GMRES (General Minimal Residual)

A Krylov subspace method to solve

$$Ax = b$$

$n \times n$
n is very large

Find the best solution in the Krylov subspace (in the least squares sense)

$$\min_{x \in \mathcal{K}_k} \|b - Ax\|_2$$

$$\mathcal{K}_k := \text{span}\{b, Ab, \ldots, A^{k-1}b\}$$

Krylov subspace (typically of dimension k)

Arnoldi process

Procedure to create an orthonormal basis for $\mathcal{K}_k$.

Suppose an orthonormal basis

$$\{q_1, q_2, \ldots, q_{k-1}\}$$

for $\mathcal{K}_{k-1}$ is already computed. Then

$$\mathcal{K}_k = \text{span}\{q_1, \ldots, q_{k-1}, A\underbrace{(A^{k-2}b)}_{\in \text{span}\{q_1, \ldots, q_{k-1}\}}\}$$

$$= \text{span}\{q_1, \ldots, q_{k-1}, Aq_{k-1}\}$$

Note $Aq_1, \ldots, Aq_{k-2} \in \mathcal{K}_{k-1}$

Assumption: $\dim \mathcal{K}_k = k$.

Generation of an orthonormal basis $\{q_1, \dots q_j\}$ for $\mathcal{K}_j$ $\quad j = 1, 2, \dots, k$

Using Gram-Schmidt

Circles - to be computed

$\mathcal{K}_1$ — $\quad q_1 = b / \|b\|_2$

$\mathcal{K}_2$ — $\quad A q_1 = h_{11} q_1 + h_{21} q_2$

$\mathcal{K}_3$ — $\quad A q_2 = h_{12} q_1 + h_{22} q_2 + h_{32} q_3$

$\vdots$

$\mathcal{K}_k$ — $\quad A q_{k-1} = h_{1(k-1)} q_1 + h_{2(k-1)} q_2 + \dots + h_{k(k-1)} q_k$

$$\underbrace{[A q_1 \;\; A q_2 \;\; \dots \;\; A q_{k-1}]}_{n \times k-1} \left(= A \, [q_1 \; q_2 \dots q_{k-1}] \right)$$

$$=$$

$$\underbrace{[q_1 \;\; q_2 \;\; \dots \;\; q_k]}_{\substack{Q_k := \\ n \times k}} \underbrace{\begin{bmatrix} h_{11} & h_{12} & \dots & h_{1(k-1)} \\ h_{21} & h_{22} & & h_{2(k-1)} \\ & h_{32} & \ddots & \vdots \\ 0 & & \ddots & h_{(k-1)(k-1)} \\ & & & h_{k(k-1)} \end{bmatrix}}_{\substack{H_k := \\ k \times k-1}} \quad (\text{Hessenberg})$$

$$\Longrightarrow \boxed{A Q_{k-1} = Q_k H_k}$$

Pseudocode (Arnoldi Process)

$q_1 \leftarrow b / \|b\|$

for $j = 1, \dots, k-1$

 $v \leftarrow A q_j$

 for $l = 1, \dots, j$

  $h_{lj} \leftarrow q_l^* v$

  $v \leftarrow v - h_{lj} q_l$

 end

 $h_{(j+1)j} \leftarrow \|v\|_2$

 $q_{j+1} \leftarrow v / h_{(j+1)j}$

end

return

(At termination, we have

$Q_k = [q_1 \cdots q_k]$ and H_k)

Efficient Solution of the least squares problem

$$\min_{x \in \mathcal{K}_k} \|b - Ax\|_2 = \min_{\alpha_k \in \mathbb{C}^k} \|b - A Q_k \alpha_k\|_2$$

$$= \min_{\alpha_k \in \mathbb{C}^k} \|\underbrace{b - Q_{k+1} H_{k+1} \alpha_k}_{z \in \mathbb{C}^n \,\to\, Q_{k+1}^* z}\|_2$$

see Remark (1)

$$= \min_{\alpha_k \in \mathbb{C}^k} \|\underbrace{Q_{k+1}^* b}_{\|b\|_2 e_1} - H_{k+1} \alpha_k\|_2$$

