Welcome to Ma 221! Lecture 13, Sep 22

4 axes to organize course
1) Math problem: Ax=b
2) Structure of A: coming up
3) Accuracy
   GEPP: Backward stable, rare failures
   \[
   \begin{bmatrix}
   1 & 0 & 0 & 1 \\
   0 & 1 & 1 & 0 \\
   0 & 1 & 0 & 1 \\
   0 & 1 & 0 & -1
   \end{bmatrix}
   \begin{bmatrix}
   2 \\
   4 \\
   4 \\
   8
   \end{bmatrix}
   \]
   repeated doubling can cause instability
   "Guaranteed accuracy", rare failures
   using iterative refinement
   convergence proof assumes \( \kappa(A) \cdot \epsilon \leq 1 \)
4) Fast as possible: minimize communication

Historically, LAPACK reorganized GEPP
to use BLAS3 (matmul, GEMM
triang solve \( \text{LX=B}, \text{TRSM} \))

Idea similar to induction proof for GEPP,
but do b columns at a time, apply
updates all at once to trailing matrix
For simplicity, ignore pivoting

\[ A = b \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} \cdot U_{11} & A_{12} \\ L_{21} \cdot U_{11} & A_{22} \end{bmatrix} \]

using existing alg to do \[ \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix} U_{11} \]

\[ = \begin{bmatrix} L_{11} \cdot U_{11} & L_{11} \cdot U_{12} \\ L_{21} \cdot U_{11} & A_{22} \end{bmatrix} \]

where we solved \( A_{12} = L_{11} \cdot U_{12} \) for \( U_{12} \) using TRSM

\[ = \begin{bmatrix} L_{11} & 0 \\ L_{21} & I \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & L_{21} \cdot U_{12} \end{bmatrix} \]

Schur complement = \( S \) mat mul, QEMM

Repeat on \( S \)

Often very fast, but for some combinations of \( n \) and cache size \( M \) can't choose \( b \) to minimize comm
i.e. reach \( O(\frac{n^3}{M}) \) words moved

Just as for mat mul, there is a recursive cache oblivious alg that reaches \( O(\frac{n^3}{V M}) \)
(Toledo, 1997)
High Level alg:
DO LU on left half of A
Update right half
(U at top, Schur complement at bottom)
Do LU on Schur Complement

function \([L, U] = RLU(A)\) ... Recursive LU
... assume \(A\) \(n \times m\), \(n \geq m\), \(m\) power of 2
if \(m=1\) .. one column
pivot so \(A_{11}\) largest entry, pivot rest of matrix
\(L = A / A_{11}, U = A_{11}\)
else ... write \(A = \frac{1}{2} \begin{bmatrix} m/2 & m/2 \\ m/2 & m/2 \end{bmatrix}
\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\)
\(L_1 = \begin{bmatrix} L_{11} \\ L_{12} \end{bmatrix}
\begin{bmatrix} m/2 \\ m/2 \end{bmatrix}
\)
\([L_1, U_1] = RLU(\begin{bmatrix} A_{11} \\ A_{22} \end{bmatrix})\) ... LU of left half
\(A_{12} = L_1 \cdot U_{12}\) for \(U_{12}\) ... update \(U\)
\(A_{22} = A_{22} - L_{21} \cdot U_{12}\) ... update Schur compl.
\([L_2, U_2] = RLU(A_{22})\)
\(L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}\)
\(U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}\)

Correctness by induction
Cost: Recurrences (for \(m=n\))
\(A(n) = 4\) arith ops = \(\frac{2}{3} n^3 + O(n^2)\)
same flops as usual alg
similar recurrence to matmul

\[ W(n) = \# \text{word moved} = O\left(\frac{n^3}{\log n}\right) \]

RLV: only hits lower bound for \#words moved
\# messages

To minimize \# messages:

1. Replace partial pivoting by tournament pivoting (see notes)
2. Keep partial pivoting, more complicated data structure; payoff unclear

How to use Strassen-like alg?

Can modify RLV to run in \(O(n^w)\) flops if matmul does

1. multiply \(L_{12} \cdot U_{12}\) using \(O(n^w)\) matmul
2. solve \(A_{12} = L_{11} \cdot U_{12}\) by divide-conquer invert \(L_{11}\) (not as stable as GEPP)

\[
T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}^{-1} = \begin{bmatrix} T_{11}^{-1} & -T_{11}^{-1} \cdot T_{12} \cdot T_{22}^{-1} \\ 0 & T_{22}^{-1} \end{bmatrix}
\]

Use \(O(n^w)\) matmul

recall \(L_{ii} = 1\), \(L_{ij} \leq 1\) so

should be reasonably conditioned
Where to find implementations

**Matlab:** $A \backslash b$, or $[L,U,P]=lu(A)$

$rcond$, cond2est to estimate $\kappa(A)$

**LAPACK:**

- `xGETRF : GEPP` \( x = S/D/C/Z \)
- `xGETRF2 : G-EPP recursively`
- `xGESV` solve \( Ax = b \)
- `xGESVX` iterative refinement in precision \( x \)
- `xGESVXX` iterative refinement residual in double precision of \( x \)
- `xGECON` for condition est

Many other libraries

- Scalapack
- SLATE cluster
- PLASMA
- MAGMA
- ...

---

**Exploiting Structure in \( A \)**

- A symmetric positive definite: (spd)
  - Cholesky, no pivoting, \( \frac{1}{2} \) flops of GEPP

- A symmetric only
  - save half flops, pivoting required
  - but care needed to keep symmetry
A: band matrix
   \( bw = \text{bandwidth} \)
   costs in flops drops from \( O(n^3) \) to \( (bw^2 \cdot n) \)
   space drops from \( O(n^2) \) to \( O(bw \cdot n) \)

A: sparse matrix: lots of zeros
   cost, space drop significantly
   depends on pattern of nonzeros
   many complicated algo; lots of software

A: structured matrices: dense but
   depend on \( O(n) \) parameters
   Vandermonde: \( V(i,j) = x_i^{j-1} \)
   Toeplitz: \( T(i,j) = t_{i-j} \)
   many more, discuss most common

Symmetric (Hermitian) Positive Definite
   spd or hpd for short

def: \( A \) real and spd iff \( A = A^T \)
   and \( x^T A x > 0 \ \forall x \neq 0 \ \ x \text{ real vector} \)

A complex and hpd iff \( A = A^* \)
   and \( x^H A x > 0 \ \forall x \neq 0 \ \ x \text{ complex vector} \)
Lemma: (just real case)

1) \( X \) nonsingular \( \Rightarrow \) \( A \) s.p.d. \( \iff \) \( X^T A X \) s.p.d.

pf: \( A \) s.p.d. and \( x \neq 0 \Rightarrow Xx \neq 0 \Rightarrow \)

\( 0 \neq (Xx)^T A (Xx) = x^T X^T A X x = x^T (X^T A X) x \)

\( \Rightarrow X^T A X \) s.p.d. other direction same.

2) \( A \) s.p.d. \( H = A(j:k, j:k) \)

\( H \) "principal submatrix" \( H \) s.p.d.

pf: \( A \) s.p.d. \( y \neq 0 \Rightarrow 0 \neq x = \begin{bmatrix} y \end{bmatrix} \)

\( \Rightarrow 0 \neq x^T A x = y^T H y \Rightarrow H \) s.p.d.

3) \( A \) s.p.d. \( \iff \) \( A = A^T \) and all evals \( d_i > 0 \)