3. Symmetric eigendecomposition

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- low rank matrix approximation
Eigenvalues and eigenvectors

A nonzero vector $x$ is an eigenvector of the $n \times n$ matrix $A$, with eigenvalue $\lambda$, if

$$Ax = \lambda x$$

- the matrix $\lambda I - A$ is singular and $x$ is a nonzero vector in the nullspace of $\lambda I - A$
- the eigenvalues of $A$ are the roots of the characteristic polynomial:

$$\det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + (-1)^n \det(A) = 0$$

- this immediately shows that every square matrix has at least one eigenvalue
- the roots of the polynomial (and corresponding eigenvectors) may be complex
- (algebraic) multiplicity of an eigenvalue is its multiplicity as a root of $\det(\lambda I - A)$
- there are exactly $n$ eigenvalues, counted with their multiplicity
- set of eigenvalues of $A$ is called the spectrum of $A$
Diagonal matrix

\[ A = \begin{bmatrix}
A_{11} & 0 & \cdots & 0 \\
0 & A_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_{nn}
\end{bmatrix} \]

- eigenvalues of \( A \) are the diagonal entries \( A_{11}, \ldots, A_{nn} \)
- the \( n \) unit vectors \( e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1) \) are eigenvectors:
  \[ Ae_i = A_{ii}e_i \]
- linear combinations of \( e_i \) are eigenvectors if the corresponding \( A_{ii} \) are equal

Example: \( A = \alpha I \) is a scalar multiple of the identity matrix
- one eigenvalue \( \alpha \) with multiplicity \( n \)
- every nonzero vector is an eigenvector
Similarity transformation

two matrices $A$ and $B$ are similar if

$$B = T^{-1}AT$$

for some nonsingular matrix $T$

- the mapping that maps $A$ to $T^{-1}AT$ is called a similarity transformation
- similarity transformations preserve eigenvalues:

$$\det(\lambda I - B) = \det(\lambda I - T^{-1}AT) = \det(T^{-1}(\lambda I - AT)) = \det(\lambda I - A)$$

- if $x$ is an eigenvector of $A$ then $y = T^{-1}x$ is an eigenvector of $B$:

$$By = (T^{-1}AT)(T^{-1}x) = T^{-1}Ax = T^{-1}(\lambda x) = \lambda y$$

of special interest will be orthogonal similarity transformations ($T$ is orthogonal)
Diagonalizable matrices

A matrix is *diagonalizable* if it is similar to a diagonal matrix:

\[ T^{-1}AT = \Lambda \]

for some nonsingular matrix \( T \)

- the diagonal elements of \( \Lambda \) are the eigenvalues of \( A \)
- the columns of \( T \) are eigenvectors of \( A \):

\[ A(Te_i) = T\Lambda e_i = \Lambda_{ii}(Te_i) \]

- the columns of \( T \) give a set of \( n \) linearly independent eigenvectors

not all square matrices are diagonalizable
Spectral decomposition

suppose $A$ is diagonalizable, with

$$A = T \Lambda T^{-1} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_n^T \end{bmatrix}$$

$$= \lambda_1 v_1 w_1^T + \lambda_2 v_2 w_2^T + \cdots + \lambda_n v_n w_n^T$$

this is a spectral decomposition of the linear function $f(x) = Ax$

- elements of $T^{-1}x$ are coefficients of $x$ in the basis of eigenvectors $\{v_1, \ldots, v_n\}$:

$$x = TT^{-1}x = \alpha_1 v_1 + \cdots + \alpha_n v_n \quad \text{where} \quad \alpha_i = w_i^T x$$

- applied to an eigenvector, $f(v_i) = Av_i = \lambda_i v_i$ is a simple scaling
- by superposition, we find $Ax$ as

$$Ax = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n = T \Lambda T^{-1} x$$
Exercise

recall from 133A the definition of a *circulant matrix*

\[
A = \begin{bmatrix}
  a_1 & a_n & a_{n-1} & \cdots & a_3 & a_2 \\
  a_2 & a_1 & a_n & \cdots & a_4 & a_3 \\
  a_3 & a_2 & a_1 & \cdots & a_5 & a_4 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_1 & a_n \\
  a_n & a_{n-1} & a_{n-2} & \cdots & a_2 & a_1
\end{bmatrix}
\]

and its factorization

\[
A = \frac{1}{n} W \text{diag}(Wa)W^H
\]

