Alex Lemon
EE263 (Introduction to Linear Dynamical Systems)
Homework 3
July 11, 2013

## 1 AR system identification

You are given input/output measurements

$$(u(1), y(1)), \dots, (u(N), y(N)),$$

where $u(t)$ and $y(t)$ are, respectively, the input and output of an unknown system at time $t$. You believe that the system can be approximated by an autoregressive model:

$$\hat{y}(t) = a_0 u(t) + b_1 y(t-1) + \cdots + b_n y(t-n).$$

(a) Explain how to choose the coefficients $a_0, b_1, \dots, b_n$ in order to minimize the sum of squared errors:

$$\sum_{t=n+1}^{N} (y(t) - \hat{y}(t))^2.$$

(b) Apply your method to the data given in `ar_system_identification_data.m`. Carry out the estimation for $n = 1, \dots, 35$. The relative error of the model is defined to be

$$\bar{\epsilon} = \frac{\sum_{t=n+1}^{N} (y(t) - \hat{y}(t))^2}{\sum_{t=n+1}^{N} y(t)^2}.$$

Submit a plot of the relative error versus $n$. What do you think is a good value of $n$?

(c) The file `ar_system_identification_data.m` contains a second set of data:

$$(u_{\text{cv}}(1), y_{\text{cv}}(1)), \dots, (u_{\text{cv}}(N), y_{\text{cv}}(N)).$$

Compute the estimated output sequence

$$\hat{y}_{\text{cv}}(t) = a_0 u_{\text{cv}}(t) + b_1 y_{\text{cv}}(t-1) + \cdots + b_n y_{\text{cv}}(t-n),$$

and the relative cross-validation error:

$$\bar{\epsilon}_{\text{cv}} = \frac{\sum_{t=n+1}^{N} (y_{\text{cv}}(t) - \hat{y}_{\text{cv}}(t))^2}{\sum_{t=n+1}^{N} y_{\text{cv}}^2(t)}.$$

Submit a plot of the relative cross-validation error versus $n$. Briefly discuss your results. What do you think is a good value of $n$?

## 2 Fitting a rational transfer function to frequency-response data

Consider a rational transfer function $H : \mathbb{C} \to \mathbb{C}$ of the form

$$H(s) = \frac{A(s)}{B(s)},$$

where $A(s)$ and $B(s)$ are $m$th-degree polynomials:

$$A(s) = a_0 + a_1 s + \cdots + a_m s^m, \qquad \text{and} \qquad B(s) = 1 + b_1 s + \cdots + b_m s^m.$$

The coefficients $a_0, \ldots, a_m$ and $b_1, \ldots, b_m$ are real numbers, and the variable $s$ is a complex number. We define the coefficient vectors

$$a = (a_0, \ldots, a_m) \in \mathbb{R}^{m+1}, \qquad \text{and} \qquad b = (b_1, \ldots, b_m) \in \mathbb{R}^m.$$

We are given empirical frequency-response data (that is, noisy measurements of $H$ at points on the imaginary axis):

$$(s_i = j\omega_i, h_i), \qquad i = 1, \ldots, N,$$

where $\omega_1, \ldots, \omega_N$ are nonnegative real numbers, and $h_1, \ldots, h_N$ are complex numbers. Our goal is to find coefficient vectors $a$ and $b$ such that $H(s_i) \approx h_i$. To judge the equality of this approximation, we use the mean squared error:

$$J = \frac{1}{N} \sum_{i=1}^{N} |H(s_i) - h_i|^2.$$

There is a famous heuristic for this problem, which is based on solving a sequence of linear least-squares problems. First, we express $J$ as

$$J = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{A(s_i) - h_i B(s_i)}{z_i} \right|^2, \qquad z_i = B(s_i), \qquad i = 1, \ldots, N.$$

