

LINEAR ALGEBRA NOTES

4/29

Nándor Sieben

CONTENTS

1. SYSTEMS OF LINEAR EQUATIONS
2. MATRICES OF A SYSTEM
3. GAUSS ELIMINATION
4. MATRICES
5. MATRIX OPERATIONS
6. INVERSE MATRIX
7. DETERMINANTS
8. VECTOR SPACES
9. LINEAR INDEPENDENCE
10. BASES
11. ROW, COLUMN AND NULL SPACES
12. COORDINATES
13. LINEAR TRANSFORMATIONS
14. EIGENVALUES AND EIGENVECTORS
15. DIAGONALIZATION
16. BILINEAR FUNCTIONAL
17. INNER PRODUCT
18. ORTHOGONAL BASES AND GRAM-SCHMIDT ALGORITHM
19. LEAST SQUARE SOLUTION AND LINEAR REGRESSION

1. SYSTEMS OF LINEAR EQUATIONS

linear equation: $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$

variables: x_1, \dots, x_n

coefficients: a_1, \dots, a_n

main coefficient: a_1

constant term: b

linear system: m equations, n unknowns

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + \cdots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

solution: n -tuple (x_1, \dots, x_n) satisfying all equations

consistent system: has a solution

inconsistent system: has no solution

solution set: set of all solutions

equivalent systems: have the same solution set

elementary operations on equations: make equivalent systems

(i) multiply an equation by a nonzero constant

(ii) interchange two equations

(iii) add a constant multiple of an equation to another

elimination: use elementary operations to eliminate unknowns

fact: a linear system has no solution, exactly one solution or infinitely many solutions

parameters: used to describe infinitely many solutions

homogeneous system: constant terms are 0 (consistent)

trivial solution: all variables are 0

2. MATRICES OF A SYSTEM

coefficient matrix:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \\ a_{m1} & & a_{mn} \end{bmatrix}$$

constant vector: $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ **unknown vector:** $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

augmented matrix:

$$[A \quad b] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & & \\ a_{m1} & & a_{mn} & b_m \end{bmatrix}$$

3. GAUSS ELIMINATION

elementary row operations: (ero) correspond to elementary operations on equations

- (i) multiply a row by a nonzero constant $r_i \leftarrow cr_i$
- (ii) interchange two rows $r_i \leftrightarrow r_j$
- (iii) add a multiple of a row to another row $r_i \leftarrow r_i + cr_j$

row equivalent matrices: one can be gotten from the other by elementary row operations

fact: linear systems with row equivalent augmented matrices have the same solution set

echelon matrix: the number of leading zeros is strictly increasing in each row until you get all 0 rows

Gauss elimination: use elementary row operations to get echelon form

leading entry: first nonzero entry in a row

leading (pivot) column: column containing a leading entry

leading variable: a variable corresponding to a leading column

free variable: not leading

back substitution: get solution set from echelon form

- (i) set free variables equal to parameters
- (ii) solve last nonzero equation for leading variable
- (iii) substitute into preceding equation
- (iv) continue

reduced echelon matrix:

- (i) echelon matrix
- (ii) every leading entry is 1
- (iii) every leading entry is the only nonzero entry in its column

Gauss-Jordan elimination: use elementary row operations to get reduced echelon form

fact: every matrix is row equivalent to a unique reduced echelon matrix

fact: system with square coefficient matrix A has unique solution iff A is row equivalent to I

fact: system with more unknowns than equations is inconsistent or has infinitely many solutions

4. MATRICES

matrix: rectangular array of numbers

notation: $A = [a_{ij}]$

scalar: real number

size of a matrix: $\text{size}(A) = m \times n$ if m rows and n columns

square matrix: $m = n$

diagonal matrix: $D = [d_{ij}]$ $d_{ij} = 0$ if $i \neq j$

zero matrix: O all entries o_{ij} are 0

identity matrix: $I = [\delta_{ij}]$ $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

(column) vector: has size $n \times 1$

row vector: has size $1 \times n$

n -tuple: $(a_1, \dots, a_n) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \neq [a_1 \ \cdots \ a_n]$ slightly abusive

\mathbf{R}^n : set of n -tuples, $\mathbf{R}^2 = \text{plane}$, $\mathbf{R}^3 = \text{space}$

