4. Matrix rank and inverses

- subspaces
- independence, basis, dimension
- rank of a matrix
- inverses of a matrix
- pseudo-inverse

Subspace

a nonempty subset $\mathcal{V}$ of $\mathbb{R}^m$ is a subspace if

$$\alpha x + \beta y \in \mathcal{V} \quad \text{for all vectors } x, y \in \mathcal{V} \text{ and } \alpha, \beta \in \mathbb{R}$$

- all linear combinations of elements of $\mathcal{V}$ are in $\mathcal{V}$
- $\mathcal{V}$ is nonempty and closed under scalar multiplication and vector addition
- geometrically, a subspace is a flat (plane) that passes through the origin

Examples

- $\{0\}$, $\mathbb{R}^m$
- the **span** of a set $\mathcal{S} \subseteq \mathbb{R}^m$: all linear combinations of elements of $\mathcal{S}$

  $$\text{span}(\mathcal{S}) = \{\beta_1 a_1 + \cdots + \beta_k a_k \mid a_1, \ldots, a_k \in \mathcal{S}, \beta_1, \ldots, \beta_k \in \mathbb{R}\}$$

  if $\mathcal{S} = \{a_1, \ldots, a_n\}$ is a finite set, we write $\text{span}(\mathcal{S}) = \text{span}(a_1, \ldots, a_n)$

  (the span of the empty set is defined as $\{0\}$)

Operations on subspaces

three common operations on subspaces ($\mathcal{V}$ and $\mathcal{W}$ are subspaces)

Sum and intersection (the result of these operations is a subspace)

- *intersection:*

$$\mathcal{V} \cap \mathcal{W} = \{x \mid x \in \mathcal{V}, x \in \mathcal{W}\}$$

- *sum:*

$$\mathcal{V} + \mathcal{W} = \{x + y \mid x \in \mathcal{V}, y \in \mathcal{W}\}$$

  if $\mathcal{V} \cap \mathcal{W} = \{0\}$ this is called the *direct sum* and written as $\mathcal{V} \oplus \mathcal{W}$

Orthogonal complement

$$\mathcal{V}^{\perp} = \{x \mid y^T x = 0 \text{ for all } y \in \mathcal{V}\}$$

- set of all vectors x, each of which is orthogonal to every vector in $\mathcal{V}$
- always a subspace even if $\mathcal{V}$ is not

Range of a matrix

suppose A is an $m \times n$ matrix with columns $a_1, \ldots, a_n$ and rows $b_1^T, \ldots, b_m^T$:

$$A = \left[\begin{array}{ccc} a_1 & \cdots & a_n \end{array}\right] = \begin{bmatrix} b_1^T \\ \vdots \\ b_m^T \end{bmatrix}$$

Range (*column space*): the span of the column vectors (a subspace of $\mathbb{R}^m$)

$$\begin{aligned} \mathrm{range}(A) &= \mathrm{span}(a_1, \ldots, a_n) \\ &= \{x_1 a_1 + \cdots + x_n a_n \mid x \in \mathbb{R}^n\} \\ &= \{Ax \mid x \in \mathbb{R}^n\} \end{aligned}$$

the *row space* of A is the range of A^T (a subspace of $\mathbb{R}^n$):

$$\begin{aligned} \mathrm{range}(A^T) &= \mathrm{span}(b_1, \ldots, b_m) \\ &= \{y_1 b_1 + \cdots + y_m b_m \mid y \in \mathbb{R}^m\} \\ &= \{A^T y \mid y \in \mathbb{R}^m\} \end{aligned}$$

Nullspace of a matrix

suppose A is an $m \times n$ matrix with columns $a_1, \ldots, a_n$ and rows $b_1^T, \ldots, b_m^T$:

$$A = [\ a_1 \ \cdots \ a_n\ ] = \begin{bmatrix} b_1^T \\ \vdots \\ b_m^T \end{bmatrix}$$

Nullspace: the set of all vectors x mapped into zero by A (a subspace of $\mathbb{R}^n$)

$$\begin{aligned} \mathrm{null}(A) &= \{x \in \mathbb{R}^n \mid Ax = 0\} \\ &= \{x \in \mathbb{R}^n \mid b_1^T x = \cdots = b_m^T x = 0\} = \mathrm{range}(A^T)^\perp \end{aligned}$$

