3. Symmetric eigendecomposition

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

Eigenvalues and eigenvectors

a nonzero vector x is an *eigenvector* of the $n \times n$ matrix A, with *eigenvalue* λ, 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

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:

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

- 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 are *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 Λ 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 $\{v_1, \ldots, v_n\}$:

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

- 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

$$A = \frac{1}{n} W^H \, \mathbf{diag}(Wa) W$$

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
- optimality theorems
- 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 symmetrix matrix is diagonalizable by an orthogonal similarity transformation:

$$Q^T A Q = \Lambda, \qquad Q^T Q = I$$

in the remainder of the lecture we assume that A is symmetric (and real)

Eigenvalues of a symmetric matrix are real

consider an eigenvalue λ and eigenvector x (possibly complex):

$$Ax = \lambda x, \quad x \neq 0$$

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

$$x^H A x = \sum_{i=1}^n \sum_{j=1}^n A_{ij} \bar{x}_i x_j = \sum_{i=1}^n A_{ii} |x_i|^2 + 2 \sum_{j<i} A_{ij} \operatorname{Re}(\bar{x}_i x_j)$$

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

$$A(x_{\rm re} + {\rm j} x_{\rm im}) = \lambda(x_{\rm re} + {\rm j} x_{\rm im}) \qquad \Longrightarrow \qquad A x_{\rm re} = \lambda x_{\rm re}, \quad A x_{\rm im} = \lambda x_{\rm im}$$

therefore, eigenvectors can be assumed to be real

Orthogonality of eigenvectors

suppose x and y are eigenvectors for different eigenvalues λ, μ:

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

- take inner products with x, y:

$$\lambda y^T x = y^T A x = x^T A y = \mu x^T y$$

  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

$$A = Q \Lambda Q^T \tag{1}$$

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

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

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 λ_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 $\begin{bmatrix} q_1 & V \end{bmatrix}$ orthogonal:

$$\begin{bmatrix} q_1^T \\ V^T \end{bmatrix} A \begin{bmatrix} q_1 & V \end{bmatrix} = \begin{bmatrix} q_1^T A q_1 & q_1^T A V \\ V^T A q_1 & V^T A V \end{bmatrix} = \begin{bmatrix} \lambda_1 q_1^T q_1 & \lambda_1 q_1^T V \\ \lambda_1 V^T q_1 & V^T A V \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & V^T A V \end{bmatrix}$$

- $V^T A V$ is a symmetric $m \times m$ matrix, so by the induction hypothesis,

$$V^T A V = \tilde{Q} \tilde{\Lambda} \tilde{Q}^T \quad \text{for some orthogonal } \tilde{Q} \text{ and diagonal } \tilde{\Lambda}$$

- the orthogonal matrix $Q = \begin{bmatrix} q_1 & V\tilde{Q} \end{bmatrix}$ defines a similarity that diagonalizes A:

$$Q^T A Q = \begin{bmatrix} q_1^T \\ \tilde{Q}^T V^T \end{bmatrix} A \begin{bmatrix} q_1 & V\tilde{Q} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \tilde{Q}^T V^T A V \tilde{Q} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix}$$

Spectral decomposition

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

$$\begin{aligned} A = Q\Lambda Q^T &= \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} \\ &= \sum_{i=1}^{n} \lambda_i q_i q_i^T \end{aligned}$$

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

- all nonzero vectors in $\text{span}\{q_i, \ldots, q_{i+k-1}\}$ are eigenvectors with eigenvalue λ_i
- $q_i, \ldots, q_{i+k-1}$ can be replaced with any orthonormal basis of this "eigenspace"

Inverse

a symmetric matrix is invertible if and only if all its eigenvalues are nonzero:

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

$$A^{-1} = (Q\Lambda Q^T)^{-1} = Q\Lambda^{-1}Q^T, \qquad \Lambda^{-1} = \begin{bmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_n \end{bmatrix}$$

