Chapter 4

EIGENVALUE PROBLEM

The German word eigen is cognate with the Old English word āgen, which became own in Middle English and “own” in modern English.

\[
\begin{align*}
\text{Given } & A \in \mathbb{R}^{n \times n} \\
& \text{solve } Ax = \lambda x \\
& \text{where } \lambda \in \mathbb{C} \\
& \text{and } 0 \neq x \in \mathbb{C}^n.
\end{align*}
\]

4.1 Mathematics

4.2 Reduction to Upper Hessenberg Form

4.3 The Power Method

4.4 The QR Iteration

4.5 The Lanczos Method

4.6 Other Eigenvalue Algorithms

4.7 Computing the SVD

4.1 Mathematics

The adjacency matrix of a directed graph is given by

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{bmatrix}
\]

96
where $a_{ij}$ is the number of arcs from $i$ to $j$. The number of paths of length $k$ joining $i$ to $j$ is given by $(A^k)_{ij}$. Is there a closed form expression for $A^k$? I.e., a more direct way of forming $A^k$?

### 4.1.1 Complex Values

The properties of vectors and matrices of real numbers extends to vectors $x \in \mathbb{C}^m$ and matrices $A \in \mathbb{C}^{m \times n}$.

We define the *conjugate transpose* by

$$A^H = A^T.$$  

A Hermitian matrix satisfies $A^H = A$ and a unitary matrix $U^H = U^{-1}$.

$\mathbb{C}^n$ is a *vector space*.

The inner product for $x, y \in \mathbb{C}^n$ is $x^H y$.

### 4.1.2 Eigenvalues and eigenvectors

If $Ax = \lambda x$ where $\lambda \in \mathbb{C}$ and $0 \neq x \in \mathbb{C}^n$ then $\lambda$ is an *eigenvalue* and $x$ is an *eigenvector*. Try solving for $x$ in

$$(A - \lambda I)x = 0.$$  

There exists a solution $x \neq 0$ only if $\det(A - \lambda I) = 0$. Hence the eigenvalues $\lambda$ are roots of the *characteristic polynomial*

$$p(\zeta) := \det(\zeta I - A).$$

For example for $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$

$$\det(\zeta I - A) = \det \begin{bmatrix} \zeta & -1 \\ -1 & \zeta - 1 \end{bmatrix} = \zeta^2 - \zeta - 1. \quad (4.1)$$

An eigenvalue $\lambda$ satisfies $\lambda^2 - \lambda - 1 = 0 \Rightarrow \lambda = \frac{1}{2} \pm \frac{\sqrt{5}}{2}$.

In general the characteristic polynomial has $n$ roots, counting multiplicities:

$$\det(\zeta I - A) = (\zeta - \lambda_1)(\zeta - \lambda_2) \cdots (\zeta - \lambda_n).$$  

Hence there are $n$ eigenvalues. Because $A$ is real, $p(\zeta)$ has real coefficients and thus each eigenvalue is either real or one of a complex conjugate pair.

---

1 characteristic value
Setting $\zeta = 0$ in (4.1), we get

$$\det(-A) = (-1)^n \lambda_1 \lambda_2 \cdots \lambda_n,$$

and since $\det(-A) = \det((-1)A) = (-1)^n \det A,$
we have

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n,$$

$\Rightarrow 0$ is an eigenvalue of a singular matrix.

Note that for an eigenpair $(\lambda, x)$

$$A^2 x = A(\lambda x) = \lambda A x = \lambda^2 x,$$

$$A^{-1} Ax = A^{-1} \lambda x \Rightarrow A^{-1} x = \lambda^{-1} x,$$

$$A^m x = \lambda^m x.$$ 

Moreover, if $q(\zeta)$ is a polynomial,

$$q(A)x = q(\lambda)x.$$ 

For a block triangular matrix $T$ with diagonal blocks $T_{11}, T_{22}, \ldots, T_{pp}$ we have

$$\det(\zeta I - T) = \det(\zeta I - T_{11}) \det(\zeta I - T_{22}) \cdots \det(\zeta I - T_{pp}).$$ 

Example

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\det(\zeta I - A) = \det(\zeta - 1) \det(\zeta I - \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}) = (\zeta - 1)(\zeta - 2)^2.$$ 

This property of determinants follows immediately if we can show

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det A \cdot \det C.$$ 

This is shown by considering two cases:
(i) A singular,

(ii) A nonsingular, in which case \[ \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & C \end{bmatrix}. \]

### 4.1.3 Similarity transformations

Return to the problem of a simpler form for \( A^k \).

**Definition** If \( \det X \neq 0 \), then \( X^{-1}AX =: B \) is called a *similarity transformation* and we say that \( B \) is *similar* to \( A \).

First a couple of easy-to-prove facts:

1. If \( B \) is similar to \( A \) and \( C \) is similar to \( B \), then \( C \) is similar to \( A \). Therefore similarity transformations can have several steps.

2. Similar matrices have identical eigenvalues including multiplicities.

Note that \( A^k = (XBX^{-1})^k = \cdots = XBX^{-k}X^{-1} \). Can we find a simple \( B \) which is similar to \( A \)?

For each of the eigenvalues \( \lambda_i \) we have \[ Ax_i = \lambda_i x_i \quad \text{for some } x_i \neq 0. \]

Hence \[ AX = X\Lambda \]

where \[ X = (x_1, x_2, \ldots, x_n) \text{ and } \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n). \]

If the eigenvectors can be chosen to be linearly independent, then \( X \) is nonsingular and \[ X^{-1}AX = \Lambda, \]

in which case we say that \( A \) is *diagonalizable*. Otherwise \( A \) is *defective*.

**example**

\[ A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \]

Eigenvalues are 1, 1, 1. Eigenvectors must satisfy \( (A-I)x = 0 \).

\[ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \Rightarrow x_3 = 0 \]

\[ \Rightarrow x = \begin{bmatrix} 0 \\ x_2 \\ x_3 \end{bmatrix} \]
where \(x_2\) and \(x_3\) are arbitrary but not both zero. Thus the eigenvectors corresponding to 1 lie in a 2-dimensional subspace. We do not have three linearly independent eigenvectors, so this matrix is defective.

**Theorem:** A matrix having distinct eigenvalues is diagonalizable.

**Proof for** \(n = 3\). Suppose

\[Ax_i = \lambda_i x_i, \quad x_i \neq 0, \quad i = 1, 2, 3.\]

Suppose

\[\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = 0.\]  \(\text{(4.2)}\)

We need to show \(\alpha_1 = \alpha_2 = \alpha_3 = 0\).