- this is the orthogonal complement of the row space
- gives ambiguity in x given $y = Ax$ ($y = Ax = A(x + \tilde{x})$ for any $\tilde{x} \in \mathrm{null}(A)$)
- nullspace of A^T is the orthogonal complement of $\mathrm{range}(A)$ (a subspace of $\mathbb{R}^m$)

$$\begin{aligned} \mathrm{null}(A^T) = \mathrm{range}(A)^\perp &= \{y \in \mathbb{R}^m \mid a_1^T y = \cdots = a_n^T y = 0\} \\ &= \{y \in \mathbb{R}^m \mid A^T y = 0\} \end{aligned}$$

Example

- the range is simply Ax:

$$\operatorname{range}\left(\begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 0 & -1 \end{bmatrix}\right) = \left\{ \begin{bmatrix} x_1 \\ x_1 + 2x_2 \\ -x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R} \right\}$$

- the nullspace is computed by solving $Ax = 0$:

$$\operatorname{null}\left(\begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{bmatrix}\right) = \text{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

to see this, solve for $Ax = 0$:

$$x_1 + 2x_2 - x_3 = 0 \quad \Rightarrow \quad x_1 = -2x_2 + x_3$$

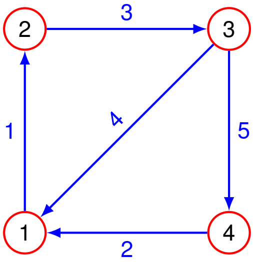
letting $x_2 = s$ and $x_3 = t$, we get:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Exercise

- directed graph with m nodes (vertices), n directed edges (arcs)
- node-arc *incidence matrix* is $m \times n$ matrix A with

$$A_{ij} = \begin{cases} 1 & \text{if edge } j \text{ points to node } i \\ -1 & \text{if edge } j \text{ points from node} i \\ 0 & \text{otherwise} \end{cases}$$



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

describe in words the subspaces $\mathrm{null}(A)$ and $\mathrm{range}(A^T)$

Outline

- subspaces
- **independence, basis, dimension**
- rank of a matrix
- inverses of a matrix
- pseudo-inverse

Linearly independent vectors

the set of vectors $a_1, \ldots, a_n$ is *linearly independent* if

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0 \quad \Longrightarrow \quad x_1 = x_2 = \cdots = x_n = 0$$

i.e., the zero vector cannot be written as a nontrivial linear combination of $a_1, \ldots, a_n$

- in matrix-vector notation with a_i the ith column of A:

$$Ax = 0 \quad \Longrightarrow \quad x = 0$$

- $a_1, \ldots, a_n$ is linearly dependent if there exist $x_1, \ldots, x_n$, not all zero, with

$$Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0$$

- a set of a single vector is linearly independent only if the vector is nonzero
- linear (in)dependence is a property of the *set* of vectors $\{a_1, \ldots, a_n\}$
- by convention, the empty set is linearly independent

Example

- the vectors

$$a_1 = (0.2, -7, 8.6), \quad a_2 = (-0.1, 2, -1), \quad a_3 = (0, -1, 2.2)$$

are linearly dependent since

$$0 = a_1 + 2a_2 - 3a_3$$

- the vectors

$$a_1 = (1, -2, 0), \quad a_2 = (-1, 0, 1), \quad a_3 = (0, 1, 1)$$

are linearly independent:

$$x_1 a_1 + x_2 a_2 + x_3 a_3 = \begin{bmatrix} x_1 - x_2 \\ -2x_1 + x_3 \\ x_2 + x_3 \end{bmatrix} = 0 \Longrightarrow x_1 = x_2 = x_3 = 0$$

Independent dimension inequality

if n vectors $a_1, a_2, \ldots, a_n$ of size m are linearly independent, then

$$n \leq m$$

i.e., the number of linearly independent vectors cannot be larger than their size

- if an $m \times n$ matrix has linearly independent columns then $m \geq n$
- if A is wide ($n > m$), then its columns are linearly dependent:
  the homogeneous equation $Ax = 0$ has nontrivial solutions ($x \neq 0$)
- if an $m \times n$ matrix has linearly independent rows then $m \leq n$
- if A is tall ($m > n$), then its rows are linearly dependent

Basis of a subspace

$\{v_1, \dots, v_k\}$ is a *basis* for the subspace $\mathcal{V}$ if two conditions are satisfied:

1. $\mathcal{V} = \text{span}(v_1, \dots, v_k)$
2. $v_1, \dots, v_k$ are linearly independent

