Lecture 8: Fast Linear Solvers (Part 6)

Nonsymmetric System of Linear Equations

- The CG method requires to A to be an $n \times n$ symmetric positive definite matrix to solve $A\boldsymbol{x} = \boldsymbol{b}$.
- If A is nonsymmetric:
  - Convert the system to a symmetric positive definite one
  - Modify CG to handle general matrices

Normal Equation Approach

The *normal equations* corresponding to $A\boldsymbol{x} = \boldsymbol{b}$ are $A^T A\boldsymbol{x} = A^T \boldsymbol{b}$

- If A is nonsingular then $A^T A$ is symmetric positive definite and the CG method can be applied to solve $A^T A\boldsymbol{x} = A^T \boldsymbol{b}$ (CG normal residual -- CGNR).
- Alternatively, we can first solve $\mathrm{A}A^T \boldsymbol{y} = \boldsymbol{b}$ for $\boldsymbol{y}$, then $\boldsymbol{x} = A^T \boldsymbol{y}$.
- **Disadvantages:**
  - Each iteration requires $A^T A$ or $\mathrm{A}A^T$
  - Condition number of $A^T A$ or $\mathrm{A}A^T$ is square of that of A. However, CG works well if condition number is small

Arnoldi Iteration

- The Arnoldi method is an orthogonal projection onto a Krylov subspace $\mathrm{K}_m(A, \boldsymbol{r}_0)$ for $n \times n$ nonsymmetric matrix A. Here $m \ll n$.
- Arnoldi reduces A to a *Hessenberg* form.

Upper Hessenberg matrix: zero entries below the first subdiagonal.

$$\begin{bmatrix} 2 & 3 & 4 & 1 \\ 2 & 5 & 1 & 9 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 3 & 2 \end{bmatrix}$$

Lower Hessenberg matrix: zero entries above the first superdiagonal.

$$\begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 5 & 1 & 0 \\ 1 & 2 & 1 & 2 \\ 3 & 4 & 3 & 2 \end{bmatrix}$$

Let H_m be a $m \times m$ Hessenberg matrix:

$$H_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & h_{m,m-1} & h_{mm} \end{bmatrix}$$

Let $(m+1) \times m$ $\overline{H}_m$ be the extended matrix of H_m:

$$\overline{H}_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & h_{m,m-1} & h_{mm} \\ 0 & \cdots & 0 & h_{m+1,m} \end{bmatrix}$$

The Arnoldi iteration produces matrices V_m, V_{m+1} and $\overline{H}_m$ for matrix A satisfying:

$$AV_m = V_{m+1}\overline{H}_m = V_m H_m + \boldsymbol{w}_m \boldsymbol{e}_m^T$$

Here V_m, V_{m+1} have orthonormal columns

$$V_m = [\boldsymbol{v}_1|\boldsymbol{v}_2|\ldots|\boldsymbol{v}_m], \quad V_{m+1} = [\boldsymbol{v}_1|\boldsymbol{v}_2|\ldots|\boldsymbol{v}_m|\boldsymbol{v}_{m+1}]$$

Arnoldi Algorithm

$$v_1 = r_0/\|r_0\|_2$$

$$\begin{aligned}
w_1 &= A\,v_1 - (A\,v_1, v_1)\,v_1, & v_2 &= w_1/\|w_1\|_2 \\
&\vdots \\
w_j &= A\,v_j - (A\,v_j, v_1)\,v_1 - \ldots - (A\,v_j, v_j)\,v_j, & v_{j+1} &= w_j/\|w_j\|_2 \\
&\vdots \\
w_m &= A\,v_m - (A\,v_m, v_1)\,v_1 - \ldots - (A\,v_m, v_m)\,v_m, & v_{m+1} &= w_m/\|w_m\|_2
\end{aligned}$$

Choose $\boldsymbol{r}_0$ and let $\boldsymbol{v}_1 = \boldsymbol{r}_0/||\boldsymbol{r}_0||$

for $j = 1, \ldots, m-1$

$\boldsymbol{w} = A\boldsymbol{v}_j - \sum_{i=1}^{j}((A\boldsymbol{v}_j)^T\boldsymbol{v}_i)\boldsymbol{v}_i$

