Outline

Key point from last lecture:

Given $A \in \mathbb{R}^{n \times n}$ (large, sparse), $\underline{b} \in \mathbb{R}^{n \times 1}$, $\underline{b} \neq \underline{0}$.

then if the grade$(\underline{b}) = m$, the minimal polynomial of $\underline{b}$ relative to A is given by

$$q(t) = t^m - \sum_{j=0}^{m-1} \alpha_j t^j \quad \ldots\ldots \quad (1)$$

with $\quad q(A)\underline{b} = \underline{0}$.

Recalling the Krylov sequence

$$\{\underline{b}, A\underline{b}, \ldots, A^{m-1}\underline{b}\}$$

the importance of (1) is that "m" is the smallest positive integer for which $A^m\underline{b}$ is expressible as a linear combination of the vectors in the Krylov sequence, i.e.,

$$A^m\underline{b} = \sum_{j=0}^{m-1} \alpha_j A^j \underline{b}.$$

In this case the solution of

$$A\underline{x} = \underline{b}$$

$$\underline{x} \in K_m(A,\underline{b}) = \text{Span}\{\underline{b}, A\underline{b}, \dots, A^{m-1}\underline{b}\}.$$

It might be that $m \ll n$!

The question is, in general, how do we determine the α_j's in (1) ?

We explored projection methods, and we will take for GMRES :

$\boxed{\text{Search Space}}$: $K_m(A, \underline{r}^{(0)})$; Constraint space : $AK_m(A, \underline{r}^{(0)})$.

$\underline{x}^{(0)}$, $\underline{r}^{(0)} = \underline{b} - A\underline{x}^{(0)}$

Arnoldi's Method

- Krylov Projection methods require a basis for the Krylov subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{r}^{(0)})$. The obvious basis

$$\left\{\mathbf{r}^{(0)}, \mathbf{A}\mathbf{r}^{(0)}, \mathbf{A}^2\mathbf{r}^{(0)}, \ldots, \mathbf{A}^{m-1}\mathbf{r}^{(0)}\right\} \tag{4}$$

  is not very attractive from a numerical point of view since the vectors tend to point more and more in the direction of the dominant eigenvector of $\mathbf{A}$ for increasing m (see the **power method** as described in MXB222).

- Instead, we generate an *orthonormal basis*. The Gram-Schmidt process takes a finite, linearly independent set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} \subset \mathbb{R}^n$ for $m \leq n$ and generates an orthonormal set of vectors that spans the same m-dimensional subspace of $\mathbb{R}^n$.

$A \in \mathbb{R}^{n \times n}$, suppose A can be diagonalised

$\{(\lambda_i, \underline{v}_i)\}_{i=1}^{n}$ is the set of eigenpairs of A.

Let $\underline{x} \in \mathbb{R}^{n \times 1}$, $\underline{x} = \sum_{j=1}^{n} c_j \underline{v}_j$, $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \dots \geq |\lambda_n|$.

$$A\underline{x} = \sum_{j=1}^{n} c_j A\underline{v}_j = \sum_{j=1}^{n} c_j \lambda_j \underline{v}_j$$

$$A^2\underline{x} = \sum_{j=1}^{n} c_j \lambda_j^2 \underline{v}_j \quad \dots \quad A^k\underline{x} = \sum_{j=1}^{n} c_j \lambda_j^k \underline{v}_j$$

$$\therefore \quad A^k\underline{x} = c_1 \lambda_1^k \underline{v}_1 + \cancel{\lambda_1^k \sum_{j=1}^{n} c_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \underline{v}_j} \to 0$$

Hence, $\lim_{k\to\infty} A^k \underline{x} = \lim_{k\to\infty} c_1 \lambda_1^k \underline{v}_1 + \underline{0}$

should be $j=2$

Gram–Schmidt Process

- Recall from MXB201 that the Gram-Schmidt process is described by the following steps:

$$\mathbf{w}_1 = \mathbf{u}_1\,, \qquad \mathbf{v}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|_2}\,, \quad \mathcal{V}_1 = \text{span}\{\underline{v}_1\}$$

$$\underline{u}_2 - P_{\mathcal{V}_1}(\underline{u}_2)$$

$$\mathbf{w}_2 = \mathbf{u}_2 - \underline{\left(\mathbf{v}_1^T\mathbf{u}_2\right)\mathbf{v}_1}\,, \qquad \mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|_2}\,, \quad \mathcal{V}_2 = \text{span}\{\underline{v}_1, \underline{v}_2\}$$

$$\mathbf{w}_3 = \mathbf{u}_3 - \left\{\left(\mathbf{v}_1^T\mathbf{u}_3\right)\mathbf{v}_1 + \left(\mathbf{v}_2^T\mathbf{u}_3\right)\mathbf{v}_2\,,\right\} \qquad \mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|_2}\,, \quad \mathcal{V}_3 = \text{span}\{\underline{v}_1, \underline{v}_2, \underline{v}_3\}$$

