Arnoldi Iteration

Krylov subspace methods are the basis for the iterative methods for eigenvalue problems (and also for solving linear systems)

An important advantage: Krylov methods do not deal directly with A, but rather with matrix-vector products involving A.

This is particularly helpful when A is large, since matrix-vector multiplications are relatively cheap.

Preliminaries:

We define a matrix as being in Hessenberg form in the following way:

$\longrightarrow$ A is called upper-Hessenberg if $a_{ij} = 0$ for all $i > j+1$

$\longrightarrow$ A is called lower-Hessenberg if $a_{ij} = 0$ for all $j > i+1$

The Arnoldi iteration is a Krylov subspace iterative method that reduces A to upper-Hessenberg form.

Arnoldi Iteration :

For $A \in \mathbb{C}^{n \times n}$, we want to compute $A = QHQ^*$, where H is upper Hessenberg, and Q is unitary, i.e., $QQ^* = I$.

However, we suppose that n is huge; therefore, we do not try to compute the full factorization.

Instead, let us consider just the first $m \ll n$ columns of the factorization $AQ = QH$.

Therefore, on the left-hand side, we only need the matrix $Q_m \in \mathbb{C}^{n \times m}$.

$$Q_m = \left[\begin{array}{c|c|c|c} & & & \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_m \\ & & & \end{array}\right]$$

On the right-hand side, we only need the first m columns of H. More specifically, due to upper-Hessenberg structure, we only need $\tilde{H}_m$, which is the $(m+1) \times m$ upper-left section of H:

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

$\tilde{H}_m$ only interacts with the first $m+1$ columns of Q; hence we have

$$AQ_m = Q_{m+1}\tilde{H}_m$$

The m^{th} column can be written as

$$Aq_m = h_{1m}\vec{q}_1 + \cdots + h_{mm}\vec{q}_m + h_{m+1,m}\vec{q}_{m+1}.$$

Or equivalently,

$$\vec{q}_{m+1} = (Aq_m - h_{1m}\vec{q}_1 - \cdots - h_{mm}\vec{q}_m)/h_{m+1,m}$$

Remark ▽: Arnoldi iteration is just the Gram-Schmidt method that constructs h_{ij} and (orthonormal vectors q_i).

Arnoldi Iteration

1: choose b arbitrarily, then $q_1 = b/\|b\|_2$
2: **for** $m = 1, 2, 3, \ldots$ **do**
3: $\quad v = Aq_m$
4: $\quad$ **for** $j = 1, 2, \ldots, m$ **do**
5: $\quad\quad h_{jm} = q_j^* v$
6: $\quad\quad v = v - h_{jm} q_j$
7: $\quad$ **end for**
8: $\quad h_{m+1,m} = \|v\|_2$
9: $\quad q_{m+1} = v/h_{m+1,m}$
10: **end for**

This is akin to the modified Gram–Schmidt method because the updated vector v is used in line 5 (vs. the "raw vector" Aq_m)

Also, we only need to evaluate Aq_m and perform some vector operations in each iteration

Q: How do we find eigenvalues from the Arnoldi iteration?

A: Let $H_m = Q_m^* A Q_m$ be the $m \times m$ matrix obtained by removing the last row from $\tilde{H}_m$

At each step m, we compute the eigenvalues of the Hessenberg matrix H_m (This is how eigs in Python/MATLAB works)

This provides estimates for m eigenvalues/eigenvectors ($m \ll n$) called Ritz values, Ritz vectors respectively.

Just as with the power method, the Ritz values will typically converge to extreme eigenvalues of the spectrum.

WHY?

We now examine why eigenvalues of H_m approximate extreme eigenvalues of A.

Let $\mathbb{P}^m_{\text{monic}}$ denote the monic polynomials of degree m.

Theorem: The characteristic polynomial of H_m is the unique solution of the approximation $p \in \mathbb{P}^m_{\text{monic}}$ such that

$$\| p(A)\vec{b} \|_2 = \text{minimum}.$$

This theorem implies that Ritz values, i.e., eigenvalues of H_m, are the roots of the optimal polynomial

$$p^* = \underset{p \in \mathbb{P}^m_{\text{monic}}}{\operatorname{argmin}} \| p(A)\vec{b} \|_2$$

Now, let's consider what p^* should look like in order to minimize $\| p(A)\vec{b} \|_2$.

We can illustrate the important ideas with a simple case. Suppose that A has only $m \ll n$ distinct eigenvalues, and $\vec{b} = \sum_{i=1}^{m} \alpha_i \vec{v}_i$, where $\vec{v}_i$ is an eigenvector corresponding to λ_i.

Then, for $p \in \mathbb{P}^m_{\text{monic}}$, we have

$$p(x) = c_0 + c_1 x + c_2 x^2 + \ldots + x^m$$

for some coefficients $c_0, c_1, \ldots, c_{m-1}$.

Applying this polynomial to a matrix A gives

$$\begin{aligned} p(A)\vec{b} &= (c_0 I + c_1 A + c_2 A^2 + \ldots + A^m)\vec{b} \\ &= \sum_{i=1}^{m} (c_0 I + c_1 A + c_2 A^2 + \ldots + A^m)\alpha_i \vec{v}_i \end{aligned}$$

$$= \sum_{i=1}^{m} \left(c_0 + c_1 \lambda_i + c_2 \lambda_i^2 + \dots + \lambda_i^m \right) \alpha_i \vec{v}_i$$

$$= \sum_{i=1}^{m} \alpha_i p(\lambda_i) \vec{v}_i$$

Then the polynomial $p^* \in \mathbb{P}^m_{\text{monic}}$ with roots at $\lambda_1, \lambda_2, \dots, \lambda_m$ minimizes $\| p(A)\vec{b} \|_2$, since $\| p^*(A)\vec{b} \|_2 = 0$

This means that the Ritz values after m iterations are then exactly the m distinct eigenvalues of A.