**Time Series Analysis : The Cross-Correlation Function -** Timothy R. Derrick

* The concept of cross-correlation has been developed in two distinct fields signal processing and statistics.

**: cross-correlation는 신호처리와 통계 두 분야에 의해 정의된다.**

* In the realm of statistics, cross-correlation functions provide a measure of association between signals.

The Pearson product-moment correlation coefficient is simply a normalized version of a cross-correlation.

**: cross-correlation을 통해 규명하는 것은 관계성이며, 통계에 기반한다.**

**Pearson correlation의 계수는 cross-correlation의 *정규화* 형태이다.**

**Time Series Analysis**

* Positions, velocities, accelerations, forces can be collected at high rates for increaseingly longer periods of time

: **현대과학에서 위치, 속도, 가속도, 힘 등은 점점 더 긴 주기동안 높은 비율로 수집할 수 있다.**

* 3가지 통상적 시계열 분석 방법.

1. **First permanent discrete points on the occur could be identified, and there magnitude and/or the time at which the points occur could be noted.**

**: 이산시점이나 크기, 특정 기점에 대한 시간역을 통해 분석.**

1. **Second, the entire curve could be used to calculate a variable such as the average.**

**: 전반적인 데이터의 변화에 대하여 예를 들어, 평균과 같은 ‘변량’을 책정**

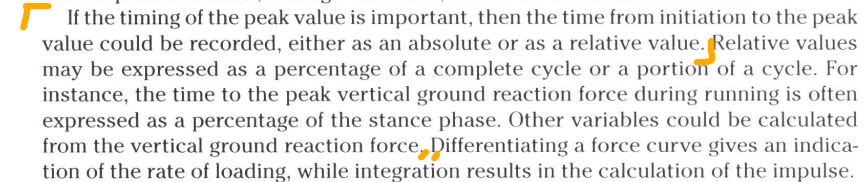
1. **Transform**

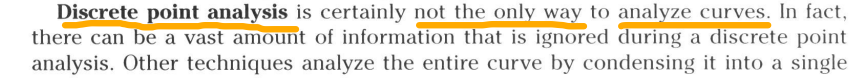
**An example is a differentiation transformation that one would use to calculate velocity from position data ---.**

**: 한 예로, 위치 데이터(위상) 차의 변화 속도**

**(단서 : Peak velocities, average velocities ---)**

**본문참조 ;**

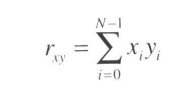
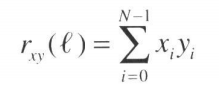
****

****

**Defining the Cross-Correlation Function**

**: 2개의 시계열 각각의 피크 값들 간의 정보들을 측정**

**: 두 시계열 데이터 간의 포인트끼리의 곲과 이의 합을 기준으로 관계의 정도를 책정**

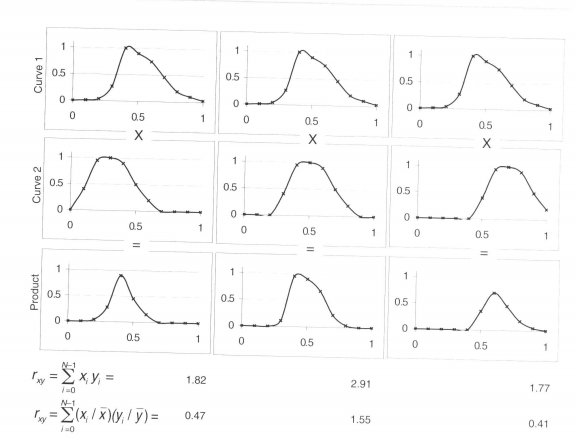
**** ****

**N : number of data points in each data series**

**X\_t : ith data point of the first data series**

**Y\_t ; ith data point of the second data series**

**R\_xy : correlation**

****

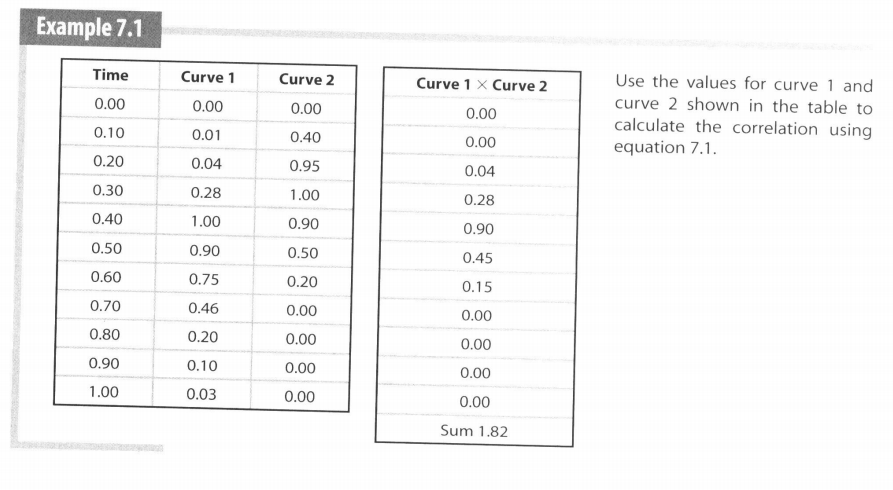
***The middle column shows the greatest degree of linear relationship between curve 1 and curve 2.***

* The curves tend to rise and fall at the same time.
* The cross correlation value of 2.91 quantifies this similarity.
* Time shifting curve 2 to the left or to the right tends to reduce the cross-correlation (1.82 and 1.77, respectively)
* Substracting the mean value of each time series has the effect of accentuaating cross-correlation vlaues because times at which the curves have the opposite sign reduce the cross-correlation value.

: **곡선이 반대신호를 가질 때에 cross correlation 은 감소되는 이유로,**

**각 시계열 데이터의 평균값을 빼는 것은 cross-correlation 값을 강조하는 효과가 있다.**

* Substracting the mean value also gives significance to a negative cross-correlation.

****

**The data should be wrapped only if they are circular.**

**For instance, if an activity such as steady state cycling was being investigated and data were collected from top dead center(TDC) of one stroke to TDC of the next stroke of the same leg, then the data could be considered circular and wrapping may be appropriate.**

**:** 동일한 시간차(leg)에 대해, TDC는 wrapping과 순환을 정의하기에 적합하다,

반면, 단순 위상에 대한 해석은, 완전 주기에 대한 정도를 설명할 뿐, wrapping과 순환을 정의하기에 부적합 하다.

* Unmatched data points에 대해 해석해내지 못한다면, 그만큼 leg가 증가함에 따른 cross-correlation은 낮아지게 되고, correlation에 기초한 특정 데이터 포인트는 감소

**The cross-correlation depends on the units of x and y, and therefore it is difficult to compare cross-correlations from different data sets.**

* 위와 같은 이유로, 데이터 셋 전체에 대한 상관관계를 규명하는 것보다 부분에 대한 상관관계 규명이 적합한 것은 아닐지 고민이 됨. **From 내생각**
* In order to prevent a reduction in the sum of the products with increasing lags and to make the cross correlation unitless, most application use “normalize” equation.