$\mathbf{R}^{m \times n}$: set of $m \times n$ matrices, $\mathbf{R}^{n \times 1}$ is identified with \mathbf{R}^n

basic unit vectors: $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ (1 in j -th position), column vectors of $I = [e_1 \ \cdots \ e_n]$

column vectors: $A = [c_1 \ \cdots \ c_n]$

5. MATRIX OPERATIONS

matrix addition: $A + B = [a_{ij} + b_{ij}]$ if A, B have the same size

matrix subtraction: $A - B = [a_{ij} - b_{ij}]$

scalar multiplication: $cA = [ca_{ij}]$

negative matrix: $-A = (-1)A$

properties:

$A + B = B + A$ commutative

$A + (B + C) = (A + B) + C$ associative

$c(A + B) = cA + cB$ distributive

$(c + d)A = cA + dA$ distributive

$(cd)A = c(dA)$ associative

matrix multiplication: $C = AB$, $\text{size}(C) = m \times n$, $\text{size}(A) = m \times p$, $\text{size}(B) = p \times n$

$c_{ij} = \sum_{k=1}^p a_{ik}b_{kj} = (i\text{-th row of } A) \cdot (j\text{-th column of } B)$

properties:

$A(BC) = (AB)C$ associative

$A(B + C) = AB + AC$ distributive

$(A + B)C = AC + BC$ distributive

$c(AB) = (cA)B = A(cB)$

warning:

$AB \neq BA$ in general

$AB = AC \not\Rightarrow B = C$

$AB = O \not\Rightarrow A = O$ or $B = O$

transpose: $A^T = [b_{ij}]$ where $b_{ij} = a_{ji}$

properties:

$(A^T)^T = A$

$(A + B)^T = A^T + B^T$

$(cA)^T = cA^T$

$(AB)^T = B^T A^T$

fact: product of diagonal matrices is diagonal

matrix form of linear system: $Ax = b$, $A = [a_{ij}]$, $x = (x_1, \dots, x_n)$, $b = (b_1, \dots, b_n)$

linear combination: of objects v_i is a finite sum of scalar multiples of the objects $\sum_{i=1}^n c_i v_i$, $c_i \in \mathbf{R}$

span: of objects v_i is the set of linear combinations of the objects $\text{span}\{v_1, \dots, v_n\} = \{\sum_{i=1}^n c_i v_i \mid c_i \in \mathbf{R}\}$

fact: solution set of homogeneous system is the span of particular solutions (one for each parameter)

6. INVERSE MATRIX

A invertible: $\exists B$ such that $AB = BA = I$

B is the **inverse** of A (A is also the inverse of B)

properties:

invertible \Rightarrow square

inverse is unique if exists, notation A^{-1}

$(A^{-1})^{-1} = A$

$(AB)^{-1} = B^{-1}A^{-1}$

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

if A is invertible then $Ax = b$ has unique solution $x = A^{-1}b$

fact: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible iff $ad \neq bc$, $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

elementary matrix: $I \xrightarrow{\text{ero}} E$ single elementary row operation

properties:

$$I \xrightarrow{\text{ero}} E \text{ implies } A \xrightarrow{\text{ero}} EA \text{ equivalently } [I \quad A] \xrightarrow{\text{ero}} [E \quad EA]$$

$$I \xrightarrow{\text{ero}} E^{-1} \text{ inverse ero}$$

fact: A invertible iff A row equivalent to I

fact: A, B row equivalent iff $A = E_1 \cdots E_n B$, for E_i elementary

algorithm for A^{-1} : $[A \quad I] \xrightarrow{\text{ero's}} [I \quad A^{-1}]$

more generally $[A \quad B] \xrightarrow{\text{ero's}} [I \quad A^{-1}B]$

7. DETERMINANTS

notation: $A = [a_{ij}] \ n \times n$

1×1 matrix: $\det [a] = a$

2×2 matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

notation: A_{ij} = submatrix after deletion of i -th row and j -th column

ij -th cofactor of A : $C_{ij} = (-1)^{i+j} \det A_{ij}$

chess board rule: $\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} (-1)^{i+j}$

inductive definition: $\det A = \sum_{j=1}^n a_{1j} C_{1j}$
 cofactor expansion along first row

cofactor expansion:

along i -th row $\det A = \sum_{j=1}^n a_{ij} C_{ij}$

along j -th column $\det A = \sum_{i=1}^n a_{ij} C_{ij}$

elementary row operations: $A \xrightarrow{\text{ero}} B$

$r_i \leftarrow cr_i$: $\det B = c \cdot \det A$

$r_i \leftrightarrow r_j$: $\det B = -\det A$

$r_i \leftarrow r_i + cr_j$: $\det B = \det A$

properties:

A triangular implies $\det(A) = a_{11} \cdots a_{nn}$

$\det I = 1$

$r_i = r_j$ implies $\det A = 0$

$\det kA = k^{\text{size}(A)} \det A$

$\det A^T = \det A$

$\det(AB) = \det A \cdot \det B$

$\det(A^{-1}) = \frac{1}{\det(A)}$

A invertible iff $\det A \neq 0$

Cramer's rule: $\det A \neq 0$ implies solution of $Ax = b$ is

$$x_i = \frac{\det A_i}{\det A} \text{ where } A_i \text{ comes from } A \text{ after replacing } i\text{-th column by } b$$

adjoint of A : $\text{adj} A = [C_{ij}]^T$ transpose of matrix of cofactors

adjoint formula for inverse: $A^{-1} = \frac{\text{adj} A}{\det A}$

8. VECTOR SPACES

vector space: set V of vectors with vector addition and scalar multiplication satisfying
for all $u, v, w \in V$ and $c, d \in \mathbf{R}$

- i) $u + v = v + u$
- ii) $(u + v) + w = u + (v + w)$
- iii) $\exists \underline{0} \in V, u + \underline{0} = u$
- iv) $\exists -u \in V, u + (-u) = \underline{0}$
- v) $c(u + v) = cu + cv$
- vi) $(c + d)u = cu + du$
- vii) $c(du) = (cd)u$
- viii) $1u = u$

examples: \mathbf{R}^n , $\mathbf{R}^{m \times n}$, \mathbf{P} polynomials, \mathbf{P}_n polynomials with degree less than n , sequences, sequences converging to 0, functions on \mathbf{R} , $C(\mathbf{R})$ continuous functions on \mathbf{R} , solutions of homogeneous systems

subspace of V : subset W of V that is a vector space with same operations

proper subspace of V : subspace but not $\{\underline{0}\}$ and not V

examples:

- $W = \{0\}$ and $W = V$, subspaces of V
- $W =$ lines through origin, subspace of $V = \mathbf{R}^2$
- $W =$ planes through origin, subspace of $V = \mathbf{R}^3$
- $W =$ diagonal $n \times n$ matrices, subspace of $V = \mathbf{R}^{n \times n}$
- $W = \text{span}\{v_1, \dots, v_n\}$, subspace of V where $v_1, \dots, v_n \in V$
- $W =$ convergent sequences, subspace of $V =$ sequences
- $W =$ continuous functions on \mathbf{R} , subspace of $V =$ functions on \mathbf{R}

fact: subset W of V is a subspace of V iff

- nonempty:** $W \neq \emptyset$
- closed under addition:** $\forall u, v \in W, u + v \in W$
- closed under scalar multiplication:** $\forall c \in \mathbf{R} \forall u \in W, cu \in W$

9. LINEAR INDEPENDENCE

v_1, \dots, v_n **linearly independent:** $\sum_{i=1}^n c_i v_i = \underline{0}$ implies $\forall i, c_i = 0$

linearly dependent: not independent

parallel vectors: one is scalar multiple of the other

notation $u \parallel v$

properties:

- u, v linearly independent iff $u \parallel v$
- vectors are dependent iff one of them is linear combination of the others
- subset of linearly independent set is linearly independent
- columns of matrix A are independent iff $AX = 0$ has only trivial solution
- columns of square matrix A are independent iff A invertible iff $\det A \neq 0$
- v_1, \dots, v_n independent, $v_{n+1} \notin \text{span}\{v_1, \dots, v_n\}$ implies v_1, \dots, v_{n+1} independent
- v_1, \dots, v_n independent, $\sum_{i=1}^n c_i v_i = \sum_{i=1}^n d_i v_i$ implies $\forall i, c_i = d_i$
- rows of row echelon matrix are independent
- leading columns of echelon matrix are independent

10. BASES

S spans W : $\text{span}S = W$

S is a spanning set of W

basis of V : linearly independent spanning set of V

maximal independent set in V

minimal spanning set of V

standard bases $E = \{e_1, \dots, e_n\}$ for V :

$\{(1, 0), (0, 1)\}$ for \mathbf{R}^2

$\{1, x, x^2\}$ for \mathbf{P}_3

$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ for $\mathbf{R}^{2 \times 2}$

properties:

$\text{span}S = V, T \subseteq V, |T| > |S|$ implies T dependent

all bases of V has same number of vectors

dimension of V : $\dim V = \text{number of vectors in a basis of } V$

examples:

$\dim \mathbf{R}^n = n$

$\dim \{0\} = 0$

$\dim \mathbf{R}^{m \times n} = mn$

$\dim \mathbf{P}_n = n$

$\dim \mathbf{P} = \infty$

$\dim(\text{span}\{u\}) = 1$

properties:

W proper subspace of V implies $\dim W < \dim V$

independent subset of V can be extended to a basis of V

spanning set of V contains a basis of V

11. ROW, COLUMN AND NULL SPACES

notation: $\text{size}A = m \times n$

row space of A : $\text{Row}A = \text{subspace of } \mathbf{R}^m \text{ spanned by rows of } A$

row rank of A : $\dim \text{Row}A$

column space of A : $\text{Col}A = \text{subspace of } \mathbf{R}^n \text{ spanned by columns of } A$

column rank of A : $\dim \text{Col}A$

algorithm for basis of $\text{Row}A$:

(i) reduce A to echelon form B

(ii) take nonzero row vectors of B

algorithm for basis of $\text{Col}A$:

(i) reduce A to echelon form B

(ii) take columns of A corresponding to leading columns of B

fact: row rank A equals column rank A

rank A : this common value

null space of A : $\text{Null}A = \{x \mid Ax = 0\} = \text{solution set of homogeneous system, subspace of } \mathbf{R}^n$

properties:

A, B row equivalent implies $\text{Row}A = \text{Row}B$

A, B row equivalent implies columns of A and columns of B have the same dependence relations

$Ax = b$ consistent iff $b \in \text{Col}A$

$\text{rank}A + \dim \text{Null}A = n$

12. COORDINATES

notation: $B = \{b_1, \dots, b_n\}$, $D = \{d_1, \dots, d_n\}$ bases for V , $E = \{e_1, \dots, e_n\}$ standard basis for V

fact: each $v \in V$ can be written uniquely as $v = c_1b_1 + \dots + c_nb_n$

coordinates of v in basis B : $[v]_B = (c_1, \dots, c_n) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ if $v = \sum_{i=1}^n c_i b_i$

huge fact: $v \mapsto [v]_B : V \rightarrow \mathbf{R}^n$ is an isomorphism (\mathbf{R}^n are the 'only' finite dimensional vector spaces)

transition matrix from basis B to basis D : $T_B^D = [[b_1]_D \ \dots \ [b_n]_D]$ square matrix

properties:

$$\begin{aligned} [v]_D &= T_B^D [v]_B \\ T_B^D &= (T_D^B)^{-1} \\ T_B^D &= T_E^D T_B^E = (T_D^E)^{-1} T_B^E \\ [T_D^E \ T_B^E] &\overset{\text{cross}}{\mapsto} [I \ T_B^D] \end{aligned}$$

algorithm for finding a basis for $W = \text{span}\{v_1, \dots, v_n\}$ in V :

- (i) find a bases B for V (use standard if possible)
- (ii) put the coordinates of the v_i 's as rows (columns) for a matrix A
- (iii) find a basis for the rowspace (columnspace) of A
- (iv) use this basis as coordinates to build the basis of W

13. LINEAR TRANSFORMATIONS

notation: $B = \{b_1, \dots, b_m\}$ basis for V , $D = \{d_1, \dots, d_n\}$ basis for W , E standard basis for V

linear transformation: $L : V \rightarrow W$ such that for all $u, v \in V$, $\alpha \in \mathbf{R}$

$$L(u + v) = L(u) + L(v) \text{ additive}$$

$$L(\alpha u) = \alpha L(u) \text{ multiplicative}$$

kernel: $\ker L = \{v \in V \mid L(v) = \mathbf{0}\}$

range: $\text{ran} L = \{L(v) \mid v \in V\}$
 $\text{ran} L = \text{span}\{Lb_1, \dots, Lb_m\}$

