CS412: Introduction to Numerical Methods

MIDTERM #2 — 2:30PM - 3:45PM, Thursday, 04/23/2015

Instructions: This exam is a **closed book** and **closed notes** exam, i.e., you are not allowed to consult any textbook, your class notes, homeworks, or any of the handouts from us. You are not permitted to refer to any other material either (including, of course, online material). No use of computers, cell phones, etc. is permitted.

Name	
University ID	

Part #1	
Part #2	
Part #3	
Part #4	
Part #5	
TOTAL	

1. [24% = 4 questions × 6% each] MULTIPLE CHOICE SECTION. Circle or underline the correct answer (or answers). No justification is required for your answer(s).

   (a) Which of the following statements regarding the cost of methods for solving an $n \times n$ linear system $Ax = b$ are true?

      i. The cost of computing the LU factorization is generally proportional to n^2.

      ii. The cost of backward substitution on a dense upper triangular matrix is generally proportional to n^2.

      iii. If a matrix A has no more than 3 non-zero entries per row, the cost of each iteration of the Jacobi method is proportional to n.

   (b) Which of the following statements regarding numerical methods are true?

      i. If the local error of an integration rule scales like $O(h^d)$, the global error will be $O(h^{d+1})$.

      ii. With a second order accurate rule, if we increase the number of points in the integration rule by 10, we should expect the error to decrease by approximately a factor of 20.

      iii. If a method computes the integral of polynomials up to order d exactly, then the global error is on the order of $O(h^{d+1})$.

   (c) Which of the following statements regarding methods for solving initial value problems are true?

      i. Every step of an explicit method is very inexpensive, but we may need to keep the maximum time step Δt small to obtain a reasonable solution.

      ii. If a differential equation has unstable solutions, using an implicit method will guarantee convergence to the correct solutions, where explicit methods would diverge away from the real solution.

      iii. Implicit methods can be used to solve systems of ordinary differential equations, while explicit methods only work with individual differential equations (with just one unknown function).

   (d) Which of the following methods can be used for solving the system $Ax = b$, where A is a square $n \times n$ matrix?

      i. LU factorization with full pivoting.

      ii. QR factorization.

      iii. System of normal equations.

      iv. Gauss-Seidel method.

      v. Jacobi method.

2. [18% = 3 questions × 6% each] SHORT ANSWER SECTION. Answer each of the following questions in no more than 2-3 sentences.

(a) Consider the following matrix A whose LU factorization we wish to compute using Gaussian elimination:

$$A = \begin{bmatrix} 4 & -8 & 1 \\ 6 & 5 & 7 \\ 0 & -10 & -3 \end{bmatrix}$$

What will be the initial pivot element if (no explanation required)

- No pivoting is used?
  Answer: 4
- Partial pivoting is used?
  Answer: 6
- Full pivoting is used?
  Answer: -10

(b) State one defining property of a *singular* matrix A. Suppose that the linear system $Ax = b$ has two distinct solutions x and y. Use the property you gave to prove that A must be singular.

Answer: A matrix A is singular when the system $Ax = 0$ has a solution other than $x = 0$. If $Ax = b$ and $Ay = b$, subtracting the two equations gives $A(x - y) = 0$. Since x and y are two distinct vectors, $x - y \neq 0$, implying that the matrix is singular.

(c) Mention one advantage of the Gauss-Seidel algorithm over the Jacobi algorithm and one disadvantage.

Answer:

- Advantage: Gauss-Seidel algorithm has a faster convergence rate compared to the Jacobi algorithm.
- Disadvantage: Gauss-Seidel algorithm is difficult to parallelize compared to the Jacobi algorithm.

3. [10%] Consider the function $f(x) = x^2$ in the interval $[a, a+nh]$. Derive the error when numerically integrating f with interval spacing h using the rectangle rule. Assuming that $nh = \mathsf{constant}$ (the length of the interval), comment on the order of acccuracy. Use the equalities $\sum_{i=0}^{n-1} i = n(n-1)/2$ and $\sum_{i=0}^{n-1} i^2 = n(n-1)(2n-1)/6$.

