AMS526: Numerical Analysis I (Numerical Linear Algebra for Computational and Data Sciences)
Lecture 21: Arnoldi and Lanczos Iterations

Xiangmin Jiao

Stony Brook University

Outline

Krylov Subspace Methods

- Given A and b, Krylov subspace

$$\{b, Ab, A^2 b, \ldots, A^{k-1} b\}$$

	linear systems	eigenvalue problems
Hermitian	CG	Lanczos
Nonhermitian	GMRES, BiCG, etc.	Arnoldi

- CG, GMRES etc. are Krylov subspace methods for solving sparse linear systems (later)
- Lanczos and Arnoldi iterations are Krylov subspace methods for reduction to Hessenberg form

Review: Reduction to Hessenberg Form

- General A: First convert to *upper-Hessenberg* form, then to upper triangular

- Hermitian A: First convert to *tridiagonal* form, then to diagonal

- In general, phase 1 is direct and requires $O(n^3)$ flops

Arnoldi Iteration

- The Arnoldi iteration reduces a general, nonsymmetric matrix A to Hessenberg form by similarity transformation $A = QHQ^*$
- It is analogous to Gram-Schmidt-style iteration instead of Householder reflections
- Let $Q_k = [q_1 \mid q_2 \mid \cdot \mid q_k]$ be $n \times k$ matrix with first k columns of Q and $\tilde{H}_k$ be $(k+1) \times k$ upper-left section of H, i.e., $\tilde{H}_k = H_{1:k+1,1:k}$
- Consider first k columns of $AQ = QH$, or $AQ_k = QH_{:,1:k} = Q_{k+1}\tilde{H}_k$

$$
\underbrace{\begin{bmatrix} \\ \\ A \\ \\ \\ \end{bmatrix}}_{A}
\underbrace{\begin{bmatrix} \mid & & \mid \\ q_1 & \cdots & q_k \\ \mid & & \mid \end{bmatrix}}_{Q_k}
=
\underbrace{\begin{bmatrix} \mid & & \mid \\ q_1 & \cdots & q_{k+1} \\ \mid & & \mid \end{bmatrix}}_{Q_{k+1}}
\underbrace{\begin{bmatrix} h_{11} & \cdots & h_{1k} \\ h_{21} & & \vdots \\ & \ddots & \\ & & h_{k+1,k} \end{bmatrix}}_{\tilde{H}_k}
$$

- Question: How do we choose q_1?

Arnoldi Algorithm

- Start with a random q_1, then determine q_2 and $\tilde{H}_1$, and so on
- The kth columns of $AQ_k = Q_{k+1}\tilde{H}_k$ can be written as

$$Aq_k = h_{1k}q_1 + \cdots + h_{kk}q_k + h_{k+1,k}q_{k+1}$$

where $h_{ik} = q_i^* Aq_k$.

Algorithm: Arnoldi Iteration
    given random nonzero b, let $q_1 = b/\|b\|$
    for $k = 1, 2, 3, \ldots$
           $v = Aq_k$
           for $j = 1$ **to** k
                $h_{jk} = q_j^* v$
                $v = v - h_{jk}q_j$
           $h_{k+1,k} = \|v\|$
           $q_{k+1} = v/h_{k+1,k}$

- Question: What if q_1 happens to be an eigenvector?

QR Factorization of Krylov Matrix

- The vector q_j from Arnoldi are orthonormal bases of successive Krylov subspaces

$$\mathcal{K}_k = \left\langle b, Ab, \ldots, A^{k-1}b \right\rangle = \langle q_1, q_2, \ldots, q_k \rangle \subseteq \mathbb{C}^n,$$

assuming $h_{k+1,k} \neq 0$

- Q_k is reduced QR factorization $K_k = Q_k R_k$ of Krylov matrix

$$K_k = \begin{bmatrix} b & Ab & \cdots & A^{k-1}b \end{bmatrix}$$

- However, K_k and R_k are not formed explicitly; forming them explicitly would be unstable and can suffer from overflow and underflow

Projection onto Krylov Subspaces

- Arnoldi process computes projections of A onto successive Krylov subspaces

$$H_k = Q_k^* A Q_k$$

because $A Q_k = Q_{k+1} \tilde{H}_k$, $\tilde{H}_k = Q_{k+1}^* A Q_k$, and $H_k = \tilde{H}_{1:k,1:k}$

- H_k can be interpreted as orthogonal projection of A onto $\mathcal{K}_k$ in the basis $\{q_1, q_2, \ldots, q_k\}$, restricting mapping $A : \mathbb{C}^n \to \mathbb{C}^n$ to $H_k : \mathcal{K}_k \to \mathcal{K}_k$. This kind of projection is known as *Rayleigh-Ritz procedure*