Multiply \(\text{(4.2)}\) by \((A - \lambda_2 I)(A - \lambda_3 I)\),

\[(A - \lambda_2 I)(A - \lambda_3 I)x_i = (A - \lambda_2 I)(\lambda_i - \lambda_3)x_i = (\lambda_i - \lambda_2)(\lambda_i - \lambda_3)x_i,
\]

and \(\text{(4.2)}\) becomes

\[\alpha_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)x_1 = 0.\]

Distinctness \(\Rightarrow\) \(\alpha_1 = 0\). \(\square\)

**Example** Recall that \[
\begin{bmatrix}
 0 & 1 \\
 1 & 1
\end{bmatrix}
\]
has eigenvalues \(\frac{1}{2} + \frac{\sqrt{5}}{2}, \frac{1}{2} - \frac{\sqrt{5}}{2}\). The corresponding eigenvectors are \[
\begin{bmatrix}
-\frac{1}{2} + \frac{\sqrt{5}}{2} \\
1
\end{bmatrix}, \begin{bmatrix}
-\frac{1}{2} - \frac{\sqrt{5}}{2} \\
1
\end{bmatrix}.
\]

So \[
\begin{bmatrix}
0 & 1 \\
1 & 1
\end{bmatrix} = X\begin{bmatrix}
\frac{1}{2} + \frac{\sqrt{5}}{2} & 0 \\
0 & \frac{1}{2} - \frac{\sqrt{5}}{2}
\end{bmatrix}X^{-1}
\]

where \(X = \begin{bmatrix}
-\frac{1}{2} + \frac{\sqrt{5}}{2} & -\frac{1}{2} - \frac{\sqrt{5}}{2} \\
1 & 1
\end{bmatrix}\), so

\[
\begin{bmatrix}
0 & 1 \\
1 & 1
\end{bmatrix}^k = \frac{1}{\sqrt{5}}\begin{bmatrix}
\left(\frac{1+\sqrt{5}}{2}\right)^{k-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{k-1} & \left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k \\
\left(\frac{1+\sqrt{5}}{2}\right)^{k-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{k-1} & \left(\frac{1+\sqrt{5}}{2}\right)^k + \left(\frac{1-\sqrt{5}}{2}\right)^k
\end{bmatrix}.
\]

For large \(k\), \(A^k\) is dominated by the largest eigenvalue(s). Hence the spectral radius

\[\rho(A) := \max_{1 \leq i \leq n} |\lambda_i(A)|.\]

**Jordan canonical form.** Any matrix can be transformed by means of a similarity transformation to the form \(\text{diag}(J_1, J_2, \ldots, J_m)\) where each basic Jordan block \(J_k\) has the form

\[
J_k = \begin{bmatrix}
\lambda_k & 1 \\
& \lambda_k \\
& & \ddots \\
& & & 1
\end{bmatrix}.
\]

The **Cayley-Hamilton Theorem** asserts that any matrix \(A\) satisfies \(p(A) = 0\) where \(p(\zeta)\) is the characteristic polynomial of \(A\).
4.1.4 Unitary similarity transformations

The $X$ needed to diagonalize a matrix can be very ill-conditioned, e.g., for $A = \begin{bmatrix} 1 + \varepsilon & 1 \\ 0 & 1 \end{bmatrix}$
we need $X = \begin{bmatrix} 1 & -1/\varepsilon \\ 0 & 1 \end{bmatrix}$ to get $X^{-1}AX = \begin{bmatrix} 1 + \varepsilon & 0 \\ 0 & 1 \end{bmatrix}$. (For $\varepsilon = 0$, $A$ is defective.) However, to get eigenvalues (and eigenvectors) it is sufficient to have $X^{-1}AX$ upper triangular. And it is possible to achieve this using an $X$ with $\kappa_2(X) = 1$. For a unitary matrix $U \in \mathbb{C}^{n \times n}$, $\kappa_2(U) = \|U^{-1}\|_2\|U\|_2 = \|U^H\|_2\|U\|_2 = 1$.

**THEOREM Schur decomposition** If $A \in \mathbb{C}^{n \times n}$, then there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ such that $R = U^H AU$ is upper triangular.

**Proof.** By induction on $n$. True for $n = 1$. Assume true for $n - 1$. We will seek unitary $P$ such that

$$P^H AP = \begin{bmatrix} \lambda & b^H \\ 0 & B \end{bmatrix}.$$ 

for some $\lambda$, $b^H$, $B$. In other words we want

$$P^H AP e_1 = \lambda e_1$$

or equivalently

$$A(Pe_1) = \lambda(Pe_1).$$

Let $\lambda$ be an eigenvalue of $A$ and $x$ an eigenvector with $\|x\|_2 = 1$. We are done if we can choose unitary $P$ so that $Pe_1 = x$. Equivalently we want $P^H x = e_1$. A generalized Householder reflection

$$P^H = I - 2 \frac{vv^H}{v^H v}, \quad v = x - e_1$$

can be shown to work. To guard against division by zero, choose $x$ so that $e_1^T x \leq 0$. There is unitary $V$ such that $V^H BV =: R$ is upper triangular. Let

$$U = P \begin{bmatrix} 1 \\ V \end{bmatrix}.$$ 

then

$$U^H AU = \begin{bmatrix} \lambda & b^H \\ 0 & R \end{bmatrix}. \quad \Box$$

The construction in this proof is known as deflation: it depends on being able to find one eigenvalue of a matrix. If $A$ is real and has only real eigenvalues then $U$ and $R$ may be chosen to be real. If $A$ is real but has some imaginary eigenvalues (which must occur in pairs $\lambda, \bar{\lambda}$), then neither $U$ nor $R$ can be real. Unfortunately complex arithmetic is expensive: double the storage, four times the computation. In the case of factoring polynomials we can avoid complex arithmetic by allowing the factors to be either linear or quadratic with two imaginary roots $\lambda$ and $\bar{\lambda}$. Analogously real matrices have a real Schur decomposition:
THEOREM *Real Schur decomposition* If \( A \in \mathbb{R}^{n \times n} \), then there exists an orthogonal matrix \( Q \in \mathbb{R}^{n \times n} \) such that \( R = Q^T AQ \) is block upper triangular with diagonal blocks that either are 1 by 1 or are 2 by 2 with imaginary eigenvalues.

### 4.1.5 Symmetric matrices

*Theorem* If \( A \) is real symmetric, then there exists a real orthogonal \( Q \) such that \( Q^T AQ = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \), \( \lambda_i \in \mathbb{R} \).

*Proof.* Consider the real Schur decomposition
\[
R = Q^T AQ.
\]
Then
\[
R^T = Q^T A^T Q = Q^T AQ = R.
\]
Hence \( R \) is symmetric block diagonal with blocks that either are 1 by 1 or are symmetric and 2 by 2 with imaginary eigenvalues. However, a 2 by 2 symmetric matrix cannot have imaginary eigenvalues, so \( R \) must be diagonal. \( \square \)

\[
AQ = QA \quad \text{and} \quad A(Qe_i) = (Qe_i)\lambda_i
\]

\( Qe_i \) is an eigenvector, and \( \lambda_i \) is eigenvalue. These eigenvectors form an orthonormal set. The eigenvalues are real. If \( A \) is s.p.d., the eigenvalues are positive.