Answer: Consider the interval $[a+ih, a+(i+1)h]$. The rectangle rule gives the following approximation for the intergral:

$$I_i = (a+ih)^2 \cdot h = a^2h + 2ah^2i + h^3i^2$$

Hence, the value of the computed integral is:

$$I = \sum_{i=0}^{n-1} I_i = \sum_{i=0}^{n-1}(a^2h + 2ah^2i + h^3i^2) = a^2h\sum_{i=0}^{n-1} 1 + 2ah^2\sum_{i=0}^{n-1} i + h^3\sum_{i=0}^{n-1} i^2$$

Using the given equalities gives,

$$\begin{aligned} I &= a^2nh + 2ah^2 \cdot \frac{n(n-1)}{2} + h^3 \cdot \frac{n(n-1)(2n-1)}{6} \\ &= a^2nh + ah^2n^2 - ah^2n + \frac{h^3n(n-1)(2n-1)}{6} \end{aligned}$$

The exact answer is:

$$I_{\mathsf{analytic}} = \frac{(a+nh)^3}{3} - \frac{a^3}{3} = a^2nh + an^2h^2 + \frac{n^3h^3}{3}$$

Thus, the error is:

$$I - I_{\mathsf{analytic}} = -ah^2n + \frac{h^3n(n-1)(2n-1)}{6} - \frac{n^3h^3}{3} = -ah^2n + \frac{nh^3}{6} - \frac{n^2h^3}{2}$$

Since $nh = \mathsf{constant}$, the error becomes

$$I - I_{\mathsf{analytic}} = O(h) + O(h^2) = O(h)$$

4. [22%] Let A be a rectangular $m \times n$ matrix with linearly independent columns. The *nullspace* of A (denoted as $\mathsf{Null}(A)$) is defined as the set of vectors x, such that $Ax = 0$. A fundamental theorem in linear algebra states that any vector $x \in \mathbb{R}^m$ can be written as $x = x_1 + x_2$, where $x_1 \in \mathsf{Null}(A^T)$ and x_2 lies in the column space of A, i.e., there exists a vector y such that $x_2 = Ay$. Consider the reduced QR decomposition of A.

   (a) [10%] Show that $P_0 = I - QQ^T$ is the *projection matrix* onto the nullspace of A^T, i.e., $P_0 x \in \mathsf{Null}(A^T)$ for all $x \in \mathbb{R}^m$.

   (b) [6%] Show that for every $x \in \mathbb{R}^n$, we have

$$||Ax - b||_2^2 = ||A(x - x_0)||_2^2 + ||Ax_0 - b||_2^2$$

   where x_0 is the least squares solution of $Ax = b$.

   (c) [6%] Show that the minimum value for the 2-norm of the residual is $||P_0 b||_2$ and is attained when x is equal to the least squares solution.

Answer:

(a) From the fundamental theorem, we know that any vector $x \in \mathbb{R}^m$ can be written as $x = x_1 + x_2$, where x_1 is in the nullspace of A^T and x_2 is in the column space of A. For x_1, this means that $A^T x_1 = 0$. Since the columns of A are linearly independent, the reduced QR decomposition is defined and

$$A^T x_1 = 0 \Rightarrow R^T Q^T x_1 = 0 \Rightarrow Q^T x_1 = 0$$

since R is nonsingular. On the other hand, x_2 belongs to the column space of A, therefore it can be written as $x_2 = Ay = QRy$, where $y \in \mathbb{R}^n$. Thus, the action of P_0 on x amounts to

$$P_0 x = (I - QQ^T)(x_1 + x_2) = x_1 + x_2 - QQ^T QRy = x_1 + x_2 - QRy = x_1$$

This implies that P_0 is the projection matrix onto the nullspace of A^T.

(b) We have

$$\begin{aligned}
||Ax - b||_2^2 &= ||A(x - x_0) + (Ax_0 - b)||_2^2 \\
&= [A(x - x_0) + (Ax_0 - b)]^T [A(x - x_0) + (Ax_0 - b)] \\
&= ||A(x - x_0)||_2^2 + ||Ax_0 - b||_2^2 + 2(x - x_0)^T A^T (Ax_0 - b) \\
&= ||A(x - x_0)||_2^2 + ||Ax_0 - b||_2^2 + 2(x - x_0)^T (A^T A x_0 - A^T b) \\
&= ||A(x - x_0)||_2^2 + ||Ax_0 - b||_2^2
\end{aligned}$$

The last equality follows because x_0 satisfies the normal equations.

(c) From the above equation, we have that the minimum value for $||Ax - b||_2$ is attained for $x = x_0$, since the term $||Ax_0 - b||_2$ does not depend on x. The least squares solution is given as $Rx_0 = Q^T b$. Thus,

$$\begin{aligned} ||Ax_0 - b||_2 &= ||QRR^{-1}Q^T b - b||_2 = ||QQ^T b - b||_2 = ||(QQ^T - I)b||_2 \\ &= ||(I - QQ^T)b||_2 = ||P_0 b||_2 \end{aligned}$$

Intuitively, this means that the least squares solution annihilates the component of the residual in the column space of A and the minimum value for the residual is exactly the component of b that is *not* contained in the column space of A.