- eigenvectors of A^{-1} are the eigenvectors of A
- eigenvalues of A^{-1} are reciprocals of eigenvalues of A

Spectral matrix functions

Integer powers

$$A^k = (Q\Lambda Q^T)^k = Q\Lambda^k Q^T, \qquad \Lambda^k = \mathbf{diag}(\lambda_1^k, \ldots, \lambda_n^k)$$

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

Square root

$$A^{1/2} = Q\Lambda^{1/2}Q^T, \qquad \Lambda^{1/2} = \mathbf{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$$

- 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,

$$\exp(A) = Q \exp(\Lambda) Q^T, \qquad \exp(\Lambda) = \mathbf{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})$$

Range, nullspace, rank

eigendecomposition with nonzero eigenvalues placed first in Λ:

$$A = Q\Lambda Q^T = \left[\begin{array}{cc} Q_1 & Q_2 \end{array}\right] \left[\begin{array}{cc} \Lambda_1 & 0 \\ 0 & 0 \end{array}\right] \left[\begin{array}{c} Q_1^T \\ Q_2^T \end{array}\right] = Q_1 \Lambda_1 Q_1^T$$

diagonal entries of Λ_1 are the nonzero eigenvalues of A

- columns of Q_1 are an orthonormal basis for $\mathrm{range}(A)$
- columns of Q_2 are an orthonormal basis for $\mathrm{null}(A)$
- this is an example of a full-rank factorization (page 1.32): $A = BC$ with

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

- 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 Λ_1 are the nonzero eigenvalues of A

- pseudo-inverse follows from page 1.39 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

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

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

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

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

$$\mathrm{trace}(Q \Lambda Q^T) = \mathrm{trace}(Q^T Q \Lambda) = \mathrm{trace}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

Frobenius norm

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

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

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

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

Proof:

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

- 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:

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

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

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

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

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

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**
- optimality theorems
- low rank matrix approximation

Quadratic forms

the eigendecomposition is a useful tool for problems involving quadratic forms

$$f(x) = x^T A x$$

- 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
- in this basis, the quadratic form is *separable* (variables are decoupled)
- the orthogonal change of variables preserves inner products and norms:

$$\|y\| = \|Q^T x\| = \|x\|$$

Maximum and minimum value

consider the following optimization problems with variable x

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & x^T x = 1 \end{array} \qquad \begin{array}{ll} \text{minimize} & x^T A x \\ \text{subject to} & x^T x = 1 \end{array}$$

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

$$\begin{array}{ll} \text{maximize} & \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\ \text{subject to} & y_1^2 + \cdots + y_n^2 = 1 \end{array} \qquad \begin{array}{ll} \text{minimize} & \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\ \text{subject to} & y_1^2 + \cdots + y_n^2 = 1 \end{array}$$

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

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

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

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

Exercises

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

$$\max_{x \neq 0} \frac{x^T Ax}{x^T x}, \qquad \min_{x \neq 0} \frac{x^T Ax}{x^T x}$$

Exercise 2: solve the optimization problems

$$\begin{array}{ll} \text{maximize} & x^T Ax \\ \text{subject to} & x^T x \leq 1 \end{array} \qquad \begin{array}{ll} \text{minimize} & x^T Ax \\ \text{subject to} & x^T x \leq 1 \end{array}$$

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

matrix property	condition on eigenvalues
positive definite	$\lambda_n > 0$
positive semidefinite	$\lambda_n \geq 0$
indefinite	$\lambda_n < 0$ and $\lambda_1 > 0$
negative semidefinite	$\lambda_1 \leq 0$
negative definite	$\lambda_1 < 0$

- λ_1 and λ_n denote the largest and smallest eigenvalues:

$$\lambda_1 = \max_{i=1,\ldots,n} \lambda_i, \qquad \lambda_n = \min_{i=1,\ldots,n} \lambda_i$$

