Arnoldi Method Theory

Professor Tim Moroney

School of Mathematical Sciences, QUT

Review: The Arnoldi process

$\beta = \|b\|, \qquad v_1 = b/\beta$

For $j = 1, 2, \dots m$

$$h_{ij} = \langle Av_j, v_i \rangle, \qquad i = 1, 2, \dots, j$$

$$w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j}$$

End

At any step m the columns of $V_m = [v_1 \; v_2 \cdots v_m]$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2b, ..., A^{m-1}b\}$.

We have also generated the matrix $\overline{H}_m$ whose $(i, j)th$ entry is h_{ij}.

Review: Arnoldi relation 1

$$Av_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m$$

$$A \left[\; v_j \; \right] = \left[\; v_1 \;\middle|\; v_2 \;\middle|\; \cdots \;\middle|\; v_{j+1} \;\middle|\; \cdots \;\middle|\; v_{m+1} \; \right] \begin{bmatrix} h_{1,j} \\ h_{2,j} \\ \vdots \\ h_{j+1,j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$AV_m(:,j) = V_{m+1}\overline{H}_m(:,j), \qquad j = 1, \ldots, m$$

Altogether,

$$\boxed{AV_m = V_{m+1}\overline{H}_m}$$

Review: Arnoldi relation 2

$$AV_m = V_{m+1}\overline{H}_m$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{ccccc} & & H_m & & \\ \hline 0 & 0 & \cdots & 0 & h_{m+1,m} \end{array}\right]$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline h_{m+1,m}\, {e_m}^T \end{array}\right]$$

$$\boxed{AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T}$$

Review: Arnoldi relation 3

$$AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T$$

$$\begin{aligned} V_m^* A V_m &= V_m^* V_m H_m + h_{m+1,m}\, V_m^*\, v_{m+1}\, {e_m}^T \\ &= I_m H_m + h_{m+1,m}\, 0\, {e_m}^T \\ &= H_m \end{aligned}$$

$$\boxed{V_m^* A V_m = H_m}$$

Review: The Arnoldi relations

$$
\begin{aligned}
AV_m &= V_{m+1}\overline{H}_m \\
AV_m &= V_m H_m + h_{m+1,m}\, v_{m+1}\, e_m{}^T \\
V_m^* A V_m &= H_m
\end{aligned}
$$

- A is $n \times n$, where n is typically large
- V_m is $n \times m$: m columns of n-dimensional vectors
- V_{m+1} is $n \times (m+1)$: it's just V_m but one iteration further along
- $\overline{H}_m$ is $(m+1) \times m$ (rectangular) upper Hessenberg
- H_m is $m \times m$ (square) upper Hessenberg: it's just $\overline{H}_m$ without the bottom row

For eigenvalue problems H_m is the relevant matrix.

Review: Arnoldi method

The projection of A onto $\mathcal{K}_m$ is simply

$$\boxed{A_{\mathcal{K}_m} = H_m}$$

We compute eigenvalues $\theta_i^{(m)}$ and eigenvectors $y_i^{(m)}$ of H_m

$$H_m\, y_i^{(m)} = \theta_i^{(m)} y_i^{(m)}$$

This is a small, m-dimensional eigenproblem, easy to solve.

Then map back to n-dimensions with

$$u_i^{(m)} = V_m y_i^{(m)}$$

The pair $(\theta_i^{(m)}, u_i^{(m)})$ are the Ritz values and Ritz vectors, respectively.

Ideally they are good approximations of true eigenvalue/eigenvector pairs of A.

Review: Arnoldi residual

$$
\begin{aligned}
A\,u^{(m)} - \theta^{(m)}\,u^{(m)} &= h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\
\left\|A\,u^{(m)} - \theta^{(m)}\,u^{(m)}\right\| &= \left\|h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)}\right\| \\
&= \left|h_{m+1,m}\right|\,\left\|v_{m+1}\right\|\,\left|{e_m}^T\,y^{(m)}\right| \\
&= \left|h_{m+1,m}\right|\,\left|{e_m}^T\,y^{(m)}\right| \\
&= \left|h_{m+1,m}\right|\,\left|{y^{(m)}}_m\right|
\end{aligned}
$$

Note the "lucky breakdown" in the Arnoldi process occurs when $h_{m+1,m} = 0$, so the Ritz values and vectors would in fact be true eigenvalues and eigenvectors of A in that case.

