6. Matrix inverses

- linear independence
- left and right inverse
- pseudo-inverse
- matrices with orthonormal columns
- condition of linear systems

Linearly independent vectors

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

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

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

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

- list $a_1, \ldots, a_n$ 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

$$a_1 = \begin{bmatrix} 0.2 \\ -7 \\ 8.6 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -0.1 \\ 2 \\ -1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ -1 \\ 2.2 \end{bmatrix}$$

are linearly dependent

- 0 can be expressed as a nontrivial linear combination of a_1, a_2, a_3:

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

- a_1 can be expressed as a linear combination of a_2, a_3:

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

  (and similarly a_2 and a_3)

Example

the vectors

$$a_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

are linearly independent:

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

holds only if $x_1 = x_2 = x_3 = 0$

Independent dimension inequality

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

$$n \leq m$$

- if an $m \times n$ matrix has linearly independent columns then $m \geq n$
- if A is wide ($n > m$), the 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 ($m > n$), its rows are linearly dependent

Basis

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

- any n-vector b 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

- 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
- we see 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

Outline

- linear independence
- **left and right inverse**
- pseudo-inverse
- matrices with orthonormal columns
- condition of linear systems

Left and right inverse

for $m \times n$ matrix A we distinguish to types of inverses

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 $a_i \neq 0$

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}
- A is *invertible* if its inverse exists
- invertible matrices must be square

Properties

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

Column and row independence

Left inverse: a matrix is left-invertible iff its **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 indep.-dimension inequality)

Right inverse: A is right-invertible iff its **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

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
3. A is right-invertible
4. the rows of A are linearly independent

a square matrix A satisfying the above is called *nonsingular* ($\equiv$ 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

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix} \quad \text{with } t_i \neq t_j \text{ for } i \neq j$$

the Vandermonde matrix is nonsingular

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 $Ax = b$

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

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

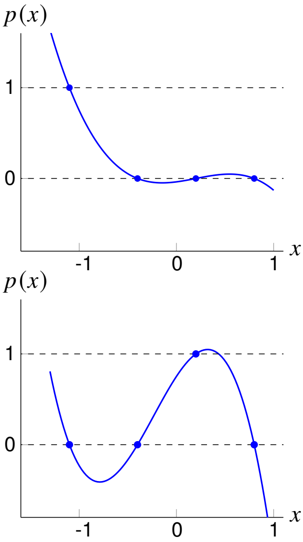
- so, *e.g.*, c_1 is not very sensitive to b_1 or b_4

- 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

- linear independence
- left and right inverse
- **pseudo-inverse**
- matrices with orthonormal columns
- condition of linear systems

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:

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

  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

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

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

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

- 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

1. A is left-invertible
2. the columns of A are linearly independent
3. A^TA is nonsingular

(1 $\Longrightarrow$ 2 from page 6.13, 2 $\Leftrightarrow$ 3 from page 6.22, 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
3. AA^T is nonsingular

Outline

- linear independence
- left and right inverse
- pseudo-inverse
- **matrices with orthonormal columns**
- condition of linear systems

Orthonormal vectors

a set of real m-vectors $a_1, a_2, \ldots, a_n$ is *orthonormal* if

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- the vectors have unit norm: $\|a_i\| = 1$ (normalized)
- they are mutually orthogonal: $a_i^T a_j = 0$ if $i \neq j$

Examples

- standard unit n-vectors $e_1, \ldots, e_n$
- the three vectors

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

Orthonormal expansion

if $a_1, \ldots, a_n$ are orthonormal, then they are lin. indep., hence basis for $\mathbb{R}^n$

- therefore, for any n-vector x,

$$x = \beta_1 a_1 + \cdots + \beta_n a_n \quad \text{for some unique } \beta_i$$

  this is called *orthonormal expansion* of x (in the orthonormal basis)

