Introduction to Krylov Subspace Methods

- Define $k$-dimensional Krylov subspace matrix
  \[ K_k = \begin{bmatrix} x_0 & Ax_0 & \cdots & A^{k-1}x_0 \end{bmatrix} \]

Krylov subspace methods seek to best use the information in $K_k$ to solve eigenvalue problems (or linear systems/least squares problems).

- Show that $K_n^{-1}AK_n$ is a companion matrix $C$:
  
  Letting $k_n^{(i)} = A^{i-1}x$, we observe that
  
  \[
  AK_n = \begin{bmatrix} Ak_1^{(1)} & \cdots & Ak_1^{(n-1)} & Ak_1^{(n)} \end{bmatrix} = \begin{bmatrix} k_1^{(2)} & \cdots & k_1^{(n)} & Ak_1^{(n)} \end{bmatrix},
  \]

  therefore premultiplying by $K_n^{-1}$ transforms the first $n - 1$ columns of $AK_n$ into the last $n - 1$ columns of $I$,

  \[
  K_n^{-1}AK_n = \begin{bmatrix} K_n^{-1}k_1^{(2)} & \cdots & K_n^{-1}k_1^{(n)} & K_n^{-1}Ak_1^{(n)} \end{bmatrix} = \begin{bmatrix} e_2 & \cdots & e_n & K_n^{-1}Ak_1^{(n)} \end{bmatrix}
  \]
Krylov Subspaces

Given $QR = K_k$, we obtain an orthonormal basis for the Krylov subspace,

$$\mathcal{K}_k(A, x_0) = \text{span}(Q) = \{\rho(A)x_0 : \text{deg}(\rho) < k\}$$

Consider whether $k - 1$ steps of power iteration starting from $x_0$ lead to an approximation in the Krylov subspace, also consider QR (subspace) iteration:

The approximation obtained from $k - 1$ steps of power iteration starting from $x_0$ is given by the Rayleigh-quotient of $y = A^kx_0$. This vector is within the Krylov subspace, $y \in \mathcal{K}_k(A, x_0)$. 
Given $QR = K_k$, we obtain an orthonormal basis for the Krylov subspace and $H_k = Q^T AQ$ which minimizes $||AQ - QH||_2$:

The solution to the linear least squares problem $QX \approx AQ$ is $X = Q^T AQ = H$

$H_k$ is Hessenberg, because the companion matrix $C_k$ is Hessenberg:

$$H_k = Q^T AQ = RK_k^{-1} A K_k R^{-1} = RC_k R^{-1}$$
Rayleigh-Ritz Procedure

- The eigenvalues/eigenvectors of $H_k$ are the Ritz values/vectors:

$$H_k = XD X^{-1}$$

*eigenvalue approximations based on Ritz vectors $X$ are given by $QX$*

- The Ritz vectors and values are the *ideal approximations* of the actual eigenvalues and eigenvectors based on only $H_k$ and $Q$:

*The Ritz value with greatest magnitude $\lambda_{\text{max}}(H)$ will be the maximum Rayleigh quotient of any vector in $\mathcal{K}_k = \text{span}((Q))$,*

$$\max_{x \in \text{span}(Q)} \frac{x^T A x}{x^T x} = \max_{y \neq 0} \frac{y^T Q^T AQ y}{y^T y} = \max_{y \neq 0} \frac{y^T H y}{y^T y} = \lambda_{\text{max}}(H),$$

*the quality of the approximation can also be shown to be optimal for other eigenvalues/eigenvectors.*
Arnoldi Iteration

- Arnoldi iteration computes $H$ directly using the recurrence $q_i^T A q_j = h_{ij}$:

  We have that

  $$q_i^T A q_j = q_i^T (Q H_n Q^T) q_j = e_i H_n e_j = h_{ij}$$

- After each matrix-vector product, orthogonalization is done with respect to each previous vector:

  Given $u_j = A q_j$, compute $h_{ij} = q_i^T u_j$ for each $i \leq j$, forming a column of the $H$ matrix at a time
Lanczos Iteration

- Lanczos iteration provides a method to reduce a symmetric matrix to tridiagonal matrix:

  *Arnoldi iteration on a symmetric matrix, will result in an upper-Hessenberg matrix $H$ as before, except that it must also be symmetric, since*

  $$H^T = (Q^T A Q)^T = Q^T A^T Q = Q^T A B Q = H,$$

  *which implies that $H$ must be tridiagonal.*

- After each matrix-vector product, it suffices to orthogonalize with respect to two previous vectors:

  *Since $h_{ij} = 0$ if $|i - j| > 1$, given $u_j = Aq_j$, it suffices to compute only $h_{jj} = q_j^T u_j$ and $h_{j-1,j} = h_{j,j-1} = q_{j-1}^T q_j.$*
Cost Krylov Subspace Methods

- Consider a matrix with $m$ nonzeros, what is the cost of a matrix-vector product?
  
  $m$ multiplications and at most $m$ additions

- How much does it cost to orthogonalize the vector at the $k$th iteration?
  
