HW 8: Learning Theory II (580.692)

Instructor: René Vidal, E-mail: rvidal@cis.jhu.edu
Grader: Mihály Petreczky, E-mail: mihaly@cis.jhu.edu

Due 11/30/06 beginning of the class

1. Read Chapter 10 and 11 of GPCA book. Go to *http://www.vision.jhu.edu/gpcabook/* and submit all the typos you find as well as suggestions you may have to improve the quality and/or readability of the material. **You will receive credit for each interesting typo or suggestion you submit**.

2. **Observable Canonical Form**

   A linear time invariant system with state space representation

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k \end{aligned} \tag{1}$$

   where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times n}$, is said to be in *observer canonical form* if

$$A = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^T.$$

   (a) Show that

$$C(zI - A)^{-1} = [z^{n-1}, z^{n-2}, \dots, z, 1]/d(z) \tag{2}$$

   where $d(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$.

   (b) Consider the following ARX model

$$y_{k+1} = -a_1 y_k - a_2 y_{k-1} - \cdots - a_n y_{k-n+1} + b_1 u_k + b_2 u_{k-1} + \cdots + b_n u_{k-n+1} \tag{3}$$

   Consider the polynomials $d(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ and $n(z) = b_1 z^{n-1} + \cdots b_{n-1}z + b_n$. Recall that the rational expression $H(z) = \frac{n(z)}{d(z)}$ is called *the transfer function* of (3). Show that there exists a state-space model in observer canonical with the transfer function $H(z)$.
   Hint. Use the fact that the transfer function of (1) is $C(zI - A)^{-1}B$ and equation (2).

   (c) Prove that for any observable state-space representation (A, B, C) there exists a nonsingular transformation T such that the representation (TAT^{-1}, TB, CT^{-1}) is in the observable canonical form.
   Hint. Assume that $T = [t_1^T, \dots, t_n^T]^T$ and notice that if $\hat{A} = TAT^{-1}$, $\hat{B} = TB$ and $\hat{C} = CT^{-1}$, then

$$\hat{A} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} t_1 A \\ t_2 A \\ \vdots \\ t_n A \end{bmatrix} \quad \text{and} \quad \hat{C}T = C$$

   Use the definition of the observable canonical form.

3. **Design of Luenberger Observers using the Ackermann formula**
Assume that the state-space representation (A, B, C) is observable. Let $\alpha(z) = z^n + \beta_1 z^{n-1} + \cdots \beta_n$ be an arbitrary polynomial.

   (a) Assume that (A, B, C) is in observer canonical form. Find a matrix L such that $\det(zI - (A - LC)) = \alpha(z)$. Compute L explicitly as a function of $a_1, \ldots, a_n$ and $\beta_1, \ldots, \beta_n$.

   (b) Let now (A, B, C) be an arbitrary observable state-space representation, not necessarily in observer canonical form. Denote by $\mathcal{O}$ the observability matric of (A, B, C). Show that $L = \alpha(A)\mathcal{O}^{-1}[0, 0, 0, \cdots, 1]^T$ is such that $\det(zI - (A - LC)) = \alpha(z)$. The above expression for L is called the *Ackermann formula*.

4. **The Kalman Filter**
Consider the following stochastic linear state-space model.

$$\begin{aligned} x_{k+1} &= Ax_k + Bw_k \\ y_k &= Cx_k \end{aligned} \tag{4}$$

Assume that x_0, w_k are jointly Gaussian, zero-mean random variables. Assume that x_0, w_k are mutually independent for all k and w_l, w_k are independent for all $k \neq l$. Assume that $w_k \sim N(0, I)$.

   (a) Show that x_k and y_k are Gaussian random variables that are independent of w_l for all $l \geq k$.

   (b) Denote by $x_{k|k-1} = E[x_k | y_1, \ldots, y_{k-1}]$ the conditional expectation of x_k given $y_1, \ldots, y_{k-1}$.
   Denote by $x_{k-1|k-1} = E[x_{k-1} | y_1, \ldots, y_{k-1}]$ the conditional expectation of x_{k-1} given $y_1, \ldots, y_{k-1}$.
   Denote by $\Sigma_{k|k-1} = E[(x_k - x_{k|k-1})(x_k - x_{k|k-1})^T]$ the variance of $x_k - x_{k|k-1}$.
   Denote by $\Sigma_{k-1|k-1} = E[(x_{k-1} - x_{k-1|k-1})(x_{k-1} - x_{k-1|k-1})^T]$ the variance of $x_{k-1} - x_{k-1|k-1}$.