$\boldsymbol{v}_{j+1} = \boldsymbol{w}/||\boldsymbol{w}||_2$

endfor

$$V_m = \begin{pmatrix} | & | & & | \\ | & | & & | \\ v_1 & v_2 & \vdots & v_m \\ | & | & & | \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n\times m}$$

$$\bar{H}_m = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1m} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2m} \\ & h_{32} & h_{33} & \dots & h_{3m} \\ & & \ddots & \ddots & \vdots \\ & & & h_{m,m-1} & h_{mm} \\ & & & & h_{m+1,m} \end{pmatrix} = \begin{pmatrix} (A\,v_1, v_1) & (A\,v_2, v_1) & (A\,v_3, v_1) & \dots & (A\,v_m, v_1) \\ (A\,v_1, v_2) & (A\,v_2, v_2) & (A\,v_3, v_2) & \dots & (A\,v_m, v_2) \\ & (A\,v_2, v_3) & (A\,v_3, v_3) & \dots & (A\,v_m, v_3) \\ & & \ddots & \ddots & \vdots \\ & & & (A\,v_{m-1}, v_m) & (A\,v_m, v_m) \\ & & & & (A\,v_m, v_{m+1}) \end{pmatrix}$$

- $V_m^T V_m = I_{m\times m}$.
- If Arnoldi process breaks down at mth step, $\boldsymbol{w}_m = \mathbf{0}$ is still well-defined but not v_{m+1}, and the algorithm stop.
- In this case, the last row of $\overline{H}_m$ is set to zero, $h_{m+1,m} = 0$

Theorem. The Arnoldi procedure generates a reduced QR factoriztion of the Krylov matrix
$\mathrm{K}_m = [\boldsymbol{r}_0|A\boldsymbol{r}_0|A^2\boldsymbol{r}_0| \ldots |A^{k-1}\boldsymbol{r}_0]$ in the form
$\mathrm{K}_m = V_m R_m,$
with R_m being a triangular matrix $R_m \in R^{m\times m}$.
Furthermore,

$$V_m^T A V_m = H_m.$$

Remark:

$A\boldsymbol{v}_j = \sum_{i=1}^{j+1} h_{ij}\boldsymbol{v}_i,$ for $j = 1, \ldots, m-1$

$AV_m = V_{m+1}\overline{H}_m = V_m H_m + \boldsymbol{w}_m \boldsymbol{e}_m^T$

Stable Arnoldi Algorithm

Choose $\boldsymbol{x}_0$ and let $\boldsymbol{v}_1 = \boldsymbol{x}_0/||\boldsymbol{x}_0||$.

for $j = 1, \ldots, m$

$\quad \boldsymbol{w} = A\boldsymbol{v}_j$

$\quad$ **for** $i = 1, \ldots, j$

$\quad\quad h_{ij} =< \boldsymbol{w}, \boldsymbol{v}_i >$

$\quad\quad \boldsymbol{w} = \boldsymbol{w} - h_{ij}\boldsymbol{v}_i$

$\quad$ **endfor**

$\quad h_{j+1,i} = ||\boldsymbol{w}||_2$

$\quad \boldsymbol{v}_{j+1} = \boldsymbol{w}/h_{j+1,i}$

endfor

Generalized Minimum Residual (GMRES) Method

Let the Krylov space associated with $A\boldsymbol{x} = \boldsymbol{b}$ be $\mathrm{K}_k(A, \boldsymbol{r}_0) = span\{\boldsymbol{r}_0, A\boldsymbol{r}_0, A^2\boldsymbol{r}_0, \dots, A^{k-1}\boldsymbol{r}_0\}$, where $\boldsymbol{r}_0 = \boldsymbol{b} - A\,\boldsymbol{x}_0$ for some initial guess $\boldsymbol{x}_0$.

The kth ($k \geq 1$) iteration of GMRES is the solution to the least squares problem:

$$minimize_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathrm{K}_k} ||\boldsymbol{b} - A\boldsymbol{x}||_2, \text{ i.e.}$$

$$\text{Find } \boldsymbol{x}_k \in \boldsymbol{x}_0 + \mathrm{K}_k \text{ such that } ||\boldsymbol{b} - A\boldsymbol{x}_k||_2 = min_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathrm{K}_k} ||\boldsymbol{b} - A\boldsymbol{x}||_2$$

If $\boldsymbol{x} \in \boldsymbol{x}_0 + \mathrm{K}_k$, then $\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{j=0}^{k-1} \gamma_j A^j \boldsymbol{r}_0$.
So $\boldsymbol{b} - A\boldsymbol{x} = \boldsymbol{b} - A\boldsymbol{x}_0 - \sum_{j=0}^{k-1} \gamma_j A^{j+1} \boldsymbol{r}_0 = \boldsymbol{r}_0 - \sum_{j=1}^{k} \gamma_{j-1} A^j \boldsymbol{r}_0$.

