DN2222 Applied Numerical Methods - part 2: Numerical Linear Algebra

Lecture 5 Singular Value Decomposition (cont) & Eigenvalues

2012-11-15

Note!

Next Lecture (F6) is Monday 19/11 at 13-15 in room ??? (NA seminar room has only 20 seats?)

Next Lecture (F7) is Wednesday 21/11 at 10-12 in room E32

Next Lab Help Session is Friday 23/11 at 8-10 in the NA seminar room

---

Def of eigen-values and -vectors, (Dp140)

- The polynomial $p(\lambda) = det(A - \lambda I)$ is called the characteristic polynomial of A. The roots of $p(\lambda) = 0$ are the eigenvalues of A (D: Def 4.1)
- The characteristic polynomial of an $n \times n$ matrix A is of degree n and thus have n roots.
- A non-zero vector x satisfying $Ax = \lambda x$ is a (right) eigenvector for the eigenvalue λ. (D: Def 4.2)
- A non-zero vector y satisfying $y^*A = \lambda y^*$ is a left eigenvector for the eigenvalue λ. (D: Def 4.2)

General

- Algorithms for eigenvalues problems can roughly be divided into two categories: *direct* and *iterative* methods.
- Since determining eigenvalues is always an iterative method, by direct is meant methods that converges within a certain number of iterations. They usually cost $O(n^3)$ and are independent of the matrix entries.
- Iterative methods are usually used for sparse matrices, where the matrix-vector multiplication is relatively cheap. Convergence rate depends strongly on the matrix entries.

- Most algorithms will involve transforming the matrix A into a simpler, or *canonical* form, from which it is easy to calculate the eigen-values and eigen-vectors.
- Let S be any nonsingular matrix. Then A and $B = S^{-1}AS$ are *similar* matrices and S is a similarity transformation. (D: Def 4.3)
- Let $B = S^{-1}AS$, so A and B are similar. Then A and

B have the same eigenvalues, and x (or y) is a right (or left) eigenvector of A iff $S^{-1}x$ (or S^*y) is a right (or left) eigenvector of B.

- Let $B = S^{-1}AS$, so A and B are similar. Then A and B have the same eigenvalues, and x (or y) is a right (or left) eigenvector of A iff $S^{-1}x$ (or S^*y) is a right (or left) eigenvector of B.

PROOF1: Using that $det(XY) = det(X) \cdot det(Y)$ for any square matrices, we have
$det(A - \lambda I) = det(S^{-1}(A - \lambda I)S) = det(B - \lambda I)$

PROOF2: $Ax = \lambda x$ hold iff $S^{-1}ASS^{-1}x = \lambda S^{-1}x$ or $B(S^{-1}x) = \lambda(S^{-1}x)$

Quiz

- Which are the two simplest matrix forms for determining eigenvalues?
- To avoid complex numbers we might consider block triangular matrices. Why?

General ideas

- The two simplest matrix forms for determining eigenvalues are diagonal and triangular!
- To avoid complex numbers we might consider block triangular matrices. Since for real matrix elements - any complex valued eigenvalues comes in pairs - 2x2 and 1x1 blocks are useful.
- The two most common canonical forms are the *Jordan form* and the *Schur form*.

- Given A there exists a nonsingular S such that $S^{-1}AS = J$, where J is in Jordan canonical form. This means that J is block diagonal, with $J = diag(J_{n_1}(\lambda_1), J_{n_2}(\lambda_2), \ldots J_{n_k}(\lambda_k))$ and the $n_i \times n_i$ matrix

$$J_{n_i} = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

  J is unique, up to permutations of the blocks.

- Each $J_m(\lambda)$ is called a *Jordan block* with eigenvalue λ of algebraic multiplicity m.
- If $n_i = 1$ and that λ_i is an eigenvalue of only that block, λ_i is called a simple eigenvalue.
- If all $n_i = 1$, J is diagonal and A is diagonalizable, otherwise it is called defective.
- A defective matrix does not have n eigenvectors.

- In invariant subspace of A is a subspace $X \epsilon R^n$ such that $x \epsilon X \to Ax \epsilon X$

- The Jordan form tells everything about a matrix: eigenvalues, eigenvectors and invariant subspaces. But it is bad to compute for 2 numerical reasons! 1. It is sensitive to round-off errors. 2. It cannot be computed stably in general.