\(W\) is the discrete Fourier transform matrix (\(Wa\) is the DFT of \(a\)) and

\[
W^{-1} = \frac{1}{n} W^H
\]

what is the spectrum of \(A\)?
Outline

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- low rank matrix approximation
Symmetric eigendecomposition

eigenvalues/vectors of a symmetric matrix have important special properties

- all the eigenvalues are real
- the eigenvectors corresponding to different eigenvalues are orthogonal
- a symmetric matrix is diagonalizable by an orthogonal similarity transformation:

\[ Q^T A Q = \Lambda, \quad Q^T Q = I \]

in the remainder of the lecture we assume that \( A \) is symmetric (and real)
Eigenvalues of a symmetric matrix are real

consider an eigenvalue \( \lambda \) and eigenvector \( x \) (possibly complex):

\[
Ax = \lambda x, \quad x \neq 0
\]

- inner product with \( x \) shows that \( x^H Ax = \lambda x^H x \)
- \( x^H x = \sum_{i=1}^{n} |x_i|^2 \) is real and positive, and \( x^H Ax \) is real:

\[
x^H Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \bar{x}_i x_j = \sum_{i=1}^{n} A_{ii} |x_i|^2 + 2 \sum_{j<i} A_{ij} \text{Re}(\bar{x}_i x_j)
\]

- therefore \( \lambda = (x^H Ax)/(x^H x) \) is real
- if \( x \) is complex, its real and imaginary part are real eigenvectors (if nonzero):

\[
A(x_{\text{re}} + jx_{\text{im}}) = \lambda(x_{\text{re}} + jx_{\text{im}}) \quad \implies \quad Ax_{\text{re}} = \lambda x_{\text{re}}, \quad Ax_{\text{im}} = \lambda x_{\text{im}}
\]

therefore, eigenvectors can be assumed to be real
Orthogonality of eigenvectors

suppose $x$ and $y$ are eigenvectors for different eigenvalues $\lambda$, $\mu$:

$$Ax = \lambda x, \quad Ay = \mu y, \quad \lambda \neq \mu$$

- take inner products with $x$, $y$:

$$\lambda y^T x = y^T Ax = x^T Ay = \mu x^T y$$

second equality holds because $A$ is symmetric

- if $\lambda \neq \mu$ this implies that

$$x^T y = 0$$
Eigendecomposition

every real symmetric $n \times n$ matrix $A$ can be factored as

$$A = Q\Lambda Q^T$$

(1)

- $Q$ is orthogonal
- $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal, with real diagonal elements
- $A$ is diagonalizable by an orthogonal similarity transformation: $Q^T AQ = \Lambda$
- the columns of $Q$ are an orthonormal set of $n$ eigenvectors: write $AQ = Q\Lambda$ as

$$A \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 q_1 & \lambda_2 q_2 & \cdots & \lambda_n q_n \end{bmatrix}$$
Proof by induction

- the decomposition (1) obviously exists if \( n = 1 \)
- suppose it exists if \( n = m \) and \( A \) is an \((m + 1) \times (m + 1)\) matrix
- \( A \) has at least one eigenvalue (page 3.2)
- let \( \lambda_1 \) be any eigenvalue and \( q_1 \) a corresponding eigenvector, with \(|q_1| = 1\)
- let \( V \) be an \((m + 1) \times m\) matrix that makes the matrix \([q_1 \ V]\) orthogonal:

\[
\begin{bmatrix}
q_1^T \\
V_T
\end{bmatrix}
A
\begin{bmatrix}
q_1 \\
V
\end{bmatrix}
= 
\begin{bmatrix}
q_1^TAq_1 & q_1^TA\ V \\
V^TAq_1 & V^TA\ V
\end{bmatrix}
= 
\begin{bmatrix}
\lambda_1 q_1^Tq_1 & \lambda_1 q_1^T\ V \\
\lambda_1 V^Tq_1 & V^TA\ V
\end{bmatrix}
= 
\begin{bmatrix}
\lambda_1 & 0 \\
0 & V^TA\ V
\end{bmatrix}
\]

- \( V^TA\ V \) is a symmetric \( m \times m \) matrix, so by the induction hypothesis,

\[
V^TA\ V = \tilde{Q}\tilde{\Lambda}\tilde{Q}^T
\]

for some orthogonal \( \tilde{Q} \) and diagonal \( \tilde{\Lambda} \)

- matrix \( Q = [q_1 \ V\tilde{Q}] \) is orthogonal and defines a similarity that diagonalizes \( A \):

\[
Q^TAQ = 
\begin{bmatrix}
q_1^T \\
\tilde{Q}^TV_T
\end{bmatrix}
A
\begin{bmatrix}
q_1 \\
V\tilde{Q}
\end{bmatrix}
= 
\begin{bmatrix}
\lambda_1 & 0 \\
0 & \tilde{Q}^TV^TA\ V\tilde{Q}
\end{bmatrix}
= 
\begin{bmatrix}
\lambda_1 & 0 \\
0 & \tilde{\Lambda}
\end{bmatrix}
\]

