CS 450: Numerical Anlaysis

Lecture 12

Chapter 4 – Eigenvalue Problems

Krylov Subspace Methods and Applications of Eigenvalue Problems

Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

March 1, 2018

Introduction to Krylov Subspace Methods

- Define k-dimensional Krylov subspace matrix
  $\boldsymbol{K}_k = \begin{bmatrix} \boldsymbol{x_0} & \boldsymbol{A}\boldsymbol{x_0} & \cdots & \boldsymbol{A}^{k-1}\boldsymbol{x_0} \end{bmatrix}$
  Krylov subspace methods seek to best use the information in $\boldsymbol{K}_k$ to solve eigenvalue problems (or linear systems/least squares problems).
- Show that $\boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n$ is a companion matrix $\boldsymbol{C}$:
  Letting $\boldsymbol{k}_n^{(i)} = \boldsymbol{A}^{i-1}\boldsymbol{x}$, we observe that

$$\boldsymbol{A}\boldsymbol{K}_n = \begin{bmatrix} \boldsymbol{A}\boldsymbol{k}_n^{(1)} & \cdots & \boldsymbol{A}\boldsymbol{k}_n^{(n-1)} & \boldsymbol{A}\boldsymbol{k}_n^{(n)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{k}_n^{(2)} & \cdots & \boldsymbol{k}_n^{(n)} & \boldsymbol{A}\boldsymbol{k}_n^{(n)} \end{bmatrix},$$

  therefore premultiplying by $\boldsymbol{K}_m^{-1}$ transforms the first $n-1$ columns of $\boldsymbol{A}\boldsymbol{K}_n$ into the last $n-1$ columns of $\boldsymbol{I}$,

$$\begin{aligned} \boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n &= \begin{bmatrix} \boldsymbol{K}_n^{-1}\boldsymbol{k}_n^{(2)} & \cdots & \boldsymbol{K}_n^{-1}\boldsymbol{k}_n^{(n)} & \boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{k}_n^{(n)} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n & \boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{k}_n^{(n)} \end{bmatrix} \end{aligned}$$

Krylov Subspaces

- Given $\boldsymbol{QR} = \boldsymbol{K}_k$, we obtain an orthonormal basis for the Krylov subspace,

$$\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{x}_0) = span(\boldsymbol{Q}) = \{\rho(\boldsymbol{A})\boldsymbol{x}_0 : deg(\rho) < k\}$$

- Consider whether $k-1$ steps of power iteration starting from $\boldsymbol{x_0}$ 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 $\boldsymbol{x}_0$ is given by the Rayleigh-quotient of $\boldsymbol{y} = \boldsymbol{A}^k\boldsymbol{x}_0$. This vector is within the Krylov subspace, $\boldsymbol{y} \in \mathcal{K}_k(\boldsymbol{A}, \boldsymbol{x}_0)$.

Krylov Subspace Methods

- Given $\boldsymbol{QR} = \boldsymbol{K}_k$, we obtain an orthonormal basis for the Krylov subspace and $\boldsymbol{H}_k = \boldsymbol{Q}^T\boldsymbol{AQ}$ which minimizes $||\boldsymbol{AQ} - \boldsymbol{QH}||_2$:

  The solution to the linear least squares problem $\boldsymbol{QX} \cong \boldsymbol{AQ}$ is $\boldsymbol{X} = \boldsymbol{Q}^T\boldsymbol{AQ} = \boldsymbol{H}$

- $\boldsymbol{H}_k$ is Hessenberg, because the companion matrix $\boldsymbol{C}_k$ is Hessenberg:

$$\boldsymbol{H}_k = \boldsymbol{Q}^T\boldsymbol{AQ} = \boldsymbol{RK}_k^{-1}\boldsymbol{AK}_k\boldsymbol{R}^{-1} = \boldsymbol{RC}_k\boldsymbol{R}^{-1}$$

Rayleigh-Ritz Procedure

- The eigenvalues/eigenvectors of $\boldsymbol{H}_k$ are the *Ritz values/vectors*:

$$\boldsymbol{H}_k = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$$

eigenvalue approximations based on Ritz vectors $\boldsymbol{X}$ are given by $\boldsymbol{Q}\boldsymbol{X}$