5. [26%] Consider the elimination matrix $M_k = I - m_k e_k^T$ and its inverse $L_k = I + m_k e_k^T$ used in the LU decomposition process, where

$$m_k = (0, \dots, 0, m_{k+1}^{(k)}, \dots, m_n^{(k)})^T$$

and e_k is the kth column of the identity matrix. Let $P^{(ij)}$ be the permutation matrix that results from swapping the i-th and j-th rows (or columns) of the identity matrix.

(a) [6%] Show that if $i, j > k$ then $L_k P^{(ij)} = P^{(ij)}(I + P^{(ij)} m_k e_k^T)$.

(b) [10%] Recall that the matrix L resulting from performing Gaussian elimination with partial pivoting is given by

$$L = P_1 L_1 \dots P_{n-1} L_{n-1}$$

where the permutation matrix P_i permutes row i with some row i' where $i < i'$. Show that L can be rewritten as

$$L = P_1 \dots P_{n-1} L_1^P \dots L_{n-1}^P$$

where $L_k^P = I + (P_{n-1} \dots P_{k+1} m_k) e_k^T$.

(c) [10%] Show that $L_1^P \dots L_{n-1}^P$ is lower triangular.

Answer:

(a) The matrix $m_k e_k^T$ has non-zero elements only on the kth column, in the positions corresponding to rows $(k+1)$ through n. Additionally, $m_k e_k^T P^{(ij)}$ is the result of swapping the ith and jth columns of $m_k e_k^T$, which are both zero. Thus, $m_k e_k^T P^{(ij)} = m_k e_k^T$. Using this result, we have

$$\begin{aligned} (I + m_k e_k^T) P^{(ij)} &= P^{(ij)} + m_k e_k^T P^{(ij)} \\ &= P^{(ij)} + m_k e_k^T \\ &= P^{(ij)} + P^{(ij)} P^{(ij)} m_k e_k^T \\ &= P^{(ij)}(I + P^{(ij)} m_k e_k^T) \end{aligned}$$

The third equality follows because $P^{(ij)} P^{(ij)} = I$, the identity matrix.

(b) Let q_k be a vector containing non-zero entries only in the positions $(k+1)$ through n. Then, using part (a), we have

$$(I + q_k e_k^T) P_i = P_i (I + P_i q_k e_k^T) = P_i (I + \hat{q}_k e_k^T)$$

where the vector $\hat{q}_k = P_i q_k$ also has non-zero entries in the positions $(k+1)$ through n. Consequently, in the product $P_1 L_1 \dots P_{n-1} L_{n-1}$, we can "propagate" each permutation matrix P_i (in increasing order of the index i) to the left of all matrices L_k with $k \le i$ while changing each matrix L_k according to the equation above (multiplying its second term with P_i from the left). For example,

$$\begin{aligned}
P_1L_1P_2L_2P_3L_3 &= P_1(I+m_1e_1^T)P_2(I+m_2e_2^T)P_3(I+m_3e_3^T)\\
&= P_1P_2(I+P_2m_1e_1^T)(I+m_2e_2^T)P_3(I+m_3e_3^T)\\
&= P_1P_2(I+P_2m_1e_1^T)P_3(I+P_3m_2e_2^T)(I+m_3e_3^T)\\
&= P_1P_2P_3(I+P_3P_2m_1e_1^T)(I+P_3m_2e_2^T)(I+m_3e_3^T)\\
&= P_1P_2P_3L_1^PL_2^PL_3^P
\end{aligned}$$

where $L_k^P = I + (P_{n-1}\dots P_{k+1}m_k)e_k^T$. For the complete solution, this argument should be extended to an arbitrary n via mathematical induction.

(c) Each matrix L_k^P can be written as $L_k^P = I + \hat{q}_k e_k^T$, where $\hat{q}_k = P_{n-1}\dots P_{k+1}m_k$, like m_k, only has non-zero entries in the positions $(k+1)$ through n. Furthermore,

$$\begin{aligned}
L_1^PL_2^P\dots L_{n-1}^P &= (I+\hat{q}_1e_1^T)(I+\hat{q}_2e_2^T)\dots(I+\hat{q}_{n-1}e_{n-1}^T)\\
&= I+\hat{q}_1e_1^T+\hat{q}_2e_2^T+\dots+\hat{q}_{n-1}e_{n-1}^T
\end{aligned}$$

since $e_i^T\hat{q}_j = 0$ for $i < j$, causing all the cross-terms $(\hat{q}_ie_i^T)(\hat{q}_je_j^T)$ in the original product to vanish (for $i < j$). Since each term $\hat{q}_ie_i^T$ contributes non-zero entries only below the diagonal, the entire matrix $I+\hat{q}_1e_1^T+\hat{q}_2e_2^T+\dots+\hat{q}_{n-1}e_{n-1}^T$ is lower triangular.