Alex Lemon
EE263 (Introduction to Linear Dynamical Systems)
Problem Session 3
July 3, 2013

1 Estimation with sensor offset and drift

Suppose we are given measurements

$$y_i = a_i^\mathsf{T} x + v_i, \qquad i = 1, \ldots, m,$$

where $y_i \in \mathbb{R}$ is the ith measurement, $x \in \mathbb{R}^n$ is an unknown vector of parameters, and $v_i \in \mathbb{R}$ is the ith measurement error. We assume that the ith measurement is taken at time $t = iT$: that is, the measurements are taken every T seconds starting at time $t = T$. You can assume that the measurement matrix

$$A = \begin{bmatrix} a_1^\mathsf{T} \\ \vdots \\ a_m^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

is skinny and full rank. We usually assume (often implicitly) that the measurement errors v_i are random, unpredictable, small, and centered around zero. In such cases, least-squares estimation works well. However, sometimes the measurement error included predictable components. For example, each sensor measurement might have a (common) offset or bias, as well as a term that grows linearly with time (called drift). We model this situation as

$$v_i = \alpha + \beta i T + w_i,$$

where α is the sensor bias, β is the drift coefficient, and w_i is random, unpredictable, small, and centered around zero. We assume that α and β are unknown, but the same for all measurements. Explain how to simultaneously estimate the parameter vector $x \in \mathbb{R}^n$, the offset $\alpha \in \mathbb{R}$, and the drift coefficient $\beta \in \mathbb{R}$. State any assumptions that are needed for your method to work, and give a simple example of a sensor matrix A for which your assumptions do not hold.

2 Identifying a system from input/output data

Suppose $y = Ax + v$, where $x \in \mathbb{R}^n$ is the input, $y \in \mathbb{R}^m$ is the output, $A \in \mathbb{R}^{m \times n}$ is the sensor matrix, and $v \in \mathbb{R}^m$ is measurement noise. Suppose we are given N input/output pairs

$$(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)}).$$

(a) Explain how to choose A in order to minimize

$$J = \sum_{k=1}^{N} \|Ax^{(k)} - y^{(k)}\|^2.$$

State any assumptions that are needed for your method to work.

(b) Apply your method to the data defined in `system_identification_data.m`. Report the average relative approximation error:

$$\frac{1}{N}\sum_{k=1}^{N}\frac{\|Ax^{(k)}-y^{(k)}\|}{\|y^{(k)}\|}.$$

3 Vector time series modeling

Consider a vector time-series $y(1),\ldots,y(T)\in\mathbb{R}^n$. For example, the components of $y(t)$ might represent measurements of n different quantities at time t. We assume that the $y(t)$ are related by a first-order recursion:

$$y(t+1)=Ay(t)+v(t),\qquad t=1,\ldots,T-1,$$

where $A\in\mathbb{R}^{n\times n}$ is a model parameter, and $v(t)\in\mathbb{R}^n$ is a disturbance that is small, random and centered around zero. Such a model is very common, and has several names: it is sometimes called a first-order vector autoregressive model, denoted VAR(1); other people call this a Gauss-Markov model, meaning a linear system driven by noise. If we knew the parameter A, then we could use $y(t-1)$ to predict $y(t)$:

$$\hat{y}(t)=Ay(t-1),\qquad t=2,\ldots,T,$$

where our prediction simply ignores the disturbance because $v(t)$ is assumed to be small, random and centered around zero. Then, the prediction error is given by

$$e(t)=\hat{y}(t)-y(t),\qquad t=2,\ldots,T.$$

Finally, we will assume that A is lower triangular, so that $y_i(t+1)$ only depends on $y_j(t)$ for $j=1,\ldots,i$. (This structural assumption would have to come from the particular situation that we are modeling: that is, how the n different quantities that we measure at time $t+1$ are related to the corresponding quantities at time t.)

(a) Explain how to find the lower-triangular matrix A that minimizes the mean squared prediction error:

$$\frac{1}{T-1}\sum_{t=2}^{T}\|e(t)\|^2.$$

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

- `T`, the length of the time series
- `n`, the number of quantities in each vector measurement
- `y`, the matrix of measurements (each column is one vector measurement)

Apply your method to this instance of the problem. Report your value of A and the corresponding mean squared prediction error. Plot $y_i(t)$ and $\hat{y}_i(T)$ on the same set of axes for $i=1,\ldots,n$. Also give the mean squared value of y:

$$\frac{1}{T}\sum_{t=1}^{T}\|y(t)\|^2,$$

and $\hat{y}(T+1)=Ay(T)$.