H_{k+1}: $(k+1)\times k$, Hessenberg

Remarks

(1) Letting $z := b - Q_{k+1} H_{k+1} d_k \in Col(Q_{k+1})$ and β s.t. $z = Q_{k+1}\beta$, we have

$$\|z\|_2 = \sqrt{(\beta_1 q_1 + \cdots + \beta_{k+1} q_{k+1})^* (\beta_1 q_1 + \cdots + \beta_{k+1} q_{k+1})}$$

$$= \sqrt{|\beta_1|^2 + \cdots + |\beta_{k+1}|^2} = \|\beta\|_2$$

and

$$\|Q_{k+1}^* z\|_2 = \|Q_{k+1}^* Q_{k+1} \beta\|_2$$

$$= \|\beta\|_2 = \|z\|_2$$

(2) The last LSP on page 3 is $(k+1)\times k$ and can be solved by computing the QR factor. of H_{k+1} at a cost of $O(k^2)$. (using Householder reflectors or rotators)

Pseudocode (GMRES) (Given τ, e.g. $\tau = 10^{-12}$ a tolerance for termination)

$\rho \leftarrow \|b\|_2$, $\quad q_1 \leftarrow b/\rho$, $\quad k \leftarrow 1$

while $\rho > \tau \|b\|_2$

$\quad v \leftarrow A q_k$

$\quad$ for $j = 1, \ldots, k$

$\quad\quad h_{jk} \leftarrow q_j^* v$

$\quad\quad v \leftarrow v - h_{jk} q_j$

$\quad$ end

$\quad h_{(k+1)k} \leftarrow \|v\|_2$

$\quad q_{k+1} \leftarrow v / h_{(k+1)k}$

$\quad$ Solve the LSP $\min\limits_{\alpha \in \mathbb{C}^k} \big\| \|b\| e_1 - \underbrace{H_{k+1}}_{(k+1)\times k \text{ Hessenberg}} \alpha \big\|_2$

$\quad$ Let ρ the minimal value, α_k the minimizer of LSP

$\quad k \leftarrow k+1$

end

$x \leftarrow \underbrace{[q_1 \ldots q_{k-1}]}_{Q_{k-1}} \alpha_{k-1}$ (unless $k=1$ at termination, in which case set $x = 0$)

return

Minimization Property of GMRES

Let $x^{(k)}$ be the minimizer of

$\|b - Ax\|_2$ over $x \in \mathcal{K}_k$.

Every $x \in \mathcal{K}_k$ can be written as

$$x = \sum_{j=0}^{k-1} \alpha_j A^j b$$

for some $\alpha_0, \alpha_1, \ldots, \alpha_{k-1} \in \mathbb{C}$.

Hence,

$$b - Ax = b - \sum_{j=0}^{k-1} \alpha_j A^{j+1} b$$

$$= \underbrace{\left(I_n - \sum_{j=0}^{k-1} \alpha_j A^{j+1}\right)}_{p(A)} b$$

where $p(z) := 1 - \sum_{j=0}^{k-1} \alpha_j z^{j+1}$.

<u>THM</u>

$$\|b - Ax^{(k)}\|_2 = \min_{p \in \mathcal{P}_k^0} \|p(A)\, b\|_2$$

where

$$\mathcal{P}_k^0 := \{ p : \mathbb{C} \to \mathbb{C},\ p(z) = 1 + \alpha_0 z + \ldots + \alpha_{k-1} z^{k} \mid \alpha_0, \ldots, \alpha_{k-1} \in \mathbb{C} \}$$

Consequences

For any $\tilde{p} \in P_k^\circ$, we have

$$\| b - A x^{(k)} \|_2 \leq \| \tilde{p}(A) \|_2 \| b \|_2 .$$

Suppose A has n linearly independent eigenvectors so that

$$A v_j = \lambda_j v_j \quad j = 1, \dots, n$$

$$\Longrightarrow A [v_1 \dots v_n] = \underbrace{[v_1 \dots v_n]}_{V} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\Lambda}$$

$$\Longrightarrow A = V \Lambda V^{-1}$$

But then

$$\tilde{p}(A) = I + \sum_{j=1}^{k} \alpha_j A^j$$

$$= V V^{-1} + \sum_{j=1}^{k} \alpha_j V \Lambda^j V^{-1}$$

$$= V \begin{bmatrix} 1 + \sum_{j=1}^{k} \alpha_j \lambda_1^j & & \\ & \ddots & \\ & & 1 + \sum_{j=1}^{k} \alpha_j \lambda_n^j \end{bmatrix} V^{-1}$$

