Linear Algebra

Lecture Notes

Rostyslav Hryniv

Ukrainian Catholic University
Business Analytics and Computer Science Programmes

4^{th} term
Spring 2020

Lecture 11. Symmetric matrices and quadratic forms

Outline

What did we learn last time?

- Eigenvalue λ ($\in \mathbb{C}$) and corresponding eigenvector $\mathbf{x}$ ($\neq \mathbf{0}$) of an $n \times n$ matrix A:

$$A\mathbf{x} = \lambda \mathbf{x}$$

- Every such A has at most n eigenvalues
  (exactly n if counted with multiplicities)
- If A has n distinct EV's $\lambda_1, \ldots, \lambda_n$, then the corresponding EVc's $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent
- form P with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$; then

$$P^{-1}AP = D := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

- Thus: in its eigenbasis, A just multiplies by λ_j in the direction $\mathbf{v}_j$
  $\Longrightarrow$ functions of A easy to calculate; e.g.
  - A^n to solve the difference equations
  - e^{tA} to solve the linear differential equation
- However: not all A are diagonalizable!
  E.g. Jordan blocks are not: too few eigenvectors

2×2 examples

Definition

Spectrum of a square matrix A is the set of all its eigenvalues

The matrices

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

have different spectral properties:

- A is a rotation of the plane by $\pi/2 \implies$ no real eigenvalues
- $\det(B - \lambda I) = \lambda^2 - 1 = 0 \implies \lambda_{1,2} = \pm 1$
- eigenvectors $\mathbf{v}_{1,2} = (1, \pm 1)^\top$ are pairwise orthogonal
- $P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is orthogonal
- thus $P^{-1}AP = P^\top AP = \operatorname{diag}(1, -1)$

General 2×2 symmetric matrices

The main reason is that B above is symmetric, i.e., $B^\top = B$

For general symmetric 2×2 matrix $B = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$:

- $p(\lambda) = \lambda^2 - (a+d)\lambda + (ad - b^2) \implies$
  $$r_{1,2} = \tfrac{1}{2}\left[(a+d) \pm \sqrt{(a-d)^2 + 4b^2}\right] \quad \text{are real}$$
- $r_1 = r_2 \iff a = d \quad \text{and} \quad b = 0 \implies B = aI$;
  then $\mathbf{v}_1$ and $\mathbf{v}_2$ can be chosen orthogonal;
- if $r_1 \neq r_2$ and $\mathbf{v}_1$, $\mathbf{v}_2$ are the EVc's, then
  $$(r_1 - r_2)\mathbf{v}_1^T\mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2 - \mathbf{v}_1^T A\mathbf{v}_2 = \mathbf{v}_1^T A^T\mathbf{v}_2 - \mathbf{v}_1^T A\mathbf{v}_2 = 0$$
  so that $\mathbf{v}_1^T\mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \implies \mathbf{v}_1, \mathbf{v}_2$ are orthogonal
- set $\mathbf{w}_i := \mathbf{v}_i / \|\mathbf{v}_i\|$ and $P = (\mathbf{w}_1 \ \mathbf{w}_2)$; then P is orthogonal and
  $$P^TAP = P^{-1}AP = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix}.$$

General symmetric matrices

All symmetric $n \times n$ matrices possess most of the desirable properties:

- all their eigenvalues are real;
- there are enough eigenvectors (generalized eigenvectors never occur)
- the eigenvectors are orthogonal to each other

Theorem

Let A be an $n \times n$ symmetric matrix. Then

(a) *all EV's are real;*

(b) *EVc's corresponding to distinct EV's are orthogonal;*

(c) *even if A has multiple EV's, $\exists$ a nonsingular P whose columns $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are EVc's of A s.t.*

  (i) *$\mathbf{w}_1, \ldots, \mathbf{w}_n$ are mutually orthonormal;*

  (ii) *$P^{-1} = P^T$ (i.e., P is orthogonal);*

  (iii) *$P^{-1}AP = P^TAP = \mathrm{diag}\{r_1, \ldots, r_n\}$.*

Proof:

- Allowing non-real objects requires replacing $\mathbb{R}^n$ with $\mathbb{C}^n$:
  - scalar product is $\mathbf{x}^*\mathbf{y} := (\overline{\mathbf{x}})^\top \mathbf{y} = \overline{x_1}y_1 + \cdots + \overline{x_n}y_n$
  - norm is $\|\mathbf{x}\| = \sqrt{\mathbf{x}^*\mathbf{x}} = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{1/2}$
- EV's real: $A\mathbf{v} = \lambda\mathbf{v} \implies A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$

$$\overline{\lambda}\mathbf{v}^*\mathbf{v} = (\overline{\lambda}\overline{\mathbf{v}})^\top\mathbf{v} = (A\overline{\mathbf{v}})^\top\mathbf{v} = \overline{\mathbf{v}}^\top(A\mathbf{v}) = \lambda\mathbf{v}^*\mathbf{v}$$

  therefore, $(\overline{\lambda} - \lambda)\mathbf{v}^*\mathbf{v} = 0 \implies \lambda$ real
- EVc's orthogonal: $A\mathbf{v}_j = r_j\mathbf{v}_j \implies$

$$(r_1 - r_2)\mathbf{v}_1^*\mathbf{v}_2 = (A\mathbf{v}_1)^*\mathbf{v}_2 - \mathbf{v}_1^*A\mathbf{v}_2 = \mathbf{v}_1^*A^\top\mathbf{v}_2 - \mathbf{v}_1^*A\mathbf{v}_2 = 0$$

  so that $\mathbf{v}_1^*\mathbf{v}_2 = 0 \implies \mathbf{v}_1, \mathbf{v}_2$ are orthogonal
- no Jordan blocks, ie, no generalized eigenvectors:

$$(A - rI)\mathbf{v} = \mathbf{0}, \qquad (A - rI)\mathbf{w} = \mathbf{v} \implies$$
$$\mathbf{v}^*\mathbf{v} = \left((A - rI)\mathbf{w}\right)^*\mathbf{v} = \mathbf{w}^*(A - rI)\mathbf{v} = \mathbf{w}^*\mathbf{0} = 0$$

- therefore, enough eigenvectors even for repeated EV's

Symmetric vs Hermitian

- A symmetric $\iff$
  - A has real entries
  - $A^T = A$
- A is Hermitian $\iff$
  - A has complex entries
  - $A^* := (\overline{A})^\top = A$ (i.e., $a_{ij} = \overline{a_{ji}}$); A^* is the adjoint of A

Example (A Hermitian, B not Hermitian)

$$A = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

Note that every real symmetric matrix is also Hermitian!

Hermitian matrices have the same properties:

- EV's real
- EVc's orthogonal (although may have complex entries!)
- no generalized EVc's
- thus diagonalizable!

Example

$$A = \begin{pmatrix} 4 & 3 & -1 \\ 3 & 4 & -1 \\ -1 & -1 & 8 \end{pmatrix}$$

has eigenvalues $r = 1$ (visual), $r = 6$ (column sums), and $r = 9$ (trace rule). The corresponding normalized eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{u}_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{u}_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix},$$

and the orthogonal matrix Q is

$$Q = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \end{pmatrix}$$

Finally, $Q^{-1} = Q^T$ and $Q^T A Q = \mathrm{diag}(1, 6, 9)$

Diagonalize the matrix

$$A = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$$

with EV's $r = 1, 1, 1$, and 5. The eigenvectors $\mathbf{v} = (u_1, u_2, u_3, u_4)^T$ corresponding to $r = 1$ satisfy $u_1 + u_2 + u_3 + u_4 = 0$ and can be taken

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Under Gram–Schmidt orthogonalization they produce

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{w}_3 = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1 \end{pmatrix}$$