Let $k$ denote the iteration counter, and $a^{(k)}$, $b^{(k)}$ and $z_i^{(k)}$ denote the values of $a$, $b$ and $z_i$, respectively, in iteration $k$. Let $A^{(k)}(s)$ and $B^{(k)}(s)$ denote

$$A^{(k)}(s) = a_0^{(k)} + a_1^{(k)} s + \cdots + a_m^{(k)} s^m, \qquad \text{and} \qquad B^{(k)}(s) = 1 + b_1^{(k)} s + \cdots + b_m^{(k)} s^m.$$

respectively. In each iteration, we first update our estimates of the $z_i$:

$$z_i^{(k+1)} = B^{(k)}(s_i), \qquad i = 1, \ldots, N.$$

Then, we choose $a^{(k+1)}$ and $b^{(k+1)}$ in order to minimize

$$\frac{1}{N} \sum_{i=1}^{N} \left| \frac{A^{(k+1)}(s_i) - h_i B^{(k+1)}(s_i)}{z_i^{(k+1)}} \right|^2.$$

We can start the process with $z_i^{(1)} = 1$ for $i = 1, \ldots, N$, and we can terminate the process when successive iterates are very close (that is, $a^{(k+1)} \approx a^{(k)}$ and $b^{(k+1)} \approx b^{(k)}$). Several things can go wrong with this method. For example, we can run into a situation where $z_i^{(k)} = 0$, or the least-squares problem used to update our estimates of $a$ and $b$ may not have a unique solution. However, you may ignore such issues in this problem.

(a) Explain how to compute the updates $a^{(k+1)}$ and $b^{(k+1)}$. Be sure to keep in mind that the coefficients $a_0, \ldots, a_m$ and $b_1, \ldots, b_m$ are real numbers, whereas many other quantities in this problem are complex.

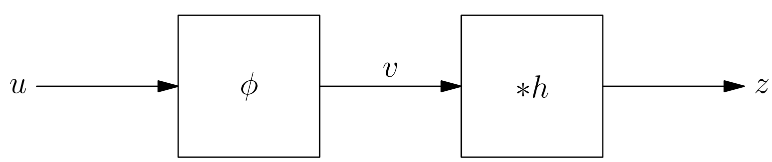
(b) Apply this method to the data in `rational_transfer_function_data.m`. Terminate the algorithm when

$$\left\| \begin{bmatrix} a^{(k+1)} - a^{(k)} \\ b^{(k+1)} - b^{(k)} \end{bmatrix} \right\| \leq 10^{-6}.$$

Report the final coefficient estimates $a^{(k)}$ and $b^{(k)}$, and the associated value of $J$. Plot $J$ versus the iteration number $k$, using a logarithmic scale for $J$, and a linear scale for $k$. For the first and final coefficient estimates, and the given data, plot $|H(j\omega)|$ versus $\omega$, using a logarithmic scale for $|H(j\omega)|$, and a linear scale for $\omega$. (Note that the first coefficient estimate corresponds to $k = 2$.)

## 3 Designing a nonlinear equalizer from input/output data

Consider the discrete-time system shown below, which consists of a memoryless nonlinearity $\phi$, followed by a convolution filter with finite impulse response $h$. The scalar signal $u$ is the input, and the scalar signal $z$ is the output.