Symmetric eigendecomposition 3.12
Spectral decomposition

the decomposition (1) expresses $A$ as a sum of rank-one matrices:

$$A = Q\Lambda Q^T = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- the matrix–vector product $Ax$ is decomposed as

$$Ax = \sum_{i=1}^{n} \lambda_i q_i (q_i^T x)$$

- $(q_1^T x, \ldots, q_n^T x)$ are coordinates of $x$ in the orthonormal basis $\{q_1, \ldots, q_n\}$
- $(\lambda_1 q_1^T x, \ldots, \lambda_n q_n^T x)$ are coordinates of $Ax$ in the orthonormal basis $\{q_1, \ldots, q_n\}$
Non-uniqueness

some freedom exists in the choice of $\Lambda$ and $Q$ in the eigendecomposition

$$A = Q\Lambda Q^T = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

Ordering of eigenvalues

diagonal $\Lambda$ and columns of $Q$ can be permuted; we will assume that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Choice of eigenvectors

suppose $\lambda_i$ is an eigenvalue with multiplicity $k$: $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+k-1}$

- nonzero vectors in $\text{span}\{q_i, \ldots, q_{i+k-1}\}$ are eigenvectors with eigenvalue $\lambda_i$
- $q_i, \ldots, q_{i+k-1}$ can be replaced with any orthonormal basis of this “eigenspace”
a symmetric matrix is invertible if and only if all its eigenvalues are nonzero:

- inverse of $A = Q\Lambda Q^T$ is

$$A^{-1} = (Q\Lambda Q^T)^{-1} = Q\Lambda^{-1}Q^T,$$

$$\Lambda^{-1} = \begin{bmatrix}
\frac{1}{\lambda_1} & 0 & \cdots & 0 \\
0 & \frac{1}{\lambda_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\lambda_n}
\end{bmatrix}$$

- eigenvectors of $A^{-1}$ are the eigenvectors of $A$

- eigenvalues of $A^{-1}$ are reciprocals of eigenvalues of $A$
Spectral matrix functions

Integer powers

\[ A^k = (Q\Lambda Q^T)^k = Q\Lambda^k Q^T, \quad \Lambda^k = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k) \]

- negative powers are defined if \( A \) is invertible (all eigenvalues are nonzero)
- \( A^k \) has the same eigenvectors as \( A \), eigenvalues \( \lambda_i^k \)

Square root

\[ A^{1/2} = Q\Lambda^{1/2} Q^T, \quad \Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}) \]

- defined if eigenvalues are nonnegative
- a symmetric matrix that satisfies \( A^{1/2} A^{1/2} = A \)

Other matrix functions: can be defined via power series, for example,

\[ \exp(A) = Q \exp(\Lambda) Q^T, \quad \exp(\Lambda) = \text{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}) \]
Range, nullspace, rank

eigendecomposition with nonzero eigenvalues placed first in $\Lambda$:

$$A = Q\Lambda Q^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1\Lambda_1 Q_1^T$$

diagonal entries of $\Lambda_1$ are the nonzero eigenvalues of $A$

- columns of $Q_1$ are an orthonormal basis for $\text{range}(A)$
- columns of $Q_2$ are an orthonormal basis for $\text{null}(A)$
- this is an example of a full-rank factorization (page 1.27): $A = BC$ with

$$B = Q_1, \quad C = \Lambda_1 Q_1^T$$

- rank of $A$ is the number of nonzero eigenvalues (with their multiplicities)
Pseudo-inverse

we use the same notation as on the previous page

\[
A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1 \Lambda_1 Q_1^T
\]

diagonal entries of \( \Lambda_1 \) are the nonzero eigenvalues of \( A \)

- pseudo-inverse follows from page 1.36 with \( B = Q_1 \) and \( C = \Lambda_1 Q_1^T \)
- the pseudo-inverse is \( A^\dagger = C^\dagger B^\dagger = (Q_1 \Lambda_1^{-1}) Q_1^T \):

\[
A^\dagger = Q_1 \Lambda_1^{-1} Q_1^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}
\]

- eigenvectors of \( A^\dagger \) are the eigenvectors of \( A \)
- nonzero eigenvalues of \( A^\dagger \) are reciprocals of nonzero eigenvalues of \( A \)
- range, nullspace, and rank of \( A^\dagger \) are the same as for \( A \)
Trace

the trace of an $n \times n$ matrix $B$ is the sum of its diagonal elements

$$\text{trace}(B) = \sum_{i=1}^{n} B_{ii}$$

- **transpose:** $\text{trace}(B^T) = \text{trace}(B)$
- **product:** if $B$ is $n \times m$ and $C$ is $m \times n$, then

$$\text{trace}(BC) = \text{trace}(CB) = \sum_{i=1}^{n} \sum_{j=1}^{m} B_{ij}C_{ji}$$

