Math 364 - Assignment #2

Due date: September 15, 2005

1 Textbook Exercises

Cleve Moler's book Chapter 1 pages 43–54:

#1.6
#1.41 (Hints: (a) is an extra credit problem, (e) you can either use primespiral(n,c) and "count" the number of prime numbers in the plot or use the resulting P matrix in [S,P] = primespiral(n,c) and use imagesc(P) for easy counting.).

2 Other Exercises

A) The electricity accounts of residents in a very small town are calculated as follows:

- if 500 units or less are used the cost is 2 cents per unit;
- if more than 500, but not more than 1000 unist are used, the cost is $10 for the first 500 units and then 5 cents for every unit in excess of 500;
- if more than 1000 units are used; the cost is $35 for the first 1000 units plus 10 cents for every unit in excess of 1000;
- in addition, a basci service fee of $5 is charged, no matter how much electricity is used.

Write a program which enters the following five consumptions into a vector, and uses a **for** loop to calculate and display the toatl charge for each one: 200, 500, 700, 1000, 1500. (Answers: $\$9, \$15, \$25, \$40, \$90$)

B) Reproduce the attached 2-pages document using LaTeX.

2.2.4 Probability Distribution Assumptions

Statistical inference in graphical models involves computing conditional probability distributions of the hidden nodes given the values of the observed nodes plus the exogenous variables. A complete description of a time series is provided by the joint distribution function which is hard to find if the random variables are not jointly normally distributed. Fortunately, the Markov properties of the model allow us to express the joint probability of a sequence of T states and outputs replicated N times as the product of the marginals as follows (the superscripts distinguish replicates):

$$\prod_{j=1}^{N} P(\{x\}_j, \{y\}_j) = \prod_{j=1}^{N} \left(P(x_1^{(j)}) \prod_{t=1}^{T-1} P(x_{t+1}^{(j)} | x_t^{(j)}, h_t^{(j)}) \prod_{t=1}^{T} P(y_t^{(j)} | x_t^{(j)}, u_t^{(j)}) \right) \tag{2.29}$$

Statistical inference in graphical models also involves computing marginal probabilities which are required when computing the likelihood. The Gaussian assumptions about the noise and initial distributions define the conditional distribution of the states and the observables (for a given replicate) by:

$$P(x_{t+1}|x_t, u_t) \sim N(Ax_t + Bh_t, Q) \tag{2.30}$$

$$P(y_t|x_t, u_t) \sim N(Cx_t + Du_t, R) \tag{2.31}$$

Here Q, a KxK matrix, and R, a pxp matrix, are the state and observation noise covariances respectively, and they are assumed to be nonsingular. Hence, the multivariate density functions for a given replicate can be written as

$$P(x_{t+1}|x_t, h_t) = \frac{e^{\{-\frac{1}{2}(x_{t+1} - Ax_t - Bh_t)' Q^{-1} (x_{t+1} - Ax_t - Bh_t)\}}}{(2\pi)^{K/2} |Q|^{1/2}} \tag{2.32}$$

$$P(y_t|x_t, u_t) = \frac{e^{\{-\frac{1}{2}(y_t - Cx_t - Du_t)' R^{-1} (y_t - Cx_t - Du_t)\}}}{(2\pi)^{p/2} |R|^{1/2}} \tag{2.33}$$

where $|\cdot|$ denotes the determinant. We also will assume $P(x_1)$ is Gaussian (x_0, Q_1), then

$$P(x_1) = \frac{e^{\{-\frac{1}{2}(x_1 - x_0)' Q_1^{-1} (x_1 - x_0)\}}}{(2\pi)^{K/2} |Q_1|^{1/2}} \tag{2.34}$$

These assumptions form the basis for solving problems involving statistical inference for time series

[91]. Incorporating replicates, the expressions for the double products in (2.29) have the form

$$\prod_{j=1}^{N} P(x_1^{(j)}) = \frac{e^{\{-\frac{1}{2} \quad {}_{j=1}^{N} (x_1^{(j)} - x_0)' Q_1^{-1} (x_1^{(j)} - x_0)\}}}{(2\pi)^{NK/2} |Q_1|^{N/2}}$$

$$\prod_{j=1}^{N} \prod_{t=1}^{T-1} P(x_{t+1}^{(j)} | x_t^{(j)}, h_t^{(j)}) = \frac{e^{\{-\frac{1}{2} \quad {}_{j=1}^{N} \quad {}_{t=1}^{T-1} (x_{t+1}^{(j)} - Ax_t^{(j)} - Bh_t^{(j)})' Q^{-1} (x_{t+1}^{(j)} - Ax_t^{(j)} - Bh_t^{(j)})\}}}{(2\pi)^{N(T-1)K/2} |Q|^{N(T-1)/2}}$$

$$\prod_{j=1}^{N} \prod_{t=1}^{T} P(y_t^{(j)} | x_t^{(j)}, u_t^{(j)}) = \frac{e^{\{-\frac{1}{2} \quad {}_{j=1}^{N} \quad {}_{t=1}^{T} (y_t^{(j)} - Cx_t^{(j)} - Du_t^{(j)})' R^{-1} (y_t^{(j)} - Cx_t^{(j)} - Du_t^{(j)})\}}}{(2\pi)^{NTp/2} |R|^{NT/2}}$$

In the application of the EM algorithm (see next section), we will need to maximize the expected log-likelihood (conditional on the entire observation sequence), or minimize minus twice the expected log likelihood. From the joint probability equation (2.36), this is equivalent to the minimization of:

$$\begin{aligned} -2 \sum_{j=1}^{N} \log P(\{x\}_j, \{y\}_j) &= \sum_{j=1}^{N} (x_1^{(j)} - x_0)' Q_1^{-1} (x_1^{(j)} - x_0) \\ &+ \sum_{j=1}^{N} \sum_{t=1}^{T-1} (x_{t+1}^{(j)} - Ax_t^{(j)} - Bh_t^{(j)})' Q^{-1} (x_{t+1}^{(j)} - Ax_t^{(j)} - Bh_t^{(j)}) \\ &+ \sum_{j=1}^{N} \sum_{t=1}^{T} (y_t^{(j)} - Cx_t^{(j)} - Du_t^{(j)})' R^{-1} (y_t^{(j)} - Cx_t^{(j)} - Du_t^{(j)}) \\ &+ N(T-1) \log |Q| + NT \log |R| \\ &+ NT(p+K) \log(2\pi) + N \log |Q_1| \end{aligned}$$

Now, taking the conditional expectation given all the observables up to time T of the expression above we have that the term on the right hand side expressed as a function $\mathcal{W}$ of the parameters is