### 4.1.6 Matrix norms

**THEOREM** The spectral norm

\[
\|A\|_2 = \sqrt{\rho(A^T A)}.
\]

*Proof.*

\[
\|A\|_2 = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x}.
\]

Since \( A^T A \) is real symmetric, there exists a real orthogonal \( Q \) and real diagonal \( \Lambda \) such that
\[
Q^T (A^T A) Q = \Lambda.
\]

So
\[
\|A\|_2^2 = \max_{x \neq 0} \frac{x^T Q A Q^T x}{x^T x}.
\]

102
Let $y = Q^T x$:

$$
\|A\|_2^2 = \max_{y \neq 0} \frac{y^T A y}{y^T y} = \max_{1 \leq i \leq n} \lambda_i \quad \text{(exercise)}
= \rho(A^T A). \quad \square
$$

If $A$ is symmetric

$$
\|A\|_2 = \rho(A^2)^{1/2} = \rho(A).
$$

**THEOREM**

$$
\|A^T\|_2 = \|A\|_2.
$$

**Proof.** It is enough to show that $\rho(AA^T) = \rho(A^T A)$, and for this it is enough to show that if $\lambda \neq 0$ then

$$
\lambda \text{ is an eigenvalue of } AA^T \iff \lambda \text{ is an eigenvalue of } A^T A.
$$

Thus let $AA^T x = \lambda x \lambda \neq 0$, $x \neq 0$. Then $A^T A(A^T x) = \lambda (A^T x)$. $\lambda \neq 0$ $\Rightarrow A^T x \neq 0$, which means $\lambda$ is an eigenvalue of $A^T A$. Similarly, a nonzero eigenvalue of $A^T A$ is an eigenvalue of $AA^T$. $\square$

### 4.1.7 * Gerschgorin disks

Useful information about eigenvalues is often provided by

**THEOREM (Gerschgorin)**

$$
\lambda(A) \subset D_1 \cup D_2 \cup \cdots \cup D_n
$$

where

$$
D_i = \{z : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \} \quad \text{Gerschgorin disk.}
$$

**Example**

$$
A = \begin{bmatrix}
1 & 0.1 & 0.1 \\
0.1 & 1.1 & 0.1 \\
0.1 & 0.1 & 2
\end{bmatrix} \quad D_1 : |z - 1| \leq 0.2 \\
D_2 : |z - 1.1| \leq 0.2 \\
D_3 : |z - 2| \leq 0.2
$$

complex plane complex plane complex plane
Proof.

\[ \lambda \in \lambda(A) \Rightarrow Ax = \lambda x \text{ for some } x \neq 0. \]

\[ (\lambda - a_{ii})x_i = \sum_{j \neq i} a_{ij}x_j. \]

\[ |\lambda - a_{ii}|x_i| \leq \sum_{j \neq i} |a_{ij}||x_j| \leq \|x\|_\infty \sum_{j \neq i} |a_{ij}|. \]

Choose \( i \) so that \( |x_i| = \|x\|_\infty \), and divide by this. \( \square \)

THEOREM If \( k \) Gerschgorin disks of the matrix \( A \) are disjoint from the other disks, then exactly \( k \) eigenvalues of \( A \) lie in the union of the \( k \) disks. \( \square \)

Example (continued). Two eigenvalues lie in \( D_1 \cup D_2 \); one lies in \( D_3 \). Since similarity transformations do not change \( \lambda(A) \), consider \( D^{-1} = \text{diag}(1, 1, \varepsilon) \) and

\[ D^{-1}AD = \begin{bmatrix}
1 & 0.1 & 0.1/\varepsilon \\
0.1 & 1.1 & 0.1/\varepsilon \\
0.1\varepsilon & 0.1\varepsilon & 2
\end{bmatrix}. \]

\( D_3 : |z - 2| \leq 0.2\varepsilon \)

contains an eigenvalue as long as \( D_3 \) and \( D_1 \cup D_2 \) are disjoint:

\[ 2 - 0.2\varepsilon > 1.2 + 0.1/\varepsilon \]

e.g., \( \varepsilon = 0.2 \). Hence \( |\lambda - 2| \leq 0.04 \) for some \( \lambda \in \lambda(A) \).

Review questions

1. For \( A \in \mathbb{C}^{m \times n} \) what is the meaning of \( A^H \)?
2. What is a Hermitian matrix?
3. What is a unitary matrix?
4. What is the inner product of \( x \) and \( y \) for \( x, y \in \mathbb{C}^n \)?
5. Define an eigenvalue and eigenvector.
6. What is the characteristic polynomial of a matrix?
7. Define the multiplicity of an eigenvalue.
8. What can we say about the eigenvalues of a real matrix?
9. What is the relationship between the determinant of a matrix and its eigenvalues?
10. Given eigenvalues and eigenvectors of a matrix $A$ what can we say about those of $f(A)$ where $f(\zeta)$ is a polynomial? where $f(\zeta) = \zeta^{-1}$?

11. What can we say about the determinant of a block triangular matrix with diagonal blocks $T_{11}, T_{22}, \ldots, T_{pp}$?

12. What can we say about the eigenvalues of a block triangular matrix with diagonal blocks $T_{11}, T_{22}, \ldots, T_{pp}$?

13. Define a similarity transformation.

14. Show that being similar is an *equivalence relation*.

15. Show that similar matrices have the same eigenvalues including multiplicities.

16. What does it mean for a matrix to be diagonalizable?

17. What condition on the eigenvalues is sufficient for a matrix to be diagonalizable? Prove it.

18. What is a defective matrix? Give an example.

19. What condition on the eigenvalues of a matrix is sufficient for it to be diagonalizable?

20. How can one express the $k$th power of a matrix in terms of the $k$th power of scalars?

21. Define the spectral radius of a matrix?

22. What is a basic Jordan block?

23. What is a Jordan canonical form?

24. What can be proved about Jordan canonical forms?

25. What is the Cayley-Hamilton Theorem?

26. What is the condition number of a unitary matrix?

27. What is a Schur decomposition?

28. How many linearly independent eigenvectors must an $n$ by $n$ matrix have?

29. What is the name of the process called by which we “factor out” an eigenvalue and eigenvector of a matrix?

30. Under what condition can a Schur decomposition consist of real matrices?

31. To what extent can we factor a polynomial with real coefficients into lower degree polynomials with real coefficients?
32. What is a real Schur decomposition of a real matrix?
33. What form does a real Schur decomposition of a real symmetric matrix take?
34. What can we say about the eigenvalues and eigenvectors of a real symmetric matrix?
35. What can be said about the eigenvalues and eigenvectors of a Hermitian matrix?
36. What is the relation between a consistent norm of a matrix and its eigenvalues?
37. Define the spectral norm and give a formula for it.
38. What is the spectral norm of a symmetric matrix?
39. What is the spectral norm of the transpose of a matrix?