- **eigenvalues:** the trace of a symmetric matrix is the sum of the eigenvalues

$$\text{trace}(Q\Lambda Q^T) = \text{trace}(Q^TQ\Lambda) = \text{trace}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$
Frobenius norm

recall the definition of *Frobenius norm* of an $m \times n$ matrix $B$:

$$
\|B\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} B_{ij}^2} = \sqrt{\text{trace}(B^T B)} = \sqrt{\text{trace}(BB^T)}
$$

• this is an example of a *unitarily invariant* norm: if $U, V$ are orthogonal, then

$$
\|UBV\|_F = \|B\|_F
$$

*Proof:*

$$
\|UBV\|_F^2 = \text{trace}(V^T B^T U^T UBV) = \text{trace}(VV^T B^T B) = \text{trace}(B^T B) = \|B\|_F^2
$$

• for a symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$,

$$
\|A\|_F = \|Q\Lambda Q^T\|_F = \|\Lambda\|_F = \left(\sum_{i=1}^{n} \lambda_i^2\right)^{1/2}
$$
2-Norm

recall the definition of 2-norm or spectral norm of an $m \times n$ matrix $B$:

$$\|B\|_2 = \max_{x \neq 0} \frac{\|Bx\|}{\|x\|}$$

• this norm is also unitarily invariant: if $U$, $V$ are orthogonal, then

$$\|UBV\|_2 = \|B\|_2$$

**Proof:**

$$\|UBV\|_2 = \max_{x \neq 0} \frac{\|UBVx\|}{\|x\|} = \max_{y \neq 0} \frac{\|UBy\|}{\|VTy\|} = \max_{y \neq 0} \frac{\|By\|}{\|y\|} = \|B\|_2$$

• for a symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$,

$$\|A\|_2 = \|Q\Lambda Q^T\|_2 = \|A\|_2 = \max_{i=1,\ldots,n} |\lambda_i| = \max\{\lambda_1, -\lambda_n\}$$

Symmetric eigendecomposition 3.21
Exercises

Exercise 1

suppose $A$ has eigendecomposition $A = Q\Lambda Q^T$; give an eigendecomposition of

$$A - \alpha I$$

Exercise 2

what are the eigenvalues and eigenvectors of an orthogonal projector

$$A = UU^T \quad \text{(where } U^T U = I)$$

Exercise 3

the condition number of a nonsingular matrix is defined as

$$\kappa(A) = \|A\|_2\|A^{-1}\|_2$$

express the condition number of a symmetric matrix in terms of its eigenvalues
Outline

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- low rank matrix approximation
Quadratic forms

the eigendecomposition is a useful tool for problems that involve quadratic forms

\[ f(x) = x^T Ax \]

• substitute \( A = Q \Lambda Q^T \) and make an orthogonal change of variables \( y = Q^T x \):

\[ f(Qy) = y^T \Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \]

• \( y_1, \ldots, y_n \) are coordinates of \( x \) in the orthonormal basis of eigenvectors

• the orthogonal change of variables preserves inner products and norms:

\[ \|y\|_2 = \|Q^T x\|_2 = \|x\|_2 \]
**Maximum and minimum value**

consider the optimization problems with variable $x$

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad x^T x = 1
\end{align*}
\]

\[
\begin{align*}
\text{minimize} & \quad x^T A x \\
\text{subject to} & \quad x^T x = 1
\end{align*}
\]

change coordinates to the spectral basis ($y = Q^T x$ and $x = Q y$):

\[
\begin{align*}
\text{maximize} & \quad \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\
\text{subject to} & \quad y_1^2 + \cdots + y_n^2 = 1
\end{align*}
\]

\[
\begin{align*}
\text{minimize} & \quad \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\
\text{subject to} & \quad y_1^2 + \cdots + y_n^2 = 1
\end{align*}
\]

- maximization: $y = (1, 0, \ldots, 0)$ and $x = q_1$ are optimal; maximal value is

\[
\max_{\|y\|=1} \max_{\|x\|=1} x^T A x = \max_{\|y\|=1} \left( \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \right) = \lambda_1 = \max_{i=1,\ldots,n} \lambda_i
\]

