Fast direct sparse solvers

Zhe Chen

Courant, NYU

December 2nd, 2020

Review for Ch.20

- Laplace equation $-\Delta u = 0$ on a unit square domain $[0, 1]^d$
- LU factorization of the permuted discrete system $P^*AP = LU$
- $N = n^d, d = 2, 3$

Table 1: Complexity of LU

d	Simple ordering	Nested dissection ordering
2	$O(N^2)$	$O(N^{3/2})$
3	$O(N^{7/3})$	$O(N^2)$

Only work for moderate size problems!

Goal for Ch.21

1. Linear or near-to-linear complexity, i.e. Fast direct sparse solvers (FDSS)
2. High level view of the general ideas
3. Heuristic understanding of when it will work

FDSS

Main cost in computing LU of the Schur complement (in the root node).
Schur Complement behaves as a Dirichelet-to-Neumann operator, so it's rank-structured.
Consider it as electrostatic problem.

$$[\mathsf{Au}](k) = (\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{s}})) + (\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{e}})) + (\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{n}})) + (\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{w}}))$$

The Schur complement is rank structured

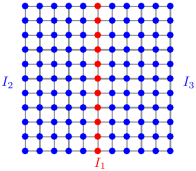

Figure 1: Top level of nested dissection

Schur Complement: $S_1 = A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1} - A_{1,3}A_{3,3}^{-1}A_{3,1}$

Linear system: $\begin{bmatrix} A_{3,3} & 0 & A_{3,1} \\ 0 & A_{2,2} & A_{2,1} \\ A_{1,3} & A_{1,2} & A_{1,1} \end{bmatrix} \begin{bmatrix} u_3 \\ u_2 \\ u_1 \end{bmatrix} = \begin{bmatrix} f_3 \\ f_2 \\ f_1 \end{bmatrix}$

Eliminate points in I_2:

$$\begin{bmatrix} A_{3,3} & A_{3,1} \\ A_{1,3} & A_{1,3} \end{bmatrix} \begin{bmatrix} u_3 \\ u_1 \end{bmatrix} = \begin{bmatrix} f_3 \\ f_1 - A_{1,2}A_{2,2}^{-1}f_2 + A_{1,2}A_{2,2}^{-1}A_{2,1}u_1 \end{bmatrix}$$

Hierarchical-off-diagonal-low-ranked

$$S_{\alpha,\beta} = S\left(J_\alpha, J_\beta\right) = A\left(I_\alpha, I_2\right) A\left(I_2, I_2\right)^{-1} A\left(I_2, I_\beta\right)$$

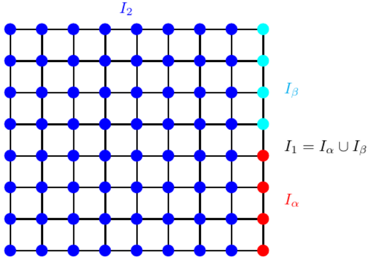

Figure 2: Illustration for off-diagonal on the top level

Compressibility

σ_j is the jth singular value of $S_{\alpha,\beta}$

- **Dependence on the problem size**

n	32	64	128	256	512	1024
k	8	10	12	14	16	18

Table 2: 5-pt stencil, precision $\epsilon = 1e - 12$, k is the numerical ranks

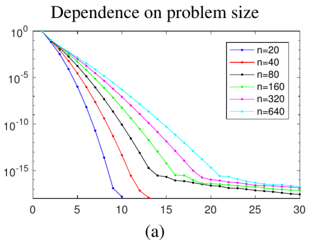

Figure 3: σ_j/σ_1 versus j

Compressibility

- **Convection term** $[Au](\boldsymbol{x}) = -\Delta u(\boldsymbol{x}) + b\left(\frac{\partial u(\boldsymbol{x})}{\partial x_1} + \frac{\partial u(\boldsymbol{x})}{\partial x_2}\right)$

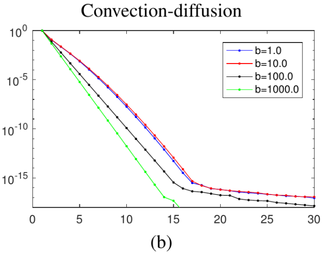

Figure 4: σ_j/σ_1 versus j, 5-pt stencil for the Laplace operator and centered differences for the convection term