- condition 1 means that every $x \in \mathcal{V}$ can be expressed as

$$x = \beta_1 v_1 + \cdots + \beta_k v_k$$

- condition 2 means that the coefficients $\beta_1, \dots, \beta_k$ are unique:

$$\left.\begin{array}{l} x = \beta_1 v_1 + \cdots + \beta_k v_k \\ x = \gamma_1 v_1 + \cdots + \gamma_k v_k \end{array}\right\} \Longrightarrow (\beta_1 - \gamma_1)v_1 + \cdots + (\beta_k - \gamma_k)v_k$$

$$\Longrightarrow \beta_1 = \gamma_1, \quad \dots, \quad \beta_k = \gamma_k$$

Basis of $\mathbb{R}^n$

any set of n linearly independent n-vectors $a_1, \ldots, a_n$ is a *basis* for $\mathbb{R}^n$

- any $b \in \mathbb{R}^n$ can be expressed as a linear combination of them:

$$b = \beta_1 a_1 + \cdots + \beta_n a_n \quad \text{for some} \quad \beta_1, \ldots, \beta_n$$

  and these coefficients are unique

- formula above is called *expansion* of b in the $a_1, \ldots, a_n$ basis
- example: $e_1, \ldots, e_n$ is a basis, expansion of b is

$$b = b_1 e_1 + \cdots + b_n e_n$$

Example: single period loans

- consider cash flows over n periods
- define the single-period loan cash flow n-vectors as

$$l_i = \begin{bmatrix} 0_{i-1} \\ 1 \\ -(1+r) \\ 0_{n-i-1} \end{bmatrix}, \quad i = 1, \ldots, n-1,$$

  where $r \geq 0$ is the per-period interest rate

- l_i represents a $\$1$ loan in period i, paid back in period $i+1$ with interest r
- scaling l_i changes the loan amount
- vectors $e_1, l_1, \ldots, l_{n-1}$ are a basis of $\mathbb{R}^n$

Example: single period loans

- to see this observe

$$\beta_1 e_1 + \beta_2 l_1 + \cdots + \beta_n l_{n-1} = \begin{bmatrix} \beta_1 + \beta_2 \\ \beta_3 - (1+r)\beta_2 \\ \vdots \\ \beta_n - (1+r)\beta_{n-1} \\ -(1+r)\beta_n \end{bmatrix} = 0$$

- working backward gives $\beta_1 = \cdots = \beta_n = 0$
- this means that any cash flow n-vector c can be expressed as

$$c = \alpha_1 e_1 + \alpha_2 l_1 + \cdots + \alpha_n l_{n-1}$$

- it can be shown that

$$\alpha_1 = c_1 + \frac{c_2}{1+r} + \cdots + \frac{c_n}{(1+r)^{n-1}}$$

  is the net present value (NPV) of the cash flow, with interest rate r

- this means that any cash flow can be replicated as an income in period 1 equal to its NPV, plus a linear combination of one-period loans at interest rate r

Extension of dimension inequality

- let $\{v_1, \ldots, v_k\}$ be a basis for a subspace $\mathcal{V} \subseteq \mathbb{R}^m$
- if $a_1, \ldots, a_n$ are linearly independent vectors in $\mathcal{V}$, then $n \leq k$
- this improves the dimension inequality ($n \leq m$) on page 4.10

Proof

- each a_i can be expressed as a linear combination of the basis vectors:

$$a_1 = Bx_1, \quad a_2 = Bx_2, \quad \ldots, \quad a_n = Bx_n$$

for some k-vectors $x_1, \ldots, x_n$, where B is the $m \times k$ matrix $B = [v_1 \ \cdots \ v_k]$

- the k-vectors $x_1, \ldots, x_n$ are linearly independent:

$$\begin{aligned} \beta_1 x_1 + \cdots + \beta_l x_n = 0 &\Longrightarrow B(\beta_1 x_1 + \cdots + \beta_l x_n) = \beta_1 a_1 + \cdots + \beta_l a_n = 0 \\ &\Longrightarrow \beta_1 = \cdots = \beta_n = 0 \end{aligned}$$

- by the dimension inequality of page 4.10, this implies $n \leq k$

Dimension of a subspace

- every basis of a subspace $\mathcal{V}$ contains the same number of vectors
- this number is called the *dimension* of $\mathcal{V}$, denoted $\dim(\mathcal{V})$

