CS 370 Winter 2018: Assignment 4

Due April 3, 5 pm.

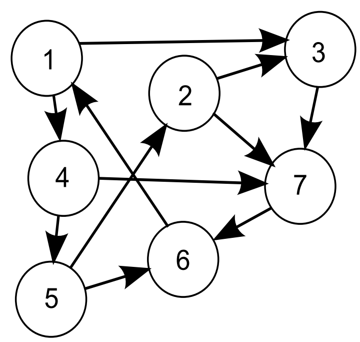
Instructor: G. Labahn   Office: DC3629   e-mail: *glabahn@uwaterloo.ca*
Lectures: MWF 8:30, 11:30   Office Hours: Tues 11:00-12:00

Instructor: Y. Li   Office: DC3623   e-mail: *yuying@uwaterloo.ca*
Lectures: MWF 1:30   Office Hours: Thurs 2:00-3:00

Web Site: cs370 piazza

Your assignment should be handed in electronically on UW Learn. The submission should include one pdf file containing the assignment answers and the cover sheet and all the m-files required to run the code, with no folder structure (not zipped).

1. **(20 marks)** Consider the small web given by



   (a) Construct the Google matrix M for this web.
   (b) Run the PageRank algorithm for 15 iterations to find a ranking vector $\vec{x}$. Use $\alpha = 0.85$.
   (c) Verify that the ranking vector $\vec{x}$ satisfies $M\vec{x} = \vec{x}$, up to at least 3 significant digits.

   You can use either Maple or Matlab for this question.

2. **(10 marks)** A matrix $Q = [q_{ij}]$ is a positive Markov matrix if $0 < q_{ij} < 1$ and $\sum_i q_{ij} = 1$. Show that the Google matrix

$$M = \alpha(P + \frac{1}{R}\mathbf{e}\ \mathbf{d}^T) + (1-\alpha)\frac{1}{R}\mathbf{e}\ \mathbf{e}^T$$

   is a positive Markov matrix.

3. **(10 marks)** Suppose a square, $n \times n$ nonsingular matrix A has already been factored

$$A = LU$$

   where L is unit lower triangular and U is upper triangular. Show how to use this factorization to give a quadratic time procedure for solving

$$AA^T \cdot \vec{x} = \vec{b}$$

4. **(10 marks)** Find the $PA = LU$ factorization by hand using row pivoting with maximal pivot for the following matrix:

$$A = \begin{bmatrix} 2.0 & 2.0 & 2.0 \\ -4.0 & -2.0 & 6.0 \\ 2.0 & -1.0 & 4.0 \end{bmatrix}.$$

Use this factorization to solve $Ax = b$ where

$$b = \begin{bmatrix} 2.0 \\ 4.0 \\ 7.0 \end{bmatrix}.$$

5. **(20 marks)** Computing a natural spline with interpolation points having Δx_i equal for all i requires solving a linear system $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 2 & 1 & 0 & & \cdots & 0 & 0 \\ 1 & 4 & 1 & & & 0 & 0 \\ 0 & 1 & 4 & 1 & & 0 & 0 \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & & 1 & 4 & 1 \\ 0 & 0 & 0 & & & 1 & 2 \end{bmatrix}_{n\times n}$$

and $\vec{x}$ is the vector of derivatives of the spline. In this question you will investigate the solution of the linear system and determine its cost.

(a) The factors of $A = LU$ have the form

$$L = \begin{bmatrix} 1 & 0 & & \cdots & 0 & 0 \\ l_2 & 1 & 0 & & & 0 \\ 0 & l_3 & 1 & & & \\ & & \ddots & \ddots & & \\ 0 & 0 & & & 1 & 0 \\ 0 & 0 & & & l_n & 1 \end{bmatrix} \qquad U = \begin{bmatrix} d_1 & 1 & & \cdots & 0 & 0 \\ 0 & d_2 & 1 & & & 0 \\ 0 & 0 & d_3 & 1 & & \\ & & & \ddots & \ddots & \\ 0 & 0 & & & d_{n-1} & 1 \\ 0 & 0 & & & 0 & d_n \end{bmatrix}$$

Find the recurrence equations which determine l_i and d_i.

(b) What is the cost (that is, $O(n^k)$, for some k) of finding an LU decomposition of A? Justify your answer.

(c) What would the cost be to compute a natural spline when all the Δx_i are the same? Justify your answer.

6. **(30 marks)**

(a) Write a MATLAB function to compute the page rank of a webgraph G.

```
function [p, iter] = PageRank(G, alpha)
```

which determines the pagerank for a network using the iterative method described in the course notes. The inputs are the (sparse) adjacency matrix G, representing the directed graph of the network, and the weight alpha. (The adjacency matrix is defined as G(i; j) = 1 if there is a link from page j pointing to page i, otherwise G(i; j) = 0.) The output is the vector `p` of page ranks, and `iter` is the number of iterations that were required for the computation. Use a tolerance value of 10^{-7}.

Your function must take advantage of the sparsity of G. Avoid using additional loops (within the iteration loop) or creating full matrices (see Sec. 7.6 in the course notes).

(b) Repeat question 1 but using the above tolerance value as the stopping criterion. That is, write a MATLAB script to construct the adjacency matrix G and compute the pageranks for the web shown above using your PageRank function with $\alpha = 0.85$. Use the bar command to plot the pagerank scores. Use the **spy** command to graph the sparsity pattern of G. Your plots should have labels and titles. What is the order of importance of the nodes according to your results?

(c) A connectivity matrix G and a list of URLs U are provided in *uwaterloo.mat*. The data represents a network of 500 pages and was generated starting from the website www.math.uwaterloo.ca. Write a script to load the data and compute the pageranks, with $\alpha = 0.85$. As in part (b) use the **spy** command to graph the sparsity pattern of G.

Use the following code to obtain the final ranking order and list the top twenty results.

```
[y I] = sort(p, 'descend');
for n = 1:min(length(I),20)
    disp([num2str(n) ': ' U{I(n)}]);
end
```

(d) Experiment with the *uwaterloo.mat* data using the following varied values of α namely: $\{0.15, 0.35, 0.55, 0.75, 0.95\}$. Report the number of iterations in each case. What do you notice about the relationship between α and the number of iterations? Explain.