- minimization: $y = (0, 0, \ldots, 1)$ and $x = q_n$ are optimal; minimal value is

\[
\min_{\|y\|=1} \min_{\|x\|=1} x^T A x = \min_{\|y\|=1} \left( \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \right) = \lambda_n = \min_{i=1,\ldots,n} \lambda_i
\]
Exercises

**Exercise 1:** find the extreme values of the *Rayleigh quotient* \((x^T Ax)/(x^T x)\), i.e.,

\[
\max_{x \neq 0} \frac{x^T Ax}{x^T x}, \quad \min_{x \neq 0} \frac{x^T Ax}{x^T x}
\]

**Exercise 2:** solve the optimization problems

- maximize \(x^T Ax\) subject to \(x^T x \leq 1\)
- minimize \(x^T Ax\) subject to \(x^T x \leq 1\)

**Exercise 3:** show that (for symmetric \(A\))

\[
\|A\|_2 = \max_{i=1,\ldots,n} |\lambda_i| = \max_{\|x\|=1} |x^T Ax|
\]
# Sign of eigenvalues

<table>
<thead>
<tr>
<th>matrix property</th>
<th>condition on eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive definite</td>
<td>$\lambda_n &gt; 0$</td>
</tr>
<tr>
<td>positive semidefinite</td>
<td>$\lambda_n \geq 0$</td>
</tr>
<tr>
<td>indefinite</td>
<td>$\lambda_n &lt; 0$ and $\lambda_1 &gt; 0$</td>
</tr>
<tr>
<td>negative semidefinite</td>
<td>$\lambda_1 \leq 0$</td>
</tr>
<tr>
<td>negative definite</td>
<td>$\lambda_1 &lt; 0$</td>
</tr>
</tbody>
</table>

- $\lambda_1$ and $\lambda_n$ denote the largest and smallest eigenvalues:

\[
\lambda_1 = \max_{i=1,...,n} \lambda_i, \quad \lambda_n = \min_{i=1,...,n} \lambda_i
\]

- properties in the table follow from

\[
\lambda_1 = \max_{\|x\|=1} x^T Ax = \max_{x \neq 0} \frac{x^T Ax}{x^T x}, \quad \lambda_n = \min_{\|x\|=1} x^T Ax = \min_{x \neq 0} \frac{x^T Ax}{x^T x}
\]
Ellipsoids

if $A$ is positive definite, the set

$$\mathcal{E} = \{ x \mid x^T A x \leq 1 \}$$

is an ellipsoid with center at the origin

$$\frac{1}{\sqrt{\lambda_1}} q_1$$

$$\frac{1}{\sqrt{\lambda_n}} q_n$$

after the orthogonal change of coordinates $y = Q^T x$ the set is described by

$$\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \leq 1$$

this shows that:

- eigenvectors of $A$ give the principal axes
- the width along the principal axis determined by $q_i$ is $2/\sqrt{\lambda_i}$
Exercise

give an interpretation of \( \text{trace}(A^{-1}) \) as a measure of the size of the ellipsoid

\[
\mathcal{E} = \{ x \mid x^T A x \leq 1 \}
\]
Max–min characterization of eigenvalues

as an extension of the maximization problem on page 3.24, consider

\[
\begin{align*}
\text{maximize} & \quad \lambda_{\text{min}}(X^TAX) \\
\text{subject to} & \quad X^TX = I \\
\end{align*}
\]

(2)

the variable \(X\) is an \(n \times k\) matrix, for some given value of \(k\) between 1 and \(n\)

- \(\lambda_{\text{min}}(X^TAX)\) denotes the smallest eigenvalue of the \(k \times k\) matrix \(X^TAX\)
- for \(k = 1\) this is the problem on page 3.24: \(\lambda_{\text{min}}(x^TxAx) = x^TxAx\)

**Solution:** from the eigendecomposition \(A = QQ^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T\)

- the optimal value of (2) is the \(k\)th eigenvalue \(\lambda_k\) of \(A\)
- an optimal choice for \(X\) is formed from the first \(k\) columns of \(Q\):

\[
X = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix}
\]

this is known as the *Courant–Fischer min–max theorem*
Proof of the max–min characterization

we make a change of variables \( Y = Q^T X \):

maximize \( \lambda_{\min}(Y^T \Lambda Y) \)
subject to \( Y^T Y = I \)

we also partition \( \Lambda \) as

\[
\Lambda = \begin{bmatrix}
\Lambda_1 & 0 \\
0 & \Lambda_2
\end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix}
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_k
\end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix}
\lambda_{k+1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{bmatrix}
\]

we show that the matrix \( \hat{Y} = \begin{bmatrix} I \\ 0 \end{bmatrix} \) is optimal

