DN2222
Applied Numerical Methods
- part 2:
Numerical Linear Algebra

Lecture 3
Direct Methods for Sparse Matrices

2012-11-06

Rep: Gaussian Elimination

- The basic algorithm for solving a linear system of equations, $Ax = b$, is Gaussian Elimination.
- The operation count for GE is $\sim n^3/3$.
- Time is not always proportional to flops. By doing operations and data acquisition in the right order, or even in parallel, time can be saved.
- When doing LU-factorization, we normally get fill in, ie L and U are less sparse than A.

Rep: Gaussian Elimination, cont.

- Even if A is non-singular, GE might fail without pivoting.
- Gaussian Elimination with re-ordering the rows is called GE with *partial pivoting*, GEPP. (Very common.)
- Gaussian Elimination with re-ordering of both rows and columns is called GE with *complete pivoting*, GECP. (Less common.)

Rep: Gaussian Elimination, cont.

- Renumbering the variables x_i, corresponds to simultaneous permutations of rows and columns in the matrix, $P^T AP$.
- Multiplication from the left with a permutation matrix interchanges the rows.
- Multiplication from the right with a permutation matrix interchanges the columns.

Quiz

While solving $Ax = b$ do I need to renumber the variables x_i

- ... if the columns in A are interchanged?
- ... if the rows in A are interchanged?

Which is/are the correct alternative/s?

1) No, not in any.
2) Only when shifting columns.
3) Only when shifting rows.
4) Yes both times.

Example (Demmel p45)

- $A = \begin{pmatrix} 1.00 \cdot 10^{-4} & 1 \\ 1 & 1 \end{pmatrix}$ makes $L = \begin{pmatrix} 1 & 0 \\ 1.00 \cdot 10^{+4} & 1 \end{pmatrix}$
- Then $U = \begin{pmatrix} 1 \cdot 10^{-4} & 1 \\ 0 & 1 - 1 \cdot 10^{+4} \end{pmatrix}$

  But on a 3-digit-machine we get $U = \begin{pmatrix} 1.00 \cdot 10^{-4} & 1.00 \\ 0.00 & -1.00 \cdot 10^{+4} \end{pmatrix}$
- Then $LU = \begin{pmatrix} 1.00 \cdot 10^{-4} & 1.00 \\ 1.00 & 0.00 \end{pmatrix}$ (NB: $\kappa(A)_\infty \approx 4$)

Demmel p45, cont.

- If $b = (2,3)^T$ then $x = A^{-1}b \approx (1.00, 2.00)^T$ but with the rounded U we get $\tilde{x} = U^{-1}(L^{-1}b) \approx (3.00, 2.00)^T$

Demmel p45, cont.

- This is called numerical instability.
- The computed solution x is worthless.
- $\kappa(L)$ and $\kappa(U) \gg \kappa(A)$ is a warning.
- Here: $\kappa_\infty(A) = \kappa_\infty(LU) \approx 4$ but $\kappa_\infty(L) \cdot \kappa_\infty(U) \approx 10^8 \cdot 10^8 = 10^{16}$.
- The problem is well-posed since $\kappa(A)$ is low.
- With GEPP we get an accurate and reliable solution.

Special Linear Systems

- For some linear systems, a solution can be obtained even more efficiently the Gaussian Elimination:
- Symmetric positive definite matrices (SPD)
- Symmetric indefinite matrices
- Band matrices
- General sparse matrices
- Dense matrices depending on a few parameters ($< n^2$)

Banded matrices

- The L and U matrices keep the band structure.
- Without pivoting, L has the lower bandwidth b_L and U has the lower bandwidth b_U (D Prop 2.3)
- With pivoting, L has the lower bandwidth $1 + b_L$ and U has the lower bandwidth $b_U + b_L$ (D Prop 2.4)
- Banded matrices are very common in numerical applications, typically meshes with neighbour interaction.

Real symmetric positive definite matrices

- SPD matrices can be factorized into $A = LL^T$,
  by the Cholesky factorization.
- Cholesky only requires half the flops of GE.
- Pivoting is not necessary for Cholesky to be numerically stable.
- Cholesky factorization is the cheapest way to test if a matrix is SPD.

Band matrices are common! (Demmel p81)

- Using FDM on a linear 2nd order 1D ODE
  $y''(x) - p(x) * y'(x) - q(x) * y(x) = r(x)$
  leads to a non-symmetric, tridiagonal matrix.
- For a small enough grid-size, the matrix is positive definite.
- From Gerschgorin Theorem we see that all eigenvalues lie in circles centered at $1 + h^2 q_i$ with radius 1.

Example, cont.

- The similarity transformation $\tilde{A} = DAD^{-1}$ makes $\tilde{A}$ symmetric. $\tilde{A}$ and A have the same eigenvalues.
- Thus Cholesky can be used.
- GE and Cholesky on band matrices are available in LAPACK routines like `ssbsv, sspsv` and `sptsv`.
- Note: A turns almost singular as $h \to 0$.

Sparse Matrices

- A *sparse matrix* has a lot of zero elements.
- If it is very sparse, algorithms avoiding the zero elements pay off.
- Often iterative methods are used for sparse matrices.
- For very special matrices, fast direct methods can be found.
- Vandermonde, Hilbert and Toeplitz are such matrices.