      i. Show that $x_{k|k-1} = Ax_{k-1|k-1}$ and $\Sigma_{k|k-1} = A\Sigma_{k-1|k-1}A^T + BB^T$

      ii. Show that $y_{k|k-1} = E[y_k | y_1, \ldots, y_{k-1}] = Cx_{k|k-1}$

      iii. Let $\widetilde{y}_{k-1|k-2} = y_{k-1} - y_{k-1|k-2}$. Show that $x_{k-1|k-1} = E[x_{k-1} | y_1, \ldots, y_{k-2}, \widetilde{y}_{k-1|k-2}]$ and that $y_1, \ldots, y_{k-2}$ and $\widetilde{y}_{k-1|k-2}$ are independent

      iv. Show that $\Sigma^y_{k-1|k-2} = E[\widetilde{y}_{k-1|k-2}\widetilde{y}^T_{k-1|k-2}] = C\Sigma_{k-1|k-2}C^T$.

      v. Show that $x_{k-1|k-1} = x_{k-1|k-2} + (\Sigma_{k-1|k-2})C^T(\Sigma^y_{k-1|k-2})^{-1}\widetilde{y}_{k-1|k-2}$.

   (c) Show that

$$x_{k|k-1} = Ax_{k-1|k-2} + A\Sigma_{k-1|k-2}C^T[C\Sigma_{k-1|k-2}C^T]^{-1}(y_{k-1} - Cx_{k-1|k-2})$$

   and

$$\Sigma_{k|k-1} = A[\Sigma_{k-1|k-2} - \Sigma_{k-1|k-2}C^T[C\Sigma_{k-1|k-2}C^T]^{-1}C\Sigma_{k-1|k-2}]A^T + BB^T$$

The last two equations are known as the *Kalman Filter* equations. The objective of the Kalman filter is to obtain an optimal estimate of the state at time k, x_k, from measurements of the output up to time $k-1$, $y_0, \ldots, y_{k-1}$. Since you have proven that x_k is a Gaussian random variable, then x_k is completely determined by its mean $x_{k|k-1}$ and its covariance $\Sigma_{k|k-1}$. The equations of the Kalman filter provide a closed form expression for the mean and the covariance. The Kalman filter has been the basic estimation algorithm in almost all real applications involving dynamical systems, e.g., cruise control in airplanes.

5. **Recursive Identification**

Consider the multiple-input multiple-output ARX system

$$y_{k+1} = -a_1 y_k - a_2 y_{k-1} - \cdots - a_n y_{k-n+1} + b_1 u_k + b_2 u_{k-1} + \cdots + b_n u_{k-n+1}. \tag{5}$$

Assume that measurements y_i, u_i are available for $i = 1, \ldots, N$

   (a) Assume that A, B, C, R are real matrices with A and R nonsingular. Prove that

$$(A + BRC)^{-1} = A^{-1} - A^{-1}B(R^{-1} + CA^{-1}B)^{-1}CA^{-1}$$

   The equation above is known as the *matrix inversion lemma*.

(b) Let $\theta = [a_1, \dots, a_n, b_1, \dots, b_n]^T$, $\phi_k = [y_k, y_{k-1}, \dots, y_{k-n+1}, u_k, \dots, u_{k-n+1}]$, $X(N) = [\phi_n^T, \phi_{n+1}^T, \dots, \phi_{N-1}^T]^T$ and $Y(N) = (y_{n+1}^T, y_{n+2}^T, \dots, y_N^T)$. Show that

  (a) $Y(N) = X(N)\theta$

  (b) Let $\hat{y}_{k+1}(\theta) = \phi_k^T\theta$ for all $k \geq n+1$. Define $e(N,\theta) = \sum_{k=n+1}^{N} ||y_k - \hat{y}_k(\theta)||_2^2$. Show that

$$\theta_N = (X(N)^T X(N))^{-1} X^T(N) Y(N)$$

  minimizes $e(N,\theta)$, i.e. for all θ, $e(N,\theta) \geq e(N,\theta_N)$.

(c) Denote by $P_N = (X(N)^T X(N))^{-1}$.

  (a) Show that $P_{N+1}^{-1} = P_N^{-1} + \phi_N^T \phi_N$

  (b) Show that

$$P_{N+1} = P_N - P_N \phi_N^T [\phi_N P_N \phi_N^T + I]^{-1} \phi_N P_N$$

  Hint. Use the matrix inversion lemma.

  (c) Show that

$$\theta_{N+1} = \theta_N + P_{N+1}\phi_N^T (y_{N+1} - \phi_N \theta_N)$$

6. **Modeling and Segmentation of Dynamic Textures**

(a) Use the function `dytex.m` to learn a dynamical model for the sequence of water in the class webpage using different orders for the dynamical model. Use the function `synth.m` to synthesize a new video sequence using the learnt model for different initial conditions. Generate a video with the original sequence and the synthesized sequence for the order and initial condition giving the best results.

(b) Apply GPCA to the ocean-steam sequence in the course webpage to segment the two dynamic textures. Use 4,5,6 principal components and choose the one giving the best results.