Proof: consider two bases of $\mathcal{V}$

$$\{v_1, \ldots, v_k\}, \quad \{w_1, \ldots, w_l\}$$

from previous page,

- $l \leq k$, because $w_1, \ldots, w_l$ are linearly independent and $\{v_1, \ldots, v_k\}$ is a basis
- $k \leq l$ because $v_1, \ldots, v_k$ are linearly independent and $\{w_1, \ldots, w_l\}$ is a basis

therefore $k = l$

Completing a basis

let $\mathcal{V}$ be a subspace in $\mathbb{R}^m$

- suppose $\{v_1, \ldots, v_j\} \subset \mathcal{V}$ is a linearly independent set (possibly empty)
- then there exists a basis of $\mathcal{V}$ of the form $\{v_1, \ldots, v_j, v_{j+1}, \ldots, v_k\}$
- we complete the basis by adding the vectors $v_{j+1}, \ldots, v_k$

Proof

- if $\{v_1, \ldots, v_j\}$ is not already a basis, its span is not $\mathcal{V}$
- then there exist vectors in $\mathcal{V}$ that are not linear combinations of $v_1, \ldots, v_j$
- choose one of those vectors, call it v_{j+1}, and add it to the set
- the set $\{v_1, \ldots, v_{j+1}\}$ is a linearly independent subset of $\mathcal{V}$ with $j + 1$ elements
- repeat this process until it terminates
- it terminates because a linearly independent set in $\mathbb{R}^m$ has at most m elements

Consequence: every subspace of $\mathbb{R}^m$ has a basis

Outline

- subspaces
- independence, basis, dimension
- **rank of a matrix**
- inverses of a matrix
- pseudo-inverse

Rank of a matrix

Rank: the *rank* of a matrix is the dimension of its range

$$\mathrm{rank}(A) = \dim(\mathrm{range}(A))$$

this is also the maximal number of linearly independent columns

Example: a 4×4 matrix with rank 3

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 3 & 1 \\ -1 & 2 & 0 & 0 \\ 1 & -1 & 3 & 0 \\ -1 & 2 & 0 & 1 \end{bmatrix}$$

- $\{a_1\}$ is linearly independent (a_1 is not zero)
- $\{a_1, a_2\}$ is linearly independent
- $\{a_1, a_2, a_3\}$ is linearly dependent: $a_3 = 6a_1 + 3a_2$
- $\{a_1, a_2, a_4\}$ is a basis for $\mathrm{range}(A)$: linearly independent and spans $\mathrm{range}(A)$

Rank of transpose

the rank of a matrix is equal to the rank of its transpose:

$$\mathrm{rank}(A^T) = \mathrm{rank}(A)$$

(proof given on next page)

- column space (range) of a matrix has the same dimension as its row space:

$$\dim(\mathrm{range}(A^T)) = \dim(\mathrm{range}(A))$$

- max no. of linearly independent columns = max no. of linearly independent rows

Proof

let $A = [a_1 \ \ldots \ a_n]$ be an $m \times n$ matrix with $\mathrm{rank}(A) = r$

- then there are r linearly independent column vectors, denoted by $b_1, \ldots, b_r$
- we have $a_i = Bc_i$ for some vector c_i where $B = [b_1 \ \cdots \ b_r]$ is an $m \times r$ matrix
- in matrix notation, $A = BC$ where $C = [c_1 \ \cdots \ c_n]$ is an $r \times n$ matrix
- implies that every row of A is a linear combination of the rows of C and

$$\mathrm{rank}(A^T) \le r = \mathrm{rank}(A)$$

- now suppose A has p linearly independent rows
- repeating similar argument for A^T, we get

$$\mathrm{rank}(A) \le p = \mathrm{rank}(A^T)$$

the previous two inequalities hold only when $p = r$, *i.e.*, $\mathrm{rank}(A) = \mathrm{rank}(A^T)$

Full rank matrices

for any $m \times n$ matrix

$$\mathrm{rank}(A) \leq \min\{m, n\}$$

Full rank: A has full rank if $\mathrm{rank}(A) = \min\{m, n\}$

Full column rank: A has full column rank if $\mathrm{rank}(A) = n$

- A has linearly independent columns
- must be tall or square

Full row rank: A has full row rank if $\mathrm{rank}(A) = m$

- A has linearly independent rows
- must be wide or square

Rank of product

let A be $m \times p$ matrix and B is $p \times n$ matrix, then,

$$\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$$

Proof

- we have $\mathrm{range}(AB) \subseteq \mathrm{range}(A)$, so

$$\mathrm{rank}(AB) \leq \mathrm{rank}(A)$$

- using this, we also have

$$\mathrm{rank}(AB) = \mathrm{rank}(B^T A^T) \leq \mathrm{rank}(B^T) = \mathrm{rank}(B)$$