• for this matrix

\[
\hat{Y}^T \Lambda \hat{Y} = \begin{bmatrix} I \\ 0 \end{bmatrix}^T \begin{bmatrix}
\Lambda_1 & 0 \\
0 & \Lambda_2
\end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix} = \Lambda_1, \quad \lambda_{\min}(\hat{Y}^T \Lambda \hat{Y}) = \lambda_{\min}(\Lambda_1) = \lambda_k
\]

• on the next page we show that \( \lambda_{\min}(Y^T \Lambda Y) \leq \lambda_k \) if \( Y \) is \( n \times k \) with \( Y^T Y = I \)
Proof of the max–min characterization

• on page 3.24, we have seen that

\[ \lambda_{\text{min}}(Y^T \Lambda Y) = \min_{\|u\|=1} u^T (Y^T \Lambda Y) u \]

• if \( Y \) has \( k \) columns, there exists \( v \neq 0 \) such that \( Yv \) has \( k - 1 \) leading zeros:

\[
Yv = \begin{bmatrix}
Y_{11} & \cdots & Y_{1k} \\
\vdots & \ddots & \vdots \\
Y_{k-1,1} & \cdots & Y_{k-1,k} \\
Y_{k1} & \cdots & Y_{kk} \\
\vdots & \ddots & \vdots \\
Y_{n1} & \cdots & Y_{nk}
\end{bmatrix}
\begin{bmatrix}
v_1 \\
v_k
\end{bmatrix}
= \begin{bmatrix}
0 \\
\vdots \\
y_k \\
y_n
\end{bmatrix}
\]

• if \( Y^T Y = I \) and we normalize \( v \), then \( \|Yv\| = \|v\| = 1 \) and

\[
(Yv)^T \Lambda (Yv) = \lambda_k y_k^2 + \cdots + \lambda_n y_n^2 \leq \lambda_k( y_k^2 + \cdots + y_n^2) = \lambda_k
\]

• this shows that

\[ \lambda_{\text{min}}(Y^T \Lambda Y) = \min_{\|u\|=1} u^T (Y^T \Lambda Y) u \leq v^T (Y^T \Lambda Y)v \leq \lambda_k \]
Min–max characterization of eigenvalues

the minimization problem on page 3.24 can be extended in a similar way:

\[
\begin{align*}
\text{minimize} & \quad \lambda_{\text{max}}(X^TAX) \\
\text{subject to} & \quad X^TX = I
\end{align*}
\]

(3)

the variable \(X\) is an \(n \times k\) matrix

- \(\lambda_{\text{max}}(X^TAX)\) denotes the largest eigenvalue of the \(k \times k\) matrix \(X^TAX\)
- for \(k = 1\) this is the minimization problem on page 3.24: \(\lambda_{\text{max}}(x^TAx) = x^TAx\)

Solution: from the eigenvalue decomposition \(A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T\)

- the optimal value of (3) is eigenvalue \(\lambda_{n-k+1}\) of \(A\)
- an optimal choice of \(X\) is formed from the last \(k\) columns of \(Q\):

\[
X = \begin{bmatrix}
q_{n-k+1} & \cdots & q_{n-1} & q_n
\end{bmatrix}
\]

this follows from the max–min characterization on page 3.29 applied to \(-A\)
**Exercises**

**Exercise 1:** suppose $B$ is an $m \times m$ principal submatrix of $A$, for example,

$$B = \begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1m} \\
A_{21} & A_{22} & \cdots & A_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
A_{m1} & A_{m2} & \cdots & A_{mm}
\end{bmatrix}, \quad (4)$$

and denote the $m$ eigenvalues of $B$ by $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$

show that

$$\mu_1 \leq \lambda_1, \quad \mu_2 \leq \lambda_2, \quad \ldots, \quad \mu_m \leq \lambda_m$$

($\lambda_1, \ldots, \lambda_m$ are the first $m$ eigenvalues of $A$)

**Exercise 2:** consider the matrix $B$ in (4) with $m = n - 1$; show that

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n$$

this is known as the eigenvalue **interlacing theorem**
Eigendecomposition of covariance matrix

