IV. Covariance Analysis
Autocovariance

Remember that when a stochastic process has time values that are interdependent, then we can characterize that interdependency by computing the autocovariance function.

This function compares the time series with a replica of itself shifted in time. It takes on a continuous form for continuous time series and a discrete form for discrete time series. It provides an indication of the degree to which the amplitude of the time series at one time relates to or can be inferred from its amplitude at another time.
The autocovariance function receives its name by being an extension of the statistical covariance measure for random variables $X$ and $Y$. The covariance is derived from the variance, and has the same form in terms of the second central moment. The variance of random variable $X$ is the expected value of $(X-\mu_x)^2$, and the covariance of random variables $X$ and $Y$ is the expected value of the product $(X-\mu_x)(Y-\mu_y)$, where $\mu_x$ and $\mu_y$ are the mean values of $X$ and $Y$.

\[
cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)]
\]
Suppose that $X = X(t)$ and $Y = X(t + \tau)$.

The covariance of $X(t)$ and $X(t + \tau)$ is then a function of their time separation (or lag), $\tau$.

Because the covariance is that of an individual time series, it is called an autocovariance.

To simplify the discussion, we will assume that the ensemble mean of $x(t)$ is zero.
The autocovariance function of a stationary stochastic process with zero mean is:

\[ C_{XX}(\tau) = E[(X(t))(X(t + \tau))] \]

This definition says that we get the expected value by taking the mean product of the time series and its shifted replica.
By assuming ergodicity we can estimate the autocovariance function as the expected value (mean) over time for one realization. We can express this estimate as the limit of the temporal mean as $T$ increases to infinity:

$$C_{xx}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} X(t)X(t + \tau)dt$$
It is evident that computing the autocovariance function of a time series involves 3 operations:
(1) shift the replica with respect to the time series
(2) multiply each value of the time series and that of the shifted replica
(3) average the cross-products
Fig. 24. A. Comparison of an EEG shifted with respect to itself in time, by an amount $\tau$. B. Same but for the EEGs from two different regions.
Fig. 25. Computation of points of the autocorrelogram of a 10 c/sec sine wave for selected values of delay. The baseline of zero correlation is approximately at the midpoint of the chart paper. The correlograms in this and in the subsequent illustrations were computed by means of an analog correlator especially designed for brain potentials. (Barlow et al. 1954, 1955)
Estimating the Autocovariance Function

In general, the autocovariance function cannot be computed exactly for finite length (T) time series. A major problem is that the shift operation causes loss of information when the number of cross-products decreases with increasing lag. We now consider how to estimate the autocovariance function.
First consider the case of a periodic time series. In this simple case, there is no problem with the shift operation. Let us assume that T corresponds to one period. The shifted replica extends outside of T, but because the time series is periodic, the points outside of T extend continuously from the time points within T. (The operation is continuous at the boundaries.) Thus, the autocovariance function can be computed exactly by utilizing these extended points. For a periodic time series:

$$C_{xx}(\tau) = \frac{1}{T} \int_{0}^{T} X(t)X(t + \tau)dt$$

where $-\infty < \tau < \infty$
Now consider aperiodic time series. The autocovariance is:

\[ C_{xx}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} X(t)X(t + \tau)\,dt \]

In this definition, \(X(t)\) must be infinitely long, and \(T\) and \(\tau\) can also be infinitely long. However, in practice, the time series has finite length \(T\). The circular autocovariance estimator is a way that has been proposed to estimate the autocovariance function for finite length \(T\). This estimator makes the simplifying (erroneous) assumption that \(T\) represents one period of an infinite periodic function. By this assumption, the same equation may be used for aperiodic time series as for periodic time series:

\[ C_{xx}(\tau) = \frac{1}{T} \int_{0}^{T} X(t)X(t + \tau)\,dt \]
Using the circular autocovariance estimator creates a further problem. Namely, the operation is no longer continuous at the boundaries. An alternative is to shrink the analysis window as $\tau$ increases. The sample autocovariance estimator (Jenkins & Watts, 1968) is given by:

$$\hat{C}_{XX}(\tau) = \frac{1}{T} \int_{0}^{T-|\tau|} X(t)X(t+\tau)dt, \quad 0 \leq |\tau| \leq T$$

This estimator makes explicit the fact that $X(t) = 0$ outside $(0, T)$.

Jenkins & Watts (1968, pp. 174-180) discuss the fact that the division by $T$ gives this estimator less error than division by $T - |\tau|$. 
In practice, T must be chosen according to the particular problem under study. If T is too limited, appreciable statistical fluctuation in the autocovariance function occurs. On the other hand, T must not be so long as to include time points that are obviously nonstationary.

In practice also, we deal with the discrete rather than the continuous case. This leads to the finite, discrete definition of the autocovariance function:

\[
\hat{C}_{XX}(k) = \frac{1}{N} \sum_{i=0}^{N-k} X(i)X(i+k)
\]

where \(x(i \Delta t)\) are discrete values of \(X\) at sample points \(i\), \(T=N \Delta t\), and \(\tau=k \Delta t\). We can assume that \(\Delta t = 1\) without any loss of generality.
In practice, the autocovariance function is computed over a range of $\tau$ from 0 to some maximum value that cannot be larger than $T$. Since the number of products that contribute to the estimate of $C_{XX}(\tau)$ from the sample autocovariance estimator decreases as $\tau$ increases, the operation is usually terminated at a point between $T/20$ and $T/5$. 
The autocovariance is an even function:

\[ C_{xx}(\tau) = C_{xx}(-\tau) \]

Its maximum value occurs at \( \tau = 0 \)

If we consider \( \tau = 0 \) in the definition of autocovariance, we get the variance of the time series (we are still considering the mean of \( x(t) \) to be zero).

\[ C_{xx}(\tau) = E[(X(t))(X(t + \tau))] \]
We can normalize the autocovariance function by dividing it by the variance.