$$= V \tilde{p}(\Lambda) V^{-1}$$

Notation: $\Lambda(A)$ - the set of eigenvalues of A

For any $\tilde{p} \in P_k^\circ$,

$$\text{(x)} \quad \| b - A x^{(k)} \|_2 \leq \| V \|_2 \| V^{-1} \|_2 \| \tilde{p}(\Lambda) \|_2 \| b \|_2$$

$$= \kappa_2(V) \left\{ \max_{z \in \Lambda(A)} | \tilde{p}(z) | \right\} \| b \|_2$$

Corollary

(Assuming A has n linearly independent eigenvectors)

Suppose $A \in \mathbb{C}^{n \times n}$ has k distinct eigenvalues. Then $x^{(k)}$ satisfies $Ax^{(k)} = b$.

Proof

Denoting the distinct eigenvalues of A with $\lambda_1, \ldots, \lambda_k$, and letting

$$\tilde{p}(z) := \prod_{j=1}^{k} (\lambda_j - z) / \lambda_j \in \mathcal{P}_k^0 ,$$

we have

$$|\tilde{p}(\lambda_j)| = 0 \qquad j = 1, \ldots, k$$

$$\Longrightarrow \max_{z \in \Lambda(A)} |\tilde{p}(z)| = 0$$

From (x) on page 2

$$\Longrightarrow \|b - Ax^{(k)}\|_2 = 0 \qquad \Longrightarrow Ax^{(k)} = b \qquad \square$$

Hence, GMRES terminates after at most n iterations.

Preconditioning

Let $M \in \mathbb{C}^{n \times n}$ be invertible.

Left preconditioning

$$Ax = b$$

$$\Longleftrightarrow$$

$$(*) \quad MAx = Mb$$

Apply GMRES to $(*)$, letting $\hat{x}^{(k)}$ be the best solution in $\mathcal{K}_k$.

$$\|Mb - MA\hat{x}^{(k)}\|_2 \leqslant \|\tilde{p}(MA)\|_2 \, \|Mb\|_2$$

for any $\tilde{p} \in \mathcal{P}_k^0$, in particular for

$$\tilde{p}(z) := (1-z)^k \in \mathcal{P}_k^0 .$$

Consequently,

$$\|Mb - MA\hat{x}^{(k)}\|_2 \leqslant \|I - MA\|_2^k \, \|Mb\|_2$$

Choose M so that

$$\|I - MA\|_2 < 1$$

is as small as possible.

* We roughly aim $M \approx A^{-1}$
* But M has to be cheap to compute.

Some simple choices

(1) $M = D^{-1}$

(2) $M = (L+D)^{-1}$

(especially if A is diagonally dominant)

Here,

$$A = L + D + U$$

such that L is $n \times n$ lower triangular part of A, D is $n \times n$ diagonal part, U is $n \times n$ upper triangular part.

e.g.

$$A = \begin{bmatrix} 3 & 4 & 5 \\ 2 & 2 & 1 \\ 3 & 1 & -4 \end{bmatrix} \qquad L = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 3 & 1 & 0 \end{bmatrix}$$

$$U = \begin{bmatrix} 0 & 4 & 5 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Right preconditioning

$$Ax = b$$

$$\Longleftrightarrow$$

$$(**) \quad AM\hat{x} = b \qquad (\text{where } x = M\hat{x})$$

Apply GMRES to $(**)$, choose M so that

$$\| I - AM \|_2 < 1$$

is small.

Polynomial Preconditioning

$M = P(A)$ and such that

$$\| I - P(A) A \|_2$$

is small.

Suppose A has n linearly independent eigenvectors, indeed

$$A = V \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} V^{-1}.$$

Then we have

$$\| I - p(A) A \|_2 = \left\| V \begin{bmatrix} 1-\lambda_1 p(\lambda_1) & & \\ & \ddots & \\ & & 1-\lambda_n p(\lambda_n) \end{bmatrix} V^{-1} \right\|_2 .$$

Hence, choose $q(z) := 1 - z p(z)$ such that

(i) $|q(\lambda_j)| \approx 0 \quad j = 1, \ldots, n$, equivalently

(ii) $p(\lambda_j) \approx 1/\lambda_j \quad j = 1, \ldots, n$.