$$\underline{u}_3 - P_{\mathcal{V}_2}(\underline{u}_3)$$

$$\vdots \qquad\qquad \vdots$$

$$\mathbf{w}_m = \mathbf{u}_m - \sum_{k=1}^{m-1}\left(\mathbf{v}_k^T\mathbf{u}_m\right)\mathbf{v}_k\,, \qquad \mathbf{v}_m = \frac{\mathbf{w}_m}{\|\mathbf{w}_m\|_2}$$

$$- P_{\mathcal{V}_{m-1}}(\underline{u}_m)$$

- Note that generation of the m^{th} basis vector is equivalent to:

$$\mathbf{w}_m = \mathbf{u}_m - \mathbf{V}_{m-1}\mathbf{V}_{m-1}^T\mathbf{u}_m\,, \qquad \mathbf{v}_m = \frac{\mathbf{w}_m}{\|\mathbf{w}_m\|_2}$$

where $\mathbf{V}_{m-1} = [\mathbf{v}_1, \ldots, \mathbf{v}_{m-1}]$.

$$\underline{\omega}_m = \underline{u}_m - \sum_{k=1}^{m-1} \underbrace{\underline{v}_k \underline{v}_k^T}_{\text{outer Product}} \underline{u}_m$$

$$= \Big[I_n - \underbrace{\sum_{k=1}^{m-1} \underline{v}_k \underline{v}_k^T}_{V_{m-1} V_{m-1}^T} \Big] \underline{u}_m$$

$$V_{m-1} = \left[\underline{v}_1 \,\vdots\, \underline{v}_2 \,\vdots\, \cdots \,\vdots\, \underline{v}_{m-1} \right] \begin{bmatrix} \underline{v}_1^T \\ \cdots \\ \underline{v}_2^T \\ \cdots \\ \vdots \\ \cdots \\ \underline{v}_{m-1}^T \end{bmatrix}_{V_{m-1}^T}$$

$$= \sum_{k=1}^{m-1} \underline{v}_k \underline{v}_k^T .$$

$$\therefore \quad \underline{\omega}_m = \left(I_n - V_{m-1} V_{m-1}^T \right) \underline{u}_m$$

$$\underline{v}_m = \frac{\underline{\omega}_m}{\|\underline{\omega}_m\|_2} .$$

Gram–Schmidt Process

- The resulting vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ form an orthonormal set, that is, the vectors are orthogonal to one another ($\mathbf{v}_i^T \mathbf{v}_j = 0$ for all $i \neq j$) and each vector has unit length ($\|\mathbf{v}_i\|_2 = 1$ for all $i = 1, \ldots, m$).

Arnoldi's Method

- *Arnoldi's method* or the *Arnoldi process* makes use of the Gram-Schmidt process to generate an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ for the Krylov subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{r}^{(0)})$.
- The *Arnoldi process* produces the following decomposition:

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T \tag{5}$$

or equivalently:

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_{m+1}\overline{\mathbf{H}}_m\,, \qquad \overline{\mathbf{H}}_m = \left[\begin{array}{c} \mathbf{H}_m \\ \hdashline \underbrace{h_{m+1,m}\mathbf{e}_m^T} \end{array}\right] \tag{6}$$

where the columns of $\mathbf{V}_m = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m] \in \mathbb{R}^{n\times m}$ and $\mathbf{V}_{m+1} = [\mathbf{V}_m, \mathbf{v}_{m+1}] \in \mathbb{R}^{n\times(m+1)}$ form an orthonormal basis for $\mathcal{K}_m(\mathbf{A}, \mathbf{r}^{(0)})$ and $\mathcal{K}_{m+1}(\mathbf{A}, \mathbf{r}^{(0)})$, respectively.

$$\mathcal{K}_m(A, \underline{r}^{(0)}) = \text{span}\{\underline{r}^{(0)}, A\underline{r}^{(0)}, \ldots, A^{m-1}\underline{r}^{(0)}\}.$$

Apply GRAM-SCHMIDT PROCESS to the sequence $\{\underline{r}^{(0)}, A\underline{r}^{(0)}, \ldots, A^{m-1}\underline{r}^{(0)}\}$ to generate O.N.B. $\{\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_m\}$

<u>step 1</u>: $\underline{v}_1 = \dfrac{\underline{r}^{(0)}}{\beta_0}$, $\beta_0 = \|\underline{r}^{(0)}\|_2$, $\cancel{\underline{v}_1 = \dfrac{\underline{\omega}_1}{\|\underline{\omega}_1\|_2}}$

$$\mathcal{K}_1(A, \underline{r}^{(0)}) = \text{span}\{\underline{r}^{(0)}\} = \text{span}\{\underline{v}_1\}.$$

