Time-varying signals: cross- and auto-correlation, correlograms

NEU 466M

Spring 2020

Statistical measures

- We first considered simple statistical measures for single variables (mean, variance).
- We next considered measures for the relationship between two (stationary) random variables (covariance, Pearson's correlation coefficient; regression).
- Extension to *K* variables: pairwise relationships (covariance matrix).
- Now, extension to time-series: relationships between different time-varying signals.

Time-series data

$$\{\cdots g_{t-1}, g_t, g_{t+1} \cdots\}$$

g: a temporally varying signal sampled at discrete intervals

$$\{\cdots h_{t-1}, h_t, h_{t+1} \cdots\}$$

h: another time-varying signal, sampled at the same times

How many variables? In *g* alone?

Time-series data

$$\{\cdots g_{t-1}, g_t, g_{t+1} \cdots\}$$

g: a temporally varying signal sampled at discrete intervals

- Time-series not independent sampling.
- Could think of response at each time point as separate (though typically not independent) variable.
- If length(*g*) = *T*, then *T* variables in *g*.
- Same time-point in repetitions of the series from same initial condition: multiple samples of that variable.

Finding structure between time-series

$$\{\cdots g_{t-1}, g_t, g_{t+1} \cdots\}$$

g: a temporally varying signal sampled at discrete intervals

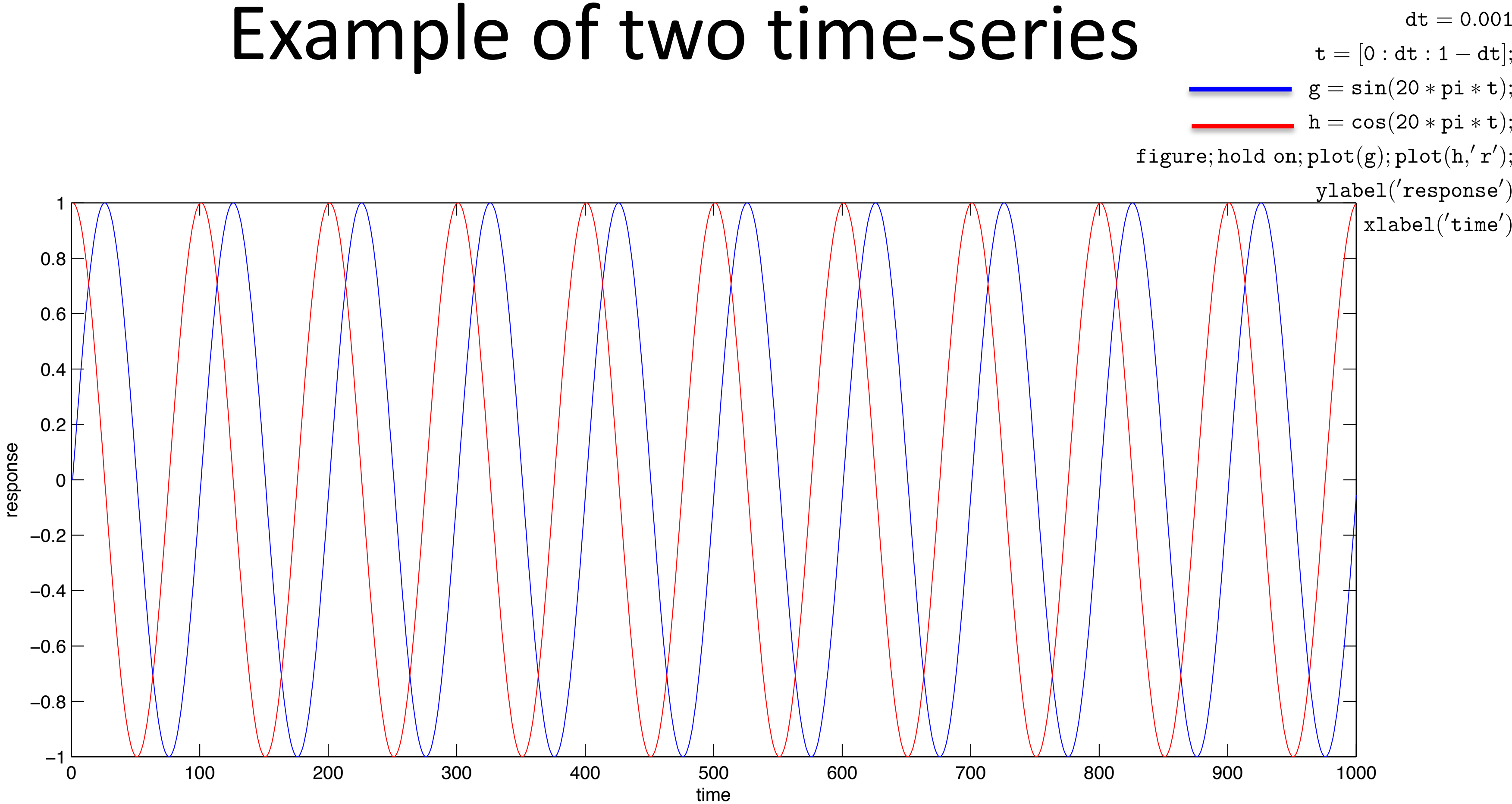
$$\{\cdots h_{t-1}, h_t, h_{t+1} \cdots\}$$