Spectral decomposition

Spectral decomposition of a symmetric matrix

Let A be $n \times n$ symmetric with

- eigenvalues $\lambda_1, \dots, \lambda_n$
- normalized pairwise orthogonal eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$

Then

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^\top$$

In other words,

$$A\mathbf{x} = \lambda_1 \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \lambda_n \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n$$

and

$$\mathbf{x}^\top A\mathbf{x} = \langle A\mathbf{x}, \mathbf{x} \rangle = \lambda_1 |\langle \mathbf{x}, \mathbf{v}_1 \rangle|^2 + \cdots + \lambda_n |\langle \mathbf{x}, \mathbf{v}_n \rangle|^2$$

Minimax properties

- Assume A is nonnegative, ie, $\mathbf{x}^T A\mathbf{x} \geq 0$
- Order EV's of A: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$; then
$$\lambda_1 = \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^T A\mathbf{x}}{\|\mathbf{x}\|^2} = \max\{\mathbf{x}^T A\mathbf{x} \mid \|\mathbf{x}\| = 1\}$$
- With $\mathbf{v}_1, \ldots, \mathbf{v}_k$ the normalized EVc's for $\lambda_1, \ldots, \lambda_k$:
$$\lambda_{k+1} = \max\{\mathbf{x}^T A\mathbf{x} \mid \mathbf{x} \perp S_k, \|\mathbf{x}\| = 1\},$$
with S_k the spectral subspace $S_k := \mathrm{ls}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$
- $$\lambda_{k+1} = \min_{L_k:\dim L_k = k} \max\{\mathbf{x}^T A\mathbf{x} \mid \mathbf{x} \perp L_k, \|\mathbf{x}\| = 1\}$$
- λ_{k+1} is the largest EV for $P_k A P_k$, where
$$P_k = I - (\mathbf{v}_1\mathbf{v}_1^T + \cdots + \mathbf{v}_k\mathbf{v}_k^T)$$
is the orthogonal projection onto $\mathbb{R}^n \ominus S_k$

Quadratic forms

A quadratic form Q on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ s.t.

$$Q(\mathbf{x}) = Q(x_1, x_2, \ldots, x_n) = \sum a_{ij} x_i x_j$$

Q.f. are in $1-1$ correspondence with symmetric matrices via

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

Example

$$\begin{aligned} Q(\mathbf{x}) &= x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3 \\ &= (x_1 \ x_2 \ x_3) \begin{pmatrix} 1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \end{aligned}$$

Hessian of a function

Quadratic forms naturally arise as follows

Recall the Taylor expansion around x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + o((x - x_0)^2)$$

For a function $f : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{x}_0) &+ (\mathbf{x} - \mathbf{x}_0)^\top \nabla f(\mathbf{x}_0) \\ &+ \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top H(f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2) \end{aligned}$$

where ∇f is the gradient of f and $H(f)$ is the Hessian of f, i.e., the $n \times n$ matrix with entries

$$h_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$H(f)$ is a symmetric matrix!

Change of variables

After the change of variables $\mathbf{x} = P\mathbf{y}$,

$$Q(\mathbf{x}) = Q(P\mathbf{y}) = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T(P^T AP)\mathbf{y}$$

so that Q is associated with $P^T AP$ in the variables $\mathbf{y}$

Remark (Matrices vs quadratic forms)

*The **matrix** A (not quadratic form Q) in the variables $\mathbf{y}$ becomes $P^{-1}AP$! However, for an orthogonal matrix P both $P^T AP$ and $P^{-1}AP$ are the same, which gives the following theorem*

Theorem (Principal axes theorem)

If A is a symmetric matrix, then there is an orthogonal change of variables $\mathbf{x} = P\mathbf{y}$ that transforms the quadratic form $\mathbf{x}^\top A\mathbf{x}$ into

$$\mathbf{x}^\top A\mathbf{x} = \mathbf{y}^\top D\mathbf{y} = r_1 y_1^2 + \cdots + r_n y_n^2,$$

in which $r_1, \ldots, r_n$ are the eigenvalues of A corresponding to the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ that form the successive columns of P.

