332:525 – Optimum Signal Processing
Computer Experiment 2 – Due February 10, 2011

1. Consider the following AR(1), first-order, autoregressive signal model with a time-varying parameter:

$$y_n = a(n)y_{n-1} + \epsilon_n \tag{1}$$

where ϵ_n is zero-mean, unit-variance, white noise. The filter parameter $a(n)$ can be tracked by the following adaptation equations (which are equivalent to the exact recursive least-squares order-1 adaptive predictor):

$$\begin{aligned} R_0(n) &= \lambda R_0(n-1) + \alpha\, y_{n-1}^2 \\ R_1(n) &= \lambda R_1(n-1) + \alpha\, y_n y_{n-1} \\ \hat{a}(n) &= \frac{R_1(n)}{R_0(n)} \end{aligned}$$

where $\alpha = 1 - \lambda$. The two filtering equations amount to sending the quantities y_{n-1}^2 and $y_n y_{n-1}$ through an exponential smoother. To avoid possible zero denominators, initialize R_0 to some small positive constant, $R_0(-1) = \delta$, such as $\delta = 10^{-3}$.

(a) Show that $\hat{a}(n)$ satisfies the recursion:

$$\hat{a}(n) = \hat{a}(n-1) + \frac{\alpha}{R_0(n)} y_{n-1} e_{n/n-1}, \qquad \text{where} \quad e_{n/n-1} = y_n - \hat{a}(n-1)y_{n-1}$$

where $e_{n/n-1}$ is referred to as the a priori estimation (prediction) error.

(b) Using Eq. (1), generate a data sequence y_n, $n = 0, 1, \dots, N-1$ using the following time varying coefficient, sinusoidally switching from a positive value to a negative one (the synthesis filter switches from lowpass to highpass):

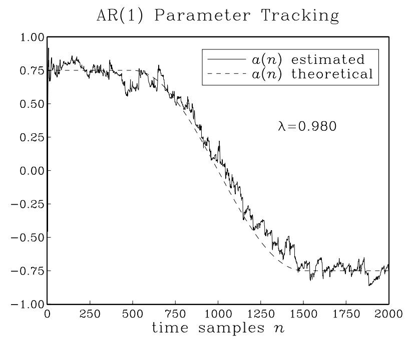
$$a(n) = \begin{cases} 0.75, & 0 \le n \le N_a - 1 \\ 0.75\cos\left(\pi \dfrac{n - N_a}{N_b - N_a}\right), & N_a \le n \le N_b \\ -0.75, & N_b + 1 \le n \le N - 1 \end{cases}$$

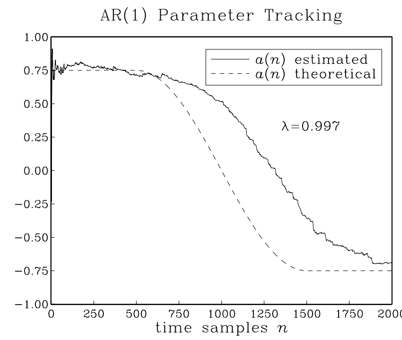
Use the following numerical values:

$$N_a = 500, \quad N_b = 1500, \quad N = 2000$$

Determine the estimated $\hat{a}(n)$ and plot it versus n together with the theoretical $a(n)$ using the parameter value $\lambda = 0.980$. Repeat using the value $\lambda = 0.997$. Comment on the tracking capability of the algorithm versus the accuracy of the estimate.

(c) Study the sensitivity of the algorithm to the choice of the initialization parameter δ.





2. Next, consider an AR(2), second-order, model with time-varying coefficients:

$$y_n = -a_1(n)y_{n-1} - a_2(n)y_{n-2} + \epsilon_n \tag{2}$$

If the coefficients were stationary, then the theoretical Wiener solution for the prediction coefficients a_1 and a_2 would be:

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = -\begin{bmatrix} R_0 & R_1 \\ R_1 & R_0 \end{bmatrix}^{-1} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = -\frac{1}{R_0^2 - R_1^2} \begin{bmatrix} R_0R_1 - R_1R_2 \\ R_0R_2 - R_1^2 \end{bmatrix} \tag{3}$$

For a time-varying model, the coefficients can be tracked by replacing the theoretical autocorrelation lags with their recursive, exponentially smoothed, versions:

$$\begin{aligned} R_0(n) &= \lambda R_0(n-1) + \alpha\, y_n^2 \\ R_1(n) &= \lambda R_1(n-1) + \alpha\, y_n y_{n-1} \\ R_2(n) &= \lambda R_2(n-1) + \alpha\, y_n y_{n-2} \end{aligned}$$