h: another time-varying signal, sampled at the same times

How about trying previously seen statistical measures?

$$\text{compute } cov(g, h)$$

Example of two time-series



$$C(g,h) = 0, \langle gh \rangle \approx 0$$

Correct but unsatisfying:

g, h similarly time-varying functions: the same function with a π/2 shift.

Definition: cross-correlation function

$$C_{g,h}(n) = \sum_{m=-\infty}^{\infty} g^*(m)h(m+n)$$

$n = 0:$

$$\begin{matrix} [\cdots g^*_{-3} & g^*_{-2} & g^*_{-1} & g^*_0 & g^*_1 & g^*_2 & g^*_3 \cdots] \\ | & | & | & | & | & | & | \\ [\cdots h_{-3} & h_{-2} & h_{-1} & h_0 & h_1 & h_2 & h_3 \cdots] \end{matrix}$$

take time-by-time product, add all terms

Definition: cross-correlation function

$$C_{g,h}(n) = \sum_{m=-\infty}^{\infty} g^*(m)h(m+n)$$

$n = 1:$

$$[\cdots g^*_{-3} \quad g^*_{-2} \quad g^*_{-1} \quad g^*_0 \quad g^*_1 \quad g^*_2 \quad g^*_3 \cdots]$$

$$\longleftarrow [\cdots h_{-2} \quad h_{-1} \quad h_0 \quad h_1 \quad h_2 \quad h_3 \quad h_4 \cdots]$$

h – tape shifted leftwards by 1

Measure of relatedness at different time shifts

$$C_{g,h}(n) = \sum_{m=-\infty}^{\infty} g^*(m)h(m+n)$$

- The cross-correlation function is a measure of covariance between *g, h* at different relative time-shifts (for zero-mean or mean-subtracted signals).
- How is *g* at any time (linearly) related to *h* *n* time-steps away?