- hence, the inequality holds

Rank-r matrix in factored form

let A be an $m \times n$ matrix with rank r, then it can be factored as

$$A = BC$$

- B is $m \times r$ with linearly independent columns (full column rank)
- C is $r \times n$ with linearly independent rows (full row rank)

- this is called a *full-rank factorization* of A
- proof shown on next page
  - can also be shown using QR factorization or SVD (discussed later)
- we will see that the converse is true (shown on page 4.45)

Proof

let $A = [a_1 \ \ldots \ a_n]$ be an $m \times n$ matrix with $\mathrm{rank}(A) = r$

- then there are r linearly independent column vectors, denoted by $b_1, \ldots, b_r$
- so $a_i = Bc_i$ for some vector c_i where $B = [b_1 \ \cdots \ b_r]$ is an $m \times r$ matrix
- in matrix notation, $A = BC$ where $C = [c_1 \ \cdots \ c_n]$ is an $r \times n$ matrix
- observe that $r = \mathrm{rank}(A) = \mathrm{rank}(BC) \leq \mathrm{rank}(C) \leq r$
- so we must have $\mathrm{rank}(C) = r$

Low-rank matrix

an $m \times n$ matrix has *low rank* if

$$\mathrm{rank}(A) \ll \min\{m, n\}$$

if $r = \mathrm{rank}(A) \ll \min\{m, n\}$, a factorization

$$A = BC \quad \text{(with } B \in \mathbb{R}^{m\times r} \text{ and } C \in \mathbb{R}^{r\times n}\text{)}$$

gives an efficient representation of A

- memory: B and C have $r(m+n)$ entries, compared with mn for A
- fast matrix-vector product: computing $y = Ax$ as

  $$y = B(Cx)$$

  require $2r(m+n)$ flops compared with $2mn$ for general product $y = Ax$

Low-rank approximation

(approximate) low-rank representations

$$A \approx BC$$

are useful in many applications

Singular value decomposition (SVD)

finds the best approximation (in Frobenius norm or 2-norm) of a given rank

Heuristic algorithms

less expensive than SVD but not guaranteed to find a best approximation

Optimization algorithms

can handle certain constraints on B, C (for example, entries must be nonnegative)

Outline

- subspaces
- independence, basis, dimension
- rank of a matrix
- **inverses of a matrix**
- pseudo-inverse

Left and right inverse

let A be an $m \times n$ matrix

Left inverse: X is a *left inverse* of A if

$$XA = I$$

A is *left-invertible* if it has at least one left inverse

Right inverse: X is a *right inverse* of A if

$$AX = I$$

A is *right-invertible* if it has at least one right inverse

Immediate properties

- a left or right inverse of an $m \times n$ matrix must have size $n \times m$
- X is a left (right) inverse of A if and only if X^T is a right (left) inverse of A^T

Examples

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- A is left-invertible; the following matrices are left inverses:

$$\frac{1}{9}\begin{bmatrix} -11 & -10 & 16 \\ 7 & 8 & -11 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1/2 & 3 \\ 0 & 1/2 & -2 \end{bmatrix}$$

- B is right-invertible; the following matrices are right inverses:

$$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- for n-vector a ($n \times 1$ matrix), $x = (1/a_i)e_i^T$ is left inverse for any i with $a_i \neq 0$

Column and row independence

Left inverse: A is left-invertible iff $\mathrm{rank}(A) = n$ (*columns* are linearly independent)

- to see this: $CA = I$ then

$$Ax = 0 \Longrightarrow C(Ax) = (CA)x = x = 0$$

- the converse is also true (shown later)
- left-invertible matrices are tall or square (by dimension inequality)

Right inverse: A is right-invertible iff $\mathrm{rank}(A) = m$ (*rows* are linearly independent)

- A is right-invertible if and only if A^T is left-invertible:

$$AX = I \Longleftrightarrow (AX)^T = I \Longleftrightarrow X^T A^T = I$$

- hence, A is right-invertible if and only if its rows are linearly independent
- right-invertible matrices are wide or square

Inverse

if A has a left **and** a right inverse, then they are equal and unique:

$$XA = I, \quad AY = I \quad \Longrightarrow \quad X = X(AY) = (XA)Y = Y$$

- in this case, we call $X = Y$ the inverse of A, denoted A^{-1} ($AA^{-1} = A^{-1}A = I$)
- A is called *invertible* if its inverse exists
- invertible matrices must be square

Properties

- $(AB)^{-1} = B^{-1}A^{-1}$ (provided inverses exist)
- $(A^T)^{-1} = (A^{-1})^T$ (denoted A^{-T})
- negative matrix powers: $(A^{-1})^k = (A^k)^{-1}$ (denoted A^{-k})
- with $A^0 = I$, identity $A^kA^l = A^{k+l}$ holds for any integers k, l