(a) Using Eq. (2), generate a non-stationary data sequence y_n by driving the second-order model with a unit-variance, zero-mean, white noise signal ϵ_n and using the following theoretical time-varying coefficients:

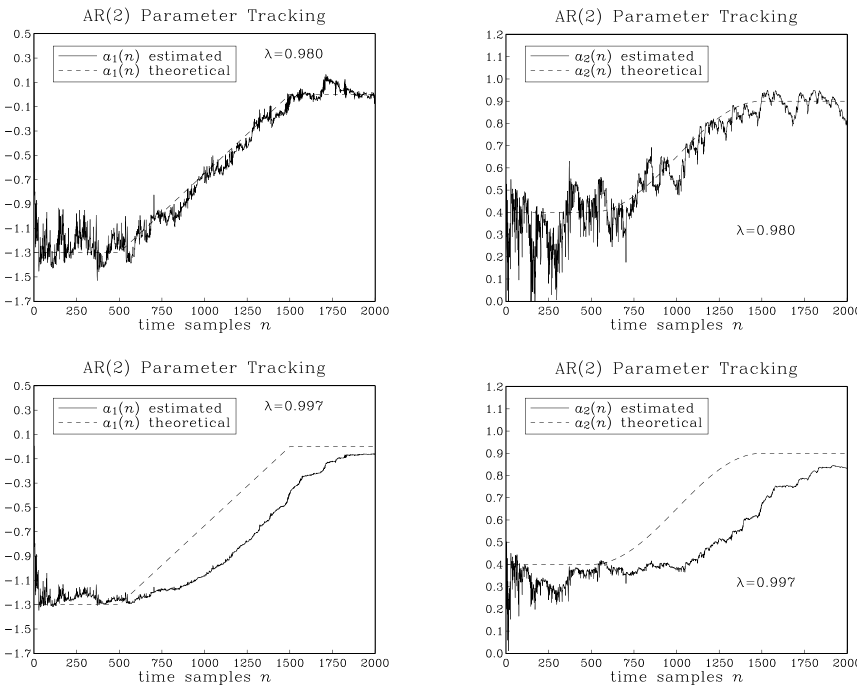
$$a_1(n) = \begin{cases} -1.3, & 0 \le n \le N_a - 1 \\ 1.3\,\dfrac{n - N_b}{N_b - N_a}, & N_a \le n \le N_b \\ 0, & N_b + 1 \le n \le N - 1 \end{cases}$$

$$a_2(n) = \begin{cases} 0.4, & 0 \le n \le N_a - 1 \\ 0.65 - 0.25\cos\left(\pi\dfrac{n - N_a}{N_b - N_a}\right), & N_a \le n \le N_b \\ 0.9, & N_b + 1 \le n \le N - 1 \end{cases}$$

Thus, the signal model for y_n switches continuously between the synthesis filters:

$$B(z) = \frac{1}{1 - 1.3z^{-1} + 0.4z^{-2}} \quad \Rightarrow \quad B(z) = \frac{1}{1 + 0.9z^{-2}}$$

(b) Compute the adaptive coefficients $\hat{a}_1(n)$ and $\hat{a}_2(n)$ using the two forgetting factors $\lambda = 0.980$ and $\lambda = 0.997$. Plot the adaptive coefficients versus n, together with the theoretical time-varying coefficients and discuss the tracking capability of the adaptive processor.



3. Next, we will apply the adaptive method of part-2 to some real data. The file `sunspots.dat` contains the yearly mean number of sunspots for the years 1700–2008. To unclutter the resulting graphs, we will use only the data for the last 200 years, over 1809–2008. These can be read into MATLAB as follows:

```
Y = loadfile('sunspots.dat');
i = find(Y(:,1)==1809);
y = Y(i:end,2);                  % number of sunspots
N = length(y);                   % here, N=200
m = mean(y); y = y-m;            % zero-mean data
```

where the last line determines the mean of the data block and subtracts it from the data. The mean m will be restored at the end.