Cross-correlation function for finite-length signals

$$\{g_1, \cdots, g_N\}$$
$$\{h_1, \cdots, h_N\}$$

g, h: time-series of length N

$$C_{g,h}(n) = \sum_{m=1}^{N-n} g^*(m)h(m+n)$$

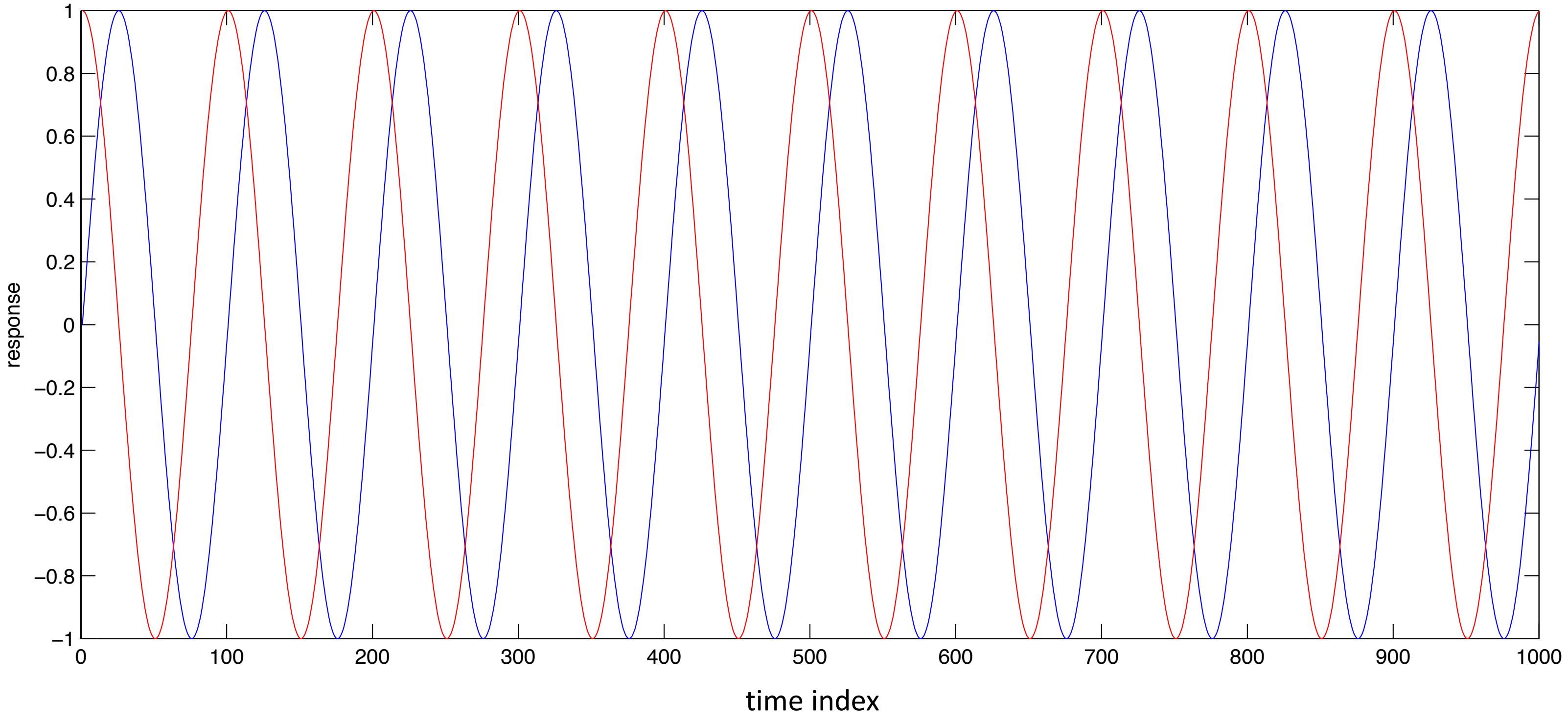
average over
(N-|n|) terms

Total length of cross-correlation: *2N-1*
Zero-shifted entry: *N*

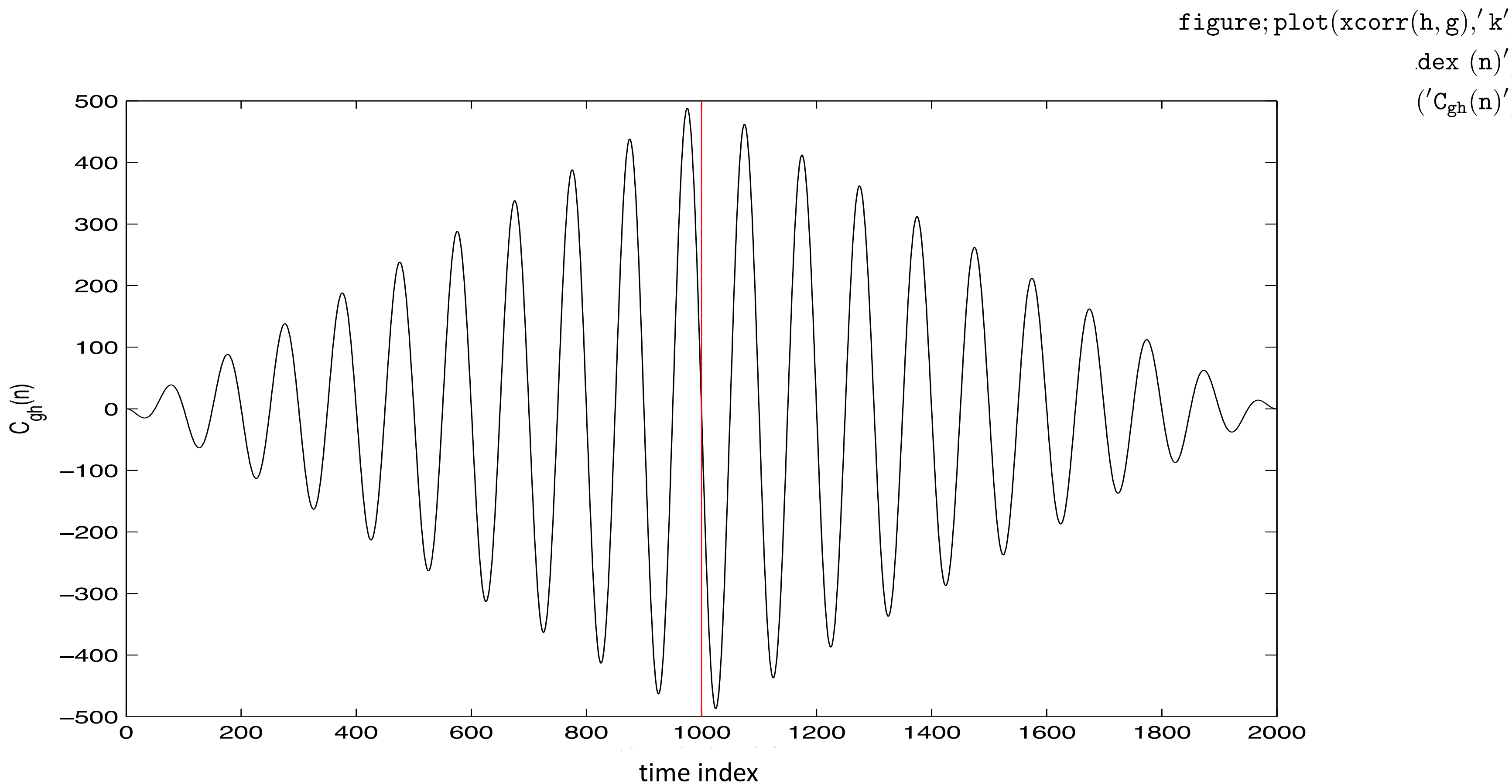
Properties of the cross-correlation

- $C_{gh}(n) \neq C_{hg}(n)$: does not commute (contrast with covariance).
- In fact, $C_{gh}(n) = C_{hg}(-n)$: shifting *h* to *right* relative to *g:* equivalent to shifting *g* to *left* relative to *h*. (Same plot, flipped time axis.)
- Ordering matters: tells which signal leads the other.