Examples

- inverse of identity is simply the identity $I^{-1} = I$
- $A = \mathrm{diag}(a_1, \ldots, a_n)$ has inverse $A = \mathrm{diag}(\frac{1}{a_1}, \ldots, \frac{1}{a_n})$ if and only if $a_i \neq 0$
- 2×2 matrix A is invertible if and only $A_{11}A_{22} \neq A_{12}A_{21}$

$$A^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

- a non-obvious example:

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 2 \\ -3 & -4 & -4 \end{bmatrix}, \qquad A^{-1} = \frac{1}{30} \begin{bmatrix} 0 & -20 & -10 \\ -6 & 5 & -2 \\ 6 & 10 & 2 \end{bmatrix}$$

verified by checking $AA^{-1} = I$ or $A^{-1}A = I$

Nonsingular matrix

for a **square** matrix A the following properties are equivalent

1. A is left-invertible
2. the columns of A are linearly independent
   - implies $\mathrm{null}(A) = \{0\}$, and since A is square, $\mathrm{range}(A) = \mathbb{R}^n$
3. A is right-invertible
4. the rows of A are linearly independent

- a square matrix A satisfying the above is called *nonsingular* (≡ invertible)
  - these properties imply that A is invertible if and only if $\mathrm{rank}(A) = n$
- otherwise, the matrix is *singular* (≡ non-invertible)

Proof

we show 1 $\Rightarrow$ 2 $\Rightarrow$ 3 $\Rightarrow$ 4 $\Rightarrow$ 1

- if A is left-invertible with left inverse B, then

$$Ax = 0 \quad \Longrightarrow \quad BAx = 0 \quad \Longrightarrow \quad x = 0$$

  so the columns of A are linearly independent

- if columns of A are l.i., then they form a basis for $\mathbb{R}^n$ and there exist solutions to

$$Ax_1 = e_1, \ldots, Ax_n = e_n \quad \Longrightarrow \quad AX = I$$

  hence, A is right-invertible

- apply same argument to A^T to show if A is right-invertible then its rows are l.i.

- apply same argument to A^T to show if A has l.i. rows then A is left-invertible

Examples

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

- A is nonsingular because its columns are linearly independent:

$$x_1 - x_2 + x_3 = 0, \quad -x_1 + x_2 + x_3 = 0, \quad x_1 + x_2 - x_3 = 0$$

  is only possible if $x_1 = x_2 = x_3 = 0$

- B is singular because its columns are linearly dependent:

$$Bx = 0 \text{ for } x = \mathbf{1} = (1, 1, 1, 1)$$

Example: Vandermonde matrix

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix}$$

the Vandermonde matrix is nonsingular if $t_i \neq t_j$ for $i \neq j$

Proof

- $Ax = 0$ implies $p(t_1) = p(t_2) = \cdots = p(t_n) = 0$ where

  $$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

  $p(t)$ is a polynomial of degree $n-1$ or less

- for $x \neq 0$, $p(t)$ can not have more than $n-1$ distinct real roots

- therefore $p(t_1) = \cdots = p(t_n) = 0$ is only possible if $x = 0$

Linear equations and matrix inverses

Left inverse: if X is a left inverse of A, then

$$Ax = b \quad \Longrightarrow \quad x = XAx = Xb$$

- if there is a solution, it must be equal to Xb
- if $A(Xb) \neq b$, then there is no solution

Right inverse: if X is a right inverse of A, then

$$x = Xb \quad \Longrightarrow \quad Ax = AXb = b$$

- there is at least one solution: $x = Xb$ for any b
- there can be other solutions

Inverse: if A is invertible, then $x = A^{-1}b$ is the *unique* solution to $Ax = b$

Linear equations and rank

for $A \in \mathbb{R}^{m\times n}$, the linear equation $Ax = b$ has a solution if and only if

$$\operatorname{rank} A = \operatorname{rank}[A \;\; b]$$

- implies that $b \in \operatorname{range}(A)$ (system called *consistent*):

  b can be expressed as a linear combination of columns of A

- no solution exists if $b \notin \operatorname{range}(A)$ (inconsistent)

