Linear Algebra

Lecture Notes

Rostyslav Hryniv

Ukrainian Catholic University
Business Analytics and Computer Science Programmes

4^{th} term
Spring 2020

Lecture 12. Special matrices

Outline

What did we learn last time?

- Symmetric matrices are orthogonally diagonalizable:
  - there is an orthogonal matrix P and a diagonal D such that
$$P^{-1}AP = P^\top AP = D \iff A = PDP^\top$$
  - columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of P are eigenvectors of A
  - diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ of D are eigenvalues of A
- Then
$$\begin{aligned} A\mathbf{x} = PDP^\top\mathbf{x} &= \begin{pmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^\top\mathbf{x} \\ \vdots \\ \mathbf{u}_n^\top\mathbf{x} \end{pmatrix} \\ &= \lambda_1\mathbf{u}_1\mathbf{u}_1^\top\mathbf{x} + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^\top\mathbf{x} \end{aligned}$$
- This is the Spectral Theorem:
$$\boxed{A = \lambda_1\mathbf{u}_1\mathbf{u}_1^\top + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^\top}$$
- With symmetric A there is associated a quadratic form
$$Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$$
- Positive/negative definiteness of Q depends on $\lambda_1, \ldots, \lambda_n$

Symmetric A with prescribed spectrum

Symmetric A with prescribed spectrum

The Spectral Theorem guarantees there is symmetric A with prescribed eigenvalues and eigenvectors: if

- if $\lambda_1, \ldots, \lambda_n$ are real,
- if $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthonormal basis,
- then

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

Example

Eigenvectors corresponding to distinct eigenvalues are always orthogonal, so there is no symmetric matrix with eigenvectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

unless $\lambda_1 = \lambda_3$

Positive definiteness

Recall A is positive definite if

$$Q(\mathbf{x}) := \mathbf{x}^\top A\mathbf{x} > 0 \qquad \text{for all nonzero } \mathbf{x}$$

A is positive definite if and only if

- all eigenvalues of A are positive
- or if $A = B^\top B$ for a nonsingular B

Definition

Principal minors of A are determinants of submatrices A_k of A formed by first k columns and rows, for $k = 1, \ldots, n$.

Example (Principal minors of A)

$$A = \begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 4 \end{pmatrix}$$

are 2, -3, and $\det A = -15$.

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;
$D = \mathrm{diag}(d_1, \dots, d_n)$ with pivots d_k

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

□

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_3 of opposite sign!

Remark (Negative definiteness)

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

Why quadratic forms?

Most tasks of LA are related to the equation

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

- it is solvable iff $\mathbf{b}$ is in the column space of A
- if not solvable, look for the best approximate solution minimizing

$$F(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}) = \mathbf{x}^\top (A^\top A)\mathbf{x} - 2\mathbf{x}^\top (A^\top \mathbf{b}) + \|\mathbf{b}\|^2$$

- the point of minimum is the least squares solution; it satisfies

$$\operatorname{grad} F(\mathbf{x}) = 0 \qquad \Longleftrightarrow \qquad 2(A^\top A)\mathbf{x} - 2A^\top \mathbf{b} = 0$$

Conjugate gradient method

- CGM is a method of solving $A\mathbf{x} = \mathbf{b}$ with positive definite A
- The equation possesses a unique solution $\mathbf{x}_*$
- Used e.g. when A is large and sparse
- A introduces the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_A := \langle A\mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A\mathbf{y}$

Definition

$\mathbf{u}$ and $\mathbf{v}$ are conjugate iff $\langle \mathbf{u}, \mathbf{v} \rangle_A = 0$

- Choose n mutually conjugate vectors $\mathbf{p}_1, \ldots, \mathbf{p}_n$
- they form a basis of $\mathbb{R}^n$, thus $\mathbf{x}_* = \sum \alpha_i \mathbf{p}_i$
- $\langle \mathbf{p}_k, \mathbf{x}_* \rangle_A = \alpha_k \langle \mathbf{p}_k, \mathbf{p}_k \rangle_A$, but also $\langle \mathbf{p}_k, \mathbf{x}_* \rangle_A = \langle \mathbf{p}_k, A\mathbf{x}_* \rangle = \langle \mathbf{p}_k, \mathbf{b} \rangle$
- thus

$$\alpha_k = \frac{\langle \mathbf{p}_k, \mathbf{b} \rangle}{\langle \mathbf{p}_k, \mathbf{p}_k \rangle_\mathbf{A}}$$

- this gives a unique solution $\mathbf{x}_* = \sum \alpha_i \mathbf{p}_i$

Anti-symmetric matrices

Definition

A matrix A is called anti-symmetric (resp. skew-Hermitian) if

$$A^\top = -A \qquad (\text{resp. } A^* = -A)$$

Lemma

For a real matrix A,
A is anti-symmetric $\iff$ iA is Hermitian

Proof.