- multiplying by a_i^T on left, we have $\beta_i = a_i^T x$ and hence

$$x = (a_1^T x) a_1 + \cdots + (a_n^T x) a_n$$

Matrix with orthonormal columns

$A \in \mathbb{R}^{m\times n}$ has orthonormal columns if its Gram matrix is the identity matrix:

$$A^TA = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

there is no standard short name for "matrix with orthonormal columns"

- A is left-invertible with left inverse A^T
- A has linearly independent columns: $Ax = 0 \Longrightarrow A^TAx = x = 0$
- A is tall or square: $m \geq n$
- if A is tall $m > n$, then A has no right inverse; in particular

$$AA^T \neq I$$

Matrix-vector product

if $A \in \mathbb{R}^{m\times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- preserves inner products:

$$(Ax)^T(Ay) = x^T A^T A y = x^T y$$

- preserves norms:

$$\|Ax\| = \left((Ax)^T(Ax)\right)^{1/2} = (x^T x)^{1/2} = \|x\|$$

- preserves distances: $\|Ax - Ay\| = \|x - y\|$

- preserves angles:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \angle(x, y)$$

Orthogonal matrix

a **square** real matrix with orthonormal columns is called *orthogonal*

Nonsingularity: if A is orthogonal, then

- A is invertible, with inverse A^T:

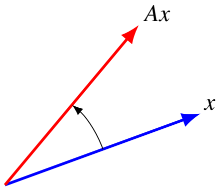
$$\left.\begin{array}{l} A^T A = I \\ A \text{ is square} \end{array}\right\} \quad \Longrightarrow \quad A A^T = I$$

- A^T is also an orthogonal matrix
- rows of A are orthonormal (have norm one and are mutually orthogonal)

Example: rotation in a plane

rotation matrices are orthogonal

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



Rotation in a coordinate plane in $\mathbb{R}^n$: for example,

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

describes a rotation in the (x_1, x_3) plane in $\mathbb{R}^3$

Example: permutation matrices

- permutation matrix is square with exactly one entry of each row/column is one
- let $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ be a *permutation* (reordering) of $(1, 2, \ldots, n)$
- permutation matrix A,

$$A_{i\pi_i} = 1, \quad A_{ij} = 0 \text{ if } j \neq \pi_i$$

  is orthogonal

- Ax is a permutation of the elements of x: $Ax = (x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n})$

Proof

- $A^T A = I$ because A has one element equal to one in each row and column

$$(A^T A)_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

- $A^T = A^{-1}$ is the inverse permutation matrix

Example: permutation on $\{1, 2, 3, 4\}$

$$(\pi_1, \pi_2, \pi_3, \pi_4) = (2, 4, 1, 3)$$

- corresponding permutation matrix and its inverse

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A^{-1} = A^T = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- A^T is permutation matrix associated with the permutation

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4) = (3, 1, 4, 2)$$

Product of orthogonal matrices

if $A_1, \ldots, A_k$ are orthogonal matrices and of equal size, then the product

$$A = A_1 A_2 \cdots A_k$$

is orthogonal:

$$\begin{aligned} A^T A &= (A_1 A_2 \cdots A_k)^T (A_1 A_2 \cdots A_k) \\ &= A_k^T \cdots A_2^T A_1^T A_1 A_2 \cdots A_k \\ &= I \end{aligned}$$

Linear equation with orthogonal matrix

linear equation with orthogonal coefficient matrix A of size $n \times n$

$$Ax = b$$

solution is

$$x = A^{-1}b = A^T b$$

- can be computed in $2n^2$ flops by matrix-vector multiplication
- cost is less than order n^2 if A has special properties; for example,

	order
permutation matrix:	0 flops
plane rotation:	1 flops

Outline

- linear independence
- left and right inverse
- pseudo-inverse
- matrices with orthonormal columns
- **condition of linear systems**