- properties in the table follow from

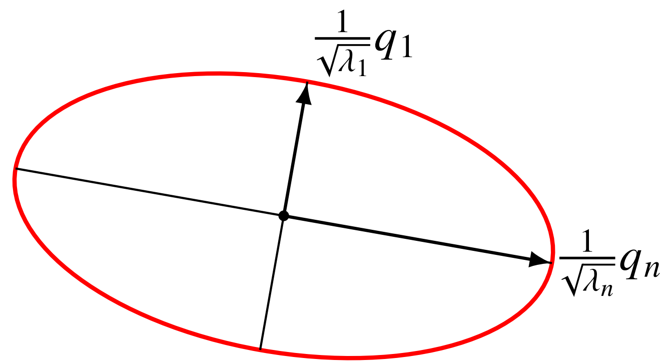
$$\lambda_1 = \max_{\|x\|=1} x^T A x = \max_{x \neq 0} \frac{x^T A x}{x^T x}, \qquad \lambda_n = \min_{\|x\|=1} x^T A x = \min_{x \neq 0} \frac{x^T A x}{x^T x}$$

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



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 $\mathrm{trace}(A^{-1})$ as a measure of the size of the ellipsoid

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

Eigendecomposition of covariance matrix

- suppose x is a random n-vector with mean μ, covariance matrix Σ
- Σ is positive semidefinite with eigendecomposition

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

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

- y has zero mean and covariance matrix Λ:

$$\mathbf{E}(yy^T) = Q^T\,\mathbf{E}((x-\mu)(x-\mu)^T)Q = Q^T\Sigma Q = \Lambda$$

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

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

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

Multivariate normal distribution

multivariate normal (Gaussian) probability density function

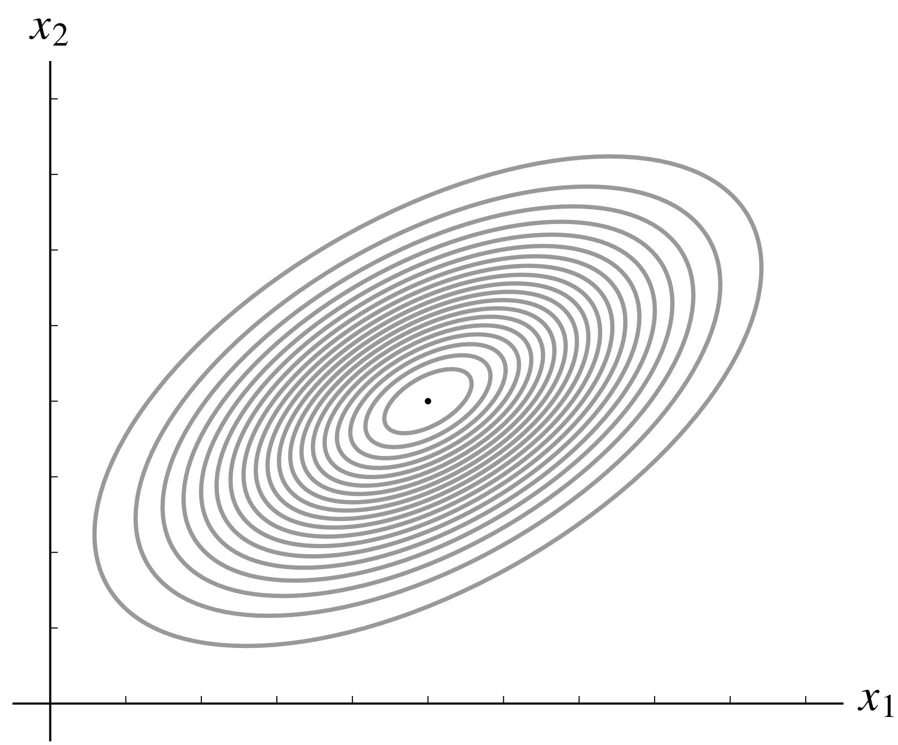
$$p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

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 Σ are $\lambda_1 = 2$, $\lambda_2 = 1$,