Arnoldi theory

Our goal will be to show that the Arnoldi method minimises the expression

$$\|p_m(A)b\|$$

over all monic polynomials of degree m.

Specifically, the characteristic polynomial of H_m, $p_m(z) = \det(zI - H_m)$ is precisely the minimiser.

To whatever extent the Arnoldi method is useful for finding eigenvalues, it is a side-effect of this result.

We have a few steps to build up first.

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

$$b = \beta v_1 = \beta V_m e_1$$

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

$$b = \beta v_1 = \beta V_m e_1$$

$$Ab = A(\beta V_m e_1) = \beta A V_m e_1$$

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

$$b = \beta v_1 = \beta V_m e_1$$

$$\begin{aligned} Ab = A(\beta V_m e_1) &= \beta A V_m e_1 \\ &= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T)e_1 \end{aligned}$$

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

$$b = \beta v_1 = \beta V_m e_1$$

$$\begin{aligned} Ab = A(\beta V_m e_1) &= \beta A V_m e_1 \\ &= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) e_1 \\ &= \beta V_m H_m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T e_1 \end{aligned}$$

Formulas for A^k b

The theoretical tool involves $p_m(A)b$, a monic polynomial of A times b.

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

So we'd better work out what Ab, A^2b, A^3b, etc. look like.

We emphasise that this is a theoretical tool. Arnoldi does not actually form these products.

$$b = \beta v_1 = \beta V_m e_1$$

$$\begin{aligned} Ab = A(\beta V_m e_1) &= \beta A V_m e_1 \\ &= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) e_1 \\ &= \beta V_m H_m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T e_1 \\ &= \beta V_m H_m e_1 \end{aligned}$$

Formulas for A^k b

$$A^2 b = A(\beta V_m H_m e_1) = \beta A V_m H_m e_1$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m e_1
\end{aligned}
$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} e_m{}^T) H_m e_1 \\
&= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} e_m{}^T (H_m e_1)
\end{aligned}
$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} e_m{}^T) H_m e_1 \\
&= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} e_m{}^T (H_m e_1) \\
&= \beta V_m H_m^2 e_1
\end{aligned}
$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} e_m{}^T) H_m e_1 \\
&= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} e_m{}^T (H_m e_1) \\
&= \beta V_m H_m^2 e_1
\end{aligned}
$$

$$
A^3 b = A(\beta V_m H_m^2 e_1) = \beta A V_m H_m^2 e_1
$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m e_1 \\
&= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T (H_m e_1) \\
&= \beta V_m H_m^2 e_1
\end{aligned}
$$

$$
\begin{aligned}
A^3 b = A(\beta V_m H_m^2 e_1) &= \beta A V_m H_m^2 e_1 \\
&= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m^2 e_1
\end{aligned}
$$

Formulas for A^k b

$$\begin{aligned} A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\ &= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m e_1 \\ &= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T (H_m e_1) \\ &= \beta V_m H_m^2 e_1 \end{aligned}$$

$$\begin{aligned} A^3 b = A(\beta V_m H_m^2 e_1) &= \beta A V_m H_m^2 e_1 \\ &= \beta (V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m^2 e_1 \\ &= \beta V_m H_m^3 e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T (H_m^2 e_1) \end{aligned}$$

Formulas for A^k b

$$
\begin{aligned}
A^2 b = A(\beta V_m H_m e_1) &= \beta A V_m H_m e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m e_1 \\
&= \beta V_m H_m^2 e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T (H_m e_1) \\
&= \beta V_m H_m^2 e_1
\end{aligned}
$$

$$
\begin{aligned}
A^3 b = A(\beta V_m H_m^2 e_1) &= \beta A V_m H_m^2 e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) H_m^2 e_1 \\
&= \beta V_m H_m^3 e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T (H_m^2 e_1) \\
&= \beta V_m H_m^3 e_1
\end{aligned}
$$

Formulas for A^k b

$$
\begin{aligned}
A^{m-1}b &= A(\beta V_m {H_m}^{m-2} e_1) \\
&= \beta A V_m {H_m}^{m-2} e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) {H_m}^{m-2} e_1 \\
&= \beta V_m {H_m}^{m-1} e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T ({H_m}^{m-2} e_1) \\
&= \beta V_m {H_m}^{m-1} e_1
\end{aligned}
$$

