f \star g)[n] \equiv \sum_{m=-\infty}^{\infty} \bar{f}[m]\, g[m+n]\,. \tag{2}$$

The cross-correlation is similar in nature to the convolution $(f \circ g)$ of two functions f and g:

$$(f \circ g)(t) \equiv \int_{-\infty}^{+\infty} f(\tau)\, g(t-\tau)\, d\tau, \tag{3}$$

where one sees that convolution and cross-correlation are equivalent for symetrical functions $g(\tau) = g(-\tau)$. One can also think of the convolution as equivalent to the cross-correlation of $f(t)$ and $g(-t)$.

As an illustration, consider two real functions f and g differing only by an unknown shift along the t axis. The cross-correlation slides the g function along the t axis, calculating the integral of their product at each position. When the functions match, the value of $(f \star g)$ is maximized.

[1] it may differ in other fields

2 Properties

- The cross-correlation of functions $f(t)$ and $g(t)$ is equivalent to the convolution of $\bar{f}(-t)$ and $g(t)$.
- If f is Hermitian[2], then convolutions and cross-correlation are equivalent.
- $f \star f$ is maximum at the origin.
- The cross-correlation satisfies

$$\mathcal{F}\{f \star g\} = \overline{\mathcal{F}\{f\}} \cdot \mathcal{F}\{g\}, \tag{4}$$

  where $\mathcal{F}$ denotes the Fourier transform.

This last property will be used to computed efficiently our autocorrelation functions.

3 Efficient computation

Brute force algorithm from Eq. 2 is order n^2. We use property of Eq. 4 to have the much better order $n \log(n)$. The algorithm is:

$$(f \star g) = \mathrm{FFT}^{-1}\left\{\mathrm{FFT}\{f\}\,\overline{\mathrm{FFT}\{f\}}\right\}. \tag{5}$$

Since f is a real function in our problem, Eq. 5 may be rewritten as

$$(f \star g) = \mathrm{FFT}^{-1}\left\{|\mathrm{FFT}\{f\}|^2\right\}, \tag{6}$$

where $|z|^2$ denotes the absolute square[3] of x.

4 In Mathematica

Mathematica has a built-in function to compute autocorrelations: CorrelationFunction. Nevertheless, in Mathematica, it is defined as the normalized covariance function, which is different from definition 1.

5 How to use our Fortran code

5.1 Input file

Input files must look like the following,

$$f_x^{(i)}[t] \qquad f_y^{(i)}[t] \qquad f_z^{(i)}[t]$$

[2] The real part of f is an even function and the imaginary part of f is an odd function

[3] $|z|^2 = z\bar{z}$

$$
\begin{array}{rrr}
f_x^{(i+1)}[t] & f_y^{(i+1)}[t] & f_z^{(i+1)}[t] \\
 & \cdots & \\
f_x^{(\ldots)}[t] & f_y^{(\ldots)}[t] & f_z^{(\ldots)}[t] \\
f_x^{(N)}[t] & f_y^{(N)}[t] & f_z^{(N)}[t] \\
f_x^{(1)}[t+1] & f_y^{(1)}[t+1] & f_z^{(1)}[t+1] \\
f_x^{(2)}[t+1] & f_y^{(2)}[t+1] & f_z^{(2)}[t+1] \\
\cdots & \cdots & \cdots
\end{array}
$$

where N is the number of atoms in the supercell, and t is the discrete index of timesteps. For a given timestep, each atom is considered. The timestep is then incremented.

References

[1] `http://en.wikipedia.org/wiki/Autocorrelation`

[2] `http://en.wikipedia.org/wiki/Cross-correlation`

[3] `http://mathworld.wolfram.com/Autocorrelation.html`