Arnoldi's Method

For large sparse eigenvalue problems

$A \in \mathbb{C}^{n\times n}$ - n very large
eigenvalues sought

Terminology and Notation

Krylov subspace

$$\mathcal{K}_k = \text{span}\{b, Ab, \ldots, A^{k-1}b\}$$

Krylov matrix

$$\underbrace{K_k}_{n\times k,\ n\gg k} = [b \ \ Ab \ \ \ldots \ \ A^{k-1}b]$$

$\underbrace{Q_k}_{n\times k} = [q_1 \ \ldots \ q_k]$ matrix whose columns form an orthonormal basis for $\mathcal{K}_k$

Description of the method

(i) Compute Q_k by Arnoldi process.

(ii) Compute the eigenvalues of

$$\tilde{H}_k = Q_k^* A Q_k$$

Meaning of $Q_k^* A Q_k$

(1) Restriction of domain of $x \longmapsto Ax$ (to $\mathcal{K}_k$)

$$Q_k^* A \underbrace{Q_k \alpha}_{v} = Q_k^* A v$$

$\in \mathcal{K}_k$

(2) Projection of range of $x \longmapsto Ax$ (onto $\mathcal{K}_k$)

$$Q_k^* A Q_k \alpha = Q_k^* \underbrace{(Q_k Q_k^*)}_{\text{orthogonal projection onto } \mathcal{K}_k} (A Q_k \alpha)$$

representation in coordinates relative to Q_k

Suppose $v \in \mathcal{K}_k$. Then $v = Q_k \alpha$ $\exists \alpha$.

coordinates relative to Q_k

$$Q_k^* v = Q_k^* Q_k \alpha = \alpha$$

Eigenvalues $\theta_1, \ldots, \theta_k$ of $\tilde{H}_k = Q_k^* A Q_k$ are called the Ritz values.

They turn out to be good estimates for the eigenvalues of A.

A simplification on $\tilde{H}_k$

Recall that Arnoldi process generates

$$\underset{n\times n}{A}\ \underset{n\times k}{Q_k} = \underset{n\times(k+1)}{Q_{k+1}}\ \underbrace{\underset{(k+1)\times k}{H_{k+1}}}_{\text{Hessenberg}}$$

$$\underbrace{Q_k^* A Q_k}_{\tilde{H}_k} = Q_k^* Q_{k+1} H_{k+1}$$

$$= \left[\begin{array}{ccc|c} 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{array}\right] \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ h_{21} & \ddots & \vdots \\ 0 & \ddots & h_{kk} \\ \vdots & & \\ 0 & & h_{(k+1)k} \end{bmatrix}$$

$\qquad\qquad k\times k \qquad k\times 1$

$$= \begin{bmatrix} h_{11} & & h_{1k} \\ h_{21} & \ddots & \\ & \ddots & \\ & h_{k(k-1)} & h_{kk} \end{bmatrix} = H_{k+1}(1:k, 1:k)$$

Pseudocode (Arnoldi's Method)

$q_1 \leftarrow b / \|b\|_2$

for $j = 1, \ldots, k$

  $v \leftarrow A q_j$

  for $\ell = 1, \ldots j$

    $h_{\ell j} \leftarrow q_j^* v$

    $v \leftarrow v - h_{\ell j} q_j$

  end

  $h_{(j+1) j} \leftarrow \|v\|_2$

  $q_{j+1} \leftarrow v / h_{(j+1)j}$

end

$\theta_1, \ldots, \theta_k \leftarrow$ eigenvalues of $H(1:k, 1:k)$

computed for instance by the QR algorithm

return

Convergence

THM Suppose K_n is full rank. The polynomial

$$p_*(z) := (z-\theta_1)(z-\theta_2) \ldots (z-\theta_k)$$

is the unique monic polynomial in P_k^m s.t.

$$\|p_*(A)\, b\|_2 = \min_{p \in P_k^m} \|p(A)\, b\|_2$$

where

$$P_k^m := \{ p : \mathbb{C} \to \mathbb{C},\ p(z) = \alpha_0 + \alpha_1 z \ldots + \alpha_{k-1} z^{k-1} + z^k \mid \alpha_0, \ldots, \alpha_{k-1} \in \mathbb{C} \}$$

is the set of monic polynomials of degree k.