(Previous) example



Cross-correlation of example series

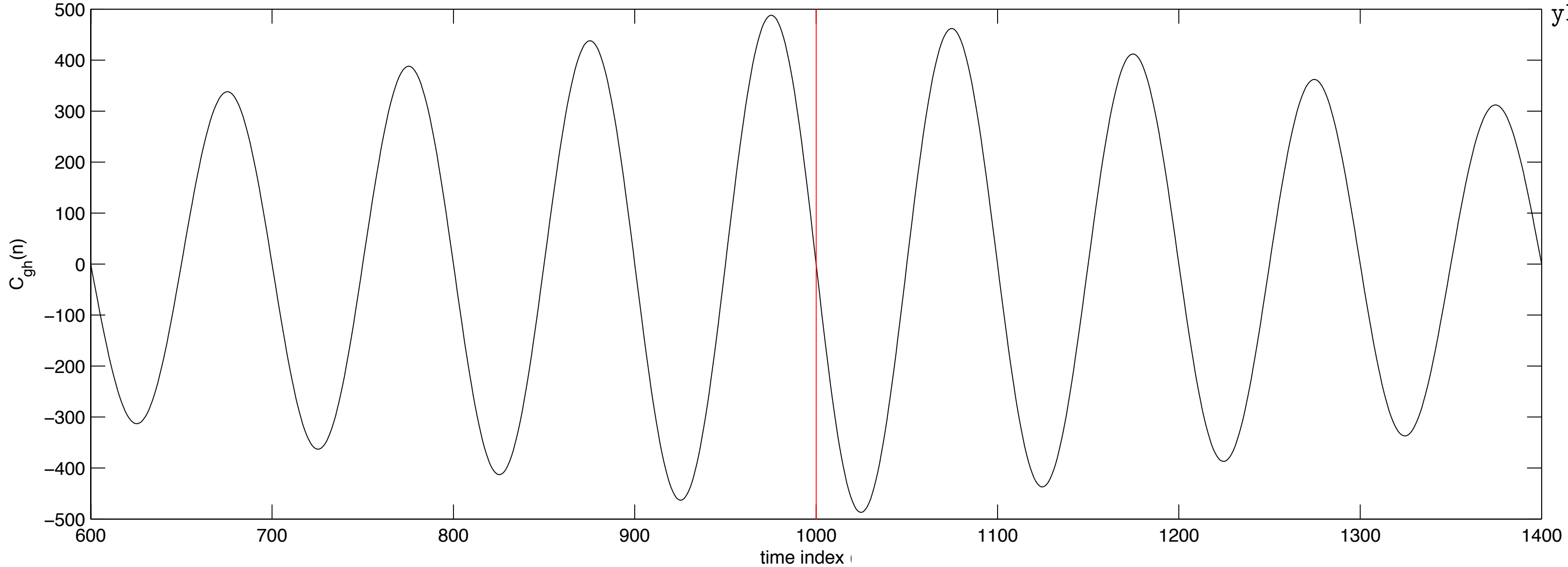


Signals of length *N = 1000*. Cross-correlation of length *2N-1.*

Cross-correlation of example series

```
figure; plot(xcorr(h, g), 'k')
xlabel('time index (n)')
ylabel('C_gh(n)')
```