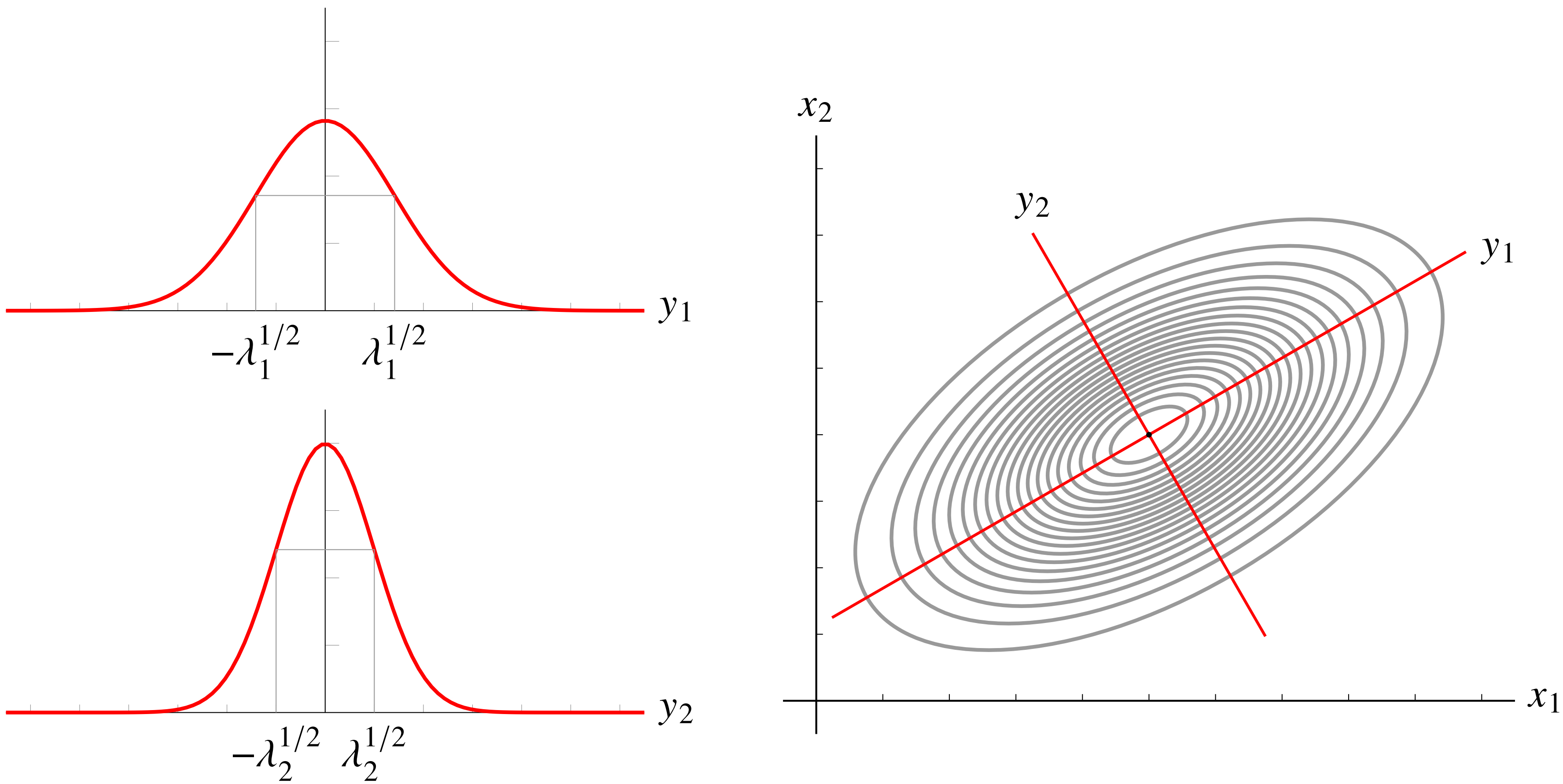
$$q_1 = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 1/2 \\ -\sqrt{3}/2 \end{bmatrix}$$