- Arnoldi iteration is useful as
  1. basis for iterative algorithms (such as GMRES, to be discussed later)
  2. technique for estimating eigenvalues of nonhermitian matrices

- Caution: eigenvalues of nonmormal matrices may have little to do with physical system, since eigenvalues of such equations are ill-conditioned. When such problems arise, the original problem is mostly likely posed improperly

Estimating Eigenvalues by Arnoldi Iteration

- Diagonal entries of H_k are Rayleigh quotients of A w.r.t. vectors q_i
- H_k is "generalized Rayleigh quotient" w.r.t Q_k, whose eigenvalues $\{\theta_j\}$ are called *Arnoldi estimates* or *Ritz values* w.r.t. $\mathcal{K}_k$ of A
- *Ritz vectors* corresponds to θ_j are $Q_k y_j$, where $H_k y_j = \theta_j y_j$
- To use Arnoldi iteration to estimate eigenvalues, compute eigenvalues of H_k at kth step
- When $k = n$, Ritz values are eigenvalues
- In general, $k \ll n$, so we can estimate only a few eigenvalues
- Which eigenvalues? Typically, it finds extreme eigenvalues first
- In many applications, extreme eigenvalues are of main interests
  - Stability analysis typically requires estimating spectral radius
  - Principal component analysis requires estimating largest eigenvalues and corresponding eigenvectors of $A^T A$

Invariance Properties of Arnoldi Iteration

Theorem

Let the Arnoldi iteration be applied to matrix $A \in \mathbb{C}^{n \times n}$ as described above.

Translation invariance. If A is changed to $A + \sigma I$ for some $\sigma \in \mathbb{C}$, and b is unchanged, then Ritz values $\{\theta_j\}$ change to $\{\theta_j + \sigma\}$.

Scale invariance. If A is changed to σA for some $\sigma \in \mathbb{C}$, and b is unchanged, then $\{\theta_j\}$ change to $\{\sigma \theta_j\}$.

Invariance under unitary similarity transformation. If A is changed to UAU^* for some unitary matrix U, and b is changed to Ub, then $\{\theta_j\}$ do not change.

In all three cases, the Ritz vectors, namely $Q_k y_k$ corresponding to eigenvectors y_j of H_k do not change under indicated transformation.

Convergence of Arnoldi Iteration

- If A has n distinct eigenvalues, Arnoldi iteration finds them all in n steps
- Under certain circumstances, convergence of some Arnoldi estimates is geometric (i.e., linear), and it accelerates in later iterations
- However, these matters are not yet fully understood
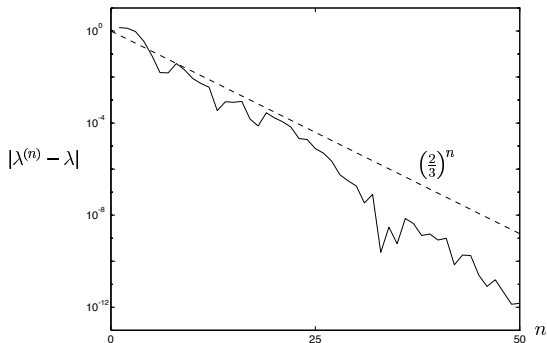
Example convergence of extreme Arnoldi eigenvalue estimation.

Outline

Lanczos Iteration for Symmetric Matrices

- For symmetric A, $\tilde{H}_k$ and H_k in Arnoldi iteration are tridiagonal
- We denote them by $\tilde{T}_k$ and T_k, respectively. Let $\alpha_k = h_{kk}$ and $\beta_k = h_{k+1,k} = h_{k,k+1}$
- $AQ_k = Q_{k+1}\tilde{H}_k$ can then be written as three-term recurrence

$$Aq_k = \beta_{k-1}q_{k-1} + \alpha_k q_k + \beta_k q_{k+1}$$

where α_i are diagonal entries and β_i are sub-diagonal entries of $\tilde{T}_k$

$$T_k = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{k-1} \\ & & & \beta_{k-1} & \alpha_k \end{bmatrix}$$