- So instead of computing $S^{-1}AS = J$, where S can be arbitrarily ill-conditioned, we will restrict S to be orthogonal (so $\kappa_2(S) = 1$) to guarantee stability:

- *The Schur canonical form*: Given A, there exists a unitary matrix Q and an upper triangular matrix T such that $Q^*AQ = T$. The eigenvalues of A are the diagonal entries of T. (D: Thm 4.2)

PROOF We use induction. It is obviously true for $n = 1$. Let λ be an eigenvalue with corresponding normalized eigenvector u. Choose $\tilde{U}$ such that $U = [u, \tilde{U}]$ is a unitary matrix. Then

$$U^* \cdot A \cdot U = \begin{bmatrix} u^* \\ \tilde{U}^* \end{bmatrix} \cdot A \cdot [\,u \quad \tilde{U}\,] = \begin{bmatrix} u^*Au & u^*A\tilde{U} \\ \tilde{U}^*Au & \tilde{U}^*A\tilde{U} \end{bmatrix}$$

But $u^*Au = u^*\lambda u = \lambda u^*u = \lambda$ and $\tilde{U}^*Au = \tilde{U}^*\lambda u = \lambda\tilde{U}^*u = 0$ and $\tilde{U}^*A\tilde{U}$ is a $(n-1) \times (n-1)$ matrix. Then

$$U^*AU = \begin{bmatrix} \lambda & x \\ 0 & \tilde{Q}\tilde{T}\tilde{Q}^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{Q} \end{bmatrix} \begin{bmatrix} \lambda & x\tilde{Q} \\ 0 & \tilde{T} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \tilde{Q}^* \end{bmatrix}$$

so $Q^*AQ = T$ with $Q = U\begin{bmatrix} 1 & 0 \\ 0 & \tilde{Q} \end{bmatrix}$, unitary as desired.

- *The real Schur canonical form:* If A is real, there exists a real orthogonal matrix V such that $V^TAV = T$ is quasi-upper triangular. This means that T is block upper triangular with 1-by-1 and 2-by-2 blocks on the diagonal. Its eigenvalues are the eigenvalues of the diagonal blocks. The 1-by-1 blocks correspond to real eigenvalues. The 2-by-2 blocks correspond to a complex conjugate pair of eigenvalues.

Computing eigenvectors from the Schur form

- Suppose $\lambda = t_{ii}$ has multiplicity 1. Write $(T - \lambda I)x = 0$ as

$$0 = \begin{bmatrix} T_{11} - \lambda I & T_{12} & T_{13} \\ 0 & 0 & T_{23} \\ 0 & 0 & T_{33} - \lambda I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} =$$

$$\begin{bmatrix} (T_{11} - \lambda I)x_1 + T_{12}x_2 + T_{13}x_3 \\ T_{23}x_3 \\ (T_{33} - \lambda I)x_3 \end{bmatrix}$$

where T_{11} is $(i-1) \times (i-1)$, $T_{22} = \lambda$ is 1×1, and T_{33} is $(n-i) \times (n-i)$, and x is split correspondingly. Since λ is simple, $(T_{33} - \lambda I)$ is nonsingular, thus $(T_{33} - \lambda I)x_3 = 0$ implies $x_3 = 0$. Choosing (arbitrarily) $x_2 = 1$ we get $x_1 = -((T_{11} - \lambda I)^{-1}T_{12}$ so

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -(T_{11} - \lambda I)^{-1}T_{12} \\ 1 \\ 0 \end{bmatrix}$$

so we only need to solve a triangular system for x_1.

Insight

- Not all matrices are diagonalizable, but we can transform any square matrix into *triangular* form by means of a unitary (or orthogonal) similarity. This is the consequence of the Schur theorem.

Multiple eigenvalues

- Multiple eigenvalues have infinite condition number.
- Eigenvalues "close to multiple" have large condition numbers, since there is a small δA such that $A + \delta A$ has multiple eigenvalues.
- Let λ be a simple eigenvalue of A with right eigenvector x and left eigenvector y, normalized so that $||x||_2 = ||y||_2 = 1$. Let $\lambda + \delta\lambda$ be the corresponding eigenvalue of $A + \delta A$. Then

$$\delta\lambda = \frac{y^* \delta A x}{y^* x} + O(||\delta A||^2)$$

$$|\delta\lambda| \leq \frac{||\delta A||}{|y^* x|} + O(||\delta A||^2)$$

so $1/|y^* x|$ is the condition number of the eigenvalue λ. (D: Thm 4.4)

- Let A be normal (ie $AA^* = A^*A$). Then $|\delta\lambda| \leq ||\delta A|| + O(||\delta A||^2)$ (D: Cor 4.1)
- Let A have all simple eigenvalues with right eigenvector x and left eigenvector y, normalized so that $||x||_2 = ||y||_2 = 1$. Then the eigenvalues of $A + \delta A$ lies in disks centered at λ_i with radius $n \cdot \frac{||\delta A||_2}{|y^* x|}$

Power method:

- Given x_0 we iterate:

```
i=0
while ...
    y=A*x;
    x=y/norm(y); % Approx eigenvector
    d=x'*A*x; % Approx eigenvalue
    i=i+1;
end;
```

- It will find the largest eigenvalue.
- The convergence rate depends on $|\lambda_2/\lambda_1|$. Even though $|\lambda_2/\lambda_1| < 1$ convergence is often slow.