Cases for solution existence

- a solution exists for *any* b iff $\operatorname{rank} A = m$ (implies $\operatorname{range}(A) = \mathbb{R}^m$)
- *unique* solution if and only if $\operatorname{rank} A = \operatorname{rank}[A \; b] = n$
  - this implies $b \in \operatorname{range}(A)$ and the columns of A are lin. indep ($\operatorname{null}(A) = \{0\}$)
  - a unique solution exists for *any* b if and only if A is nonsingular
- infinitely many solutions if and only if $\operatorname{rank} A = \operatorname{rank}[A \; b] < n$
  - this implies $b \in \operatorname{range}(A)$ and the $\operatorname{null}(A)$ is nonempty
  - infinitely many solutions for *any* b if and only if $\operatorname{rank} A = m < n$

Example: polynomial interpolation

- let's find coefficients of a cubic polynomial

$$p(t) = c_1 + c_2 t + c_3 t^2 + c_4 t^3$$

  that satisfies

$$p(-1.1) = b_1, \quad p(-0.4) = b_2, \quad p(0.1) = b_3, \quad p(0.8) = b_4$$

- write as $Ac = b$, with

$$A = \begin{bmatrix} 1 & -1.1 & (-1.1)^2 & (-1.1)^3 \\ 1 & -0.4 & (-0.4)^2 & (-0.4)^3 \\ 1 & 0.1 & (0.1)^2 & (0.1)^3 \\ 1 & 0.8 & (0.8)^2 & (0.8)^3 \end{bmatrix}$$

- (unique) coefficients given by $c = A^{-1}b$, with

$$A^{-1} = \begin{bmatrix} -0.0370 & 0.3492 & 0.7521 & -0.0643 \\ 0.1388 & -1.8651 & 1.6239 & 0.1023 \\ 0.3470 & 0.1984 & -1.4957 & 0.9503 \\ -0.5784 & 1.9841 & -2.1368 & 0.7310 \end{bmatrix}$$

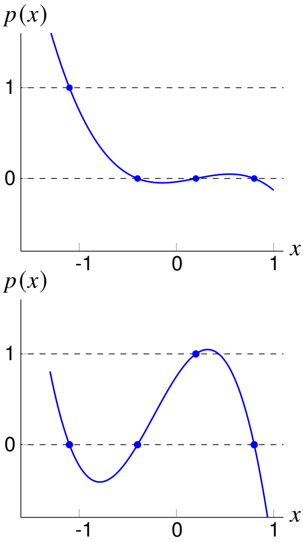
- so, *e.g.*, c_1 is not very sensitive to b_1 or b_4

- first column $A^{-1}e_1$ gives coefficients of polynomial that satisfies

$$p(-1.1) = 1, \quad p(-0.4) = 0, \quad p(0.1) = 0, \quad p(0.8) = 0$$

called (first) *Lagrange polynomial* associated with the point -1.1

Lagrange polynomials associated with points $-1.1, -0.4, 0.2, 0.8$



Outline

- subspaces
- independence, basis, dimension
- rank of a matrix
- inverses of a matrix
- **pseudo-inverse**

Gram matrix

the *Gram matrix* associated with $A = [a_1 \ \cdots \ a_m] \in \mathbb{R}^{m\times n}$ (with columns a_i) is

$$A^TA = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix}$$

Nonsingular Gram matrix: A^TA is nonsingular iff A has linearly indep. columns

- suppose A has linearly independent columns ($\mathrm{rank}(A) = n$):

  $$A^TAx = 0 \Longrightarrow x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 = 0 \Longrightarrow Ax = 0 \Longrightarrow x = 0$$

  thus A^TA is nonsingular

- assume columns of A are linearly dependent, then

  $$\text{there exists } x \neq 0,\ Ax = 0 \quad \Longrightarrow \quad A^TAx = 0$$

  therefore A^TA is singular

Pseudo-inverse of matrix with independent columns

- suppose $A \in \mathbb{R}^{m\times n}$ has linearly independent columns
- this implies that A is tall or square $(m \geq n)$

Pseudo-inverse

$$A^\dagger = (A^TA)^{-1}A^T$$

- this matrix exists, because the Gram matrix A^TA is nonsingular
- $A^\dagger$ is a left inverse of A:

$$A^\dagger A = (A^TA)^{-1}(A^TA) = I$$

- reduces to the inverse when A is square

Pseudo-inverse of matrix with independent rows

- suppose $A \in \mathbb{R}^{m \times n}$ has linearly independent rows
- this implies that A is wide or square ($m \leq n$)