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:

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

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

$$S^T A S = D, \qquad S^T B S = I$$

  for some nonsingular S and diagonal D

Algorithm: S and D can be computed is as follows

- Cholesky factorization $B = R^T R$, with R upper triangular and nonsingular
- eigendecomposition $R^{-T} A R^{-1} = Q D Q^T$, with D diagonal, Q orthogonal
- define $S = R^{-1} Q$:

$$S^T A S = Q^T R^{-T} A R^{-1} Q = \Lambda, \qquad S^T B S = Q^T R^{-T} B R^{-1} Q = Q^T Q = I$$

Optimization problems with two quadratic forms

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

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & x^T B x = 1 \end{array}$$

where A, B are symmetric and B is positive definite

- compute nonsingular S that diagonalizes A, B:

$$S^T A S = D, \qquad S^T B S = I$$

- make change of variables $x = Sy$:

$$\begin{array}{ll} \text{maximize} & y^T D y \\ \text{subject to} & y^T y = 1 \end{array}$$

- if diagonal elements of D are sorted as $D_{11} \geq \cdots \geq D_{nn}$, solution is

$$y = e_1 = (1, 0, \ldots, 0), \qquad x = Se_1, \qquad x^T A x = D_{11}$$

Outline

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- **optimality theorems**
- low rank matrix approximation

Quadratic form restricted to subspace

we consider quadratic forms $x^T Ax$ with x restricted to a subspace $\mathcal{V}$

- as before, A is symmetric, $n \times n$, with eigendecomposition

$$A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T, \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

- $\mathcal{V}$ is a k-dimensional subspace of $\mathbf{R}^n$, represented by an orthonormal basis:

$$\mathcal{V} = \{Xy \mid y \in \mathbf{R}^k\}, \qquad X^T X = I, \qquad X \in \mathbf{R}^{n\times k}$$

- eigendecomposition of $X^T AX$ characterizes the quadratic form restricted to $\mathcal{V}$
- we denote the eigendecomposition of the $k \times k$ matrix $X^T AX$ by

$$X^T AX = \sum_{i=1}^{k} \mu_i w_i w_i^T, \qquad \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$$

we are interested in how the eigenvalues $\mu_1, \ldots, \mu_k$ vary with the subspace $\mathcal{V}$

Courant–Fischer theorem

$$\begin{bmatrix} \lambda_{n-k+1} \\ \lambda_{n-k+2} \\ \vdots \\ \lambda_n \end{bmatrix} \leq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} \leq \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} \tag{2}$$

- the two inequalities hold component-wise:

$$\lambda_{n-k+1} \leq \mu_1 \leq \lambda_1, \qquad \lambda_{n-k+2} \leq \mu_2 \leq \lambda_2, \qquad \ldots, \qquad \lambda_n \leq \mu_k \leq \lambda_k$$

- right-hand inequality in (2) is an equality for $X = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix}$

  ($\mathcal{V}$ is spanned by eigenvectors of A corresponding to the first k eigenvalues)
- left-hand inequality is an equality for $X = \begin{bmatrix} q_{n-k+1} & q_{n-k+2} & \cdots & q_n \end{bmatrix}$

  ($\mathcal{V}$ is spanned by eigenvectors of A corresponding to the last k eigenvalues)

this is (one form of) the *Courant–Fischer minimax theorem*

Proof of Courant–Fischer theorem

- we prove the right-hand inequality in (2): for $1 \leq j \leq k$,

$$\mu_j \leq \lambda_j$$

- left-hand inequality follows from right-hand inequality applied to $-A$

Proof

- if we define $W_j = \begin{bmatrix} w_1 & w_2 & \cdots & w_j \end{bmatrix}$ (first j eigenvectors of $X^T AX$), then

$$\begin{aligned}
\begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_j \end{bmatrix} &= W_j^T (X^T AX) W_j \\
&= W_j^T X^T (Q \Lambda Q^T) X W_j \\
&= W_j^T X^T (\sum_{i=1}^n \lambda_i q_i q_i^T) X W_j \\
&= \sum_{i=1}^n \lambda_i (W_j^T X^T q_i)(q_i^T X W_j)
\end{aligned}$$

