Digital Communications Project 7

Correlation, Coherence, and Information Flow

1 Preface

The purpose of this project is to familiarize you with the concepts of correlation, coherence and information flow. Since MATLAB works with data stored as vectors and matrices, the aim of this project is also to learn how to program in the MATLAB environment and how to use MATLAB scripts and functions with respect to this. The communications lab is a closely-supervised hands-on course, therefor you are strongly encouraged to experiment with different MATLAB features.

Help for each MATLAB function can be displayed by entering `help <function name>`. For example `help plot` describes the `plot` function. The complete online manual can be started by entering `helpdesk` at the MATLAB command line.

2 Correlation

Correlation refers to the statistical relationship between two random variables. Cross correlation is the measure of similarity between two signals, when one of them is subjected to time lag. Correlation coefficient is a measure of linear dependence between two signals, giving values between $+1$ and -1. A positive value shows that two signals are positively correlated, zero refers to no correlation while negative value expresses negative correlation. Autocorrelation is cross correlation of a signal with its own time lagged version. Cross correlation between two time series $x(t)$ and $y(t)$ can be calculated with this expression:

$$R_{xy}(t) = \sum_{\tau=-\infty}^{\tau=+\infty} x(\tau - t)y(\tau)\,. \tag{1}$$

Correlation coefficient between two time series is a scalar value, showing the strength of the linear relationship between them. The Correlation coefficient between $x(t)$ and $y(t)$ can be calculated using given expression:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}\,. \tag{2}$$

The expression to calculate sample variance is given below:

$$s_{xy} = \frac{1}{n}\sum_{j=1}^{j=n}(x_j - \bar{x})(y_j - \bar{y})\,. \tag{3}$$

The sample mean can be calculated using:

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{j=n} x\,. \tag{4}$$

2.1 Exercise

(a) Generate the vectors `x=[9 2 6 5 8]` and `y=[12 8 6 4 10]` .

(b) Calculate the sample mean $\bar{x}$ and $\bar{y}$ for both vectors by using expression (4).

(c) Calculate the sample variance s_{xx} and s_{yy} for both vectors using expression (3). Also calculate sample covariance s_{xy}.

(d) Calculate the sample correlation coefficient r_{xx}, r_{yy} and r_{xy} using expression (2).

(e) Compare your results of the correlation coefficient with the output of the command given below:
`>>corrcoef(x,y).`

(f) Generate a vector `z=[1 3 -1 2]`.

(g) Calculate the autocorrelation vector from delay -3 to 0 using expression (1).

(h) Since correlation is similar to convolution, only difference being time reversal. We know that time reversal in time domain is equal to taking complex conjugate in frequency domain. This means that we can calculate autocorrelation of signal using its fourier transform too. Calculate the autocorrelation of the vector given in (f) using the expression:

$$Correlation = Ifft(fft(z).fft(z)^*)\,. \tag{5}$$

(i) Compare results of (g) and (h).

3 Coherence

Coherence is used to identify those variations between two signals which have similar spectral properties. Coherence (magnitude squared coherence) can be described by the expression given below:

$$Coh(w) = \frac{\mid S_{xy} \mid}{\sqrt{S_{xx}}\sqrt{S_{yy}}}\,. \tag{6}$$

Where S_{xy} is the cross power spectral density between signal x and y, while S_{xx} and S_{yy} are the auto power spectral densities of signal x and y, respectively. Cross power spectral density between x and y can be calculated as:

$$S_{xy} = fft(x).fft(y)^* \,. \tag{7}$$

In this exercise we will also use Welch's method to estimate power spectral densities.

3.1 Exercise

(a) Generate two vectors x and y using expression given below in MATLAB
```
>> fs = 1000;t = 0:1/fs:1-1/fs;
>> x = sin(2*pi*200*t)+0.5*randn(size(t));
>> y = 0.35*cos(2*pi*200*t)+0.5*randn(size(t));
```

(b) Use the MATLAB functions `cpsd` and `pwelch` to calculate the cross power spectral denstiy between the two vectors x and y and their auto power spectral density. Description about these functions can be read by typing `help <function name>`.

(c) Finally, use expression (6) to calculate and plot the coherence. Which frequency is significantly pronounced?

(d) Using the MATLAB function `mscohere` to plot the coherence and compare your results of task (c) and (d).

(e) Create a random signal x with a length of 524288 using `randn`.

(f) Pass this signal with system with transfer function h = [0 0 1 0 0 0.5] using `filter` and name the output as y1.

(g) Create another signal random signal y2 of same length as that of x.

(h) Now calculate the coherence between x and y1 and compare it with the coherence between x and y2. Take the complete signal for this calculation.

(i) Now divide all signals into windows of length 1024 and calculate the coherence for each window. Take the average of all these windows and compare your results with (h).

4 Information flow

Information flow or causality is the relation between two events, where one of them is understood as a consequence of other event. In signal processing we use the concept of information flow to estimate which event started farther back in time in relation to the other one. This can be better illustrated by following expressions:

$$x_1(t) = a_{11}x_1(t-1) + a_{12}x_2(t-1) + n_1(t) \,, \tag{8}$$

$$x_2(t) = a_{21}x_1(t-1) + a_{22}x_2(t-1) + n_2(t) \,, \tag{9}$$

Both time series are modelled as autoregressive processes with order 1, AR(1). a_{xx} are coefficients of causality and their magnitudes show the extent of the information flow between the time series.

4.1 Exercise

(a) Generate two time series of length 1000, using the following AR coefficients.
$a_{11} = 0.8$
$a_{12} = 0.2$
$a_{21} = 0.6$
$a_{22} = 0.2$

(b) Add appropriate random noise to these two time series.

(c) Construct a zero matrix K with size 999*5. Replace its first column with a vector of ones (999,1). The Second and third column should be replaced by data vectors (first 999 data points). The fourth and fifth column should be replaced by data vectors (last 999 data points).

(d) Subject this matrix K to orthogonal triangular decomposition using the MATLAB function `qr` and store output in variable R. The description of this function can be found by typing `help qr`.

(e) Extract the upper triangular part of the variable R using MATLAB function `triu` and store the output in same variable. The description of this function can be found by typing `help triu`.

(f) Extract the first three rows of the first three columns of R and store them in variable $R11$.

(g) Extract the first three rows of the fourth and fifth column of R and store them in variable $R12$.

(h) Divide $R11$ with $R12$, take its conjugate and store this in variable A.

(i) Second and third column of A contains estimated AR coefficients. Compare these results with the actual values.