Compressibility

- **Rough coefficients and large contrast ratios**
  $[\mathsf{Au}](k) = \alpha_{k,k_e}(\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{e}})) + \alpha_{k,k_{\mathrm{n}}}(\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{n}})) + \alpha_{k,k_{\mathrm{w}}}(\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{w}})) + \alpha_{k,k_s}(\mathsf{u}(k) - \mathsf{u}(k_{\mathrm{s}}))$,
  where $\alpha_{k,\ell} = 1 + c\theta_{k,\ell},\ \theta_{k,\ell} \sim U[0,1]$

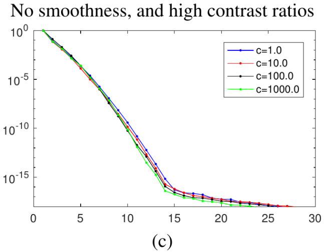

Figure 5: σ_j/σ_1 versus j, 5-pt stencil, n=400

Compressibility

- **Higher order stencils: 5,9,13 points stencil**
  Higher accuracy, thicker mesh separator, the ranks slightly increase.

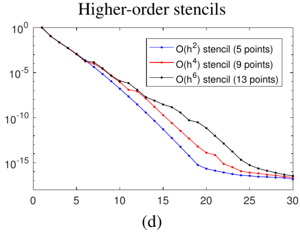

Figure 6: σ_j/σ_1 versus j

Compressibility

- **Wave number for oscillatory problems**
  Discretization for the Helmholtz operator
  $A = L - \kappa^2 I$

$1/\lambda$	0.5	5	10	15	20
k	16	17	22	27	31

Table 3: Wavelength $\lambda = 2\pi/\kappa$, $n = 512$, precision $\epsilon = 1e - 12$, $k \approx 10 + 1/\lambda$

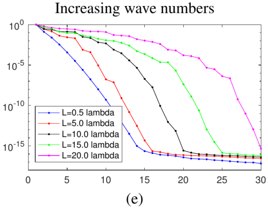

Figure 7: σ_j/σ_1 versus j

Complexity in 2D

Consider cost to compute LU of $m \times m$ Schur complement is m^{α}, $\alpha \in [1,3]$,

$$T \sim \sum_{\ell=0}^{L} 2^{\ell} m_{\ell}^{\alpha} \sim \sum_{\ell=0}^{L} 2^{\ell} 2^{-\alpha\ell/2} n^{\alpha} = n^{\alpha} \sum_{\ell=0}^{L} \left(2^{1-\alpha/2}\right)^{\ell}$$

Since $2^L \approx n^2$,

$\boxed{\alpha < 2}$ Geometric sum dominated by the *leaves*. Complexity is $O(N)$.

$\boxed{\alpha = 2}$ Every term has the same weight. Complexity is $O(n^2 L) = O(N \log N)$.

$\boxed{\alpha > 2}$ Geometric sum dominated by the *root*. Complexity is $O(n^{\alpha}) = O(N^{\alpha/2})$.

e.g. ranks scale as $log(m_{\ell})$ and then LU scales as $m_{\ell} log(m_{\ell})^2$, much better than m_{ℓ}^2

Complexity in 3D

$$T \sim \sum_{\ell=0}^{L} 2^{\ell} m_{\ell}^{\alpha} \sim \sum_{\ell=0}^{L} 2^{\ell} 2^{-2\alpha\ell/3} n^{2\alpha} = n^{2\alpha} \sum_{\ell=0}^{L} \left(2^{1-2\alpha/3}\right)^{\ell}$$

$\boxed{\alpha < 3/2}$ Geometric sum dominated by the *leaves*. Complexity is $O(N)$.

$\boxed{\alpha = 3/2}$ Every term has the same weight. Complexity is $O(n^3 L) = O(N \log N)$.

$\boxed{\alpha > 3/2}$ Geometric sum dominated by the *root*. Complexity is $O(n^{2\alpha}) = O(N^{2\alpha/3})$.

Difficulty

- Possible to parallelize but difficult
- Difficult to achieve the nested dissection ordering

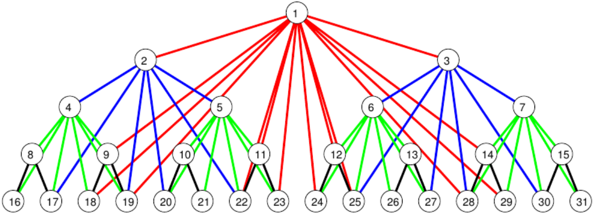

Figure 8: Nested dissection ordering