Formulas for A^k b

$$
\begin{aligned}
A^{m-1}b &= A(\beta V_m {H_m}^{m-2} e_1) \\
&= \beta A V_m {H_m}^{m-2} e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) {H_m}^{m-2} e_1 \\
&= \beta V_m {H_m}^{m-1} e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T ({H_m}^{m-2} e_1) \\
&= \beta V_m {H_m}^{m-1} e_1
\end{aligned}
$$

$$
\begin{aligned}
A^{m}b &= A(\beta V_m {H_m}^{m-1} e_1) \\
&= \beta A V_m {H_m}^{m-1} e_1 \\
&= \beta(V_m H_m + h_{m+1,m}\, v_{m+1} {e_m}^T) {H_m}^{m-1} e_1 \\
&= \beta V_m {H_m}^{m} e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1
\end{aligned}
$$

Formulas for A^k b

So for $k < m$ we have the neat result

$$\boxed{A^k b = \beta V_m {H_m}^k e_1, \qquad k < m}$$

And by linearity for all polynomials $p_{m-1}(A)$ of degree up to $m-1$

$$\boxed{p_{m-1}(A) b = \beta V_m p_{m-1}(H_m) e_1}$$

But the result for $k = m$ is not as simple:

$$\boxed{A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1}$$

Formulas for A^k b

But there is another way to formulate these results

$$A^k b = \beta V_m {H_m}^k e_1, \qquad k < m$$
$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m} v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Formulas for A^k b

But there is another way to formulate these results

$$A^k b = \beta V_m {H_m}^k e_1, \qquad k < m$$
$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m} \, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Multiply through by V_m^* to recover a neat formula up to $k = m$

Formulas for A^k b

But there is another way to formulate these results

$$A^k b = \beta V_m {H_m}^k e_1, \qquad k < m$$
$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m} v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Multiply through by V_m^* to recover a neat formula up to $k = m$

$$V_m^* A^k b = \beta V_m^* V_m {H_m}^k e_1, \qquad k < m$$
$$V_m^* A^m b = \beta V_m^* V_m {H_m}^m e_1 + \beta h_{m+1,m} V_m^* v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Formulas for A^k b

But there is another way to formulate these results

$$A^k b = \beta V_m {H_m}^k e_1, \qquad k < m$$
$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Multiply through by V_m^* to recover a neat formula up to $k = m$

$$V_m^* A^k b = \beta V_m^* V_m {H_m}^k e_1, \qquad k < m$$
$$V_m^* A^m b = \beta V_m^* V_m {H_m}^m e_1 + \beta h_{m+1,m}\, V_m^* v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

In either case the result simplifies to

$$\boxed{V_m^* A^k b = \beta {H_m}^k e_1, \qquad k \leq m}$$

And by linearity for all polynomials $p_m(A)$ of degree up to m

$$\boxed{V_m^* p_m(A) b = \beta\, p_m(H_m) e_1}$$

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

and so

$$p_m(A)b \in \mathcal{K}_{m+1}(A, b)$$

n.b. $\mathcal{K}_{m+1}$, not $\mathcal{K}_m$.

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

and so

$$p_m(A)b \in \mathcal{K}_{m+1}(A, b)$$

n.b. $\mathcal{K}_{m+1}$, not $\mathcal{K}_m$. So

$$p_m(A)b = V_{m+1}V_{m+1}^* p_m(A)b$$

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

and so

$$p_m(A)b \in \mathcal{K}_{m+1}(A, b)$$

n.b. $\mathcal{K}_{m+1}$, not $\mathcal{K}_m$. So

$$\begin{aligned} p_m(A)b &= V_{m+1}V_{m+1}^* p_m(A)b \\ &= V_{m+1}\begin{bmatrix} V_m & v_{m+1} \end{bmatrix}^* p_m(A)b \end{aligned}$$

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

and so

$$p_m(A)b \in \mathcal{K}_{m+1}(A, b)$$

n.b. $\mathcal{K}_{m+1}$, not $\mathcal{K}_m$. So

$$\begin{aligned} p_m(A)b &= V_{m+1}V_{m+1}^* p_m(A)b \\ &= V_{m+1}\begin{bmatrix} V_m & v_{m+1} \end{bmatrix}^* p_m(A)b \\ &= V_{m+1}\begin{bmatrix} V_m^* p_m(A)b \\ v_{m+1}^* p_m(A)b \end{bmatrix} \end{aligned}$$