Proof of Courant–Fischer theorem (continued)

- smallest eigenvalue μ_j of X^TAX can be expressed as

$$\mu_j = \min_{y_1^2+\cdots+y_j^2=1} (\mu_1 y_1^2 + \cdots + \mu_j y_j^2) = \min_{y_1^2+\cdots+y_j^2=1} \sum_{i=1}^{n} \lambda_i (q_i^T X W_j y)^2 \tag{3}$$

- by the dimension inequality (page 1.7) the $j-1$ linear equations

$$q_1^T X W_j y = q_2^T X W_j y = \cdots = q_{j-1}^T X W_j y = 0, \tag{4}$$

  with the j-vector y as variable, have nonzero solutions

- let $\hat{y}$ be a nonzero solution of (4), normalized to satisfy

$$1 = \|\hat{y}\|^2 = \|Q^T X W_j \hat{y}\|^2 = \sum_{i=1}^{n} (q_i^T X W_j \hat{y})^2 = \sum_{i=j}^{n} (q_i^T X W_j \hat{y})^2 \tag{5}$$

  (the second equality holds because $Q^T X W_j$ has orthonormal columns)

Proof of Courant–Fischer theorem (continued)

- since $\hat{y}_1^2 + \cdots + \hat{y}_j^2 = 1$, we have from (3)

$$
\begin{aligned}
\mu_j &= \min_{y_1^2+\cdots+y_j^2=1} \sum_{i=1}^{n} \lambda_i (q_i^T X W_j y)^2 \\
&\leq \sum_{i=1}^{n} \lambda_i (q_i^T X W_j \hat{y})^2 \\
&= \sum_{i=j}^{n} \lambda_i (q_i^T X W_j \hat{y})^2 && (\hat{y} \text{ is a solution of } (4)) \\
&\leq \lambda_j \sum_{i=k}^{n} (q_i^T X W_j \hat{y})^2 && (\lambda_j \geq \cdots \geq \lambda_n) \\
&= \lambda_j && (\text{from } (5))
\end{aligned}
$$

Rayleigh–Ritz theorem

the result on page 3.24 is a special case for $k = 1$:

$$\lambda_n \leq x^T A x \leq \lambda_1$$

for all x with $x^T x = 1$

- equality $x^T A x = \lambda_1$ holds for $x = q_1$
- equality $x^T A x = \lambda_n$ holds for $x = q_n$

this is known as the *Rayleigh–Ritz theorem*

Min–max and max–min characterization of eigenvalues

consider the optimization problems

$$\begin{array}{ll} \text{minimize} & \lambda_{\max}(X^TAX) \\ \text{subject to} & X^TX = I \end{array} \qquad \begin{array}{ll} \text{maximize} & \lambda_{\min}(X^TAX) \\ \text{subject to} & X^TX = I \end{array} \tag{6}$$

the variable X is an $n \times k$ matrix, with $1 \leq k \leq n$

- $\lambda_{\min}(X^TAX)$ and $\lambda_{\max}(X^TAX)$ are smallest and largest eigenvalue of X^TAX
- from page 3.35, an optimal solution of the maximization problem is

$$X = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix},$$

  the optimal value is $\lambda_{\min}(X^TAX) = \lambda_k$ (the kth largest eigenvalue of A)
- an optimal solution of the minimization problem is

$$X = \begin{bmatrix} q_{n-k+1} & q_{n-k+2} & \cdots & q_n \end{bmatrix},$$

  the optimal value is $\lambda_{\max}(X^TAX) = \lambda_{n-k+1}$ (the kth smallest eigenvalue of A)

