CSC349A Numerical Analysis

Lecture 13

Rich Little

University of Victoria

2023

Table of Contents I

Midterm Logistics

- The midterm is on Thursday, November 9 and is 60 minutes long
- The exam is closed book (see below regarding formula sheet)
- Only simple, scientific calculators (the ones you use for math classes) are allowed. If you bring anything programmable or with a large screen and or internet access you will not be allowed to use it.
- You can bring a single letter size (8.5 by 11) piece of paper with formulas and notes (it can be double sided)

Midterm Material

- The material covered corresponds to the end of chapter 4 to chapter 7 and the beginning of chapter 9.
- That is lectures 7 to 12 and Handouts 6 to 15.
- In terms of topics these are condition, and stability (part 1).
- Roots of equations (Bisection, Newton and Secant) and rates of convergence, Horner's method (part 2).
- Naive Gaussian Elimination (part 3).
- In addition you should study all the assignments you have completed and the corresponding problems from the sample exam questions.

Note: There is no MATLAB on the exam.

Table of Contents I

Gaussian Elimination with Partial Pivoting

- The Naive Gaussian elimination algorithm will fail if any of the pivots $a_{11}, a_{22}^{(1)}, a_{33}^{(2)}, \ldots$ is equal to 0.
- Mathematically, it works provided this does not occur.
- Algorithmically, it breaks down when the pivots are even close to 0 because of floating-point arithmetic.
- The problem occurs in the multiplier, it becomes far larger than the other entries.

Example 1

Solve the $n = 2$ linear system with the following augmented matrix using $k = 4$, $b = 10$, rounding, floating-point arithmetic.

$$\left[\begin{array}{cc|c} 0.003 & 59.14 & 59.17 \\ 5.291 & -6.13 & 46.78 \end{array}\right]$$

Example 1 continued

Analysis of the above Example

- The source of the extremely inaccurate computed solution $\hat{x}$ is the **large magnitude of the multiplier**.
- Here, 1764 is much larger than the rest of the numbers in the system.
- This number is large because the pivot, $a_{11} = 0.003$, is much smaller than the other numbers in the system.
- Consequently, in the floating-point computations of $a_{22}^{(1)}$ and $b_2^{(1)}$, the numbers -6.13 and 46.78 are so small they are lost.
- The **partial pivoting strategy** is designed to avoid the selection of small pivots.

Partial Pivoting

- At step k of forward elimination, where $1 \leq k \leq n-1$, choose the pivot to be the **largest entry in absolute value**, from
$$\begin{bmatrix} a_{kk} \\ a_{k+1,k} \\ a_{k+2,k} \\ \vdots \\ a_{n,k} \end{bmatrix}$$
- If a_{pk} is the largest (that is, $|a_{pk}| = \max_{k \leq i \leq n} |a_{ik}|$), then switch row k with row p.
- Note that $|mult| \leq 1$ for all multipliers since the denominator is always the largest value.
- Note also that switching rows does not change the final solution. It is an elementary row operation of type 3.

Partial Pivoting Pseudocode

Algorithm 1 pseudocode for partial pivoting

```
for k = 1 to n − 1 do
    p = k
    for i = k + 1 to n do
        Find largest pivot
    end for
    if p ≠ k then
        for j = k to n do
            swap a_kj and a_pj
        end for
        swap b_k and b_p
    end if
    do forward elimination
end for
do back susbstitution
```

Example 2 - Pivoting

Solve the $n = 2$ linear system with the following augmented matrix using $k = 4$, $b = 10$, rounding, floating-point arithmetic with partial pivoting.

$$\left[\begin{array}{cc|c} 0.003 & 59.14 & 59.17 \\ 5.291 & -6.13 & 46.78 \end{array}\right]$$

Example 2 - Pivoting continued

Table of Contents I

Scaling

- Section 9.4.3 on page 270 in the 8th edition of the text.
- Nothing in the Handouts on this topic.
- If the entries of maximum absolute value in different rows (equations) differ greatly, the computed solution (using floating point arithmetic and partial pivoting) can be very inaccurate.

Example 3 - Scaling

Using $k = 4$ precision, floating-point arithmetic with rounding, solve the following system by Gaussian Elimination with partial pivoting.

$$\left[\begin{array}{cc|c} 2 & 100,000 & 100,000 \\ 1 & 1 & 2 \end{array}\right]$$

Example 3 - Scaling continued

Scaling: Equilibration

We look at two ways of using **scaling** to solve this problem: (1) Equilibration and (2) Scaled Factors.

(1) Equilibration:

- Multiply each row by a nonzero constant so that the largest entry in each row of A has magnitude of 1.
- Go through example again with scaling.
- Problem with this form of scaling:
  - Introduces another source of round-off error.

Scaling: Scaled Factors

(2) Scaled Factors:

- Use the scaling factors to pick pivots but NOT actually scaling.
- Let $s_i = \max_{1 \leq j \leq n} |a_{ij}|$ for $i = 1, 2, \ldots, n$.
- Step $k = 1$: pivot is max of

$$\begin{bmatrix} |a_{11}/s_1| \\ |a_{21}/s_2| \\ \vdots \\ |a_{n1}/s_n| \end{bmatrix}$$

If $|a_{p1}/s_p|$ is the max then interchange rows 1 and p then do forward elimination step.