Back to $p_m(A)b$

For a general monic polynomial p_m, we have

$$p_m(A)b = A^m b + c_{m-1}A^{m-1}b + \ldots + c_1 Ab + c_0 b$$

and so

$$p_m(A)b \in \mathcal{K}_{m+1}(A, b)$$

n.b. $\mathcal{K}_{m+1}$, not $\mathcal{K}_m$. So

$$\begin{aligned} p_m(A)b &= V_{m+1}V_{m+1}^* p_m(A)b \\ &= V_{m+1}\begin{bmatrix} V_m & v_{m+1} \end{bmatrix}^* p_m(A)b \\ &= V_{m+1}\begin{bmatrix} V_m^* p_m(A)b \\ v_{m+1}^* p_m(A)b \end{bmatrix} \\ &= V_{m+1}\begin{bmatrix} w \\ \rho \end{bmatrix} \end{aligned}$$

Back to $p_m(A)b$

$$
\begin{aligned}
p_m(A)b &= V_{m+1}V_{m+1}^* p_m(A)b \\
&= V_{m+1}\begin{bmatrix} V_m & v_{m+1} \end{bmatrix}^* p_m(A)b \\
&= V_{m+1}V_{m+1}\begin{bmatrix} V_m^* \, p_m(A)b \\ v_{m+1}^* \, p_m(A)b \end{bmatrix} \\
&= V_{m+1}\begin{bmatrix} w \\ \rho \end{bmatrix}
\end{aligned}
$$

where

$$
w = V_m^* p_m(A)b \qquad \text{and} \qquad \rho = v_{m+1}^* p_m(A)b
$$

Now, since $V_{m+1}^* V_{m+1} = I$,

$$
\|p_m(A)b\|^2 = \left\| \begin{bmatrix} w \\ \rho \end{bmatrix} \right\|^2 = \|w\|^2 + |\rho|^2
$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m} \, {e_m}^T {H_m}^{m-1} e_1$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned} \rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b) \end{aligned}$$

since p_m is monic.

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned}\rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b)\end{aligned}$$

since p_m is monic. Take the two terms separately.
First term: involves $A^m b$, so use the result

$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned}\rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b)\end{aligned}$$

since p_m is monic. Take the two terms separately.
First term: involves $A^m b$, so use the result

$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

So

$$v_{m+1}^* A^m b = v_{m+1}^* (\beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1)$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned}\rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b)\end{aligned}$$

since p_m is monic. Take the two terms separately.
First term: involves $A^m b$, so use the result

$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

So

$$\begin{aligned}v_{m+1}^* A^m b &= v_{m+1}^* (\beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1) \\ &= \beta \boxed{v_{m+1}^* V_m} {H_m}^m e_1 + \beta h_{m+1,m} \boxed{v_{m+1}^* v_{m+1}} {e_m}^T {H_m}^{m-1} e_1\end{aligned}$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned} \rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b) \end{aligned}$$

since p_m is monic. Take the two terms separately.
First term: involves $A^m b$, so use the result

$$A^m b = \beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1$$

So

$$\begin{aligned} v_{m+1}^* A^m b &= v_{m+1}^* (\beta V_m {H_m}^m e_1 + \beta h_{m+1,m}\, v_{m+1} {e_m}^T {H_m}^{m-1} e_1) \\ &= \beta \boxed{v_{m+1}^* V_m} {H_m}^m e_1 + \beta h_{m+1,m} \boxed{v_{m+1}^* v_{m+1}} {e_m}^T {H_m}^{m-1} e_1 \\ &= \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1 \end{aligned}$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m} \, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned} \rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b) \end{aligned}$$

since p_m is monic. Take the two terms separately.

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m} \, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned} \rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b) \end{aligned}$$

since p_m is monic. Take the two terms separately.

Second term: involves $q_{m-1}(A)b$ so use the result

$$q_{m-1}(A) b = \beta V_m q_{m-1}(H_m) e_1$$

Main Lemma

For any monic polynomial p_m, ρ is fixed and given by

$$\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1$$

Proof: Begin with

$$\begin{aligned}\rho &= v_{m+1}^* p_m(A) b \\ &= v_{m+1}^* (A^m b + q_{m-1}(A) b)\end{aligned}$$

since p_m is monic. Take the two terms separately.