$(iA)^* = (\overline{iA})^\top = (-i)A^\top = iA$ $\square$

Corollary

- *Eigenvalues of anti-symmetric matrices are purely imaginary*
- *Every anti-symmetric A is orthogonally diagonalizable*

$A^* = -A \quad \iff \quad (iA)^* = iA$

Anti-symmetric matrices

Theorem

If A is anti-symmetric (skew-Hermitian), then e^{tA} is orthogonal (unitary)

Proof.

$$
\begin{aligned}
e^{tA} &= I + (tA) + \frac{1}{2}(tA)^2 + \frac{1}{3!}(tA)^3 + \dots \\
e^{-tA} &= I + (-tA) + \frac{1}{2}(-tA)^2 + \frac{1}{3!}(-tA)^3 + \dots \\
(e^{tA})^\top &= I + (tA)^\top + \frac{1}{2}(tA^\top)^2 + \frac{1}{3!}(tA^\top)^3 + \dots \\
&= I + (-tA) + \frac{1}{2}(-tA)^2 + \frac{1}{3!}(-tA)^3 + \dots
\end{aligned}
$$

$\square$

Remark

Vice versa, e^{tA} unitary implies A is skew-symmetric:

$$\frac{d}{dt}\langle e^{tA}\mathbf{x}, e^{tA}\mathbf{y}\rangle|_{t=0} = \langle A\mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{x}, A\mathbf{y}\rangle = 0$$

Orthogonal and unitary matrices

Definition

A matrix U with complex entries is unitary if

$$U^* = U^{-1}$$

Theorem (Characterization of unitary matrices)

For a square U: U is unitary $\iff$ U preserves norms, i.e. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

Proof.

$$\|U\mathbf{x}\|^2 = (U\mathbf{x})^* U\mathbf{x} = \mathbf{x}^*(U^*U)\mathbf{x}$$

- U is unitary $\implies$ $\|U\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x} = \|\mathbf{x}\|^2$
- U preserves norms $\implies$ $\mathbf{x}^*(U^*U)\mathbf{x} = \mathbf{x}^*\mathbf{x} \implies U^*U = I$ □

Remark

A unitary matrix U also preserves the scalar products:

$$(U\mathbf{x})^*(U\mathbf{y}) = \mathbf{x}^* U^* U\mathbf{y} = \mathbf{x}^*\mathbf{y}$$

Properties of unitary matrices

Theorem (Spectral properties of unitary matrices)

For a unitary matrix U

(a) *all eigenvalues λ satisfy $|\lambda| = 1$;*

(b) *eigenvectors corresponding to distinct eigenvalues are orthogonal*

Proof.

(a) If $U\mathbf{x} = \lambda\mathbf{x}$, then $\|\mathbf{x}\|^2 = \|U\mathbf{x}\|^2 = \|\lambda\mathbf{x}\|^2 = |\lambda|^2\|\mathbf{x}\|^2$, so that $|\lambda| = 1$.

(b) If also $U\mathbf{y} = \mu\mathbf{y}$, then $\mathbf{x}^*\mathbf{y} = (U\mathbf{x})^*(U\mathbf{y}) = \overline{\lambda}\mu\,\mathbf{x}^*\mathbf{y}$
$\implies \overline{\lambda}\mu = 1$ or $\mathbf{x}^*\mathbf{y} = 0$;
but $\overline{\lambda}\mu \neq 1$ for $|\lambda| = |\mu| = 1$ and $\lambda \neq \mu$ $\square$

Theorem (Spectral Theorem for unitary matrices)

A unitary matrix U is unitarily diagonalizable, i.e., there is a unitary matrix P such that

$$P^* UP = \operatorname{diag}\{e^{i\theta_1}, \dots, e^{i\theta_n}\}$$

Theorem (Cayley transform)

- *If A is a symmetric (Hermitian) matrix, then*

  $$U := (A + iI)(A - iI)^{-1}$$

  is unitary
- *If U is a unitary matrix s.t. -1 is not its eigenvalue, then*

  $$A = i(U + I)^{-1}(U - I)$$

  is Hermitian

Remark

If -1 is an eigenvalue, take $e^{i\theta} U$ instead, with a suitable θ

Direct sum

Definition

The sum $U + W$ of two subspaces U and W of a vector space V is

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U,\ \mathbf{w} \in W\}$$

This sum is called direct (denoted $U \dot{+} W$) if $U \cap W = \{\mathbf{0}\}$

Remark

$$U \dot{+} W \neq U \cup W$$

Example

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, then

$$\mathrm{ls}\{\mathbf{v}_1\} \dot{+} \mathrm{ls}\{\mathbf{v}_2\} = \mathrm{ls}\{\mathbf{v}_1, \mathbf{v}_2\}$$

Direct sums and projections

When is a sum direct?