Yule was the first to introduce the concept of an autoregressive signal model and applied it to the sunspot time series assuming a second-order model. The so-called Yule-Walker method is a block processing method in which the entire (zero-mean) data block is used to estimate the autocorrelation lags R_0, R_1, R_2 using sample autocorrelations:

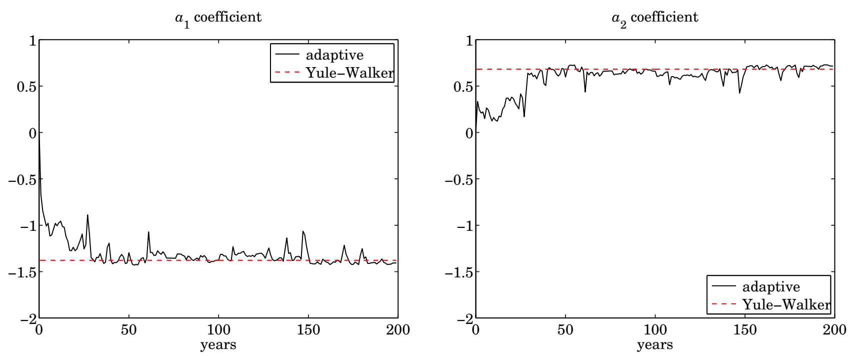
$$\hat{R}_0 = \frac{1}{N}\sum_{n=0}^{N-1} y_n^2\,, \quad \hat{R}_1 = \frac{1}{N}\sum_{n=0}^{N-2} y_{n+1}y_n\,, \quad \hat{R}_2 = \frac{1}{N}\sum_{n=0}^{N-3} y_{n+2}y_n$$

Then, the model parameters a_1, a_2 are estimated using Eq. (3):

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = -\begin{bmatrix} \hat{R}_0 & \hat{R}_1 \\ \hat{R}_1 & \hat{R}_0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{R}_1 \\ \hat{R}_2 \end{bmatrix} \qquad \text{(Yule-Walker method)}$$

(a) First, compute the values of $\hat{a}_1, \hat{a}_2$ based on the given length-200 data block.

(b) Then, apply the adaptive algorithm of the part-2 with $\lambda = 0.99$ to determine the adaptive versions $a_1(n), a_2(n)$ and plot them versus n, and add on these graphs the straight lines corresponding to the Yule-Walker estimates $\hat{a}_1, \hat{a}_2$.

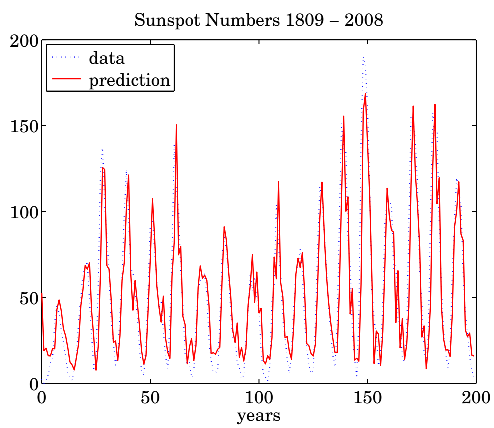


(c) At each time instant n, the value of y_n can be predicted by either of the two formulas:

$$\hat{y}_{n/n-1} = -a_1(n)y_{n-1} - a_2(n)y_{n-2}$$

$$\hat{y}_{n/n-1} = -\hat{a}_1 y_{n-1} - \hat{a}_2 y_{n-2}$$

On the same graph, plot y_n and $\hat{y}_{n/n-1}$ for the above two alternatives. The case of the adaptive predictor is shown below.



(d) Repeat the above questions using $\lambda = 0.95$ and discuss the effect of reducing λ.

(e) Apply a length-200 Hamming window w_n to the (zero-mean) data y_n and calculate the corresponding periodogram spectrum,

$$S_{\text{per}}(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} w_n y_n e^{-j\omega n}\right|^2$$

as a function of the yearly period $p = 2\pi/\omega$, over the range $2 \le p \le 20$ years. For the same p's or ω's calculate also the AR(2) spectrum using the Yule-Walker coefficients $\hat{a}_1, \hat{a}_2$:

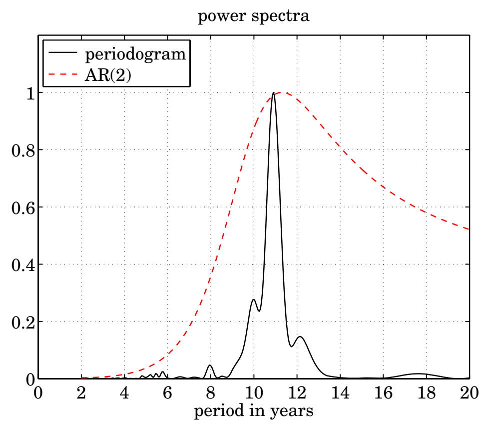
$$S_{\text{AR}}(\omega) = \frac{\sigma_\epsilon^2}{|1 + \hat{a}_1 e^{-j\omega} + \hat{a}_2 e^{-2j\omega}|^2}$$

where σ_ϵ^2 can be calculated by

$$\sigma_\epsilon^2 = \hat{R}_0 + \hat{a}_1\hat{R}_1 + \hat{a}_2\hat{R}_2$$

Normalize the spectra $S_{\rm per}(\omega), S_{\rm AR}(\omega)$ to unity maxima and plot them versus period p on the same graph. Note that both predict the presence of an approximate 11-year cycle, which is also evident from the time data.

We will revisit this example later on by applying SVD methods to get sharper peaks for the autoregressive model.



4. *Regression Lemma.* The optimum estimate and estimation error of a random vector $\mathbf{x}$ based on a vector of observations $\mathbf{y}_1$ is given by

$$\hat{\mathbf{x}}_1 = R_{xy_1}R_{y_1y_1}^{-1}\mathbf{y}_1 = E[\mathbf{x}\mathbf{y}_1^T]E[\mathbf{y}_1\mathbf{y}_1^T]^{-1}\mathbf{y}_1\,, \quad \mathbf{e}_1 = \mathbf{x}-\mathbf{x}_1$$

$$R_{e_1e_1} = E[\mathbf{e}_1\mathbf{e}_1^T] = R_{xx} - R_{xy_1}R_{y_1y_1}^{-1}R_{y_1x}$$

Suppose the observation set is enlarged by adjoining to it a new set of observations $\mathbf{y}_2$, so that the enlarged observation vector is $\mathbf{y} = \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}$. The corresponding estimate of $\mathbf{x}$ is given by

$$\hat{\mathbf{x}} = R_{xy}R_{yy}^{-1}\mathbf{y} = [R_{xy_1}, R_{xy_2}]\begin{bmatrix}R_{y_1y_1} & R_{y_1y_2}\\ R_{y_2y_1} & R_{y_2y_2}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}$$

Show that $\hat{\mathbf{x}}$ can be obtained by the following alternative expression:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_1 + R_{x\varepsilon_2}R_{\varepsilon_2\varepsilon_2}^{-1}\boldsymbol{\varepsilon}_2 \quad \text{(regression lemma)} \tag{4}$$

where $\boldsymbol{\varepsilon}_2$ is the innovations residual obtained by removing from $\mathbf{y}_2$ that part which is predictable from $\mathbf{y}_1$, that is,

$$\boldsymbol{\varepsilon}_2 = \mathbf{y}_2 - \hat{\mathbf{y}}_{2/1} = \mathbf{y}_2 - R_{y_2y_1}R_{y_1y_1}^{-1}\mathbf{y}_1$$

Hint: Let $H = R_{y_2y_1}R_{y_1y_1}^{-1}$, and prove and make use of the properties:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}I & 0\\ H & I\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix}, \quad \begin{bmatrix}R_{y_1y_1} & R_{y_1y_2}\\ R_{y_2y_1} & R_{y_2y_2}\end{bmatrix} = \begin{bmatrix}I & 0\\ H & I\end{bmatrix}\begin{bmatrix}R_{y_1y_1} & 0\\ 0 & R_{\varepsilon_2\varepsilon_2}\end{bmatrix}\begin{bmatrix}I & 0\\ H & I\end{bmatrix}^T$$

Show that the improvement in using more observations is quantified by the following result, which shows that the mean-square error is reduced:

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} \quad \Rightarrow \quad R_{ee} = R_{e_1e_1} - R_{x\varepsilon_2} R_{\varepsilon_2\varepsilon_2}^{-1} R_{\varepsilon_2 x} \tag{5}$$

where we defined,

$$R_{x\varepsilon_2} = R_{\varepsilon_2 x}^T = E[\mathbf{x}\boldsymbol{\varepsilon}_2^T]\,, \quad R_{\varepsilon_2\varepsilon_2} = E[\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2^T]$$

The above discussion assumed that the random vectors had zero mean. How are Eqs. (4) and (5) to be modified if $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2$ happen to have means $\mathbf{m}, \mathbf{m}_1, \mathbf{m}_2$, respectively?

The regression lemma is a key result in the derivation of the Kalman filter.

This part may be handed separately as a handwritten hardcopy.