Graphs & Matrices

- A *graph*, G, consists of a set of *vertices*, V, and a set of *edges*, E, connecting a pair of vertices.
- Each graph of n nodes can be described by an $n \times n$-matrix A where $a_{i,j} \neq 0$ iff there is an edge from node i to node j.
- A undirected graph corresponds to a symmetric matrix, since both $a_{i,j}$ and $a_{j,i}$ are either zero or non-zero.

Example: Choice of node numbers?

- Given one node at the center and eight more nodes on a circle perimeter.
- All outer nodes connected to their two nearest neighbours and to the center node.
- How number the nodes to give a matrix with good GE behaviour?

Graphs & Matrices (cont.)

- Renumbering the nodes will not change the structure of the graph.
- Renumbering the nodes corresponds to simultaneous permutations of rows and columns in the matrix, $P^T AP$.
- *Rep: Multiplication from the left with a permutation matrix interchanges the rows.*
- *Rep: Multiplication from the right with a permutation matrix interchanges the columns.*

Gaussian Elimination of Graphs

- When doing LU-factorization, we normally get *fill in*, ie L and U are less sparse than A.
- The instruction in GE is $a'_{i,j} = a_{i,j} - (a_{i,k}/a_{k,k}) \cdot a_{k,j}$
- We get fill in ($a'_{i,j} \neq 0$) when $a_{i,j} = 0$ and $a_{i,k}$ and $a_{k,j}$ both are non-zero.
- When eliminating k we introduce an edge between i and j if both were connected to k.

Choices of node numbering

- To find the optimal node numbering is an NP-complete problem! We will use an heuristic, reasonably effective algorithm.
- We will discuss three classes of methods: *Variable band, Nested Dissection* and *Minimum Degree.*

Variable Band: RCM

- Easy to implement.
- Also called *Skyline algorithms*.
- Elements are filled in only between the “highest” element in a column and the diagonal.
- Thin bandwidth thus gives small fill-in.
- The **Reversed Cuthill McKee** algorithm gives a reasonably low bandwidth, typically a cigar shaped band.

Graphs cont.

- A *path* between two nodes is a sequence of edges starting at i and ending at j.
- There may be several paths between two nodes i and j.
- The number of edges in the shortest path is the *distance* between i and j.
- The *diameter* is the longest distance in the graph.
- A node is *peripheral* if it is one end of a diameter.

RCM cont.

- RCM is a two phase method: First searching for peripheral nodes, by travelling through the graph. Finally numbering the nodes according to the diameter travel.
- Start with the Gibbs-Poole-Stockmeyer algorithm which builds a *level structure* for finding the diameter:
- Start in one node, this is level 1.
- The nearest neighbours are level 2.
- Each of (the new) second nearest neighbours are level 3, etc.

- If the graph is connected all nodes will be levelled.
- If there are remaining nodes when all paths are traversed, put them all in the next structure level (they could be connected through directed edges)
- Now repeat the procedure starting from a node in the last level. We will get a new level structure of at least the same height. (Why?)
- Repeat the procedure until the number of levels do not seem to rise anymore. The number of levels is (probably) the distance.

- The starting and final nodes are called *pseudo-peripheral*.
- RCM is obtained from the longest level structure, numbering the nodes level by level.
- The matrix will get a cigar like shape, open at the end.
- RCM was introduced by E Cuthill and J McKee in 1969.
- In Matlab the routine is called `symrcm`.
- Matlab and Octave do not produce the same ordering!

Example: Laplace on L-shaped region

- Assuming the follwing L-shaped 2D region and the regular 5-point stencil for the 2D Laplacian.
- Write down the matrix A using the "standard" node numbering

1	3	5	9
2	4	6	10
		7	11
		8	12

- Renumber the grid using RCM.

Nested Dissection

- ND, also called *substructuring* is based on finding a separating set of nodes dividing the graph into two parts, substructures.
- Eliminate the nodes inside each substructure first, and then take the separating set.
- The resulting matrix will have a block structure.

$$A = \begin{pmatrix} A_{11} & 0 & A_{13} \\ 0 & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

- The division into substructure can then be repeated recursively.
- Nested Dissection was introduced by Alan J George in 1973.
- For planar graphs the resulting matrix has $O(n \log n)$ nonzeros.
- In Matlab the routine is called `nested`.

Minimum Degree algorithm

- The largest possible fill-in when eliminating node k is the product of non-diagonal non-zeros in row and column k respectively.
- If r_k is the number of non-zeros in row k, and c_k is the number of non-zeros in column k, the product $m_{k,k} = (r_k - 1) \cdot (c_k - 1)$ is called the *Markowitz' costs.*
- For a symmetric matrix this is the square of the number of edges meeting node k. This number is called the *degree* of node k.

- Eliminating a node with minimum degree will give the smallest possible fill-in.
- Even though MD is a greedy algorithm it gives surprisingly good results.
- MD is derived from a method first proposed by H M Markowitz in 1957.
- In Matlab the MD routine has been replaced by an AMD routine, the symmetric approximative multiple minimum degree function, `symamd`. It seems to work faster (and better).

Some review questions:

- **Q38.** Describe how a graph defines a matrix and vice versa.
- **Q39.** How does fill-in occur when one does Gaussian elimination on a sparse matrix. Describe it in both matrix and graph terms.
- **Q40.** What is pivoting for sparsity? Describe it in both matrix and graph terms.
- **Q42.** Describe the RCM algorithm.

/$\mathcal{NC}$