Matlab: signals of length *N* have 0-shift at *N* in xcorr



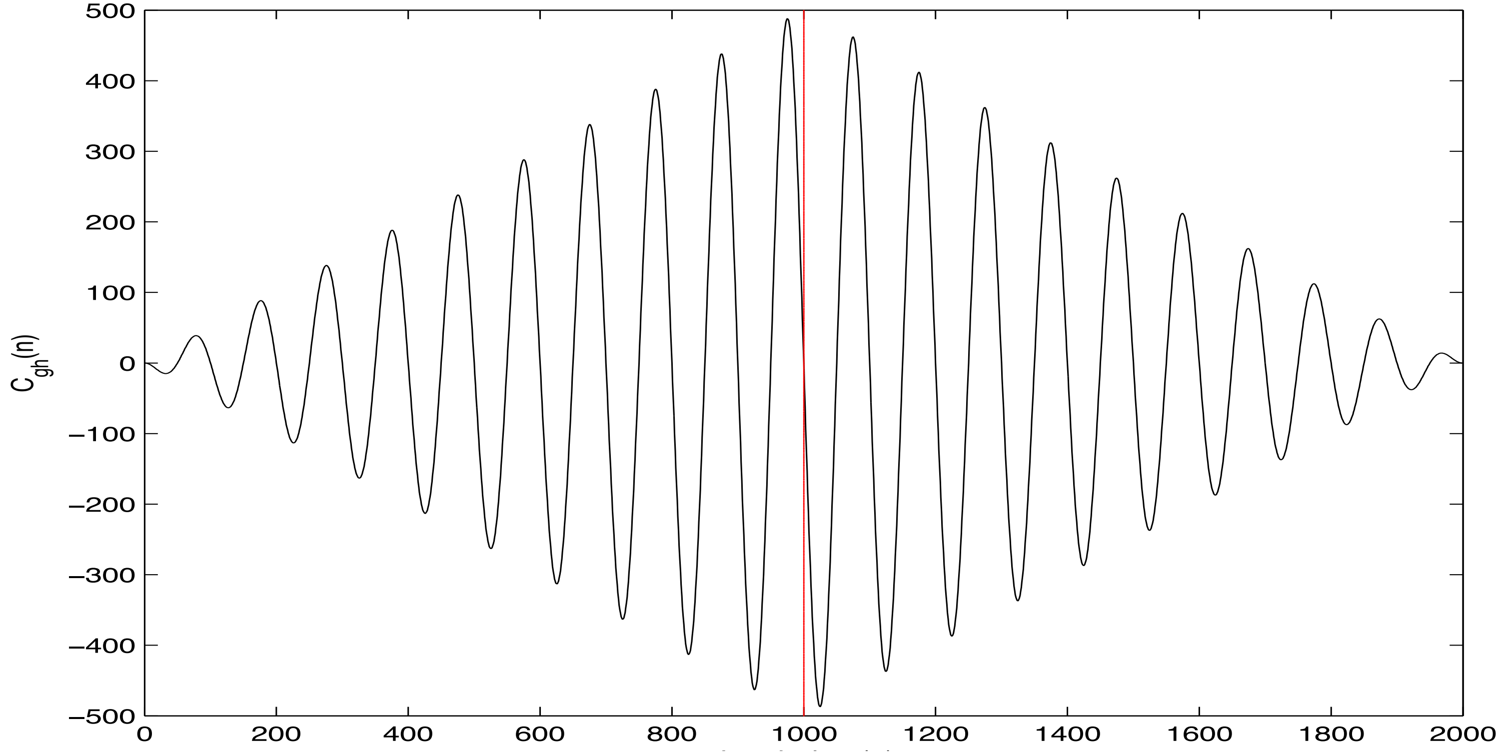
- In $C_{gh}(n)$: peak at left. Interpretation: *h* leads *g*, or *h* must be shifted right (negative *n*) to line up with *g*. In present example, cosine (*h*) leads sine (*g*) in phase.

- Multiple peaks: periodic re-alignment of cos with sin at multiples of period

- Caution! $C_{gh}(t)$ is xcorr(h,g) in Matlab: note reversed order of g, *h*!

Cross-correlation of example series

figure; plot(xcorr(h, g),'k')
.dex (n)')
('C_gh(n)')