The sum $U + W$ is direct $\iff$
every vector $\mathbf{x} \in U + W$ has a unique representation as $\mathbf{x} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in W$

Definition

Assume that $V = U \dotplus W$; then every $\mathbf{v} \in V$ can be uniquely written as $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in W$. Then the mapping

- $\mathbf{v} \mapsto \mathbf{u}$ is the projection of V onto U parallel to W
- $\mathbf{v} \mapsto \mathbf{w}$ is the projection of V onto W parallel to U

Example

If $V = U \oplus W$, then the projection of V onto U parallel to W
is the *orthogonal projection* onto U.

Projections

Lemma

The projection T of V onto U parallel to W is a linear mapping satisfying $T^2 = T$. It is uniquely fixed by the requirements that $T\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in U$ and $T\mathbf{w} = 0$ for all $\mathbf{w} \in W$.

Proof.

- Linearity of T follows from the definition of direct sum
- If $\mathbf{y} = T\mathbf{x}$, then $\mathbf{y} \in U$ and $T\mathbf{y} = \mathbf{y}$, so that $T^2\mathbf{x} = T\mathbf{x}$
- As $V = U \dot{+} W$, knowing T on U and W uniquely determines it on the whole V $\square$

Example

Projection onto $\mathbf{v}_1 = (1, 0)^\top$ parallel to $\mathbf{v}_2 = (1, 1)^\top$ is given by the matrix

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

since $A\mathbf{v}_1 = \mathbf{v}_1$ and $A\mathbf{v}_2 = \mathbf{0}$

Projections are idempotents

Definition

An $n \times n$ matrix A is called idempotent if $A^2 = A$

Corollary

If A is a matrix of the projection T of V onto U parallel to W, then A is an idempotent matrix.

Example

With any invertible B, the matrix

$$A := B \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} B^{-1}$$

is an idempotent.

Lemma

An idempotent matrix has eigenvalues: $\lambda = 0$ and/or $\lambda = 1$.

Proof.

$$A\mathbf{x} = \lambda\mathbf{x} \implies A^2\mathbf{x} = \lambda^2\mathbf{x} \implies \lambda^2 = \lambda$$

□

Idempotents are projections

Theorem

Every idempotent A is a matrix of projection of $\mathbb{R}^n$ onto the column space $\operatorname{col}(A)$ of A parallel to the nullspace $\operatorname{nul}(A)$ of A.

Proof.

The equality $\mathbf{x} = A\mathbf{x} + (\mathbf{x} - A\mathbf{x})$ shows that

$$\operatorname{col}(A) + \operatorname{nul}(A) = \mathbb{R}^n.$$

The sum is direct since if $A\mathbf{x} = (I - A)\mathbf{y}$, then

$$A\mathbf{x} = A^2\mathbf{x} = A(I - A)\mathbf{y} = \mathbf{0}.$$

□

Constructing a projection

Assume:

- $\mathbb{R}^n = U \dotplus W$;
- $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis of U;
- $\mathbf{w}_1, \ldots, \mathbf{w}_l$ is a basis of W.

Form a matrix B with columns $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_l$; then the matrix

$$A = B \operatorname{diag}\{\underbrace{1, \ldots, 1}_{k}, 0, \ldots, 0\} B^{-1}$$

is the matrix of the projection of $\mathbb{R}^n$ onto U parallel to W

Indeed, the equalities $A^2 = A$, $A\mathbf{u}_j = \mathbf{u}_j$, and $A\mathbf{w}_j = \mathbf{0}$ are verified in a straightforward manner

Orthogonal projection

Theorem (When projection is orthogonal?)

A is a matrix of an orthogonal projection if and only if A is a symmetric idemponent.

Proof.

$\Longrightarrow$ If A is a matrix of an orthogonal projection, then $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, A\mathbf{y}\rangle$$

$\Longleftarrow$ A projects onto $\mathsf{col}(A)$ parallel to $\mathsf{nul}(A)$, and it remains to show symmetry implies orthogonality of $\mathsf{col}(A)$ and $\mathsf{nul}(A)$:
$\mathbf{x} \in \mathsf{col}(A)$, $\mathbf{y} \in \mathsf{nul}(A) \implies$

$$\langle \mathbf{x}, \mathbf{y}\rangle = \langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A\mathbf{y}\rangle = 0$$

$\square$

Corollary

For any $m \times n$ matrix A with linearly independent columns the matrix $P := A(A^\top A)^{-1}A^\top$ is an orthogonal projection of $\mathbb{R}^m$ onto $\mathsf{col}(A)$

Reason: P is idempotent and symmetric; $\mathsf{nul}(P) = \mathsf{nul}(A^\top) = (\mathsf{col}(A))^\perp$