<u>step 2</u>:
$$\begin{aligned} \underline{\omega}_2 &= A\underline{v}_1 - P_{\mathcal{K}_1}(A\underline{v}_1) \\ &= A\underline{v}_1 - (\underline{v}_1^T A\underline{v}_1)\underline{v}_1 \\ &= A\underline{v}_1 - h_{11}\underline{v}_1, \quad h_{11} = \underline{v}_1^T A\underline{v}_1 \end{aligned}$$

$$\underline{v}_2 = \frac{\underline{\omega}_2}{\|\underline{\omega}_2\|_2}. \qquad \mathcal{K}_2(A, \underline{r}^{(0)}) = \text{span}\{\underline{r}^{(0)}, A\underline{r}^{(0)}\} = \text{span}\{\underline{v}_1, \underline{v}_2\}.$$

let $h_{21} = \|\underline{\omega}_2\|_2$ then

$$A\underline{v}_1 = h_{11}\underline{v}_1 + h_{21}\underline{v}_2 = [\underline{v}_1 \vdots \underline{v}_2]\begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix}$$

Set $V_1 = [\underline{v}_1]$, $V_2 = [\underline{v}_1 \vdots \underline{v}_2]$, $\overline{H}_1 = \begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix}$

then $\boxed{A V_1 = V_2 \overline{H}_1}$ ← ARNOLDI'S RELATION.

Consider, $c_1 \underline{r}^{(0)} + c_2 A \underline{r}^{(0)} = c_1 \beta_0 \underline{v}_1 + c_2 \beta_0 A \underline{v}_1$

$$= c_1\beta_0 \underline{v}_1 + c_2\beta_0 (h_{11}\underline{v}_1 + h_{21}\underline{v}_2)$$

$$= (c_1\beta_0 + c_2\beta_0 h_{11})\underline{v}_1 + c_2\beta_0 h_{21}\underline{v}_2$$

$$= b_1\underline{v}_1 + b_2\underline{v}_2, \quad b_1 = c_1\beta_0 + c_2\beta_0 h_{11}$$
$$b_2 = c_2\beta_0 h_{21}.$$

The process can be continued.

k^{th} step of process:

Given $\quad AV_{k-1} = V_k \bar{H}_{k-1}, \quad K_k(A, \underline{r}^{(0)}) = \text{span}\{\underline{v}_1, \underline{v}_2, \ldots \underline{v}_k\}$

Form $\quad \underline{\omega}_{k+1} = A\underline{v}_k - P_{K_k}(A\underline{v}_k)$

$$= A\underline{v}_k - \sum_{j=1}^{k} (\underline{v}_j^T A \underline{v}_k)\, \underline{v}_j$$

$$= A\underline{v}_k - \sum_{j=1}^{k} h_{jk}\, \underline{v}_j, \quad h_{jk} = \underline{v}_j^T A \underline{v}_k$$

Next, $\; h_{k+1,k} = \|\underline{\omega}_{k+1}\|_2, \quad \underline{v}_{k+1} = \dfrac{\underline{\omega}_{k+1}}{h_{k+1,k}} \quad (h_{k+1,k} \neq 0).$

Express $\quad A\underline{v}_k = \sum_{j=1}^{k} h_{jk}\, \underline{v}_j + h_{k+1,k}\, \underline{v}_{k+1}$

$$= [\underline{v}_1 \,\vdots\, \underline{v}_2 \,\vdots\, \cdots \,\vdots\, \underline{v}_{k+1}] \begin{pmatrix} h_{1k} \\ h_{2k} \\ \vdots \\ h_{kk} \\ \cdots \\ h_{k+1,k} \end{pmatrix}$$

$$\begin{pmatrix} h_{1k} \\ h_{2k} \\ \vdots \\ h_{kk} \end{pmatrix} = \underline{h}_k, \qquad \begin{pmatrix} \underline{h}_k \\ \cdots \\ h_{k+1,k} \end{pmatrix}$$

$$A\underline{v}_k = V_{k+1} \begin{pmatrix} \underline{h}_k \\ \cdots \\ h_{k+1,k} \end{pmatrix}$$

$$
\begin{aligned}
A \underbrace{\left[V_{k-1} \;\vdots\; \underline{v}_k \right]}_{V_k} &= \left[A V_{k-1} \;\vdots\; A \underline{v}_k \right] \\
&= \left[V_k \overline{H}_{k-1} \;\vdots\; V_{k+1} \begin{pmatrix} \underline{h}_k \\ \hline h_{k+1,k} \end{pmatrix} \right] \\
&= V_{k+1} \left(\begin{array}{c:c} \overline{H}_{k-1} & \underline{h}_k \\ \hdashline \underline{0}^T & h_{k+1,k} \end{array} \right)
\end{aligned}
$$

