CS 322 Homework 8

out: Thursday 26 April 2007
due: Fri 4 May 2007

In this homework we'll look at propagation of error in systems with $n \times n$ Hilbert matrices $\mathbf{H}_n$, which look like this:

$$\mathbf{H}_4 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix}$$

These matrices are ill-conditioned and get worse quickly as n increases.

Relevant Matlab functions for this homework include `hilb`, `cond`, `chi2pdf`, `chi2cdf`, and `chi2inv`.

Problem 1: Conditioning of linear systems.

1. One can study the propagation of error from the right-hand side to the solution in a linear system using a brute-force Monte Carlo experiment: add random errors artificially to the right hand side and see how much error they induce in the solution.

   Using the following procedure, estimate the condition numbers of $\mathbf{H}_2$ and $\mathbf{H}_5$.

   - Choose 10,000 random right-hand sides for the system $\mathbf{H}_n\mathbf{x} = \mathbf{b}$.
   - Perturb each right-hand side by a normally distributed random error with standard deviation 10^{-6}.
   - Solve the original and perturbed systems and compare the results.
   - Compare the relative errors in the right-hand sides and the solutions.

   Illustrate your work with a short Matlab transcript.

   For H_2 you should be able to estimate the true condition number pretty well. How good is your estimate for H_5 and why?

   Hint: You can do this experiment in about 5 Matlab statements.

2. Use the SVD to find a right hand side and perturbation that demonstrate error growth that achieves the bound given by the condition number.

*3. Repeat question 1, but this time perturb the matrix instead of the right-hand side. Does the condition number still bound the growth of error when it propagates from the matrix to the solution?

*4. Repeat question 2, but this time work out the worst-case perturbation to the matrix.

Problem 2: The χ^2 distribution.

1. Use Monte Carlo to estimate the probability that the norm of a point $\mathbf{x}$ drawn from a 3D Gaussian distribution with unit standard deviation has a norm greater than 2. How many samples do you need to get 3 significant figures of accuracy?
2. Use Matlab's built in χ^2 functions to compute the same number.
3. Let $\mathbf{y} = \mathbf{H}_3\mathbf{x}$. Give a 90% confidence region for $\mathbf{y}$ (specified as the axes of the ellipse) and for its first component, y_1.

Problem 3: Confidence in least squares fits.

In this problem, no fair using Monte Carlo except where specified. Specify all confidence ellipses as the axes of the ellipse.

1. The file `hw08data1.txt` contains 20 (x, y) points in which all the ys have independent Gaussian noise with $\sigma = 0.05$.
   (a) Fit the model $y = a_1 x + a_2$ to the data using linear least squares.
   (b) Give 90% confidence intervals for a_1 and a_2. Plot the lines corresponding to the four extreme combinations of parameters against the data points.
   (c) Confirm your intervals using Monte Carlo.
   (d) Give a 90% confidence ellipse for (a_1, a_2) together. Plot the lines corresponding to the ends of the principal axes of the ellipse against the data points.

*2. The file `hw08data1.txt` contains 20 (x, y) points in which all the ys have independent Gaussian noise with $\sigma = 0.01$.
   (a) Fit the model $y = a_1 e^{a_2 x}$ to the data using linear least squares.
   (b) Give 90% confidence intervals for a_1 and a_2. Plot the curves corresponding to the four extreme combinations of parameters against the data points.
   (c) Give a 90% confidence ellipse for (a_1, a_2) together. Plot the curves corresponding to the ends of the principal axes of the ellipse against the data points.
   (d) Find the 90% confidence ellipse using a Monte Carlo simulation. Why might the result be different from the previous part?