  $O(nk)$ work for $k$ inner products in Arnoldi, $O(n)$ work in Lanczos. For Arnoldi with $k$-dimensional subspace, orthogonalization costs $O(nk^2)$, matrix-vector products cost $O(mk)$, so generally desire $nk < m$. 
Restarting Krylov Subspace Methods

- In finite precision, Lanczos generally loses orthogonality, while orthogonalization in Arnoldi can become prohibitively expensive:
  - *Arnoldi cost of orthogonalization dominates if* $k > m/n$.
  - *In Lanczos, reorthogonalizing iterate to previous guesses can ensure orthogonality."
  - *Selective orthogonalization strategies control when, and even with respect to what previous columns of* $Q$, *each new iterate* $u_j = Aq_j$ *should be orthogonalized."

- Consequently, in practice low-dimensional Krylov subspace methods are constructed repeatedly using carefully selected new starting vectors:
  - *If we wish to find a particular eigenvector isolate some eigenspaces, restarting is beneficial*
    - *can orthogonalize to previous eigenvector estimates to perform deflation*
    - *can pick starting vector as Ritz vector estimate associated with desired eigenpair*
    - *given new starting vector, can discard previous Krylov subspace, which helps make storing the needed parts of* $Q$ *possible*
Convergence of Lanczos Iteration

- Cauchy interlacing theorem: eigenvalues of $H_k$, $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ with respect to eigenvalues of $A$, $\lambda_1 \geq \cdots \geq \lambda_n$ satisfy

$$\lambda_i \leq \tilde{\lambda}_i \leq \lambda_{n-k+i}$$

- Convergence to extremal eigenvalues is generally fastest:
Applications of Eigenvalue Problems: Matrix Functions

Given $A = XDX^{-1}$ how can we compute $A^k$?

$$A^2 = XDX^{-1}XDX^{-1} = XD^2X^{-1},$$
$$A^k = XD^kX^{-1}$$

What about $e^A$? $\log(A)$? generally $f(A)$?

$$e^A = I + A + A/2! + \cdots = X(I + D + D^2/2! + \cdots)X^{-1} = Xe^DX^{-1}$$
$$\log(A) = X\log(D)X^{-1}$$
$$f(A) = Xf(D)X^{-1}$$
Consider solutions to an ordinary differential equation of the form
\[
\frac{dx}{dt}(t) = Ax(t) + f(t) \quad \text{with} \quad x(0) = x_0:
\]
\[
x(t) = e^{tA}x_0 + \int_0^t e^{(t-\tau)A}f(\tau)d\tau
\]

Using \( A = XD\) allows us to compute the solution explicitly (Jordan form also suffices if \( A \) is defective):
\[
x(t) = Xe^{tD}X^{-1}x_0 + X \int_0^t e^{(t-\tau)D}X^{-1}f(\tau)d\tau
\]
Consider a more general linear differential equation of the form
\[ B \frac{dx}{dt}(t) = Ax(t) + f(t) \] with \( x(0) = x_0 \), which we can reduce to the usual form by premultiplying with \( B^{-1} \):

\[
\frac{dx}{dt}(t) = B^{-1} Ax(t) + B^{-1} f(t)
\]

However, \( B \) may not be invertible and \( B^{-1} A \) is generally nonsymmetric even when \( B^{-1} \) and \( A \) are.

If we can find \( X \) such that \( A = XD_A X^{-1} \) and \( B = XD_B X^{-1} \) we could solve this equation while preserving symmetry of \( A \) and \( B \):

\[
x(t) = e^{tB^{-1}} Ax_0 + \int_0^t e^{(t-\tau)B^{-1}} A f(\tau) d\tau
\]

\[
= e^{tXD_B^{-1}D_A X^{-1}} x_0 + \int_0^t e^{(t-\tau)XD_B^{-1}D_A X^{-1}} f(\tau) d\tau
\]

\[
= X e^{tD_B^{-1}D_A X^{-1}} x_0 + X \int_0^t e^{(t-\tau)D_B^{-1}D_A X^{-1}} f(\tau) d\tau
\]
A generalized eigenvalue problem has the form $Ax = \lambda Bx$,

$$AX = BXD$$

$$B^{-1}A = XDX^{-1}$$

When $A$ and $B$ are symmetric, if one is SPD, we can perform Cholesky on $B$, multiply $A$ by the inverted factors, and diagonalize it:

$$AX = LL^T XD$$

$$L^{-1}AL^{-T} \hat{X} = \hat{L}^T \hat{X} \hat{D}$$
For nonsingular $U, V$, $A - \lambda B = U(J - \lambda I)V^T$ where $J$ is in Jordan form.

For some unitary $P, Q$, $A = PT_AQ^H$ and $B = PT_BQ^H$ where $T_A$ and $T_B$ are triangular.
In a polynomial eigenvalue problem, we seek solutions $\lambda, x$ to

$$\sum_{i=0}^{d} \lambda^i A_i x = 0$$

Assuming for simplicity that $A_d = I$, solutions are given by solving the matrix eigenvalue problem with the block-companion matrix

$$\begin{bmatrix}
-A_{d-1} & \cdots & -A_0 \\
I & 0 & \cdots \\
& \ddots & \ddots 
\end{bmatrix}$$