$$
\boxed{A V_k = V_{k+1} \overline{H}_k}
$$

As part of induction step, need to show

$$
\begin{aligned}
K_{k+1}(A, \underline{r}^{(0)}) &= \operatorname{span}\{ \underline{r}^{(0)}, A \underline{r}^{(0)}, \dots, A^k \underline{r}^{(0)} \} \\
&= \operatorname{span}\{ \underline{v}_1, \underline{v}_2, \dots, \underline{v}_{k+1} \}.
\end{aligned}
$$

Consider $\underbrace{c_1 \underline{r}^{(0)} + c_2 A\underline{r}^{(0)} + \cdots + c_{k-1} A^{k-1}\underline{r}^{(0)}}_{\text{we would have already shown}} + c_k A^k \underline{r}^{(0)}$ (LHS)

$$= \sum_{i=1}^{k} b_i \underline{v}_i + c_k A^k \underline{r}^{(0)} = \sum_{i=1}^{k+1} \tilde{b}_i v_i$$

$\underline{r}^{(0)} = \beta_0 \underline{v}_1$, $\quad A\underline{r}^{(0)} = \beta_0 A V_1 = \beta_0 V_2 \overline{H}_1$

$A^2 \underline{r}^{(0)} = \beta_0 A V_2 \overline{H}_1 = \beta_0 V_3 \overline{H}_2 \overline{H}_1$

observe $\overline{H}_2 \overline{H}_1 = \underbrace{\underset{\overline{H}_2,\ 3\times 2}{\begin{pmatrix} \boxed{\begin{matrix} h_{11} \\ h_{21} \end{matrix}} & h_{12} \\ & h_{22} \\ 0 & h_{32} \end{pmatrix}} \underset{\overline{H}_1,\ 2\times 1}{\begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix}}} = \underline{z}_3$

Continue $A^k \underline{r}^{(0)} = \beta_0 V_{k+1} \underbrace{\overline{H}_k \overline{H}_{k-1} \cdots \overline{H}_2 \overline{H}_1}_{\underline{z}_{k+1}}$.

should include $\boxed{c_k}$

Hence, $\text{LHS} = \sum_{i=1}^{k} b_i \underline{v}_i + \beta_0 c_k V_{k+1} \underline{z}_{k+1}$

$$= V_k \underline{b} + \beta_0 \boxed{c_k} V_{k+1} \underline{z}_{k+1} = V_{k+1} \begin{pmatrix} \underline{b} + \beta_0 c_k \tilde{\underline{z}}_{k+1} \\ \hline \beta_0 c_k (\underline{z}_{k+1})_{k+1} \end{pmatrix}$$

Arnoldi's Method

which satisfies the recurrence relation:

$$\overline{\mathbf{H}}_m = \left[\begin{array}{c|c} \overline{\mathbf{H}}_{m-1} & \begin{array}{c} h_{1,m} \\ h_{2,m} \\ h_{3,m} \\ \vdots \\ h_{m,m} \end{array} \\ \hline & h_{m+1,m} \end{array}\right]$$

- The entries of $\mathbf{H}_m$ and $\overline{\mathbf{H}}_m$ are defined by the Arnoldi algorithm.

Conclusion:

$$K_{k+1}(A, \underline{r}^{(0)}) = \text{span}\{\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_{k+1}\}.$$

$$\overline{H}_m = \left(\begin{array}{c} H_m \\ \hline 0, 0, \ldots 0, h_{m+1,m} \end{array}\right) = \left(\begin{array}{c} H_m \\ \hline h_{m+1,m}\, \underline{e}_m^T \end{array}\right) \in \mathbb{R}^{m+1 \times m}.$$

$$\underline{e}_m^T = (0, 0, \ldots, 1)^T.$$

ARNOLDI RELATION:

$$V_m^T A V_m = V_{m+1} \overline{H}_m = \left(V_m \,\vdots\, \underline{v}_{m+1}\right) \left(\begin{array}{c} H_m \\ \hline h_{m+1,m}\, \underline{e}_m^T \end{array}\right)$$

$$= V_m^T V_m H_m + h_{m+1,m} V_m^T \underline{v}_{m+1} \underline{e}_m^T.$$

$$\cancel{\begin{pmatrix} \underline{v}_1^T \underline{v}_{m+1} \\ \underline{v}_2^T \underline{v}_{m+1} \\ \vdots \\ \underline{v}_m^T \underline{v}_{m+1} \end{pmatrix}} = \underline{0}$$

$$\therefore \boxed{V_m^T A V_m = H_m}.$$