Pseudo-inverse

$$A^\dagger = A^T(AA^T)^{-1}$$

- A^T has linearly independent columns
- hence its Gram matrix AA^T is nonsingular, so $A^\dagger$ exists
- $A^\dagger$ is a right inverse of A:

$$AA^\dagger = (AA^T)(AA^T)^{-1} = I$$

- reduces to the inverse when A is square

Summary

Left invertible: the following properties are equivalent for a real matrix $A \in \mathbb{R}^{m \times n}$

1. A is left-invertible
2. the columns of A are linearly independent (and $\mathrm{null}(A) = \{0\}$)
3. $A^T A$ is nonsingular

(1 $\Longrightarrow$ 2 from page 4.29, 2 $\Longleftrightarrow$ 3 from page 4.41, 3 $\Longrightarrow$ 1 since $A^\dagger$ is a left-inverse)

Right invertible: the following properties are equivalent for a real matrix A

1. A is right-invertible
2. the rows of A are linearly independent (and $\mathrm{range}(A) = \mathbb{R}^m$)
3. AA^T is nonsingular

Full rank factorization

suppose that

$$A = BC$$

- B is $m \times r$ with linearly independent columns (full column rank)
- C is $r \times n$ with linearly independent rows (full row rank)

then the matrix A has rank r

- $\mathrm{range}(A) \subseteq \mathrm{range}(B)$: each column of A is a linear comb. of columns of B
- $\mathrm{range}(B) \subseteq \mathrm{range}(A)$:

$$y = Bx \quad \Longrightarrow \quad y = B(CD)x = A(Dx)$$

  where D is a right inverse of C (for example, $D = C^\dagger$)

- therefore $\mathrm{range}(A) = \mathrm{range}(B)$ and $\mathrm{rank}(A) = \mathrm{rank}(B)$
- but the columns of B are linearly independent, so $\mathrm{rank}(A) = \mathrm{rank}(B) = r$

Pseudo-inverse of rank-deficient matrices

suppose A is $m \times n$ with rank r and full-rank factorization

$$A = BC$$

- B is $m \times r$ with linearly independent columns; its pseudo-inverse is defined as

$$B^\dagger = (B^T B)^{-1} B^T$$

- C is $r \times n$ with linearly independent rows; its pseudo-inverse is defined as

$$C^\dagger = C^T (C C^T)^{-1}$$

the **pseudo-inverse** of A is defined as

$$A^\dagger = C^\dagger B^\dagger$$

- this extends the definition of pseudo-inverse to matrices that are not full rank
- $A^\dagger$ is also known as the *Moore-Penrose (generalized) inverse*

Uniqueness

$A^\dagger = C^\dagger B^\dagger$ is unique and does not depend on the particular factorization $A = BC$

- suppose $A = \tilde{B}\tilde{C}$ is another rank factorization
- the columns of B and $\tilde{B}$ are two bases for $\mathrm{range}(A)$; therefore

$$\tilde{B} = BM \quad \text{for some nonsingular } r \times r \text{ matrix } M$$

- hence $BC = \tilde{B}\tilde{C} = BM\tilde{C}$; multiplying with $B^\dagger$ on the left shows that $C = M\tilde{C}$
- the pseudo-inverses of $\tilde{B} = BM$ and $\tilde{C} = M^{-1}C$ are

$$\tilde{B}^\dagger = (\tilde{B}^T\tilde{B})^{-1}\tilde{B}^T = M^{-1}(B^TB)^{-1}B^T = M^{-1}B^\dagger$$

  and

$$\tilde{C}^\dagger = \tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1} = C^T(CC^T)^{-1}M = C^\dagger M$$

- we conclude that $\tilde{C}^\dagger\tilde{B}^\dagger = C^\dagger M M^{-1} B^\dagger = C^\dagger B^\dagger$

Example: pseudo-inverse of diagonal matrix

- the rank of a diagonal matrix A is the number of nonzero diagonal elements
- pseudo-inverse $A^\dagger$ is the diagonal matrix with

$$(A^\dagger)_{ii} = \begin{cases} 1/A_{ii} & \text{if } A_{ii} \neq 0 \\ 0 & \text{if } A_{ii} = 0 \end{cases}$$

Example

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}, \quad A^\dagger = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 \end{bmatrix}$$

this follows, for example, from the factorization $A = BC$ with

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

References and further readings

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.* Cambridge University Press, 2018.
- L. Vandenberghe, *EE133A Lecture Notes*, University of California, Los Angeles.
- L. Vandenberghe. *EE133B Lecture Notes*, University of California, Los Angeles.
- M. T. Heath. *Scientific Computing: An Introductory Survey* (revised second edition). Society for Industrial and Applied Mathematics, 2018.