Exercises

1. Find the eigenvectors of \[
\begin{bmatrix}
1 & 2 & 4 \\
0 & 2 & 0 \\
0 & 1 & 2
\end{bmatrix}.
\]

2. Suppose \(A - 2I\) is a singular matrix. What, in terms of the matrix \(A\), is (special about) the null space of \(A - 2I\)?

3. (a) Show that if a real matrix \(A\) satisfies \(x^H A x > 0\) for all complex \(x \neq 0\), it must be symmetric positive definite.
   
   (b) Let \(A\) be a complex matrix. Show that if \(x^H A x\) is real for all complex \(x\), it must be Hermitian.

4. A matrix is idempotent if \(A^2 = A\). Show that if \(A\) is idempotent, then its eigenvalues are zeros and ones.

5. Determine the eigenvalues of \(xy^T\) where \(x\) and \(y\) are column vectors.

6. Show that \[
\begin{bmatrix}
0 & 1 \\
0 & 0
\end{bmatrix}
\] is not diagonalizable.

7. (a) Let \(A \in \mathbb{R}^{n \times n}\) have the nonreal eigenvalue \(\lambda\). Show that there exists \(X \in \mathbb{R}^{n \times 2}\), \(M \in \mathbb{R}^{2 \times 2}\) satisfying \(AX = XM\) where \(X\) has linearly independent columns and \(\lambda(M) = \{\lambda, \bar{\lambda}\}\).
   
   (b) Show that there exists an orthogonal matrix \(Q\) and a nonsingular matrix \(T\) such that
   \[
   Q^T X T^{-1} = \begin{pmatrix}
   I_2 \\
   0
   \end{pmatrix}.
   \]
(c) Show that $Q^T A Q = \begin{pmatrix} B & C \\ 0 & E \end{pmatrix}$ where $\lambda(P) = \{\lambda, \bar{\lambda}\}$. Hint: obtain the first two columns by some matrix operation.

8. Show that if $\rho(A) < 1$ then $I - A$ is nonsingular.

9. Show that a Hermitian matrix has real eigenvalues. You may use the fact that $y^H y \geq 0$ for any complex vector $y$ and that $\lambda^2 \geq 0$ implies $\lambda$ is real.

10. Show that the eigenvalues of a real symmetric positive definite matrix are positive.

11. Show that for a real diagonal matrix $D$

$$\max_{x \neq 0} \frac{x^T D x}{x^T x} = \max_{1 \leq i \leq n} d_{ii}.$$ 

12. Show that for a consistent matrix norm the condition number

$$\kappa(A) \geq \frac{\max |\lambda(A)|}{\min |\lambda(A)|}.$$ 

13. Why can we be sure that all the eigenvalues of the following matrix are real? Give a complete answer.

$$\begin{bmatrix} 2 & 0.5 & 0.2 & 0.1 \\ 0 & 4 & 0.4 & -0.3 \\ 0.3 & 0.2 & 6 & -0.1 \\ 0.1 & -0.1 & 0.1 & 8 \end{bmatrix}$$

14. Show that the spectral radius of a matrix does not exceed the norm of the matrix for any matrix norm subordinate to some vector norm.

15. What is the spectral radius of $\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}$? What is the spectral norm of this matrix?

16. from BIT “(University of Uppsala, Second Course) The eigenvalues to the matrix

$$A(\epsilon) = \begin{pmatrix} 3 & 0.5 & \epsilon & 0 \\ 0.5 & 1 & -\epsilon & 0 \\ \epsilon & -\epsilon & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

are separated for small enough $\epsilon$. Localize the eigenvalues and determine how large $\epsilon$ can be and still guarantee that the eigenvalues are separated.

*Answer:* The eigenvalues are close to 0.882, 1, 3 and 3.118. They are separated as long as $|\epsilon| \leq 0.04$.” Hint: be willing to diagonalize 2 by 2 matrices.
17. What are the possible Jordan canonical forms for a real 4 by 4 matrix two of whose eigenvalues are both equal to i?

18. Extend the construction of the Householder reflection to include any complex vector $x$. That is, construct an elementary unitary matrix $P$ such that $Px$ is a scalar multiple of $e_1$.

19. What are the possible Jordan canonical forms for a real 4 by 4 matrix two of whose eigenvalues are both equal to i?

20. Give an example of a 2 by 2 matrix $A$ such that $\rho(A) = 0$ but $\|A\|_2 = 1$.

21. Extend the construction of the Householder reflection to include any complex vector $x$. That is, construct an elementary unitary matrix $P$ such that $Px$ is a scalar multiple of $e_1$.

22. Let $c, s \in \mathbb{R}$ satisfy $c^2 + s^2 = 1$. The matrix $\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$ has eigenvalues $\lambda_1 = c + is$ and $\lambda_2 = c - is$. Calculate the eigenvector $x_1$ corresponding to $\lambda_1$.

   Calculate $x_1^T x_1$.

   Calculate $x_1^H x_1$.

23. Let $A$ be a 5 by 3 matrix with singular values 2, 1, 0. For the following questions it is enough to state the answer. What are the eigenvalues of $A^T A$?

24. Find and correct the subtle but fatal flaw in the following proof:

   **Theorem** The eigenvalues of an orthogonal matrix have modulus 1.

   **Proof.** Let $\lambda$ be an eigenvalue of an orthogonal matrix $Q$. Then $Qx = \lambda x$ and $x^T Q^T = \lambda x^T$, whence

   $x^T Q^T Q x = \lambda^2 x^T x$

   $x^T x = \lambda^2 x^T x$

   $1 = \lambda^2$

   $1 = |\lambda|$.
4.2 Reduction to Upper Hessenberg Form

In a finite number of arithmetic operations no algorithm can by similarity transformations triangularize $A$. The best we can do is

$$Q^T A Q = H = \begin{bmatrix}
\times & \times & \times & \cdots & \cdots & \times \\
\times & \times & \times & \cdots & \cdots & \times \\
\times & \times & \cdots & \cdots & \times \\
\ddots & \ddots & \ddots & \ddots & \ddots \\
\times & \times & \times & \cdots & \times \\
\times & \times & \times & \cdots & \times \\
\end{bmatrix}$$

upper Hessenberg form

We can do this column by column.