Matrix 2-norm

a matrix norm $\|\cdot\|$ is any function satisfying the properties

- nonnegative: $\|A\| \geq 0$ for all A
- positive definiteness: $\|A\| = 0$ only if $A = 0$
- homogeneity: $\|\beta A\| = |\beta|\|A\|$
- triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$

the **2-norm** or **spectral norm** is defined as

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

- the norms $\|Ax\|$ and $\|x\|$ are Euclidean norms of vectors
- no simple explicit expression, except for special A
- in MATLAB: `norm(A)`

Special cases

sometimes it is easy to maximize $\|Ax\|/\|x\|$

- zero matrix: $\|0\|_2 = 0$
- identity matrix: $\|I\|_2 = 1$
- diagonal matrix:

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix}, \quad \|A\|_2 = \max_{i=1,\ldots,n} |A_{ii}|$$

- matrix with orthonormal columns: $\|A\|_2 = 1$

General matrices: $\|A\|_2$ must be computed by numerical algorithms

Additional properties satisfied by the 2-norm

- $\|Ax\| \leq \|A\|_2\|x\|$ if the product Ax exists
- $\|AB\|_2 \leq \|A\|_2\|B\|_2$ if the product AB exists
- if A is nonsingular: $\|A\|_2\|A^{-1}\|_2 \geq 1$
- if A is nonsingular: $1/\|A^{-1}\|_2 = \min_{x\neq 0} (\|Ax\|_2/\|x\|)$
- $\|A^T\|_2 = \|A\|_2$

Other matrix norms

the *infinity-norm* is the maximum absolute row sum:

$$\|A\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$$

the *1-norm* is the maximum absolute column sum:

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$$

Example

$$A = \begin{bmatrix} 1 & 3 & 7 \\ -4 & 1.2725 & -2 \end{bmatrix}$$

we have

$$\|A\|_\infty = \max\{11, 7.2725\} = 11$$
$$\|A\|_1 = \max\{5, 4.2725, 9\} = 9$$

Condition of a set of linear equations

- assume A is nonsingular and $Ax = b$
- if we change b to $b + \Delta b$, the new solution is $x + \Delta x$ with

$$A(x + \Delta x) = b + \Delta b$$

- the change in x is

$$\Delta x = A^{-1} \Delta b$$

Condition

- well-conditioned if small Δb results in small Δx
- ill-conditioned if small Δb can result in large Δx

Example of ill-conditioned equations

$$A = \frac{1}{2}\left[\begin{array}{cc} 1 & 1 \\ 1+10^{-10} & 1-10^{-10} \end{array}\right], \quad A^{-1} = \left[\begin{array}{cc} 1-10^{10} & 10^{10} \\ 1+10^{10} & -10^{10} \end{array}\right]$$

- solution for $b = (1,1)$ is $x = (1,1)$
- change in x if we change b to $b + \Delta b$:

$$\Delta x = A^{-1}\Delta b = \left[\begin{array}{c} \Delta b_1 - 10^{10}\,(\Delta b_1 - \Delta b_2) \\ \Delta b_1 + 10^{10}\,(\Delta b_1 - \Delta b_2) \end{array}\right]$$

small Δb can lead to very large Δx

Bound on absolute error

suppose A is nonsingular and define

$$x = A^{-1}b, \quad \Delta x = A^{-1}\Delta b$$

Upper bound on $\|\Delta x\|$:

$$\|\Delta x\| \leq \|A^{-1}\|_2 \|\Delta b\|$$

- small $\|A^{-1}\|_2$ means that $\|\Delta x\|$ is small when $\|\Delta b\|$ is small
- large $\|A^{-1}\|_2$ means that $\|\Delta x\|$ can be large, even when $\|\Delta b\|$ is small
- for every A, there exists nonzero Δb such that $\|\Delta x\| = \|A^{-1}\|_2 \|\Delta b\|$

Bound on relative error

suppose in addition that $b \neq 0$; hence $x \neq 0$

Upper bound on $\|\Delta x\|/\|x\|$:

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\|_2 \|A^{-1}\|_2 \frac{\|\Delta b\|}{\|b\|}$$

- follows from $\|\Delta x\| \leq \|A^{-1}\|_2 \|\Delta b\|$ and $\|b\| \leq \|A\|_2 \|x\|$
- $\|A\|_2 \|A^{-1}\|_2$ small means $\|\Delta x\|/\|x\|$ is small when $\|\Delta b\|/\|b\|$ is small
- $\|A\|_2 \|A^{-1}\|_2$ large means $\|\Delta x\|/\|x\|$ can be much larger than $\|\Delta b\|/\|b\|$
- for every A, there exist nonzero $b, \Delta b$ such that equality holds

Condition number

the *condition number* of a nonsingular matrix A is

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

- we have $1 = \|I\| = \|A^{-1}A\| \leq \kappa(A)$
- condition number is a measure of how close a matrix is to being singular
- matrix is ideally conditioned if its condition number equals 1
- A is a well-conditioned matrix if $\kappa(A)$ is small (close to 1):
  the relative error in x is not much larger than the relative error in b
- A is badly conditioned or ill-conditioned if $\kappa(A)$ is large (nearly singular):
  the relative error in x can be much larger than the relative error in b

Example

- A is blurring matrix, nonsingular with condition number $\approx 10^9$
- we apply A to image x

blurred image

$y_1 = Ax$

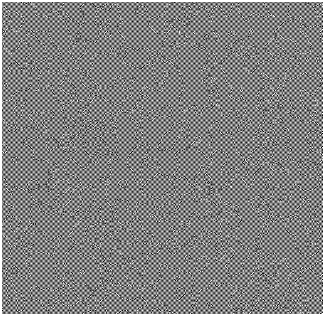
blurred and noisy image

$y_2 = Ax$+ small noise

Example

we solve $Ax = y$ for the two blurred images

$A^{-1}y_1$



$A^{-1}y_2$

- illustrates ill-conditioning of A (nearly singular)
- inverse amplifies the noise component

Residual and condition number

$$A(x + \Delta x) = b + \Delta b$$

- let $\hat{x}$ be an estimate solution of $Ax = b$
- residual $\hat{r} = b - A\hat{x}$; zero residual mean we get exact solution
- let $\Delta x = \hat{x} - x$ so $\hat{x} = x + \Delta x$
- we have

$$\Delta b = A(x + \Delta x) - b = A\hat{x} - b = -\hat{r}$$

- hence from before

$$\frac{\|\Delta x\|}{\|x\|} \le \kappa(A) \frac{\|\hat{r}\|}{\|b\|}$$

- error can be much larger than residual when condition number is large
- a small residual does not imply a small error in the solution

Example

$$A = \begin{bmatrix} 0.913 & 0.659 \\ 0.457 & 0.330 \end{bmatrix}, \quad b = \begin{bmatrix} 0.254 \\ 0.127 \end{bmatrix}$$

- consider two approximate solutions

$$\hat{x}_1 = \begin{bmatrix} 0.6391 \\ -0.5 \end{bmatrix} \quad \text{and} \quad \hat{x}_2 = \begin{bmatrix} 0.999 \\ -1.001 \end{bmatrix}$$

  the norms of their respective residuals are

$$\|\hat{r}_1\| = 6.8721 \times 10^{-5} \quad \text{and} \quad \|\hat{r}_2\| = 1.8 \times 10^{-3}$$

- $\hat{x}_1$ has smaller residual but solution is $(1, -1)$, so $\hat{x}_2$ is more accurate
- this is due to A being ill-conditioned
- in practice we cannot expect to deliver much more than a small residual

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,* Univ. of California, Los Angeles. (`http://www.seas.ucla.edu/~vandenbe/ee133a.html`)
- M. T. Heath. *Scientific Computing: An Introductory Survey* (revised second edition). Society for Industrial and Applied Mathematics, 2018.