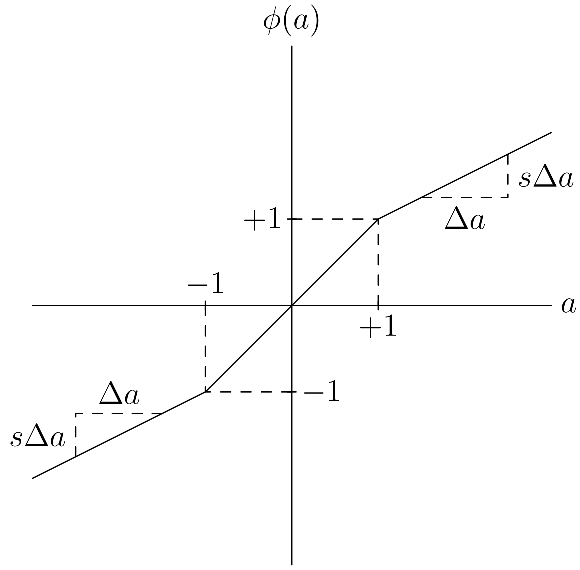

The equations describing the system are

$$z(t) = \sum_{\tau=0}^{M-1} h(\tau) v(t - \tau), \qquad v(t) = \phi(u(t)), \qquad t \in \mathbb{Z}.$$

(Note that we define the signals $u$, $v$ and $z$ for all integer times, not just nonnegative times.) The nonlinearity $\phi : \mathbb{R} \to \mathbb{R}$ has the form

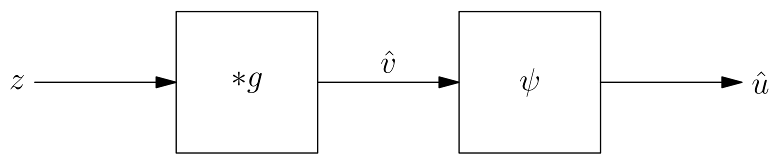
$$\phi(a) = \begin{cases} a & -1 \leq a \leq +1, \\ +1 + s(a - 1) & a > +1, \\ -1 + s(a + 1) & a < -1, \end{cases}$$

where $s > 0$ is a parameter. A sketch of the function $\phi$ is shown below.




We can think of $\phi$ as a model for a power amplifier that is nonlinear for input signals that have magnitude greater than one; $s$ is called the saturation gain of the amplifier. The convolution system represents a communication channel.

We want to design an equalizer for this system: that is, we want to find another system that takes the signal $z$ as input, and gives an output $\hat{u}$ that is an approximation of $u$. In particular, we will design an equalizer of the following form.




The equations describing the equalizer are

$$\hat{v}(t) = \sum_{\tau=0}^{M-1} g(\tau)z(t-\tau), \qquad \hat{u}(t) = \psi(\hat{v}(t)), \qquad z \in \mathbb{Z}.$$

Recall that the convolution of $g$ and $h$ is defined to be

$$(g * h)(t) = \sum_{\tau=\max\{0,t-M+1\}}^{\min\{M-1,t\}} g(\tau)h(t-\tau), \qquad t = 0, \ldots, 2M-1.$$

The equalizer will work well if $g * h \approx \delta$ (in which case $\hat{v}(t) \approx v(t)$), and $\psi = \phi^{-1}$ (in which case $\psi(\phi(a)) = a$ for all $a$). You are given some input/output data

$$(u(1), z(1)), \ldots, (u(N), z(N)),$$

and the value of $M$ (the length of finite impulse responses $h$ and $g$). You do not know the parameter $s$, the channel impulse response $h(0), \ldots, h(M-1)$, or the signals $u(t)$ or $z(t)$ for $t \leq 0$ or $t > N$.

(a) Explain how to find an estimate $\hat{s}$ of the saturation gain, and an impulse response $g(0), \ldots, g(M-1)$ that minimize the mean squared error:

$$J = \frac{1}{N-M+1} \sum_{i=M}^{N} (\hat{v}(i) - \phi(u(i)))^2.$$

(b) Apply your method to the data given in `nonlinear_equalizer_data.m`. Report your estimate $\hat{s}$, your impulse response $g(0), \ldots, g(M-1)$, and the corresponding mean squared error $J$. Submit a stem plot of $g$.

(c) Using your estimate $\hat{s}$, and your impulse response $g(0), \ldots, g(M-1)$, compute the equalized signal $\hat{u}(t)$ for $t = 1, \ldots, N$. (You may assume that $z(t) = 0$ for $t \leq 0$ when computing $\hat{u}(t)$. This assumption may lead to large equalization errors for $t = 1, \ldots, M-1$.) Plot the input signal, $u(t)$, the output signal, $z(t)$, the equalized signal, $\hat{u}(t)$, and the equalization error, $e(t) = u(t) - \hat{u}(t)$, for $t = 1, \ldots, N$.