Define: Let $\bar{p}_k$ be a kth degree polynomial such that $\bar{p}_k(0) = 1$. $\bar{p}_k$ is called a *residual polynomial*.
The set of kth degree *residual polynomial* is

$$P_k = \{\bar{p}_k | \bar{p}_k \; is \; a \; kth \; degree \; polynomial \; and \; \bar{p}_k(0) = 1\}$$

$$\boldsymbol{r} = \boldsymbol{r}_0 - \sum_{j=1}^{k} \gamma_{j-1} A^j \boldsymbol{r}_0 = \bar{p}_k(A) \boldsymbol{r}_0$$

Theorem. Let $\boldsymbol{x}_k$ be the kth GMRES iteration. Then for all $\bar{p}_k \in P_k$

$$||\boldsymbol{r}_k||_2 \leq ||\bar{p}_k(A) \boldsymbol{r}_0||_2$$

GMRES Implementation

- The kth ($k \geq 1$) iteration of GMRES is the solution to the least squares problem:

$$minimize_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathrm{K}_k} ||\boldsymbol{b} - A\boldsymbol{x}||_2$$

- Suppose we have used Arnoldi process constructed an orthogonal basis V_k for $\mathrm{K}_k(A, \boldsymbol{r}_0)$.
  - $\boldsymbol{r}_0 = \beta V_k \boldsymbol{e}_1$, where $\boldsymbol{e}_1 = (1,0,0, \dots)^T$, $\beta = ||\boldsymbol{r}_0||_2$
  - Any vector $\boldsymbol{z} \in \mathrm{K}_k(A, \boldsymbol{r}_0)$ can be written as $\boldsymbol{z} = \sum_{l=1}^{k} y_l \boldsymbol{v}_l^k$, where $\boldsymbol{v}_l^k$ is the lth column of V_k. Denote $\boldsymbol{y} = (y_1, y_2, \dots, y_k)^T \in R^k$.

$$\boldsymbol{z} = V_k \boldsymbol{y}$$

Since $\boldsymbol{x} - \boldsymbol{x}_0 = V_k \boldsymbol{y}$ for some coefficient vector $\boldsymbol{y} \in R^k$, we must have $\boldsymbol{x}_k = \boldsymbol{x}_0 + V_k \boldsymbol{y}$ where $\boldsymbol{y}$ minimizes $||\boldsymbol{b} - A(\boldsymbol{x}_0 + V_k \boldsymbol{y})||_2 = ||\boldsymbol{r}_0 - AV_k \boldsymbol{y}||_2$.

- The kth ($k \geq 1$) iteration of GMRES now is equivalent to a least squares problem in R^k, i.e.

$$minimize_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathrm{K}_k} ||\boldsymbol{b} - A\boldsymbol{x}||_2$$

$$= minimize_{y \in R^k} ||\boldsymbol{r}_0 - AV_k \boldsymbol{y}||_2$$

  - Remark: This is a linear least square problem.
  - The associate normal equation is $(AV_k)^T AV_k \boldsymbol{y} = (AV_k)^T \boldsymbol{r}_0$. But we will solve it differently.

- Let $\boldsymbol{x}_k$ be kth iterative solution of GMRES.

Define: $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k = \boldsymbol{r}_0 - A(\boldsymbol{x}_k - \boldsymbol{x}_0) = \beta V_{k+1}\boldsymbol{e}_1 - A(\boldsymbol{x}_0 + V_k\boldsymbol{y} - \boldsymbol{x}_0) = \beta V_{k+1}\boldsymbol{e}_1 - V_{k+1}\bar{H}_k\boldsymbol{y}^k = V_{k+1}(\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}^k)$

Using orthogonality of V_{k+1}:

$$minimize_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathrm{K}_k}||\boldsymbol{b} - A\boldsymbol{x}||_2$$
$$= minimize_{y\in R^k}||\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}^k||_2$$

$$minimize_{y \in R^k} ||\beta \boldsymbol{e}_1 - \bar{H}_m \boldsymbol{y}^k||_2$$

Theorem. Let $n \times k$ ($k \leq n$) matrix B be with linearly independent columns (full column rank). Let $B = QR$ be a QR factorization of B. Then for each $\boldsymbol{b} \in R^n$, the equation $B\boldsymbol{u} = \boldsymbol{b}$ has a unique least-square solution, given by $\hat{\boldsymbol{u}} = R^{-1}Q^T\boldsymbol{b}$.