Inverse power method:

- Given x_0 we iterate:

```
i=0
while ...
```

```
    y=(A-s*I)\x;
    x=y/norm(y); % App eigenvector
    d=x'*A*x; % App eigenvalue
    i=i+1;
end;
```

- It will find the largest eigenvalue of $(A - sI)^{-1}$
  ie the smallest eigenvalue of $(A - sI)$
  ie the eigenvalue of A closest to s.
- If s is very close to λ_1 the ratio $(\lambda_1 - s)/(\lambda_2 - s)$ will be very small, thus convergence is fast.

Householder algorithm:

- A *Hessenberg* matrix is upper triangular with one non-zero subdiagonal.
- If A is Hermitian (if real: symmetric) then the Hessenberg matrix will be symmetric and thus tridiagonal.
- The *Householder* algorithm transforms the matrix A into Hessenberg form with an orthogonal similarity transformation, $A = WHW^T$
- The matrix W is a product of Householder transformations (or elementary reflections) $W = H_1 H_2 \ldots H_{(n-2)}$
- An elementary reflection is a matrix, $H = I - 2uu^T$, where the vector u has $||u||_2 = 1$. An elementary reflection is both orthogonal and symmetric.
- $H_k = I - 2u_k u_k^T$ makes all elements except the $k+1$ first elements in column k of A zero. Then vector u_k is zero in the first k positions. (u_k is calculated from the last $n-k$ elements of column k of matrix A)
- With $A^{(1.5)} = H_1 A^{(1)}$ and $A^{(2)} = A^{(1.5)} H_1$ we have

$$A^{(1)} = \begin{bmatrix} x & x & x & \cdots & x \\ x & x & x & \cdots & x \\ x & x & x & \cdots & x \\ \vdots & \vdots & \vdots & & \vdots \\ x & x & x & \cdots & x \end{bmatrix}, A^{(1.5)} = \begin{bmatrix} x & x & x & \cdots & x \\ r & y & y & \cdots & y \\ 0 & y & y & \cdots & y \\ 0 & \vdots & \vdots & & \vdots \\ 0 & y & y & \cdots & y \end{bmatrix}$$

$$A^{(2)} = \begin{bmatrix} x & z & z & \cdots & z \\ r & z & z & \cdots & z \\ 0 & z & z & \cdots & z \\ 0 & \vdots & \vdots & & \vdots \\ 0 & z & z & \cdots & z \end{bmatrix} A^{(2.5)} = \begin{bmatrix} x & z & z & \cdots & z \\ r & z & z & \cdots & z \\ 0 & r & w & \cdots & w \\ 0 & 0 & \vdots & & \vdots \\ 0 & 0 & w & \cdots & w \end{bmatrix}$$

Example (Ruhe p 30, extended):

- $A = magic(4) = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix}$
- $H_k = I - 2u_k u_k^T$ with $u_{1:k} = 0, u_{k+1} = (a_{k+1,k} - \alpha)/r$ and $v_j = a_{kj}/r,\ j = k+2, k+3, \ldots, n$ with
  $\alpha = -sgn(a_{k+1,k})\sqrt{\sum_{j=k+1}^{n} a_{jk}^2}$ and

$r = \sqrt{2\alpha(\alpha - a_{k+1,k})}$ (ie u_k is constructed using the last $n-k$ components of column a_k)

- $A^{(H)} = W^T A W = \begin{bmatrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \end{bmatrix}$ with $W = H_1 H_2$.
- $H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{bmatrix}$ and $H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{bmatrix}$.
- Both H_k and W are orthogonal. But even though H_k is symmetric, W is not.

On computation efficiency:

- Even though we saw H_k as full matrices above, they are really not computed that way. Computing $H_1 a$, where a is a column of A would require n^2 multiplications.
- We use the fact that H_1 is a rank 1 matrix.
  $H_1 a = (I - 2uu^T)a = a - 2uu^T a = a - u(2u^T a)$
  $u^T a$ is a scalar, created by n multiplications. Moving up multiplication by 2 means a single multiplication. Now we have a scalar times a vector, another n multiplications. Finally subtracting two arrays, n additions. This is $2n$ operations, instead of n^2

Some review questions:

- **Q55.** What does the position of the eigenvalues in the complex plane say about the behaviour of the solution of the ODE system

$$\frac{dx}{dt} = Ax, \quad x(0) = x_0$$

- **Q56.** What is meant by two matrices being similar?
- **Q57.** Show that two similar matrices have the same set of eigenvalues. How are the eigenvectors related?