L is one-to-one (1-1): $L(u) = L(v)$ implies $u = v$

L is onto W : $\text{ran} L = W$

properties:

$$L(\mathbf{0}) = \mathbf{0}$$

$\ker L$ subspace of V

$\text{ran} L$ subspace of W

L is 1-1 iff $\ker L = \{\mathbf{0}\}$

matrix of L : $[L]_B^D = [[Lb_1]_D \ \dots \ [Lb_m]_D]$

properties:

$$\begin{aligned} [L]_B^D &= T_E^D [L]_B^E = (T_D^E)^{-1} [L]_B^E \\ [L]_B^D &= (T_D^E)^{-1} [L]_E^E T_B^E \text{ if } V = W \\ [Lv]_D &= [L]_B^D [v]_B \end{aligned}$$

R, S are similar matrices: $S = P^{-1}RP$ for some P

fact: R, S are similar iff $R = [L]_B^B$, $S = [L]_D^D$ for some $L : V \rightarrow V$ and bases B, D for V
 (P is the transition matrix)

rank of L : $\text{rank} L = \dim \text{ran} L$

properties: $M = [L]_B^D$

$$[\text{ran} L]_D = \text{Col} M$$

$$[\ker L]_B = \text{Null} M$$

$$\text{rank} L = \text{rank} M$$

$$\dim \ker L = \dim \text{null} M$$

$$\text{rank} L + \dim \ker L = \dim V$$

14. EIGENVALUES AND EIGENVECTORS

notation: $L : V \rightarrow V$ linear transformation, $A = [L]_B^B$ matrix of L , $x = [u]_B$ coordinates of u

eigenvalue problem:

transformation version $L(u) = \lambda u$, $u \neq \underline{0}$

eigenvalue: λ

eigenvector of L associated to λ : u

eigenspace associated to λ : $\ker(L - \lambda \text{id})$

matrix version $Ax = \lambda x$, $x \neq \underline{0}$

eigenvalue: λ

eigenvector of A associated to λ : x

eigenspace associated to λ : $\text{Null}(A - \lambda I)$

characteristic polynomial: $\det(A - \lambda I)$

if $A \sim B$ then $\text{charpoly}(A) = \text{charpoly}(B)$

characteristic equation: λ eigenvalue of A iff $\det(A - \lambda I) = 0$

15. DIAGONALIZATION

A diagonalizable: A similar to diagonal matrix D , $D = P^{-1}AP$

fact: $D = P^{-1}AP$ implies

$$P = [v_1 \ \cdots \ v_n]$$

$$D = [d_{ij}], d_{ij} = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$$

$$Av_i = \lambda_i v_i$$

$\{v_1, \dots, v_n\}$ is a basis of eigenvectors with associated eigenvalues in the diagonal of D

properties:

if v_1, \dots, v_n eigenvectors associated to distinct eigenvalues then they are independent

if $\text{size} A = n \times n$ and A has n distinct eigenvalues then A diagonalizable

$\lambda_1, \dots, \lambda_n$ distinct eigenvalues, B_1, \dots, B_n bases for eigenspaces implies $B_1 \cup \dots \cup B_n$ is independent

algorithm for diagonalization:

(i) solve characteristic equation to find eigenvalues

(ii) for each eigenvalue find basis of associated eigenspace

(iii) if the union of the bases is not a basis for the vectorspace than not diagonalizable

(iv) build P from the eigenvectors as columns

(v) build D from the corresponding eigenvalues

16. BILINEAR FUNCTIONAL

product of U and V : $U \times V = \{(u, v) \mid u \in U, v \in V\}$

bilinear functional on V : $f : V \times V \rightarrow \mathbf{R}$ such that for all $u, v, w \in V$ and $\alpha, \beta \in \mathbf{R}$

$$f(\alpha u + \beta v, w) = \alpha f(u, w) + \beta f(v, w)$$

$$f(w, \alpha u + \beta v) = \alpha f(w, u) + \beta f(w, v)$$

fact: Every bilinear functional f on \mathbf{R}^n is $f(u, v) = u^T A v$ for some $A \in \mathbf{R}^{n \times n}$

where $a_{ij} = f(e_i, e_j)$

The bilinear functional f can be

symmetric: $f(u, v) = f(v, u)$ for all $u, v \in V$

positive semidefinite: $f(v, v) \geq 0$ for all $v \in V$

positive definite: $f(v, v) > 0$ for all $0 \neq v \in V$

negative semidefinite: $f(v, v) \leq 0$ for all $v \in V$

negative definite: $f(v, v) < 0$ for all $0 \neq v \in V$

indefinite: neither positive nor negative semidefinite

17. INNER PRODUCT

inner product: symmetric, positive definite, bilinear functional $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbf{R}$