Second term: involves $q_{m-1}(A)b$ so use the result

$$q_{m-1}(A)b = \beta V_m q_{m-1}(H_m) e_1$$

So

$$v_{m+1}^* q_{m-1}(A) b = v_{m+1}^* \beta V_m q_{m-1}(H_m) e_1 = 0$$

Hence the result

$$\boxed{\rho = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1}$$

Putting it all together

Recall we derived

$$\|p_m(A)b\|^2 = \|w\|^2 + |\rho|^2$$

and we have now found that ρ is fixed for any monic p_m. What about w?

Putting it all together

Recall we derived

$$\|p_m(A)b\|^2 = \|w\|^2 + |\rho|^2$$

and we have now found that ρ is fixed for any monic p_m. What about w?

$$w = V_m^* p_m(A) b = \beta p_m(H_m) e_1$$

and the choice that minimises this (by making it zero!) is

$$p_m(z) = \det(zI - H_m)$$

i.e. p_m is the characteristic polynomial of H_m.

Putting it all together

Recall we derived

$$\|p_m(A)b\|^2 = \|w\|^2 + |\rho|^2$$

and we have now found that ρ is fixed for any monic p_m. What about w?

$$w = V_m^* p_m(A) b = \beta p_m(H_m) e_1$$

and the choice that minimises this (by making it zero!) is

$$p_m(z) = \det(zI - H_m)$$

i.e. p_m is the characteristic polynomial of H_m.

Conclusion: the Arnolid method generates the polynomial p_m (as the characteristic polynomial of H_m) that minimises $\|p_m(A)b\|$ over all monic polynomials of degree m, and the minimum value is

$$|\rho| = \beta h_{m+1,m}\, {e_m}^T {H_m}^{m-1} e_1 = \beta h_{m+1,m} \prod_{j=1}^{m} h_{j+1,j}.$$

Discussion

This result says nothing about eigenvalues! Evidently Arnoldi is solving a polynomial approximation problem.

Why does it produce good eigenvalue estimates?

If the "goal" is to make $\|p_m(A)b\|$ small, then a good strategy could be to make $p_m(z)$ small at the eigenvalues of A.

Remember, $p_m(A) = X \operatorname{diag}(p_m(\lambda_i))X^{-1}$ so you win by making $p_m(\lambda_i)$ as small as possible.

But p_m is the characteristic polynomial of H_m, so its roots are precisely the Ritz values θ_i: $p_m(\theta_i) = 0$.

Hence Arnoldi will arrange for the Ritz values θ_i to align with the eigenvalues λ_i as "best as possible".

Discussion

There are only m Ritz values θ_i and in practice we typically have $m \ll n$.

Arnoldi doesn't have enough degrees of freedom to make $p_m(\lambda_i) \approx 0$ for all $i = 1, \dots n$.

Which eigenvalues λ_i will it prioritise to approximate well by Ritz values?

$$p_m(A) = X \operatorname{diag}(p_m(\lambda_i)) X^{-1}$$
$$\|p_m(A)b\| \leq \|X\| \, \|\operatorname{diag}(p_m(\lambda_i))\| \, \|X^{-1}\| \, \|b\| \leq \kappa(X) \|b\| \max_i |p_m(\lambda_i)|$$

so it can't afford to have any $|p_m(\lambda_i)|$ too large.

Conclusion: Arnoldi will tend to converge to "exterior" eigenvalues first (since otherwise $p_m(\lambda_i)$ could be enormous there).

Conclusion

We have presented the Arnoldi method for approximating eigenvalues and eigenvectors of sparse matrices.

The method relies on building an onorthonormal basis for the Krylov subspace $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2b, ..., A^{m-1}b\}$.

This yields the relation (among others) $V_m^* A V_m = H_m$

We calculate eigenvalues of H_m (Ritz values), and map the associated eigenvectors $y^{(m)}$ up to n-dimensions using $u^{(m)} = V_m y^{(m)}$ (Ritz vectors).

Check residuals $|h_{m+1,m}|\,|{y^{(m)}}_m|$ to see which have converged.

All the while, the method is really optimising $\|p_m(A)b\|$ over monic polynomials of degree m: the optimal choice is $p_m(z) = \det(zI - H_m)$.

If the Ritz values are accurate, it is a side effect of this true objective.