Suppose that elements in the first $k - 1$ columns under the first subdiagonal have already been eliminated so that the matrix has the form

$$\begin{bmatrix}
H_k \\
0 & a_k \\
\end{bmatrix} \begin{bmatrix}
B_k \\
A_k \\
\end{bmatrix}$$

where $H_k$ is a $k$ by $k$ upper Hessenberg matrix and $a_k$ is an $n - k$ by 1 column vector. To leave $H_k$ unchanged, we seek an orthogonal transformation of the form

$$\begin{bmatrix}
I & 0 \\
0 & P_k^T \\
\end{bmatrix} \begin{bmatrix}
H_k \\
0 & a_k \\
\end{bmatrix} \begin{bmatrix}
B_k \\
A_k \\
\end{bmatrix} = \begin{bmatrix}
H_k \\
0 & P_k^T a_k \\
\end{bmatrix} \begin{bmatrix}
B_k P_k \\
P_k \\
\end{bmatrix}.$$

Also, we want $P_k^T a_k$ to be all zero except for its first element. This can be accomplished with a Householder reflection, for which $P_k^T = P_k$.

The cost of doing the $k$th column is as follows:

- determine $P_k = I - \beta_k v_k v_k^T$ $n - k + 1$ mults, 1 square root
- form $P_k A_k$ $2(n - k)$ mults per column times $(n - k)$ columns
- form $\begin{bmatrix} B_k \\ A_k \end{bmatrix} P_k$ $2(n - k)$ mults per row times $n$ rows
- $n - k + 1 + 2(n - k)^2 + 2n(n - k)$ mults

The total cost is $\frac{2}{3} n^3$ multiplications.

In sum with

$$P_k := \begin{bmatrix}
I & 0 \\
0 & P_k \\
\end{bmatrix}$$

we have

$$P_{n-2}(\cdots (P_2 (P_1 A P_1) P_2) \cdots) P_{n-2} =: H.$$

If $A$ is symmetric, then $H$ is symmetric.

symmetric + upper Hessenberg $\Rightarrow$ tridiagonal

Cost: $\frac{2}{3} n^3$ mults.
Review questions

1. What is an upper Hessenberg matrix?

2. How close can we come to upper triangularizing a matrix in a finite number of arithmetic operations?

3. Show how an orthogonal similarity transformation can be used to zero out all but the first two elements in the first column of a matrix. For the orthogonal matrix it is sufficient to state a previously obtained result.

4. If a square matrix $A$ is reduced by a similarity transformation to an upper Hessenberg matrix $H$ using Householder reflections $P_1, P_2, \ldots$, what is the resulting factorization of $A$?

5. What results if a symmetric matrix is reduced to upper Hessenberg form using orthogonal similarity transformations?

6. How does the cost of reducing an $n$ by $n$ matrix to upper Hessenberg form using orthogonal similarity transformations depend on $n$ for a general matrix? for a symmetric matrix?

Exercises

1. By means of an orthogonal similarity transformation make the following matrix tridiagonal:

$$
\begin{bmatrix}
4 & 2 & 1 \\
2 & 4 & 2 \\
1 & 2 & 4
\end{bmatrix}.
$$

Note that the Householder reflection for a vector $\begin{bmatrix} x \\ y \end{bmatrix}$ is given by

$$
-\text{sign}(x)(x^2 + y^2)^{-1/2} \begin{bmatrix} x \\ y \end{bmatrix}.
$$

2. Apply an orthogonal similarity transformation to the following matrix so that the (3,1) and (4,1) elements of the transformed matrix are both zero:

$$
\begin{bmatrix}
1 & 1 & 3 & 2 \\
-4 & 0 & 1 & 0 \\
0 & 0 & 2 & 1 \\
3 & 5 & 1 & 0
\end{bmatrix}
$$

You may leave your answer as a product of 4 by 4 matrices.

110
3. Reduction of a full 5 by 5 matrix $A$ to upper Hessenberg form $H$ can be accomplished by a similarity transformation

$$H = P_q \cdots P_2 P_1 A P_2 \cdots P_q$$

where each $P_k$ is a Householder reflection.

(a) How many reflections $q$ are used?

(b) In what order are the $2q$ matrix products computed? (You may use parentheses to show this.)

(c) Indicate the structure of $P_2$ using the symbols 0, 1, $\times$ for elements of $P_2$, where $\times$ represents an arbitrary value.

4. Determine an orthogonal similarity transformation that makes the following matrix tridiagonal:

$$
\begin{bmatrix}
1 & 3 & 4 \\
3 & -4 & 3 \\
4 & 3 & 1
\end{bmatrix}
$$

Hint: $(x^2 + y^2)^{-1/2} \begin{bmatrix} x & y \\ y & -x \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (x^2 + y^2)^{1/2} \\ 0 \end{bmatrix}$.

4.3 The Power Method

Eigenvalues are the roots of the characteristic polynomial $p(\lambda) := \det(\lambda I - A)$. Thus we expect no direct method; any method for finding eigenvalues must have an iterative part.

The following algorithm is unstable:

form the characteristic polynomial;
find the roots of the characteristic polynomial.

This is because finding the roots of the polynomial can be much more ill-conditioned than finding eigenvalues; i.e., the results are much more sensitive to errors in the coefficients. On the other hand the roots of $\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = 0$ are the eigenvalues of its companion matrix

$$
\begin{bmatrix}
0 & -c_0 \\
1 & -c_1 \\
1 & \ddots & -c_2 \\
\vdots & \ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots & \ddots & 0 \\
\cdots & 0 & -c_{n-2} \\
1 & \cdots & \cdots & 0 & -c_{n-1}
\end{bmatrix}
$$
It is reasonable to solve a polynomial equation by computing the eigenvalues of its companion matrix.

The power method is good for a few eigenvalues and eigenvectors. If $A$ is diagonalizable,

$$X^{-1}AX = \Lambda,$$

$$A[x_1, x_2, \ldots, x_n] = [x_1, x_2, \ldots, x_n] \text{ diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Assume there is a unique eigenvalue of maximum magnitude:

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$ (What can we say about $\lambda_1$?) Let $z_0 =$ initial guess for $x_1$. Algorithm:

$$z_1 = Az_0, \quad z_2 = Az_1, \quad \ldots, \quad z_{k+1} = Az_k$$

$$z_k = A^k z_0$$
$$= (X\Lambda X^{-1})^k z_0$$
$$= X\Lambda^k X^{-1} z_0$$
$$= X\lambda_1^k \text{ diag } \left(1, \left(\frac{\lambda_2}{\lambda_1}\right)^k, \ldots, \left(\frac{\lambda_n}{\lambda_1}\right)^k\right) X^{-1} z_0$$
$$= \lambda_1^k X \left(e_1 e_1^T + O\left(\frac{\lambda_2}{\lambda_1}\right)^k\right) X^{-1} z_0$$
$$= \lambda_1^k \left(x_1(e_1^T X^{-1} z_0) + O\left(\frac{|\lambda_2|}{\lambda_1}\right)^k\right)$$

assuming $e_1^T X^{-1} z_0 \neq 0$.

The Rayleigh quotient

$$\frac{z_k^T A z_k}{z_k^T z_k} = \lambda_1 + O\left(\frac{|\lambda_2|}{\lambda_1}\right)^k.$$ The Rayleigh quotient is the value $\lambda$ that satisfies $z_k \lambda \approx A z_k$. If $A$ is symmetric, then

$$\frac{z_k^T A z_k}{z_k^T z_k} = \lambda_1 + O\left(\frac{|\lambda_2|}{\lambda_1}\right)^{2k}.$$ The power method is good only for the largest eigenvalue in magnitude.