examples of inner products:

dot product (standard inner product) on \mathbf{R}^n : $\langle u, v \rangle = u \cdot v = \sum_{i=1}^n u_i v_i = u^T v = u^T I v$

standard inner product on $C[0, 1]$: (continuous functions on $[0, 1]$), $\langle f, g \rangle := \int_0^1 f g$

fact: every inner product on \mathbf{R}^n is $\langle u, v \rangle = u^T A v$ where A is a symmetric (therefore diagonalizable) matrix with positive eigenvalues and $a_{ij} = \langle e_i, e_j \rangle$

length (norm): $\|v\| = \sqrt{\langle v, v \rangle}$

unit vector: $\|v\| = 1$

unit vector in the direction of v : $\frac{v}{\|v\|}$

distance: $d(u, v) = \|u - v\|$

angle: $\angle(u, v) = \arccos \frac{\langle u, v \rangle}{\|u\| \cdot \|v\|}$

orthogonal: $u \perp v$ iff $\angle(u, v) = \pi/2$ iff $\langle u, v \rangle = 0$

$S = \{v_1, \dots, v_n\}$ **orthogonal:** $v_i \perp v_j$ for all i, j

fact: nonzero orthogonal vectors are independent

$S = \{v_1, \dots, v_n\}$ **orthonormal:** S is orthogonal and $\|v_i\| = 1$ for all i

Cauchy-Schwartz inequality: $|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$

Triangle inequality:

$$\|u + v\| \leq \|u\| + \|v\|$$

$$d(u, v) \leq d(u, w) + d(w, v)$$

Pythagorean theorem: $u \perp v$ implies $\|u + v\| = \|u\| + \|v\|$

orthogonal complement: $W^\perp = \{v \in V \mid v \perp w \text{ for all } w \in W\}$, W is subspace of V

properties: W is subspace of \mathbf{R}^n

W^\perp is a subspace

$$W \cap W^\perp = \{0\}$$

$W = \text{span}(S)$, $u \perp s_i$ for all i implies $u \in W^\perp$

$$(\text{Row } A)^\perp = \text{Null } A$$

$$\dim W + \dim W^\perp = n$$

(basis of W) \cup (basis of W^\perp) is basis of \mathbf{R}^n

$$(W^\perp)^\perp = W$$

18. ORTHOGONAL BASES AND GRAM-SCHMIDT ALGORITHM

fact: $\{u_1, \dots, u_n\}$ orthogonal basis for a subspace W of V , $y \in V$

$$p = \sum_{i=1}^n \frac{\langle y, u_i \rangle}{\langle u_i, u_i \rangle} u_i \in W \text{ and } q = y - p \in W^\perp$$

if $y = \tilde{p} + \tilde{q}$ such that $\tilde{p} \in W$ and $\tilde{q} \in W^\perp$ then $p = \tilde{p}$ and $q = \tilde{q}$

orthogonal projection: $\text{proj}_W y$ = the unique $p \in W$ such that $q = y - p \in W^\perp$

Gram-Schmidt algorithm: for finding an orthogonal basis $\{u_1, \dots, u_n\}$ for $\text{span}\{v_1, \dots, v_n\}$

(i) make $\{v_1, \dots, v_n\}$ independent if necessary

(ii) let $u_1 = v_1$

(ii) inductively let $u_{i+1} = v_{i+1} - \text{proj}_{\text{span}\{u_1, \dots, u_i\}} v_{i+1} = v_{i+1} - \sum_{j=1}^i \frac{\langle v_{i+1}, u_j \rangle}{\langle u_j, u_j \rangle} u_j$

fact: $W = \text{Col}(A)$, $A\beta = \text{proj}_W y$ iff $A^T A\beta = A^T y$

19. LEAST SQUARE SOLUTION AND LINEAR REGRESSION

fact: if W subspace of V , $w \in W$, $y \in V$ then $\|y - w\|$ is minimum when $w = \text{proj}_W y$

fact: $W = \text{Col } A$, $\|y - A\beta\|$ is minimum iff $A^T A\beta = A^T y$

least square regression line $ax + b$: data $\{(x_i, y_i) \mid i = 1, \dots, n\}$

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \beta = \begin{pmatrix} b \\ a \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \beta \text{ makes } \|A\beta - y\| \text{ minimum, that is, } A^T A\beta = A^T y$$