This gives the autocorrelation function:

\[ R_{xx}(\tau) = \frac{C_{xx}(\tau)}{C_{xx}(0)} \]

\[ R_{xx}(0) = 1 \]

\[ |R_{xx}(\tau)| \leq 1 \]
The Uses of Autocovariance and Autocorrelation

1. Autocovariance and autocorrelation are used to identify general features of a time series. For example, it is possible to distinguish among a pure sinusoid, narrow-band noise, and wide-band noise strictly on their appearance.

The dominant periodicity can be estimated and its stability indicated by the extent to which the amplitude is maintained with increasing $\tau$. 
Top. The autocovariance function of a pure sinusoid has peaks that appear without decrement at a period equal to that of the sinusoid. Middle. It is also periodic for narrow-band noise, but falls off with increasing lag. Bottom. For wide-band noise, there is a single peak at $\tau = 0$, and the function rapidly falls to a low value for increasing $\tau$.
2. Autocovariance and autocorrelation are used to identify signals in noise. In this example, the autocorrelation of the signal+noise identifies the sinusoidal signal, even though it is not apparent in the time series.
3. Autocovariance and autocorrelation are a means of data reduction. They do not depend on the length of the original data sample.
4. Autocovariance and autocorrelation can be used as a step toward spectral analysis.

Spectral analysis is needed to identify the frequency content of complex time series. The autocovariance and autocorrelation do not quantify the frequency content of a time series. For complex time series, it is difficult to identify the different frequency components.
**Crosscovariance**

The crosscovariance function between two different time series $x(t)$ and $y(t)$ with zero means may be defined as:

$$C_{XY}(\tau) = E[X(t)Y(t + \tau)]$$

The crosscovariance function differs from the autocovariance function only in that the delayed time series $X(t + \tau)$ is replaced by $Y(t + \tau)$. The crosscovariance function is an indication of the degree to which the amplitude of the first time series at one time relates to or can be inferred from the amplitude of the second time series at another time.
The same arguments apply to the crosscovariance as to the autocovariance. Thus, the crosscovariance estimator is:

\[
\hat{C}_{XY}(\tau) = \frac{1}{T} \int_0^{T-|\tau|} X(t)Y(t + \tau) \, dt
\]

which is of the same form as:

\[
\hat{C}_{XX}(\tau) = \frac{1}{T} \int_0^{T-|\tau|} X(t)X(t + \tau) \, dt, \quad 0 \leq |\tau| \leq T
\]
The discrete form of the definition is:

\[ \hat{C}_{XY}(k) = \frac{1}{N} \sum_{i=0}^{N-k} X(i)Y(i + k) \]

In general, the crosscovariance function is not even since it is asymmetric around \( \tau = 0 \). The maximum value does not necessarily occur at \( \tau = 0 \). The maximum occurs when \( X(t) \) and \( Y(t) \) are most in register.

Example: If \( Y(t) = X(t - \tau) \), then \( C_{XY}(\tau) = \max \).
Normalizing the crosscovariance yields the crosscorrelation function:

\[ R_{xy}(\tau) = \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0)C_{yy}(0)}} \]

where \( C_{xx}(0) \) and \( C_{yy}(0) \) are the variances of \( X(t) \) and \( Y(t) \), respectively.

Like the autocorrelation function, \( R_{xy}(\tau) \) is bounded by 1 and -1. However, \( R_{xy}(0) \) need not be equal to 1.
The uses of crosscovariance and crosscorrelation

1. The crosscovariance and crosscorrelation functions are useful for determining whether two time series contain common components, and if so, what time difference (e.g. lag or delay) exists between the common component in the two time series.
(a) If $x(t)$ and $y(t)$ are both periodic functions, only those frequencies that are common to both appear in the crosscovariance function. Phase differences between common frequencies are retained.

Fig. 27. Cross-correlogram of 2 sine waves shifted 90° in phase (left) and mixed with two independent noise signals of the same rms amplitude, respectively (right). The two independent noise signals are shown in the center. Duration of sample, $T$: 1.5 sec. (Barlow 1959)
(b) If $x(t)$ and $y(t)$ are composed of a single common frequency, then the displacement of the maximum (or minimum) value away from $\tau = 0$ corresponds to the phase difference of that frequency.

Fig. 27. Cross-correlogram of 2 sine waves shifted 90° in phase (left) and mixed with two independent noise signals of the same rms amplitude, respectively (right). The two independent noise signals are shown in the center. Duration of sample, T: 1.5 sec. (Barlow 1959)
(c) If $x(t)$ and $y(t)$ contain periodicities of different frequencies, then the crosscovariance function is 0 for all values of $\tau$. 
2. The crosscovariance function is also useful for detection of hidden common components. The hidden components may be periodic or aperiodic. For two time series that are both random, but contain a common component, a peak appears in the crosscovariance at a value of $\tau$ equal to the time difference between the components in the two time series.
Fig. 28. Cross-correlation of two random signals (of the same type as in Fig. 27) having a hidden common random component, the latter being delayed in the second signal by 40 msec with respect to the first. Duration of sample, $T$: 1.5 sec.
If there is a hidden common periodic component buried in random noise, the periodicity will appear in the crosscovariance function.