CS205 Homework #1

Problem 1

Arithmetic operations are subject to roundoff error when performed on a finite precision computer. In order to perform an operation $x\ op\ y$ on the *real* numbers x and y we deviate from the analytic result when discretizing those values to machine precision as well as when we store the resulting value.

Let $\bar{x}$ denote the discretized, floating point version of x that is stored on the computer. You may assume that

$$\bar{x} = (1+\epsilon)x$$

where ϵ is bounded as $0 \leq |\epsilon| < \epsilon_{\max}$ where $\epsilon_{\max} \ll 1$ is the machine roundoff precision.

Assume that the result of the arithmetic operation between two floating point numbers $\bar{x}$ and $\bar{y}$ is computed exactly, but when stored on the computer it is once again subject to roundoff error as

$$\overline{\bar{x}\ op\ \bar{y}} = (1+\epsilon')(\bar{x}\ op\ \bar{y})$$

where the roundoff error obeys the same bounds $0 \leq |\epsilon'| < \epsilon_{\max}$.

The relative error of a computation is defined as

$$E = \left|\frac{\text{Computed\_Result} - \text{Analytic\_Result}}{\text{Analytic\_Result}}\right|$$

Provide a bound (in terms of $\epsilon_{\max}$) for the relative error induced by the following arithmetic operations, or prove that the relative error is unbounded.

1. Subtraction, Multiplication and Division of two real numbers (for an example on addition see Heath, section 1.3.8)

2. Computing the sum $s_n = \underbrace{x + x + \cdots + x}_{n \text{ terms}}$ using the recurrence

$$s_1 = x$$
$$s_k = s_{k-1} + x$$

[Answer: $\approx n\epsilon_{\max}/2$]

3. Computing the sum $s_n = s_{2^k} = q_k = \underbrace{x + x + \cdots + x}_{n=2^k \text{ terms}}$ where $n = 2^k$ using the recurrence

$$q_0 = x$$
$$q_k = q_{k-1} + q_{k-1}$$

For (2) and (3) you may assume for simplicity that $n \ll 1/\epsilon_{\max}$.

Problem 2

Consider the elimination matrix $\mathbf{M_k} = \mathbf{I} - \mathbf{m_k e_k^T}$ and its inverse $\mathbf{L_k} = \mathbf{I} + \mathbf{m_k e_k^T}$ used in the LU decomposition process, where

$$\mathbf{m_k} = \left(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{m_{k+1}^{(k)}}, \ldots, \mathbf{m_n^{(k)}}\right)^{\mathbf{T}}$$

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

1. Show that if $i, j > k$ then $\mathbf{L_k P^{(ij)}} = \mathbf{P^{(ij)}}(\mathbf{I} + \mathbf{P^{(ij)} m_k e_k^T})$

2. Recall that the matrix $\mathbf{L}$ resulting from performing Gaussian elimination with partial pivoting is given by

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

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

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

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

3. Show that $\mathbf{L_1^P} \cdots \mathbf{L_{n-1}^P}$ is lower triangular.

Problem 3

Two vector norms $\|\mathbf{x}\|_a$ and $\|\mathbf{x}\|_b$ are called equivalent if there exist $c, d > 0$ such that $c\|\mathbf{x}\|_a \le \|\mathbf{x}\|_b \le d\|\mathbf{x}\|_a$.

1. Prove that $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ are equivalent.

2. Prove that equivalence of two vector norms implies that their induced matrix norms are also equivalent. (The definition for equivalence of matrix norms is analogous to that of vector norms, i.e there must exist $c, d > 0$ s.t. $c\|\mathbf{A}\|_a \le \|\mathbf{A}\|_b \le d\|\mathbf{A}\|_a$)