- Suppose $x$ is a random $n$-vector with mean $\mu$, covariance matrix $\Sigma$.
- $\Sigma$ is positive semidefinite with eigendecomposition

$$
\Sigma = E((x - \mu)(x - \mu)^T) = Q\Lambda Q^T
$$

define a random $n$-vector $y = Q^T(x - \mu)$

- $y$ has zero mean and covariance matrix $\Lambda$:

$$
E(yy^T) = Q^T E((x - \mu)(x - \mu)^T)Q = Q^T \Sigma Q = \Lambda
$$

- Components of $y$ are uncorrelated and have variances $E(y_i^2) = \lambda_i$
- $x$ is decomposed in uncorrelated components with decreasing variance:

$$
E(y_1^2) \geq E(y_2^2) \geq \cdots \geq E(y_n^2)
$$

the transformation is known as the *Karhunen–Loève* or *Hotelling* transform
**Multivariate normal distribution**

multivariate normal (Gaussian) probability density function

\[
p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}
\]

contour lines of density function for

\[
\Sigma = \frac{1}{4} \begin{bmatrix} 7 & \sqrt{3} \\ \sqrt{3} & 5 \end{bmatrix}, \quad \mu = \begin{bmatrix} 5 \\ 4 \end{bmatrix}
\]

eigenvalues of \( \Sigma \) are \( \lambda_1 = 2, \lambda_2 = 1 \),

\[
q_1 = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 1/2 \\ -\sqrt{3}/2 \end{bmatrix}
\]
Multivariate normal distribution

the decorrelated and de-meaned variables $y = Q^T(x - \mu)$ have distribution

$$\tilde{p}(y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\lambda_i}} \exp(-\frac{y_i^2}{2\lambda_i})$$
Joint diagonalization of two matrices

- A symmetric matrix $A$ is diagonalized by an orthogonal similarity:

$$Q^T AQ = \Lambda$$

- As an extension, if $A, B$ are symmetric and $B$ is positive definite, then

$$S^T AS = D, \quad S^T BS = I$$

for some nonsingular $S$ and diagonal $D$

**Algorithm:** $S$ and $D$ can be computed as follows

- Cholesky factorization $B = R^T R$, with $R$ upper triangular and nonsingular
- Eigendecomposition $R^{-T} AR^{-1} = QDQ^T$, with $D$ diagonal, $Q$ orthogonal
- Define $S = R^{-1} Q$:

$$S^T AS = Q^T R^{-T} AR^{-1} Q = \Lambda, \quad S^T BS = Q^T R^{-T} BR^{-1} Q = Q^T Q = I$$
Optimization problems with two quadratic forms

as an extension of the maximization problem on page 3.24, consider

\[
\begin{align*}
\text{maximize} & \quad x^T Ax \\
\text{subject to} & \quad x^T B x = 1
\end{align*}
\]

where \( A, B \) are symmetric and \( B \) is positive definite

• compute nonsingular \( S \) that diagonalizes \( A, B \):

\[
S^T AS = D, \quad S^T BS = I
\]

• make change of variables \( x = Sy \):

\[
\begin{align*}
\text{maximize} & \quad y^T D y \\
\text{subject to} & \quad y^T y = 1
\end{align*}
\]

• if diagonal elements of \( D \) are sorted as \( D_{11} \geq \cdots \geq D_{nn} \), solution is

\[
y = e_1 = (1, 0, \ldots, 0), \quad x = Se_1, \quad x^T Ax = D_{11}
\]
Outline

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- low rank matrix approximation
Low-rank matrix approximation

- low rank is a useful matrix property in many applications
- low rank is not a robust property (easily destroyed by noise or estimation error)
- most matrices in practice have full rank
- often the full-rank matrix is close to being low rank
- computing low-rank approximations is an important problem in linear algebra

on the next pages we discuss this for positive semidefinite matrices
Rank-$r$ approximation of positive semidefinite matrix

Let $A$ be a positive semidefinite matrix with $\text{rank}(A) > r$ and eigendecomposition

$$ A = Q \Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T, \quad \lambda_1 \geq \cdots \geq \lambda_n \geq 0, \quad \lambda_{r+1} > 0 $$

the best rank-$r$ approximation is the sum of the first $r$ terms in the decomposition:

$$ B = \sum_{i=1}^{r} \lambda_i q_i q_i^T $$

- $B$ is the best approximation for the Frobenius norm: for every $C$ with rank $r$,

$$ \| A - C \|_F \geq \| A - B \|_F = \left( \sum_{i=r+1}^{n} \lambda_i^2 \right)^{1/2} $$

- $B$ is also the best approximation for the 2-norm: for every $C$ with rank $r$,

$$ \| A - C \|_2 \geq \| A - B \|_2 = \lambda_{r+1} $$
Rank-$r$ approximation in Frobenius norm

the approximation problem in Frobenius norm is a nonlinear least squares problem

$$\text{minimize} \quad \|A - XX^T\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( A_{ij} - \sum_{k=1}^{r} X_{ik}X_{jk} \right)^2$$