Inverse iteration. Apply power method to $A^{-1}$ to get smallest eigenvalue in magnitude assuming $|\lambda_n| < |\lambda_{n-1}|$. More generally consider

$$(A - \sigma I)^{-1}, \quad \sigma \in \mathbb{C},$$

112
which has eigenvalues
\[ \frac{1}{\lambda_1 - \sigma}, \frac{1}{\lambda_2 - \sigma}, \ldots, \frac{1}{\lambda_n - \sigma}. \]
The largest of these is \( \frac{1}{\lambda_p - \sigma} \) where \( \lambda_p \) is closest to \( \sigma \).

\[ z_{k+1} = (A - \sigma I)^{-1} z_k \quad \text{use LU factorization} \]

Note that \( Az_{k+1} = \sigma z_{k+1} + z_k \) is available, so use

\[ \lambda_p \approx \frac{z_{k+1}^T A z_{k+1}}{z_{k+1}^T z_{k+1}} = \sigma + \frac{z_{k+1}^T z_k}{z_{k+1}^T z_{k+1}}, \]

for which one can show the error is \( O(|\lambda_p - \sigma|^k/|\lambda_q - \sigma|^k) \) where \( \lambda_q \) is the second closest eigenvalue to \( \sigma \).

To prevent overflow or underflow, form a normalized sequence:

\[ y_k = z_k/\|z_k\|_2, \]
\[ z_{k+1} = (A - \sigma I)^{-1} y_k. \]

Then

\[ \lambda_p \approx \sigma + \frac{z_{k+1}^T y_k}{z_{k+1}^T z_{k+1}}. \]

### 4.3.1 Rayleigh quotient iteration

The idea is to use the latest Rayleigh quotient estimate to continually update the eigenvalue estimate \( \lambda^* \) in inverse iteration.

\[ z_{k+1} = (A - \sigma_k I)^{-1} y_k \]

Rayleigh quotient estimate for \( A \)

\[ \sigma_{k+1} = \sigma_k + \frac{z_{k+1}^T y_k}{z_{k+1}^T z_{k+1}} \]

The order of convergence is quadratic in general

\[ \|y_{k+1} - (\pm x_p)\| = O(\|y_k - x_p\|^2) \]

and cubic for a symmetric matrix

\[ \|y_{k+1} - (\pm x_p)\| = O(\|y_k - x_p\|^3). \]

If eigenvalue \( \lambda_p < 0 \), the signs will alternate.
Review questions

1. Given a polynomial, how does one construct a matrix whose eigenvalues are the roots of the polynomial? What do we call such a matrix?

2. What is the algorithm for the power method?

3. What limitation of floating-point arithmetic is a significant danger for the power method and how is this avoided?

4. Under what condition does the power method converge and what does it converge to?

5. Assuming the power method converges, how rapidly does it converge?

6. What is the Rayleigh quotient and what does it approximate?

7. How can we generalize the power method to estimate the eigenvalue closest to some given value \( \sigma \)?

8. Assuming inverse iteration with a (fixed) shift \( \sigma \) converges, how rapidly does it converge?

9. Explain in words the idea of Rayleigh quotient iteration. What is its order of convergence in general? for a symmetric matrix?

Exercises

1. Beginning with initial guess \( \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \), do two iterations of the power method for

\[
\begin{bmatrix}
1 & 0.1 & 0.1 \\
0.1 & 1.1 & 0.1 \\
0.1 & 0.1 & 2
\end{bmatrix}.
\]

Also form the Rayleigh quotients for the zeroth and first iterates.

2. If we applied the power method to \((A - \sigma I)^{-1}\), we could change \( \sigma \) from one iteration to the next using at each iteration a new and better approximation to an eigenvalue of \( A \). There is a serious drawback, however. What is it? Secondly, what difficulty might we expect if \( \sigma \) is extremely close to an eigenvalue of \( A \)?

3. For the matrix

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{bmatrix}
\]
do two iterations of the power method beginning with \([1, 1, 1, 1, 1]^T\). (Renormalization is unnecessary.) Use the results of the two iterations to estimate the largest eigenvalue of the matrix.

### 4.4 The QR Iteration

What if we try to make \(A\) upper triangular rather than upper Hessenberg? Consider the 2 \(\times\) 2 case

\[
A = \begin{bmatrix}
a & b \\
c & d
\end{bmatrix}.
\]

Then

\[
v = \begin{bmatrix}
a + \text{sign}(a) \sqrt{a^2 + c^2} \\
c
\end{bmatrix}
\]

after much algebra

\[
P = \frac{-\text{sign}(a)}{\sqrt{a^2 + c^2}} \begin{bmatrix}
a & c \\
c & -a
\end{bmatrix}.
\]

For example,

\[
A = \begin{bmatrix}
2 & -1 \\
1 & 0
\end{bmatrix}, \quad P = \frac{-1}{\sqrt{5}} \begin{bmatrix}
2 & 1 \\
1 & -2
\end{bmatrix} =: Q_1^T \quad \text{our orthogonal matrix}
\]

\[
Q_1^T A = \frac{-1}{\sqrt{5}} \begin{bmatrix}
5 & -2 \\
0 & -1
\end{bmatrix}, \quad \text{but} \quad Q_1^T A Q_1 = \begin{bmatrix}
\frac{8}{5} & \frac{9}{5} \\
\frac{-1}{5} & \frac{2}{5}
\end{bmatrix} =: A_1.
\]

After postmultiplication it is no longer upper triangular, but the (2,1) element has been reduced. Using \(A_1\)

\[
Q_2^T = \frac{-1}{\sqrt{\frac{25}{13}}} \begin{bmatrix}
\frac{8}{5} & -\frac{1}{5} \\
\frac{-1}{5} & \frac{8}{5}
\end{bmatrix}, \quad A_2 := Q_2^T A_1 Q_2 = \begin{bmatrix}
\frac{18}{13} & \frac{25}{13} \\
\frac{-1}{13} & \frac{18}{13}
\end{bmatrix}.
\]

The (2,1) element is still smaller.

\[
Q_3^T = \frac{-1}{\sqrt{\frac{25}{13}}} \begin{bmatrix}
\frac{18}{13} & \frac{1}{13} \\
\frac{-1}{13} & \frac{18}{13}
\end{bmatrix}, \quad A_3 := Q_3^T A_2 Q_3 = \begin{bmatrix}
\frac{-23}{25} & \frac{49}{25} \\
\frac{1}{25} & -\frac{23}{25}
\end{bmatrix}.
\]

And still smaller. However convergence becomes very slow. (Eigenvalues are 1, 1.)

