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

February 23, 2018

Introduction to Krylov Subspace Methods

- Define k-dimensional Krylov subspace matrix
  $\underline{\boldsymbol{K}_k} = \underline{\begin{bmatrix} \boldsymbol{x}_0 & \boldsymbol{A}\boldsymbol{x}_0 & \cdots & \boldsymbol{A}^{k-1}\boldsymbol{x}_0 \end{bmatrix}}$

  how much information can we retain
  with k-1 matrix-vector products

- Show that $\boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n$ is a companion matrix $\boldsymbol{C}$:

$K_n^{-1} K_n = I$

$$A K_n = [A x_0, A^2 x_0, A^3 x_0 \cdots A^k x_0]$$

$$K_n [x_0, A x_0, A^2 x_0 \cdots]$$

$$K_n^{-1} A K_n = \left[\begin{matrix} 0 \\ I \end{matrix} \quad A^k x_0 \right] = [\diagdown |]$$

companion matrix

Krylov Subspaces

- Given $\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{K}_k$, we obtain an orthonormal basis for the Krylov subspace, $\mathcal{K}(\boldsymbol{A}, \boldsymbol{x}_0) = span(\boldsymbol{Q})$:

$$span(Q) = span(K_k) = \text{Krylov subspace} = \{poly(A)x_0 \text{ where } poly(A) \text{ is degree } k-1\}$$

- 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:

$$Q^T A Q = Q^T \underbrace{A K_k R^{-1}}$$

$$A K_k R^{-1}$$

$$K_k^{-1} A K_k = C$$

$$R^{-1} Q^T A Q R = C$$

$$Q^T A Q = \underbrace{R C R^{-1}} := H_k$$

upper-Hessenberg

Krylov Subspace Methods

- Given $QR = K_k$, we obtain an orthonormal basis for the Krylov subspace and $H_k = Q^T A Q$ which minimizes $||AQ - QH||_2$:

$$\underbrace{AQ}\,\underset{?}{H^{-1}} \cong Q$$

$$A = U \Sigma V^T$$

$$H^{-1} = Q^T V \Sigma^{+} U^T Q$$

$$H = Q^T U \Sigma V^T Q = Q^T A Q$$

- H_k is Hessenberg, because the companion matrix C is Hessenberg:

Rayleigh-Ritz Procedure

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

$$H_k X_k = X_k D_k$$

Ritz values ≈ largest k eig-vals of H_k, and will converge

Ritz vectors

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

Arnoldi Iteration

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

$\underbrace{}_{i\text{th column of } Q}$

$$q_i A q_j = h_{ij}$$

$$q_j A q_i = \sum_j h_{ij} \underbrace{\langle q_j, q_i \rangle}_{\text{orthonormality}} = h_{ij}$$

$$A = Q_n^T H Q$$

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

$$u_i = A q_i \qquad \left(q_0 = \text{random normalized} \right)$$

$v_i = u_i$ orthogonalized with $[q_1 \dots q_i]$

$$q_{i+1} = v_i / \|v_i\|$$

$p(A) q_0 = 0$ and $p(A)$ is degree k-1

Lanczos Iteration

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

$$Q^T A Q = T =$$

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

$u_i = A q_i$ then orthogonalize u_i to q_i and q_{i-1} and obtain for free orthogonality to $q_1 \cdots q_{i-2}$

Cost Krylov Subspace Methods

- Consider a matrix with m nonzeros, what is the cost of a matrix-vector product?

$O(m)$ $O(m+n)$ | dense $O(n^2)$

typically $m \gg n$

K_k costs $O(km)$ to compute
$O(mn)$ for K_n

- How much does it cost to orthogonalize the vector at the kth iteration?

Arnoldi kth iteration requires $O(kn)$

overall $O(k^2 n)$

matrix-vector products cost $O(mk)$
when $n \cdot k \geq m$

Restarting Krylov Subspace Methods

- In finite precision, Lanczos generally loses orthogonality, while orthogonalization in Arnoldi can become prohibitively expensive:

  for ex. in Arnoldi to compute dominant eigenvector can restart with the Ritz-vector estimate as starting guess

- Consequently, in practice low-dimensional Krylov subspace methods are constructed repeatedly using carefully selected new starting vectors:

  for example, in Lanczos, selective orthogonalization picks 'best' previous vectors to orthogonalize to

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

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

$k = n-1$

$k = n$

Spectrum: $\lambda_1 > \lambda_2 \ldots$

- Convergence to extremal eigenvalues is generally fastest:

smallest

and greatest

Applications of Eigenvalue Problems: Matrix Functions

- Given $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$ how can we compute $\boldsymbol{A}^k$?

$$A^2 = XDX^{-1}XDX^{-1} = \underline{XD^2X^{-1}}$$

$$A^k = XD^kX^{-1}$$

if X is definitive

$$f(A) = X\,f(D)\,X^{-1}$$

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

$$e^A = I + A + \frac{A^2}{2!} \cdots = X e^D X^{-1}$$

$$\sqrt{A}$$

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 $\boldsymbol{x}(0) = \boldsymbol{x_0}$:

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

SDS^{-1}

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

$$x(t) = S e^{tD} S^{-1} x_0 + \int_0^t S \underbrace{e^{(t-\tau)D}}_{} S^{-1} f(\tau) d\tau$$

Differential Equations using the Generalized Eigenvalue Problem

- Consider a more general linear differential equation of the form $\boldsymbol{B}\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x}_{\mathbf{0}}$, which we can solve by premultiplying with $\boldsymbol{B}^{-1}$,

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

Generalized Eigenvalue Problem

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

$$Ax = \lambda Bx \qquad (A - \lambda B)x = 0$$

- 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:

$$B = LL^T$$

$$L^{-1}AL^{-T} = XDX^{-1}$$

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

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