Scaling: Scaled Factors II

- Step $k = 2$: pivot is max of

$$\begin{bmatrix} |a_{22}^{(1)}/s_2| \\ |a_{32}^{(1)}/s_3| \\ \vdots \\ |a_{n2}^{(1)}/s_n| \end{bmatrix}$$

  If $|a_{q2}/s_q|$ is the max then interchange rows 2 and q then do forward elimination step.

- etc.
- Finish with back susbstitution as usual.

Example 4 - Scaled Factors

Using $k = 4$ precision, floating-point arithmetic with rounding, solve the following system by Gaussian Elimination with partial pivoting and scaling.

$$\left[\begin{array}{cc|c} 2 & 100,000 & 100,000 \\ 1 & 1 & 2 \end{array}\right]$$

Example 4 - continued

Table of Contents I

Determinant of A

The reduction of A to upper triangular form by **Naive Gaussian elimination** uses only the type 2 elementary row operation

$$E_i = E_i - factor \times E_j.$$

This row operation does not change the value of the determinant of A. That is, if no rows are interchanged then,

$$\det A = a_{11} a_{22}^{(1)} a_{33}^{(2)} \cdots a_{nn}^{(n-1)}$$

since the determinant of a triangular matrix is equal to the product of its diagonal entries.

Determinant of A II

However, if **Gaussian elimination with partial pivoting** is used, then each row interchange causes the determinant to change signs (that is, determinant is multiplied by -1.) Thus, if m row interchanges are done during the reduction of A to upper triangular form, then

$$\det A = (-1)^m a_{11} a_{22}^{(1)} a_{33}^{(2)} \cdots a_{nn}^{(n-1)}$$

As a consequence, Gaussian elimination provides us with a simple method of calculating the determinant of a matrix.

Table of Contents I

Stability of Algorithms for Solving $Ax = b$

- Given a nonsingular matrix A, a vector b and some algorithm for computing the solution of $Ax = b$, let $\hat{x}$ denote the computed solution using this algorithm.
- The computation is said to be stable if there exist small perturbations E and e of A and b, respectively, such that $\hat{x}$ is close to the exact solution y of the perturbed linear system

$$(A + E)y = b + e$$

- That is, the computed solution $\hat{x}$ is very close to the exact solution of some small perturbation of the given problem.

Known Results

- Gaussian elimination without pivoting may be unstable.
- In practice, Gaussian elimination with partial pivoting is almost always stable.
- A much more stable version of Gaussian elimination uses complete pivoting, which uses both row and column interchanges.
- However, as this algorithm is much more expensive to implement and since partial pivoting is almost always stable, complete pivoting is seldom used.

Condition of $Ax = b$

- A given problem $Ax = b$ is ill-conditioned if its exact solution is very sensitive to small changes in the data $[A|b]$.
- That is, if there exist small perturbations E and e of A and b, respectively, such that $x = A^{-1}b$ is not close to the exact solution y of the perturbed linear system

$$(A + E)y = b + e,$$

  then the linear system $Ax = b$ is ill-conditioned.
- If such perturbations E and e do not exist, then $Ax = b$ is well conditioned.
- **Example:** $n \times n$ Hilbert matrices are ill-conditioned.

Example 2 - Condition of a linear system

Recall, the linear system $Hx = b$, with $H = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$,

$b = \begin{bmatrix} 11/6 \\ 13/12 \\ 47/60 \end{bmatrix}$ and solution $x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, is ill-conditioned.

We do this by perturbing H and b as follows. Let

$(H+E) = \begin{bmatrix} 1 & 1/2 & 0.333 \\ 1/2 & 0.333 & 1/4 \\ 0.333 & 1/4 & 1/5 \end{bmatrix}$, $(b+e) = \begin{bmatrix} 1.83 \\ 1.08 \\ 0.783 \end{bmatrix}$, and

solving to get $y = \begin{bmatrix} 1.0895... \\ 0.48796... \\ 1.4910.... \end{bmatrix}$

Matrix Norms

- The *norm* of a matrix (or vector) is a measure of the "size" of the matrix.
- We denote the norm of a matrix A by $\|A\|$.
- There exist a number of different ways of calculating a norm.
  - $\|A\|_e = \sqrt{\sum_i \sum_j a_{ij}^2}$ is the Euclidean norm.
  - $\|A\|_{\text{inf}} = \max_i \sum_j |a_{ij}|$ is the uniform norm, etc.
- Turns out, for any of these forms of the norm, when solving for $Ax = b$,

$$\frac{\|x - y\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|e\|}{\|b\|}$$

Matrix Condition Number

- The condition number of a matrix, A, is given by

$$cond[A] = \|A\|\|A^{-1}\|$$

- Properties of the condition number.
  - $cond[A] \geq 1$
  - $cond[I] = 1$
- As usual, the higher the condition number the more ill-conditioned it is, but the range of well-conditioned matrices is bigger.
- For example, if consider $k = 4$, $b = 10$, floating-point, then a condition number between 1 and 100 is considered well-conditioned.

Example 3 - Condition Number

Verify Example 2 using the condition number, with uniform norms.