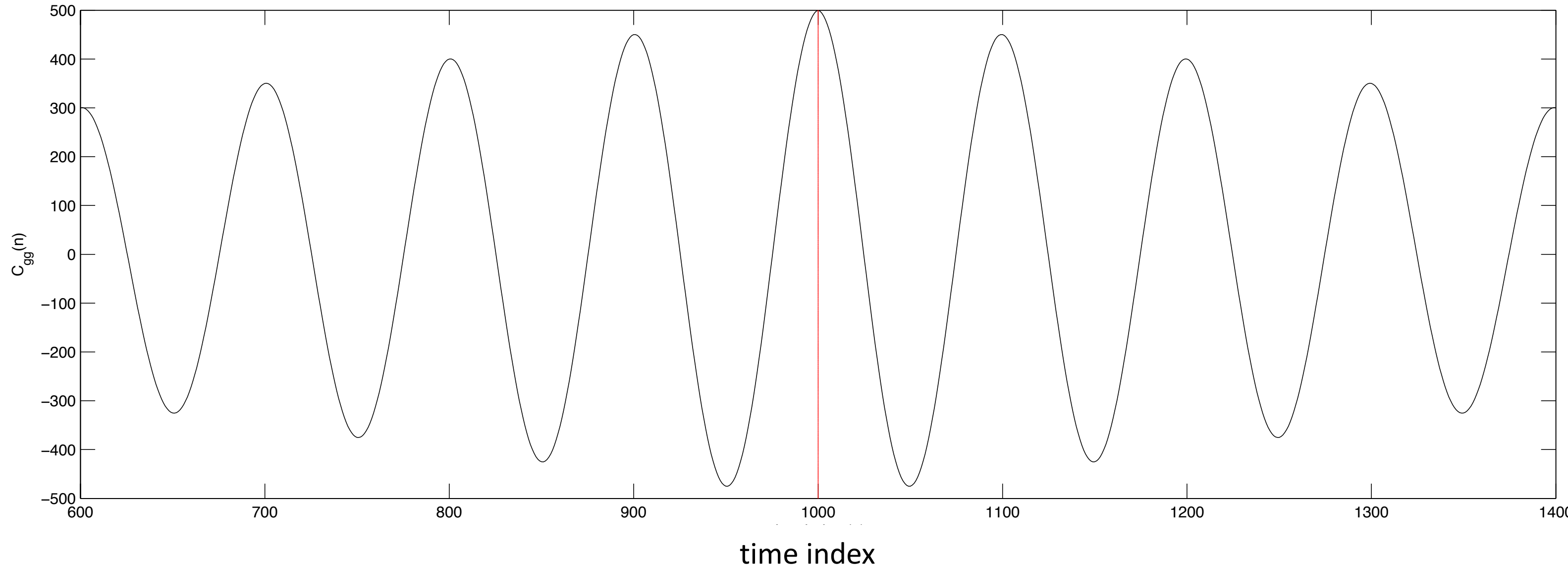


Decay in amplitude due to finite length of *g, h*: shift *n* is a sum over *N*-|*n*| terms, so amplitude will go to 0 as shift goes to *N.*

Autocorrelation function

- Special case of cross-correlation: signal correlation with itself at all time shifts.
- Commonly used to detect temporal patterns (periodic or otherwise) within noisy time-series data.
- Symmetric; central peak always at 0 time-lag.

Autocorrelation example



Overview

BACK TO A MODELING PERSPECTIVE

Time-series data

$$\{g_1, \cdots, g_N\}$$

g: a temporally varying signal sampled at discrete intervals

- If *g* is time-series of length *N*, then *N* variables within *g*.
- But then should construct *NxN* covariance matrix with *(α,β)* entry given by $cov(g_\alpha g_\beta)$, and *N(N+1)/2* distinct entries.
- Autocorrelation: only (2N-1)/2 distinct entries (1/2 because of symmetry about 0-time lag).

Autocorrelation and time-series data

What are we throwing out when studying only the autocorrelation of a time-series g (N distinct entries), compared to the NxN covariance matrix of its components (N(N+1)/2 distinct entries)?

Autocorrelation and time-series data

- One entry in cross-correlation measures relationship between terms in g at a fixed time-lag, **summed over all times.**

- Assumption of **time translation-invariance**: relationship between terms in *g* at lag *n* is similar, regardless of starting time. (E.g. whenever a cell spikes, it will tend not to spike for 2 ms: refractory period. Pattern independent of actual spike time.)

- Same assumption in cross-correlation.

Summary

- Time-series inherently more complex than random draws from an independent system.
- Look for: translation-invariant temporal patterns within and across time-series.
- Auto- and cross-correlation functions.
- Each term an average across the entire time-series: reduced noise.