For

\[
A = \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}, \quad Q_1^T = \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}, \quad \text{and} \quad Q_1^T A Q_1 = \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\]

there is no convergence. (Eigenvalues are 1, -1.)
More generally for an \( n \times n \) matrix
\[
A_0 := A, \\
A_{k+1} := (Q_{k+1}^T A_k) Q_{k+1}
\]
where \( Q_{k+1}^T \) is a sequence of \( n - 1 \) Householder reflections. As \( k \to \infty \), \( A_k \) becomes a block upper triangular matrix. The first block contains eigenvalues of largest magnitude, the second block contains eigenvalues of next to largest magnitude, \ldots, the last block contains eigenvalues of least magnitude. If \( |\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| \),
\[
A_k \to \begin{bmatrix}
\lambda_1 \times \cdots \times \\
\lambda_2 \cdots \times \\
\vdots \\
\lambda_n
\end{bmatrix}
\]
Schur form

If \( A \in \mathbb{R}^{n \times n} \), this is possible only if all eigenvalues are real. Rate of convergence depends on eigenvalue ratios, in particular \( (A_k)_{ij} = O(|\lambda_i/\lambda_j|^k) \) for \( j < i \).

Convergence can be improved by a shifted QR iteration. For example, if \( \lambda(A) = \{-1, 1, 1.1\} \), then \( \lambda(A - 0.9I) = \{-1.9, 0.1, 0.2\} \), which are much better separated. The iteration is
\[
A_{k+1} := (Q_{k+1}^T (A_k - \sigma_{k+1} I)) Q_{k+1} + \sigma_{k+1} I .
\]

Note that as before \( A_{k+1} = Q_{k+1}^T A_k Q_{k+1} \), but a different \( Q_{k+1} \) is computed. For the shifts we use approximations to eigenvalues, which can be obtained from diagonal elements of \( A_k \). In particular, we seek for \( \sigma_{k+1} \) an approximation to the smallest eigenvalue, for which the \((n, n)\) entry is a typical choice.

From before,
\[
A_1 = \begin{bmatrix}
\frac{8}{5} & \frac{9}{5} \\
-\frac{7}{5} & \frac{2}{5}
\end{bmatrix}.
\]
The last diagonal element is an approximation to the smaller eigenvalue. So we use a shift \( \sigma_2 = \frac{2}{5} \);
\[
A_1 - \sigma_2 I = \begin{bmatrix}
\frac{6}{5} & \frac{9}{5} \\
-\frac{7}{5} & \frac{2}{5}
\end{bmatrix}, \\
Q_2^T = \frac{-1}{\sqrt{\frac{37}{25}}} \begin{bmatrix}
\frac{6}{5} & \frac{1}{5} \\
-\frac{7}{5} & \frac{3}{5}
\end{bmatrix},
\]
\[
A_2 = Q_2^T (A_1 - \sigma_2 I) Q_2 + \sigma_2 I = \begin{bmatrix}
\frac{242}{185} & \frac{361}{185} \\
\frac{185}{185} & \frac{185}{185}
\end{bmatrix},
\]
which is better than without the shift. The cost of a QR iteration is \( \frac{4}{3} n^3 \) multiplications in general, but only \( 4n^2 \) multiplications for \( A_k = \) upper Hessenberg.

The shifted QR iteration preserves an upper Hessenberg form. More specifically, the following can be proved (Exercise 2):
Assume that $A_k$ is upper Hessenberg with nonzero subdiagonal elements\(^2\) and that

$$A_k - \sigma_{k+1} I = Q_{k+1} R_{k+1}$$

where $Q_{k+1}$ is orthogonal and $R_{k+1}$ is upper triangular.

Then $A_{k+1} = R_{k+1} Q_{k+1} + \sigma_{k+1} I$ is upper Hessenberg.

Thus we should choose

$$A_0 = (P_{n-2}(\cdots((P_1 A)P_1)\cdots))P_{n-2} \quad \text{upper Hessenberg.}$$

If $A$ is symmetric, then $A_k$ is symmetric and therefore \textit{tridiagonal}.

\textbf{Review questions}

1. What does one QR iteration do to the eigenvalues of a matrix?

2. In the QR iteration $A_{k+1} = Q_{k+1}^T A_k Q_{k+1}$ what is special about $Q_{k+1}$? Give a complete answer.

3. What does one QR iteration $A_{k+1} = Q_{k+1}^T A_k Q_{k+1}$ do to the eigenvectors of a matrix?

4. If $A_k$ is a sequence of square matrices produced by a QR iteration, under what assumption do the elements below the diagonal converge to zero? Under this assumption what do the diagonal elements converge to?

5. Give the formula for one iteration of a shifted QR iteration and explain how the $Q$ matrix is chosen.

6. What does one shifted QR iteration do to the eigenvalues of a matrix?

7. What preprocessing step is recommended before doing a QR iteration?

8. How does the cost of a QR iteration for an $n$ by $n$ matrix depend on $n$ for a general matrix? for an upper Hessenberg matrix?

9. Name some computationally significant properties that are preserved by a QR iteration using Householder reflections or Givens rotations.

\textbf{Exercises}

1. Show that a single QR iteration is enough to transform the square matrix $xy^T$ to a right triangular matrix. Assume $x \neq 0$ and $y_1 \neq 0$. Hint: what can you prove about a rank one upper triangular matrix?

2. Prove that the shifted QR iteration preserves an upper Hessenberg form for a matrix with nonzero subdiagonal elements. \textit{Hints:}

\(^2\)If a subdiagonal element is zero, the eigenvalue problem divides into two smaller problems.
(i) Do a \((1, n-1)\) by \((n-1, 1)\) partitioning of \(A_k - \sigma_{k+1} I\) and \(Q_{k+1}\) and an \((n-1, 1)\) by \((n-1, 1)\) partitioning of \(R_{k+1}\), and show that \(Q_{k+1}\) must be upper Hessenberg.

(ii) Do not use the fact that \(Q_{k+1}\) is orthogonal.

4.5 The Lanczos Method

The Lanczos method computes an approximation to the extreme eigenvalues of a symmetric matrix. The generalization to a nonsymmetric matrix is known as the Arnoldi method.

The most direct way of presenting this method is to consider the transformation of a symmetric matrix \(A\) to tridiagonal form

\[
AQ = QT
\]  

(4.3)

where \(Q = [q_1 q_2 \cdots q_n]\) is an orthogonal matrix and \(T\) is a symmetric tridiagonal matrix

\[
T = \begin{bmatrix}
\alpha_1 & \beta_1 & & \\
\beta_1 & \alpha_2 & \ddots & \\
& \ddots & \ddots & \beta_{n-1} \\
& & \beta_{n-1} & \alpha_n
\end{bmatrix}
\]

For large sparse matrices, most of whose elements are zeros, it is impractical to compute all of this. As a heuristic the Lanczos method computes only the upper left \(m\) by \(m\) block of \(T\), call it \(T_m\),

\[
T = \begin{bmatrix}
T_m & 0 \\
0 & \beta_m
\end{bmatrix}
\]

and computes its extreme eigenvalues as an approximation to those of \(T\).

