CS281A/Stat241A Homework Assignment 5 (due Monday November 16, 2009)

1. **(Kalman filter)** The data in file `hw5-1.data` on the course web site contains noisy measurements of the location of a particle moving in the plane, subject to gravity, random forces, and drag. Each line of the file consists of the measurements $y_t \in \mathbb{R}^2$ of the location at time $t = 1, \ldots, T$. The true location of the particle at time t is $x_t \in \mathbb{R}^2$, and its velocity is $\dot{x}_t \in \mathbb{R}^2$. The equations of motion are

$$\begin{aligned} x_{t+1} &= x_t + \dot{x}_t, \\ \dot{x}_{t+1} &= 0.98\dot{x}_t - 0.02x_t + w_t, \end{aligned}$$

where $w_t \sim N(0, 0.05I_2)$. The observations y_t have the form

$$y_t = x_t + v_t,$$

where $v_t \sim N(0, 100I_2)$. Suppose also that

$$\begin{pmatrix} x_1 \\ \dot{x}_1 \end{pmatrix} \sim N(0, 5I_4).$$

   (a) Plot the particle's true location x_t (from the file `hw5-1.true` on the web site).

   (b) Plot the observations y_t of the particle's position, on top of a plot of the true location.

   (c) Explain how to estimate the particle's initial state (that is, its position x_1 and velocity $\dot{x}_1$ at the initial time $t = 1$) from the noisy measurements $y_1, \ldots, y_T$.

   (d) Calculate the maximum a posteriori probability initial state given the data in `hw5-1.data`.

   (e) For each t, plot the vector $\tilde{x}_t$ of locations that maximize the probability $p(x_t|y_1, \ldots, y_T)$, on top of a plot of the true location. Include in the plot an arrow from $\tilde{x}_1$ in the direction of the MAP initial velocity.

2. Suppose that we have a linear state space model with Gaussian disturbances, and we wish to estimate the initial state x_1 from the noisy observations $y_1, \ldots, y_T$, as in Question 1. It seems reasonable that as T increases, later observations provide less information about the initial state. In this question, we investigate this property.

   Suppose that

$$\begin{aligned} x_t &\in \mathbb{R}^p, & & \\ y_t &\in \mathbb{R}, & & \\ x_{t+1} &= Ax_t + w_t, & w_t &\sim \mathcal{N}(0, Q), \\ y_t &= Cx_t + v_t, & v_t &\sim \mathcal{N}(0, \sigma^2). \end{aligned}$$

   (Notice that $C \in \mathbb{R}^{1\times p}$ is a row vector.)

   (a) What is the conditional distribution of y_T given x_1?

   (b) Suppose that the C vector has unit length ($\|C\| = 1$) and the matrix A is such that for all $v \in \mathbb{R}^p$, $\|A^t v\| \leq \alpha^t \|v\|$, where $\alpha < 1$. Define the $T \times p$ matrix

$$O_T = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{pmatrix}.$$

   (Matrices like O_T are called *observability matrices*: in a deterministic system, the rank of O_p characterizes whether the initial state can be inferred from subsequent observations.) Consider what happens when we start the system in two different initial states, x_1 and $\tilde{x}_1$. Give an upper bound on the KL-divergence between $p(y_T|x_1)$ and $p(y_T|\tilde{x}_1)$ as a function of $\|x_1 - \tilde{x}_1\|$, α, T, σ^2, and $\mathrm{tr}(O_T Q O_T')$.