- The Ritz vectors and values are the *ideal approximations* of the actual eigenvalues and eigenvectors based on only $\boldsymbol{H}_k$ and $\boldsymbol{Q}$:

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

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

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

Arnoldi Iteration

- Arnoldi iteration computes $\boldsymbol{H}$ directly using the recurrence $\boldsymbol{q}_i^T \boldsymbol{A} \boldsymbol{q}_j = h_{ij}$:

  We have that

$$\boldsymbol{q}_i^T \boldsymbol{A} \boldsymbol{q}_j = \boldsymbol{q}_i^T (\boldsymbol{Q} \boldsymbol{H}_n \boldsymbol{Q}^T) \boldsymbol{q}_j = \boldsymbol{e}_i \boldsymbol{H}_n \boldsymbol{e}_j = h_{ij}$$

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

  Given $\boldsymbol{u}_j = \boldsymbol{A} \boldsymbol{q}_j$, compute $h_{ij} = \boldsymbol{q}_i^T \boldsymbol{u}_j$ for each $i \leq j$, forming a column of the $\boldsymbol{H}$ 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 $\boldsymbol{H}$ *as before, except that it must also be symmetric, since*

  $$\boldsymbol{H}^T = (\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q})^T = \boldsymbol{Q}^T\boldsymbol{A}^T\boldsymbol{Q} = \boldsymbol{Q}^T\boldsymbol{A}BQ = \boldsymbol{H},$$

  which implies that $\boldsymbol{H}$ *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* $\boldsymbol{u}_j = \boldsymbol{A}\boldsymbol{q}_j$, *it suffices to compute only* $h_{jj} = \boldsymbol{q}_j^T\boldsymbol{u}_j$ *and* $h_{j-1,j} = h_{j,j-1} = \boldsymbol{q}_{j-1}^T\boldsymbol{q}_j$.

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 kth 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 stratgies control when, and even with respect to what previous columns of $\boldsymbol{Q}$, each new iterate $\boldsymbol{u}_j = \boldsymbol{A}\boldsymbol{q}_j$ 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 $\boldsymbol{Q}$ possible*

Convergence of Lanczos Iteration

- Cauchy interlacing theorem: eigenvalues of $\boldsymbol{H}_k$, $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ with respect to eigenvalues of $\boldsymbol{A}$, $\lambda_1 \geq \cdots \geq \lambda_n$ 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 $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$ how can we compute $\boldsymbol{A}^k$?

$$\begin{aligned} \boldsymbol{A}^2 &= \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1} \\ &= \boldsymbol{X}\boldsymbol{D}^2\boldsymbol{X}^{-1}, \\ \boldsymbol{A}^k &= \boldsymbol{X}\boldsymbol{D}^k\boldsymbol{X}^{-1} \end{aligned}$$

- What about $e^{\boldsymbol{A}}$? $\log(\boldsymbol{A})$? generally $f(\boldsymbol{A})$?

$$\begin{aligned} e^{\boldsymbol{A}} &= \boldsymbol{I} + \boldsymbol{A} + \boldsymbol{A}/2! + \cdots \\ &= \boldsymbol{X}(\boldsymbol{I} + \boldsymbol{D} + \boldsymbol{D}^2/2! + \cdots)\boldsymbol{X}^{-1} \\ &= \boldsymbol{X}e^{\boldsymbol{D}}\boldsymbol{X}^{-1} \\ \log(\boldsymbol{A}) &= \boldsymbol{X}\log(\boldsymbol{D})\boldsymbol{X}^{-1} \\ f(\boldsymbol{A}) &= \boldsymbol{X}f(\boldsymbol{D})\boldsymbol{X}^{-1} \end{aligned}$$

Applications of Eigenvalue Problems: Differential Equations

- Consider solutions to an ordinary differential equation of the form $\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x_0}$:

$$\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0 + \int_0^t e^{(t-\tau)\boldsymbol{A}}\boldsymbol{f}(\tau)d\tau$$

- Using $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$ permits us to compute the solution explicitly (Jordan form also suffices if $\boldsymbol{A}$ is defective):

$$\boldsymbol{x}(t) = \boldsymbol{X}e^{t\boldsymbol{D}}\boldsymbol{X}^{-1}\boldsymbol{x}_0 + \boldsymbol{X}\int_0^t e^{(t-\tau)\boldsymbol{D}}\boldsymbol{X}^{-1}\boldsymbol{f}(\tau)d\tau$$