Change of variables

Example

$$Q(\mathbf{x}) = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$$
$$= (x_1\ x_2\ x_3)\begin{pmatrix} 1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

- The eigenvalues of A are $r = 0, -3, 3$
- The eigenvectors are

$$\mathbf{v}_1 = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \qquad \mathbf{v}_2 = \frac{1}{3}\begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix}, \qquad \mathbf{v}_3 = \frac{1}{3}\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$$

- The form Q in its principal axes:

$$Q(P\mathbf{y}) = \mathbf{y}^\top(P^\top AP)\mathbf{y} = -3y_2^2 + 3y_3^2$$

Definiteness of quadratic form

Definition

A quadratic form $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ (a symmetric matrix A) is called

(a) positive definite $\iff \forall\, \mathbf{x} \neq 0$: $Q(\mathbf{x}) > 0$;

(b) positive semidefinite $\iff \forall\, \mathbf{x} \neq 0$: $Q(\mathbf{x}) \geq 0$;

(c) negative definite $\iff \forall\, \mathbf{x} \neq 0$: $Q(\mathbf{x}) < 0$;

(d) negative semidefinite $\iff \forall\, \mathbf{x} \neq 0$: $Q(\mathbf{x}) \leq 0$;

(e) indefinite $\iff \exists \mathbf{x}_\pm$ s.t. $\pm Q(\mathbf{x}_\pm) > 0$.

Example (Gram matrix)

Take a linearly independent system of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ and form a $k \times k$ matrix G with $g_{ij} := \mathbf{v}_i^\top \mathbf{v}_j$. Then the quadratic form $Q(\mathbf{c}) := \mathbf{c}^\top G\mathbf{c}$ is positive definite on $\mathbb{R}^k$:

$$Q(\mathbf{c}) = \sum_{i,j} c_i g_{ij} c_j = \sum_{i,j} c_i \mathbf{v}_i^\top \mathbf{v}_j c_j = \mathbf{w}^\top \mathbf{w} \geq 0,$$

where $\mathbf{w} := \sum_i c_i \mathbf{v}_i$; observe that $\mathbf{w} \neq \mathbf{0}$ unless $\mathbf{c} = \mathbf{0}$!

Definiteness and eigenvalues

Theorem

Let A be symmetric and $r_1, r_2, \dots, r_k$ be its EV's. Then

(a) *A is positive definite* $\iff \forall\, j = 1, \dots, k\colon r_j > 0$;

(b) *A is positive semidefinite* $\iff \forall\, j = 1, \dots, k\colon r_j \geq 0$;

(c) *A is negative definite* $\iff \forall\, j = 1, \dots, k\colon r_j < 0$;

(d) *A is negative semidefinite* $\iff \forall\, j = 1, \dots, k\colon r_j \leq 0$;

(e) *A is indefinite* $\iff \exists i, j$ *s.t.* $r_i < 0$ *and* $r_j > 0$

Proof.

Find an orthogonal P s.t. $P^T A P = D = \operatorname{diag}\{r_1, \dots, r_k\}$. In terms of $\mathbf{y} = (y_1, \dots, y_k)^T = P^{-1}\mathbf{x}$ the quadratic form Q reads

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T (P^T A P)\mathbf{y} = \mathbf{y}^T D \mathbf{y} = r_1 y_1^2 + \cdots + r_k y_k^2$$

All statements immediately follow from this representation. $\square$

Bounds on $Q(\mathbf{x})$

Assume that $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ with symmetric A and that $\lambda_{\min}$ and $\lambda_{\max}$ are the minimal and maximal eigenvalues of A. Then

$$\lambda_{\min}\|\mathbf{x}\|^2 \leq Q(\mathbf{x}) \leq \lambda_{\max}\|\mathbf{x}\|^2$$