Eigenvalue interlacing theorem

let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n-1}$ be the eigenvalues of the $(n-1) \times (n-1)$ submatrix

$$B = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1,n-1} \\ A_{21} & A_{22} & \cdots & A_{2,n-1} \\ \vdots & \vdots & & \vdots \\ A_{n-1,1} & A_{n-1,2} & \cdots & A_{n-1,n-1} \end{bmatrix}$$

- we have $B = X^T A X$ where

$$X = \begin{bmatrix} I_{n-1} \\ 0 \end{bmatrix}$$

- applying the result on page 3.35 with $k = n-1$ gives

$$\begin{bmatrix} \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \end{bmatrix} \leq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \leq \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n-1} \end{bmatrix}$$

this is known as the eigenvalue *interlacing theorem*

Exercises

give the solution of the following problems; the variable is an $n \times k$ matrix X

1.

$$\begin{array}{ll} \text{maximize} & \operatorname{trace}(X^T A X) \\ \text{subject to} & X^T X = I \end{array} \qquad \begin{array}{ll} \text{minimize} & \operatorname{trace}(X^T A X) \\ \text{subject to} & X^T X = I \end{array}$$

   recall that the trace is the sum of eigenvalues

2. assuming A is positive definite,

$$\begin{array}{ll} \text{maximize} & \det(X^T A X) \\ \text{subject to} & X^T X = I \end{array} \qquad \begin{array}{ll} \text{minimize} & \det(X^T A X) \\ \text{subject to} & X^T X = I \end{array}$$

   recall that the determinant is the product of the eigenvalues

Outline

- eigenvalues and eigenvectors
- symmetric eigendecomposition
- quadratic forms
- optimality theorems
- **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

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

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

we show that for every symmetric $n \times n$ matrix C of rank r,

$$\|A - C\|_F^2 \geq \sum_{i=1}^{r} \lambda_i^2$$

- let X be an $n \times (n - r)$ matrix with orthonormal columns that span $\mathrm{null}(C)$
- define $\tilde{X}$ as an $n \times r$ matrix that makes $\begin{bmatrix} X & \tilde{X} \end{bmatrix}$ orthogonal

$$\begin{aligned}
\|A - C\|_F^2 &= \left\| \begin{bmatrix} X^T(A-C)X & X^T(A-C)\tilde{X} \\ \tilde{X}^T(A-C)X & \tilde{X}^T(A-C)\tilde{X} \end{bmatrix} \right\|_F^2 && \text{(Frobenius norm is unitarily invariant)} \\
&\geq \|X^T(A-C)X\|_F^2 \\
&= \|X^T A X\|_F^2 && (X^T C X = 0) \\
&= \mu_1^2 + \mu_2^2 + \cdots + \mu_{n-r}^2 && \text{(if } \mu_1, \ldots, \mu_{n-r} \text{ are the eigenvalues of } X^T A X) \\
&\geq \lambda_{r+1}^2 + \lambda_{r+2}^2 + \cdots + \lambda_n^2 && \text{((2) with } k = n - r \text{ and nonnegativity of } \mu_i, \lambda_i)
\end{aligned}$$

Rank-r approximation in 2-norm

we show that for every symmetric $n \times n$ matrix C of rank r,

$$\|A - C\|_2 \geq \lambda_{r+1}$$

let X be an $n \times (n-r)$ matrix with orthonormal columns that span $\mathrm{null}(C)$

$$\begin{aligned}
\|A - C\|_2 &= \max_{\|x\|=1} |x^T(A-C)x| && \text{(exercise 3 on page 3.25)} \\
&\geq \max_{\|y\|=1} |y^T X(A-C)Xy| && (\|Xy\| = 1) \\
&= \|X^T(A-C)X\|_2 && \text{(exercise 3 on page 3.25)} \\
&= \|X^T A X\|_2 && (X^T C X = 0) \\
&= \mu_1 && \text{(2-norm of p.s.d. matrix } X^T A X \text{ is the largest eigenvalue } \mu_1) \\
&\geq \lambda_{r+1} && \text{((2) with } k = n - r)
\end{aligned}$$