Differential Equations using the Generalized Eigenvalue Problem

- Consider a more general linear differential equation of the form $\boldsymbol{B}\frac{\boldsymbol{dx}}{\boldsymbol{dt}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x_0}$, which we can reduce to the usual form by premultiplying with $\boldsymbol{B}^{-1}$:

$$\frac{\boldsymbol{dx}}{\boldsymbol{dt}}(t) = \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}^{-1}\boldsymbol{f}(t)$$

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

- If we can find $\boldsymbol{X}$ such that $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D_A}\boldsymbol{X}^{-1}$ and $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{D_B}\boldsymbol{X}^{-1}$ we could solve this equation while preserving symmetry of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\begin{aligned}
\boldsymbol{x}(t) &= e^{t\boldsymbol{B}^{-1}\boldsymbol{A}}\boldsymbol{x}_0 + \int_0^t e^{(t-\tau)\boldsymbol{B}^{-1}\boldsymbol{A}}\boldsymbol{f}(\tau)d\tau \\
&= e^{t\boldsymbol{X}\boldsymbol{D_B}^{-1}\boldsymbol{D_A}\boldsymbol{X}^{-1}}\boldsymbol{x}_0 + \int_0^t e^{(t-\tau)\boldsymbol{X}\boldsymbol{D_B}^{-1}\boldsymbol{D_A}\boldsymbol{X}^{-1}}\boldsymbol{f}(\tau)d\tau \\
&= \boldsymbol{X}e^{t\boldsymbol{D_B}^{-1}\boldsymbol{D_A}}\boldsymbol{X}^{-1}\boldsymbol{x}_0 + \boldsymbol{X}\int_0^t e^{(t-\tau)\boldsymbol{D_B}^{-1}\boldsymbol{D_A}}\boldsymbol{X}^{-1}\boldsymbol{f}(\tau)d\tau
\end{aligned}$$

Generalized Eigenvalue Problem

- A generalized eigenvalue problem has the form $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$,

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X}\boldsymbol{D}$$
$$\boldsymbol{B}^{-1}\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$$

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

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{L}\boldsymbol{L}^T\boldsymbol{X}\boldsymbol{D}$$
$$\underbrace{\boldsymbol{L}^{-1}\boldsymbol{A}\boldsymbol{L}^{-T}}_{\tilde{\boldsymbol{A}}}\underbrace{\boldsymbol{L}^T\boldsymbol{X}}_{\tilde{\boldsymbol{X}}} = \underbrace{\boldsymbol{L}^T\boldsymbol{X}}_{\tilde{\boldsymbol{X}}}\boldsymbol{D}$$

Canonical Forms Generalized Eigenvalue Problem

- For nonsingular $\boldsymbol{U}, \boldsymbol{V}$, $\boldsymbol{A} - \lambda \boldsymbol{B} = \boldsymbol{U}(\boldsymbol{J} - \lambda \boldsymbol{I})\boldsymbol{V}^T$ where $\boldsymbol{J}$ is in Jordan form
- For some unitary $\boldsymbol{P}, \boldsymbol{Q}$, $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{T}_{\boldsymbol{A}}\boldsymbol{Q}^H$ and $\boldsymbol{B} = \boldsymbol{P}\boldsymbol{T}_{\boldsymbol{B}}\boldsymbol{Q}^H$ where $\boldsymbol{T}_{\boldsymbol{A}}$ and $\boldsymbol{T}_{\boldsymbol{B}}$ are triangular

Nonlinear Eigenvalue Problem

- In a polynomial eigenvalue problem, we seek solutions $\lambda, \boldsymbol{x}$ to

$$\sum_{i=0}^{d} \lambda^i \boldsymbol{A}_i \boldsymbol{x} = \boldsymbol{0}$$

- Assuming for simplicity that $\boldsymbol{A}_d = \boldsymbol{I}$, solutions are given by solving the matrix eigenvalue problem with the block-companion matrix

$$\begin{bmatrix} -\boldsymbol{A}_{d-1} & \cdots & -\boldsymbol{A}_0 \\ \boldsymbol{I} & \boldsymbol{0} & \cdots \\ & \ddots & \ddots \end{bmatrix}$$