## 4 Fitting a model for hourly temperature

You are given a set of temperature measurements $y_t \in \mathbb{R}$ for $t = 1, \ldots, N$, where $y_t$ is the temperature (in °C). The measurements are taken hourly over one week (so $N = 7 \times 24 = 168$). You think that an appropriate model for the hourly temperature is a trend (that is, a linear function of $t$), plus a diurnal component (that is, a periodic function with period 24):

$$\hat{y}_t = at + p_t,$$

where $a \in \mathbb{R}$ and $p \in \mathbb{R}^N$ are the trend coefficient, and diurnal component, respectively. The periodicity of the diurnal component implies that

$$p_{t+24} = p_t, \qquad t = 1, \ldots, N-24.$$

(a) Explain how to choose $a$ and $p$ in order to minimize the root mean squared error fitting error.

(b) Apply your method to the data given in `hourly_temperature_data.m`. Report your estimate of $a$. Plot the observed temperature data and the fitted model on the same set of axes.

(c) *Temperature prediction.* Use your fitted model to predict the temperature for the 24-hour period after the data ends (that is, from $t = 169$ to $t = 192$). The variable `ytom` (also defined in `hourly_temperature_data.m`) gives the actual temperatures for this period. Plot the predicted and actual values on the same set of axes. Report the RMS prediction error.

## 5 State-trajectory estimation

Consider a discrete-time linear dynamical system:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(t), \\ y(t) &= Cx(t) + v(t), \end{aligned}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^p$ is the output, $w(t) \in \mathbb{R}^n$ is process noise, and $v(t) \in \mathbb{R}^p$ is measurement noise. You know the matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{p \times n}$, the input sequence $u(1), \ldots, u(T-1)$, and the output sequence $y(1), \ldots, y(T)$; you do not know $x(t)$, $w(t)$, or $v(t)$. Let $\hat{x}(t)$ denote an estimate of the state trajectory. There are two sets of residuals associated with this estimate:

$$\begin{aligned} \hat{w}(t) &= \hat{x}(t+1) - A\hat{x}(t) - Bu(t), && t = 1, \ldots, T-1, \\ \hat{v}(t) &= y(t) - C\hat{x}(t), && t = 1, \ldots, T. \end{aligned}$$

A measure of the quality of an estimate $\hat{x}(t)$ is

$$J = \sum_{t=1}^{T-1} \|\hat{x}(t+1) - (A\hat{x}(t) + Bu(t))\|^2 + \rho \sum_{t=1}^{T} \|y(t) - C\hat{x}(t)\|^2,$$

where $\rho > 0$ is a given parameter (which captures our beliefs about the relative sizes of $w(t)$ and $v(t)$).

(a) Explain how to find the state-trajectory estimate $\hat{x}(t)$ for $t = 1, \ldots, T$ that minimizes $J$. State any assumptions that are needed for your method to work.

(b) The file `state_trajectory_estimation_data.m` defines the following variables.

- `A`, `B`, and `C`, the state, input, and measurement matrices, respectively
- `m`, `n`, and `p`, the dimensions of the input, state, and output, respectively
- `rho`, the trade-off parameter
- `T`, the time horizon
- `u`, an $m \times T$ matrix, where `u(:,t)` is the input at time $t$
- `y`, a $p \times T$ matrix, where `y(:,t)` is the output at time $t$

Apply your method to this instance of the problem. Report the value of $J$ corresponding to your estimate. The file `state_trajectory_estimation_data.m` also defines a variable `x_true`, which is an $n \times T$ matrix, where `x_true(:,t)` is the true state at time $t$. Plot $x_1(t)$ and $\hat{x}_1(t)$ on the same set of axes.