- we parametrize $B$ as $B = XX^T$ with $X$ of size $n \times r$, and optimize over $X$
- this can be written in the standard nonlinear least squares form

$$\text{minimize} \quad g(x) = \|f(x)\|^2$$

with vector $x$ containing the elements of $X$ and $f(x)$ the elements of $A - XX^T$
- the first order (necessary but not sufficient) optimality conditions are

$$\nabla g(x) = 2Df(x)^Tf(x) = 0$$

- the first order optimality conditions will be derived on page 3.41; they are

$$4(A - XX^T)X = 0$$

Symmetric eigendecomposition 3.41
Solution of first order optimality conditions

\[ AX = X(X^T X) \]

- define eigendecomposition \( X^T X = U D U^T \) (\( U \) orthogonal \( r \times r \), \( D \) diagonal)
- use \( Y = XU \) and \( D \) as variables:

\[ AY = YD, \quad Y^T Y = D \]

- the \( r \) diagonal elements of \( D \) must be eigenvalues of \( A \)
- the \( r \) columns of \( Y \) are corresponding orthogonal eigenvectors
- the columns of \( Y \) are normalized to have norm \( \sqrt{D_{ii}} \)

we conclude that the solutions of the first order optimality conditions satisfy

\[ XX^T = YY^T = \sum_{i \in I} \lambda_i q_i q_i^T \]

where \( I \) is a subset of \( r \) elements of \( \{1, 2, \ldots, n\} \)
Optimal solution

among the solutions of the 1st order conditions we choose the one that minimizes

$$\|A - XX^T\|_F$$

- the squared error in the approximation is

$$\|A - XX^T\|_F^2 = \|A - \sum_{i \in I} \lambda_i q_i q_i^T\|_F^2$$

$$= \| \sum_{i \notin I} \lambda_i q_i q_i^T\|_F^2$$

$$= \sum_{i \notin I} \lambda_i^2$$

- the optimal choice for $I$ is $I = \{1, 2, \ldots, r\}$:

$$XX^T = \sum_{i=1}^{r} \lambda_i q_i q_i^T, \quad \|A - XX^T\|_F^2 = \sum_{i=r+1}^{n} \lambda_i^2$$
First order optimality

to derive the first order optimality conditions for

$$\text{minimize} \quad \|A - XX^T\|_F^2$$

we substitute $X + \delta X$, with arbitrary small $\delta X$, and linearize:

$$\|A - (X + \delta X)(X + \delta X)^T\|_F^2$$

$$= \|A - XX^T + \delta X X^T + X \delta X^T + \delta X \delta X^T\|_F^2$$

$$\approx \|A - XX^T + \delta X X^T + X \delta X^T\|_F^2$$

$$= \text{trace} \left( (A - XX^T + \delta X X^T + X \delta X^T)(A - XX^T + \delta X X^T + X \delta X^T) \right)$$

$$\approx \text{trace} \left( (A - XX^T)(A - XX^T) \right) + 2 \text{trace} \left( (\delta X X^T + X \delta X^T)(A - XX^T) \right)$$

$$= \|A - XX^T\|_F^2 + 4 \text{trace} \left( \delta X^T (A - XX^T)X \right)$$

$X$ is a stationary point if the second term is zero for all $\delta X$:

$$4(A - XX^T)X = 0$$
Rank-$r$ approximation in 2-norm

the same matrix $B$ is also the best approximation in 2-norm: if $C$ has rank $r$, then

$$\|A - C\|_2 \geq \|A - B\|_2$$

the right-hand side is

$$\|A - B\|_2 = \left\| \sum_{i=1}^{n} \lambda_i q_i q_i^T - \sum_{i=1}^{r} \lambda_i q_i q_i^T \right\|_2$$

$$= \left\| \sum_{i=r+1}^{n} \lambda_i q_i q_i^T \right\|_2$$

$$= \lambda_{r+1}$$

on the next page we show that $\|A - C\|_2 \geq \lambda_{r+1}$ if $C$ has rank $r$
Proof

- if rank($C$) = $r$, the nullspace of $C$ has dimension $n - r$
- define an $n \times (n - r)$ matrix $V$ with orthonormal columns that span null($C$)
- we use the min–max theorem on page 3.32 to bound $\|A - C\|_2$:

\[
\|A - C\|_2 = \max_{\|x\|=1} |x^T(A - C)x| \quad \text{(page 3.25)}
\geq \max_{\|x\|=1} x^T(A - C)x
\geq \max_{\|y\|=1} y^T V^T (A - C)V y \quad (\|V y\| = \|y\|)
= \max_{\|y\|=1} y^T V^T A V y \quad (V^T CV = 0)
= \lambda_{\max}(V^T A V)
\geq \lambda_{r+1} \quad \text{(page 3.32 with } k = n - r)