If $\mathbf{x} = P\mathbf{y}$ with orthogonal P reduces Q to diagonal form D, then

$$\begin{aligned} Q(\mathbf{x}) = Q(P\mathbf{y}) &= \mathbf{y}^T P^T AP\mathbf{y} \\ &= \mathbf{y}^T D\mathbf{y} = \sum_k \lambda_k |y_k|^2 \\ &\leq \lambda_{\max} \sum_k |y_k|^2 \\ &= \lambda_{\max}\|\mathbf{y}\|^2 = \lambda_{\max}\|\mathbf{x}\|^2 \end{aligned}$$

If $\mathbf{x}$ is an EVc corresponding to $\lambda_{\max}$, then

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \lambda_{\max}\mathbf{x}^T\mathbf{x} = \lambda_{\max}\|\mathbf{x}\|^2$$

Theorem

Let A be a symmetric matrix. TFAE:

(a) *A is positive definite;*

(b) *$\exists$ a nonsingular B s.t. $A = B^T B$;*

(c) *$\exists$ a nonsingular C s.t. $C^\top AC = I$.*

Proof.

Find an orthogonal P s.t. $P^T AP = D = \operatorname{diag}\{r_1, \dots, r_k\}$, i.e., $A = PDP^T$.

(a) $\implies$ (b): $r_j > 0$; take $B = D^{1/2}P^T$

(b) $\implies$ (c): take $C = B^{-1}$ and recall that $(B^T)^{-1} = (B^{-1})^T$

(c) $\implies$ (a): $\forall\ \mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} := C^{-1}\mathbf{x}$; then $\mathbf{x} = C\mathbf{y}$ and

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^\top (C^\top AC)\mathbf{y} = \mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|^2 = y_1^2 + \cdots + y_k^2 > 0.$$ $\square$

Theorem (Sylvester criterion)

A symmetric A is pos. def. $\iff$ all principal minors of A are positive

What is a Cholesky decomposition?

Recall:

- $A = LU$ (or $PA = LU$) exists for rectangular matrices
- a LU-representation is not unique!
- usually L is $m \times m$ with 1 on the main diagonal; then both L and $U = L^{-1}A$ are uniquely determined
- reason: if $A = L_1U_1 = L_2U_2$, then
  - $L_2^{-1}L_1 = U_2U_1^{-1}$
  - $L_2^{-1}L_1$ is lower-triangular with 1 on the diagonal
  - $U_2U_1^{-1}$ is upper-triangular
  - thus $L_2^{-1}L_1 = U_2U_1^{-1} = I$
- if A is nonsingular, U has nonzero diagonal
- can “factor it out” as D to get $A = LDU$ with U having 1’s on the main diagonal
- for symmetric matrices, $U = L^T$ and $A = LDL^T$
- reason: $A^T = U^TDL^T = LDU = A$ and use uniqueness

Standard Cholesky decomposition

- The standard form of Cholesky decomposition reads

$$\boxed{A = LL^T}, \qquad L \quad \text{lower-triangular}$$

- requires A positive semi-definite :

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T L^T L\mathbf{x} = (L\mathbf{x})^T L\mathbf{x} = \|L\mathbf{x}\|^2 \geq 0$$

- conversely, if A positive semi-definite, the LL^T decomposition exists
- for positive definite A, such a decomposition is unique!
- if L has diagonal S, then with $L := L_1 S$ and $D := S^2$, we get

$$A = LL^T = (L_1 S)(SL_1^T) = L_1 S^2 L_1^T = L_1 D L_1^T$$

- Thus the LDL^T decomposition is more general as it does not require positive semi-definiteness

Positive definiteness

Theorem (Definiteness and principal minors)

A is positive definite $\iff$ all its principal minors are positive

Proof.