Using Householder reflection to do QR factorization gives $\bar{H}_m = Q_{m+1}\bar{R}_m$ where $Q_{m+1} \in R^{(m+1)\times(m+1)}$ is orthogonal and $\bar{R}_m \in R^{(m+1)\times m}$ has the form $\bar{R}_m = \begin{bmatrix} R_m \\ 0 \end{bmatrix}$, where $R_m \in R^{m\times m}$ is upper triangular.

ALGORITHM 3.4.2. $\mathtt{gmresa}(x, b, A, \epsilon, kmax, \rho)$

1. $r = b - Ax$, $v_1 = r/\|r\|_2$, $\rho = \|r\|_2$, $\beta = \rho$, $k = 0$

2. While $\rho > \epsilon\|b\|_2$ and $k < kmax$ do

   (a) $k = k + 1$

   (b) for $j = 1, \ldots, k$
   $h_{jk} = (Av_k)^T v_j$

   (c) $v_{k+1} = Av_k - \sum_{j=1}^{k} h_{jk} v_j$

   (d) $h_{k+1,k} = \|v_{k+1}\|_2$

   (e) $v_{k+1} = v_{k+1}/\|v_{k+1}\|_2$

   (f) $e_1 = (1, 0, \ldots, 0)^T \in R^{k+1}$
   Minimize $\|\beta e_1 - H_k y^k\|_{R^{k+1}}$ over R^k to obtain y^k.

   (g) $\rho = \|\beta e_1 - H_k y^k\|_{R^{k+1}}$.

3. $x_k = x_0 + V_k y^k$.

“C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations” .

- $\boldsymbol{v}_j$ may become nonorthogonal as a result of round off errors.
  - $||\beta \boldsymbol{e}_1 - \overline{H}_k \boldsymbol{y}^k||_2$ which depends on orthogonality, will not hold and the residual could be inaccurate.
  - Replace the loop in Step 2c of Algorithm *gmresa* with

$$
\begin{aligned}
&v_{k+1} = Av_k \\
&\text{for } j = 1, \ldots k \\
&\quad v_{k+1} = v_{k+1} - (v_{k+1}^T v_j) v_j .
\end{aligned}
$$

We illustrate this point with a simple example from [128], doing the computations in MATLAB. Let $\delta = 10^{-7}$ and define

$$A = \begin{pmatrix} 1 & 1 & 1 \\ \delta & \delta & 0 \\ \delta & 0 & \delta \end{pmatrix}.$$

We orthogonalize the columns of A with classical Gram–Schmidt to obtain

$$V = \begin{pmatrix} 1.0000e+00 & 1.0436e-07 & 9.9715e-08 \\ 1.0000e-07 & 1.0456e-14 & -9.9905e-01 \\ 1.0000e-07 & -1.0000e+00 & 4.3568e-02 \end{pmatrix}.$$

The columns of V_U are not orthogonal at all. In fact $v_2^T v_3 \approx -.004$. For modified Gram–Schmidt

$$V = \begin{pmatrix} 1.0000e+00 & 1.0436e-07 & 1.0436e-07 \\ 1.0000e-07 & 1.0456e-14 & -1.0000e+00 \\ 1.0000e-07 & -1.0000e+00 & 4.3565e-16 \end{pmatrix}.$$

Here $|v_i^T v_j - \delta_{ij}| \leq 10^{-8}$ for all i, j.

“C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations” .

ALGORITHM 3.4.3. **gmresb**$(x, b, A, \epsilon, kmax, \rho)$

1. $r = b - Ax$, $v_1 = r/\|r\|_2$, $\rho = \|r\|_2$, $\beta = \rho$, $k = 0$

2. While $\rho > \epsilon\|b\|_2$ and $k < kmax$ do

   (a) $k = k + 1$

   (b) $v_{k+1} = Av_k$
   for $j = 1, \ldots k$

      i. $h_{jk} = v_{k+1}^T v_j$

      ii. $v_{k+1} = v_{k+1} - h_{jk}v_j$

   (c) $h_{k+1,k} = \|v_{k+1}\|_2$

   (d) $v_{k+1} = v_{k+1}/\|v_{k+1}\|_2$

   (e) $e_1 = (1, 0, \ldots, 0)^T \in R^{k+1}$
   Minimize $\|\beta e_1 - H_k y^k\|_{R^{k+1}}$ to obtain $y^k \in R^k$.

   (f) $\rho = \|\beta e_1 - H_k y^k\|_{R^{k+1}}$.

3. $x_k = x_0 + V_k y^k$.

“C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations” .