Now suppose A has k distinct eigenvalues, say $\lambda_1, \ldots, \lambda_k$ $(k \leq n)$. (Assume also A has n linearly independent eigenvectors)

Without loss of generality A can be written as

$$A = V \begin{bmatrix} \lambda_1 I_{\ell_1} & & & 0 \\ & \lambda_2 I_{\ell_2} & & \\ & & \ddots & \\ 0 & & & \lambda_k I_{\ell_k} \end{bmatrix} V^{-1}$$

where ℓ_j is the algebraic multiplicity of λ_j.

Consider the polynomial $\in \mathcal{P}_k^m$

$$\tilde{p}(z) = (z-\lambda_1)(z-\lambda_2) \cdots (z-\lambda_k),$$

observe

$$\tilde{p}(A) = (A-\lambda_1 I)(A-\lambda_2 I) \cdots (A-\lambda_k I)$$

$$= \left(V \begin{bmatrix} 0 & & & 0 \\ & (\lambda_2-\lambda_1) I_{\ell_2} & & \\ & & \ddots & \\ 0 & & & (\lambda_k-\lambda_1) I_{\ell_k} \end{bmatrix} \cancel{V^{-1}} \right) \left(\cancel{V} \begin{bmatrix} (\lambda_1-\lambda_2) I_{\ell_1} & & & 0 \\ & 0 & & \\ & & \ddots & \\ 0 & & & (\lambda_k-\lambda_2) I_{\ell_k} \end{bmatrix} V^{-1} \right)$$

$$\cdots \left(V \begin{bmatrix} (\lambda_1-\lambda_k) I_{\ell_1} & & & \\ & (\lambda_2-\lambda_k) I_{\ell_2} & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} V^{-1} \right)$$

$$= V \begin{bmatrix} 0 & & & 0 \\ & (\lambda_2-\lambda_1) I_{\ell_2} & & \\ & & \ddots & \\ 0 & & & (\lambda_k-\lambda_1) I_{\ell_k} \end{bmatrix} \begin{bmatrix} (\lambda_1-\lambda_2) I_{\ell_1} & & & \\ & 0 & & \\ & & \ddots & \\ & & & (\lambda_k-\lambda_2) I_{\ell_k} \end{bmatrix} \cdots V^{-1}$$

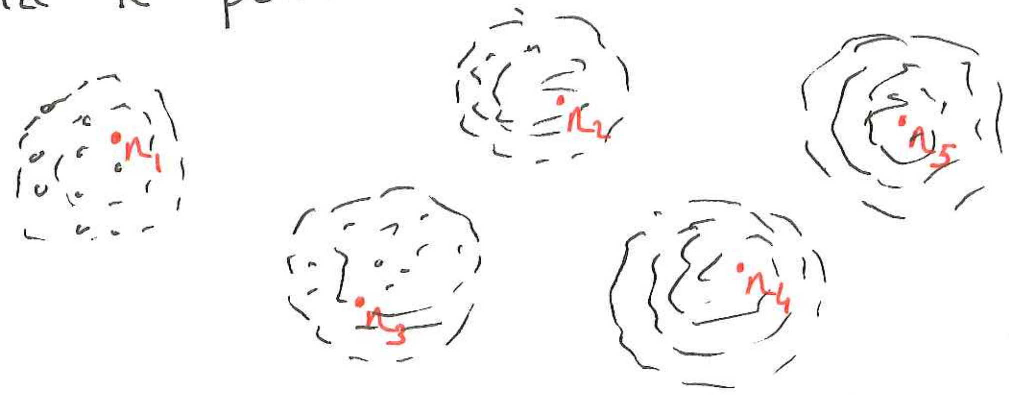
$$= 0$$

Hence, $\tilde{p}(z)$ must be the unique polynomial in $\mathcal{P}_k^m$ satisfying

$$\|\tilde{p}(A)b\|_2 = \min_{p \in \mathcal{P}_k^m} \|p(A)b\|_2 .$$

Its roots $\lambda_1, \ldots, \lambda_k$ are the Ritz values, i.e. eigenvalues of $H_k = Q_k^* A Q_k$.

Instead if A has eigenvalues clustered around k points



One ritz value in each cluster (or close to)