$\Longrightarrow$ A positive definite $\implies$ A_k are positive definite in $\mathbb{R}^k$

$\implies$ $\det A_k > 0$

$\Longleftarrow$ Use the Choleski decomposition $A = LDL^\top$:

L lower-triangular with 1 on the diagonal;

$D = \operatorname{diag}(d_1, \ldots, d_n)$ with pivots d_k

- then $A_k = L_k D_k L_k^\top$
- $\det A_k = \det D_k \implies d_1 \cdots d_k > 0$

$\square$

Remark

d_k above are not eigenvalues of A.

However, the signs of d_k and λ_k coincide.

Positive definiteness

Example (Minors vs eigenvalues)

Minors of

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

are 1, -1, and $\det A = -2 \implies A$ is not positive definite.

- $d_1 = 1$, $d_1 d_2 = -1$, $d_1 d_2 d_3 = -2$
  $\implies d_1 > 0$, $d_2 < 0$, $d_3 > 0$
- the eigenvalues of A are $\lambda_1 = 1$ (row/column sums);
  $\lambda_2 + \lambda_3 = 2$, $\lambda_2 \lambda_3 = -2$
  $\implies \lambda_2$ and λ_2 of opposite sign!

Remark (Negative definiteness)

For negative definiteness, signs of principal minors should alternate, starting from the negative one

Principal Component Analysis

Setting:

Assume $X = (x_{ij})_{n \times p}$ is the data collected from

- n observations
- p characteristics (e.g., age, education duration, salary etc)

Assuming zero means, the covariance matrix of these characteristics is

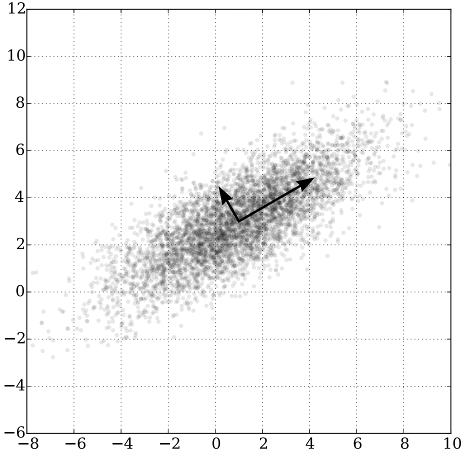
$$\Sigma = X^T X \qquad (p \times p)$$

Usually these characteristics X_j are correlated; the idea is to extract principal components, ie, orthogonal transformation C,

$$Y_j = \sum_k c_{jk} X_k$$

so that $Y_1, \dots, Y_p$ are uncorrelated and $\mathrm{Var}(Y_1)$ is the largest possible, $\mathrm{Var}(Y_2)$ the second largest possible etc.

30.09.2016 PCA.svg



What does PCA actually do?

$$
\begin{aligned}
&\mathrm{Var}(w_1 X_1 + \cdots + w_p X_p) \\
&\quad = \mathrm{Cov}(w_1 X_1 + \cdots + w_p X_p, w_1 X_1 + \cdots + w_p X_p) \\
&\quad = \sum_{jk} w_j w_k \mathrm{Cov}(X_j, X_k) = \sum_{jk} w_j \Sigma_{jk} w_k \to \max
\end{aligned}
$$

- Maximum is attained for the normalized eigenvector $\mathbf{w}$ corresponding to the largest eigenvalue σ_1
- This is the first principal component
- The normalized eigenvectors $\mathbf{w}_2, \ldots, \mathbf{w}_p$ for EV's $\sigma_2 \geq \ldots \sigma_p$ are the other principal components
- $\mathbf{w}_1, \ldots, \mathbf{w}_p$ give columns of the orthogonal matrix C
- Can only restrict to several leading principal components!

Optimization:

For $f : \mathbb{R}^n \to \mathbb{R}$

- necessary condition for local minimum at $\mathbf{x}_0$: $\nabla f(\mathbf{x}_0) = \mathbf{0}$
- sufficient condition for local minimum: in addition the Hessian be positive definite at $\mathbf{x}_0$

Then

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \nabla f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top H(f)(\mathbf{x} - \mathbf{x}_0) > f(\mathbf{x}_0)$$

for $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$