The algorithm operates by satisfying eq. (4.3) column by column. Taking the first column of this equation gives

\[
Aq_1 = \alpha_1 q_1 + \beta_1 q_2.
\]

Together with the condition that \(q_1, q_2\) be an orthonormal multiset, this constitutes \(n + 3\) equations for \(2n + 2\) unknowns. It happens that the factorization (4.3) is not unique, and we are free to choose \(q_1\) to be an arbitrary normalized vector. We are left with \(n + 2\) equations for \(n + 2\) unknowns. We obtain

\[
q_2 = (Aq_1 - \alpha_1 q_1)/\beta_1
\]
where $\alpha_1$ is obtained by the orthogonality of $q_1$, $q_2$ and $\beta_1$ by normalization of $q_2$.

Taking the $k$th column of eq. (4.3) gives

$$q_{k+1} = (Aq_k - \alpha_k q_k - \beta_{k-1} q_{k-1})/\beta_k.$$  

Orthogonality of $q_{k+1}$ with $q_j$ for $j \leq k - 1$ is automatic, orthogonality with $q_k$ determines $\alpha_k$, and normalization determines $\beta_k$.

**Review questions**

1. What distinguishes the Lanczos method from the Arnoldi method?

2. Which eigenvalues of a matrix does the Lanczos method find?

**Exercises**

1. Show that if the Lanczos algorithm is carried to completion, the last column of eq. (4.3) can be satisfied.

2. Fill in the blanks for Lanczos method:

   $q_1 = \text{arbitrary normalized vector}$;
   $p = Aq_1$;
   $\alpha_1 = \text{____;}$
   $p = p - \alpha_1 q_1$;
   $\beta_1 = \text{____;}$
   $q_2 = \text{____;}$

   for $k = 2, \ldots, n - 1$ {
   
   $p = Aq_k$;
   $\alpha_k = \text{____;}$
   $p = p - \alpha_k q_k - \beta_{k-1} q_{k-1}$;
   $\beta_k = \text{____;}$
   $q_{k+1} = \text{____;}$
   
   }

   $\alpha_n = \text{____;}$

**4.6 Other Eigenvalue Algorithms**

**4.6.1 Jacobi**

The Jacobi method is applicable to symmetric matrices. The idea is to use a Givens rotation in a similarity transformation to zero out a designated element of the matrix. To zero out
the \((i,j)\)th element of a matrix, \(j < i\), we choose a sine \(s\) and cosine \(c\) so that

\[
\begin{bmatrix}
c & s \\
-s & c
\end{bmatrix}
\begin{bmatrix}
a_{ii} & a_{ij} \\
a_{ij} & a_{jj}
\end{bmatrix}
\begin{bmatrix}
c & -s \\
s & c
\end{bmatrix} =
\begin{bmatrix}
da'_{ii} & 0 \\
0 & a'_{jj}
\end{bmatrix}.
\]

This rotation is then used to recombine the \(i\)th and \(j\)th rows and the \(i\)th and \(j\)th columns of the full matrix \(A\). For formulas for \(s\) and \(c\) see the book by Heath.

These transformations are applied in succession to all the off diagonal elements of the matrix. However, the annihilation of one pair of elements will cause previously annihilated elements to reappear. Nonetheless, the process does converge fairly quickly if it is repeated again and again.

### 4.6.2 Bisection

Let \(A\) be a symmetric tridiagonal matrix without any nonzeros in its subdiagonal. (If there is a zero, then the matrix is block diagonal and the eigenvalues of each block can be found independently.)

By constructing what is known as a Sturm sequence, it is possible to count the number of eigenvalues of \(A\) that are less than any given real number \(\sigma\). This information can be used by the bisection method to isolate eigenvalues.

The method relies upon the fact that the leading principal submatrices of \(A\) have strictly interlacing eigenvalues; more specifically, (strictly) between any two successive eigenvalues of the \(k\)th order leading principal submatrix \(A_k\) is exactly one eigenvalue of \(A_{k-1}\).

A Sturm sequence \(d_0, d_1, \ldots, d_n\) is defined by setting \(d_0 = 1\) and \(d_k = \text{det}(A_k - \sigma I)\) for \(k \geq 1\). These determinants can be calculated efficiently. The eigenvalues of \(A_k - \sigma I\), \(k = 1, 2, \ldots, n\) are strictly interlacing. We can use this fact together with the fact that \(d_k\) is the product of the eigenvalues of \(A_k - \sigma I\) to count the number of negative eigenvalues \(\nu_k\) of \(A_k - \sigma I\). Set \(\nu_0 = 0\) and suppose that \(\nu_{k-1}\) is known. Then the interlacing property implies that \(\nu_k\) is either \(\nu_{k-1}\) or \(\nu_{k-1} + 1\). If \(d_{k-1}\) and \(d_k\) are both positive or both negative, then \(\nu_k = \nu_{k-1}\). If they are of (strictly) opposite sign, then \(\nu_k = \nu_{k-1} + 1\). The case of a zero value in the Sturm sequence is left as an exercise. The result of such a calculation is a count \(\nu_n\) of the number of negative eigenvalues of \((A - \sigma I)\) or equivalently the number of eigenvalues of \(A\) which are less than \(\sigma I\).

**Exercise**

1. Explain in the cases where \(d_{k-1}\) and/or \(d_k\) is zero whether or not to increment the count of negative eigenvalues.

### 4.7 Computing the SVD

Consider the case \(A \in \mathbb{R}^{m \times n}, n \leq m\).
1. eigenvalues, eigenvectors of $A^T A, AA^T$:

\[
(A^T A)(V e_i) = \sigma_i^2 (V e_i), \\
(AA^T)(U e_i) = \begin{cases} 
\sigma_i^2 (U e_i), & i \leq n, \\
0, & i \geq n + 1.
\end{cases}
\]

2. Golub–Kahan algorithm

I. $U_0^T A V_0 = \begin{bmatrix} B \\ 0 \end{bmatrix}$

\[
B = \begin{bmatrix}
 b_{11} & b_{12} & 0 \\
 b_{22} & \ddots & \vdots \\
 0 & \ddots & b_{n-1,n} \\
 0 & & b_{nn}
\end{bmatrix}
\]

II. $U_1^T B V_1 = \Sigma_1$

Therefore

\[
A = U_0 \begin{bmatrix} U_1 & I \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V_1^T V_0^T.
\]

Review question

1. Show how to transform the problem of computing a singular value decomposition into problems of diagonalizing matrices.

Exercise

1. Let $A \in \mathbb{R}^{m \times n}$, $n < m$. Using the SVD, relate the eigenvalues of $AA^T$ to those of $A^T A$. 