- Arnoldi iteration for symmetric matrices is known as Lanczos iteration

Algorithm of Lanczos Iteration

Algorithm: Lanczos Iteration

$\beta_0 = 0$, $q_0 = 0$

given random b, let $q_1 = b/\|b\|$

for $k = 1, 2, 3, \ldots$

$\qquad v = Aq_k$

$\qquad \alpha_k = q_k v$

$\qquad v = v - \beta_{k-1} q_{k-1} - \alpha_k q_k$

$\qquad \beta_k = \|v\|$

$\qquad q_{k+1} = v/\beta_k$

- Each step consists of matrix-vector multiplication, an inner product, and a couple of vector operations
- This is particularly efficient for sparse matrices. In practice, Lanczos iteration is used to compute eigenvalues of large symmetric matrices
- Like Arnoldi iteration, Lanczos iteration is useful as
  1. basis for other iterative algorithms (such as conjugate gradient)
  2. technique for estimating eigenvalues of Hermitian matrices

Estimating Eigenvalues by Lanczos Iterations

- For symmetric matrices with evenly spaced eigenvalues, Ritz values tend to first convert to extreme eigenvalue.
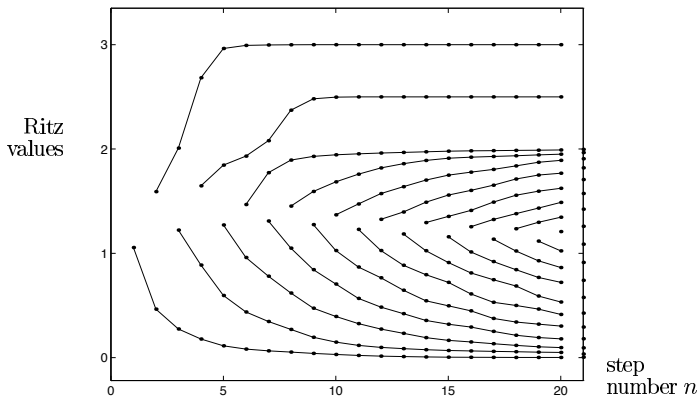

Ritz values for first 20 steps for Lanczos iteration applied to example 203×203 matrix. Convergence of extreme eigenvalues is geometric.

Effect of Rounding Errors

- Rounding errors have complex effects on Lanczos iteration and all iterations based on three-term recurrence
- Rounding errors cause loss of orthogonality of q_1, q_2, ..., q_k
  - In Arnoldi iteration, vectors q_1, q_2, ..., q_k are enforced to be orthogonal by explicit modified Gram-Schmidt orthogonalization, which suffer some but not as much loss of orthogonality
  - In Lanczos iteration, orthogonality of q_j, q_{j-1} and q_{j-2} are "enforced", but orthogonality of q_j with $q_{j-3}, \ldots, q_1$ are "automatic", based on mathematical identities
  - In practice, such mathematical identities are not accurately preserved in the presence of rounding errors
- In practice, periodic re-orthogonalization of Q_k is sometimes used to alleviate effect of rounding errors

Rounding Errors and Ghost Eigenvalues

- With rounding errors, Lanczos iteration can suffer from loss of orthogonality and can in turn lead to spurious "ghost" eigenvalues
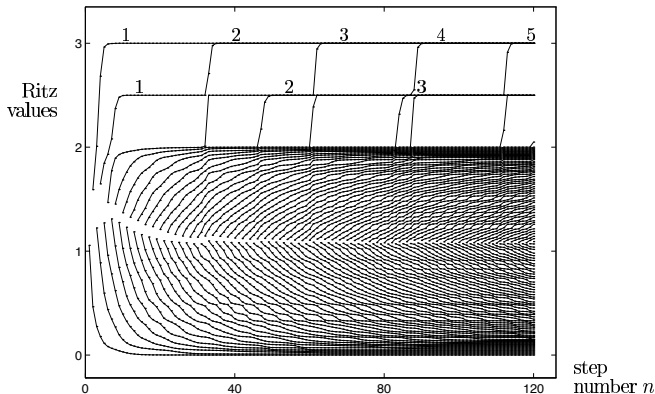
Continuation to 120 steps of Lanczos iteration. Numbers indicate multiplicities of Ritz values. 4 "ghost" copies of 3.0 and 2 "ghost" copies of 2.5 appear.

Explanation of Ghost Eigenvalues

- Intuitive explanation of ghost eigenvalues
  - Convergence of Ritz value annihilates corresponding eigenvector components in the vector being operated upon
  - With rounding errors, random noise re-introduce and excite those components again
- We cannot trust multiplicities of Ritz values as those of eigenvalues
- Nevertheless, Lanczos iteration can still be very useful in practice
  - E.g., in PCA for dimension reduction in data analysis, one needs to find leading singular values and corresponding singular vectors of A.
  - One standard approach is to apply Lanczos iteration to $A^T A$ or $A A^T$ without forming the product explicitly, and then use Ritz vectors to approximate singular vectors