Outline of Lectures

in Linear Algebra Math 320

Spring 2013

Shmuel Friedland
University of Illinois at Chicago

http://www.math.uic.edu/ ∼friedlan

Lectures updated during the course

Last update January 17, 2013

1 Main Topics of the Course

- SYSTEMS OF EQUATIONS
- VECTOR SPACES
- LINEAR TRANSFORMATIONS
- DETERMINANTS
- INNER PRODUCT SPACES
- EIGENVALUES
- JORDAN CANONICAL FORM-RUDIMENTS

Text: Jim Hefferon, *Linear Algebra*, and *Solutions*

Available for free download

ftp://joshua.smcvt.edu/pub/hefferon/book/book.pdf

ftp://joshua.smcvt.edu/pub/hefferon/book/jhanswer.pdf

Software: MatLab,Maple, Matematica.

2 Applications of Linear Algebra

- Engineering
- Biology
- Medicine
- Business
- Statistics
- Physics
- Mathematics
- Numerical Analysis

Reason: Many real world systems consist of many parts which interact linearly.

Analysis of such systems involves the notions and the tools from Linear Algebra.

3 Lecture 1

I. Systems of Linear Equations

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = b_m \end{array}$$

a. Examples

b. Solutions: Unique, Many and None (Inconsistent).

c. Graphical Examples of Systems in Two Variables

d. Equivalent Systems (have same solutions):

- Change the order of the equations
- Multiply an equation by a nonzero number
- Add (subtract) from one equation a multiple of another equation

Examples

$$\begin{aligned} x_1 &+ 2x_2 &= 5 \\ 2x_1 &+ 3x_2 &= 8 \end{aligned}$$

Subtract 2 times row from row 2.

Hefferon notation: $\rho_2 - 2\rho_1 \to \rho_2$

My notations:

$R_2 - 2R_1 \to R_2$,

$R_2 \leftarrow R_2 - 2R_1$,

$R_2 \to R_2 - 2R_1$

Obtain a new system

$$\begin{aligned} x_1 &+ 2x_2 &= 5 \\ &- x_2 &= -2 \end{aligned}$$

Find first the solution of the second equation: $x_2 = 2$.

Substitute x_2 to the first equation:

$x_1 + 2 \times 2 = 5 \Rightarrow x_1 = 5 - 4 = 1$.

Unique solution $(x_1, x_2) = (1, 2)$

e. Triangular Systems and their solutions

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\ & & & & \ldots & & a_{nn}x_n & = & b_n \end{array}$$

n equations in n unknowns with n pivots:

$a_{11} \neq 0,\ a_{22} \neq 0, \ldots a_{nn} \neq 0.$

Solve the system by back substitution from down to up:

$$x_n = \frac{b_n}{a_{nn}},$$

$$x_{n-1} = \frac{-a_{(n-1)n}x_n + b_{n-1}}{a_{(n-1)(n-1)}},$$

$$x_i = \frac{-a_{i(i+1)}x_{i+1} - \ldots - a_{in}x_n + b_i}{a_{ii}},$$

$i = n-2, \ldots, 1.$

4 Lecture 1

II. Matrix Formalism for Solving Linear Equations

a. The Coefficient Matrix of the system:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

b. The Augmented Matrix $(A|\mathrm{b}),\ (A|B)$

$$(A|b) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & | & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & | & b_2 \\ \vdots & \vdots & \vdots & \vdots & | & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & | & b_m \end{pmatrix}$$

5 Lecture 2

c. Elementary Row Operations (ERO)

- Interchange two rows

$$R_i \longleftrightarrow R_j, \quad i \neq j$$

Example: $R_2 \longleftrightarrow R_j$

- Multiply a row by a nonzero number

$$a \times R_i \longrightarrow R_i, \; a \neq 0, \quad (R_i \longrightarrow a \times R_i).$$

- Replace a row by its sum with a multiple of another row

$$R_i + a \times R_j \longrightarrow R_i, \quad (R_i \longrightarrow R_i + a \times R_j).$$

Example:

$$R_2 - 0.7R_4 \longrightarrow R_2, \quad (R_2 \longrightarrow R_2 - 0.7R_4).$$

d. Pivotal Row

e. The elementary row operations are reversible: If D is obtained from C using elementary row operations then C is obtained from D using (the inverse elementary) row operations

6 Inverse elementary row operation

$R_i \longleftrightarrow R_j, \quad i \neq j$ is inverse to itself

$\frac{1}{a} \times R_i \longrightarrow R_i, \; a \neq 0$

is the inverse of $a \times R_i \longrightarrow R_i$

$R_i - a \times R_j \longrightarrow R_i$

is the inverse of $R_i + a \times R_j \longrightarrow R_i$

Denote by E^{-1} the inverse elementary row operation

Assume that D was obtained from C by using the following sequence of k elementary row operations:
$E_k E_{k-1} \ldots E_2 E_1$

Then C is obtained from D by the elementary operations
$E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1}$

Elementary row operations on the system of linear equations performed on augmented matrices give rise to the equivalent system of equations

Two systems of linear equations are equivalent if they have the same solutions

Row Echelon Form of a matrix.

- The first nonzero entry in each row is $\mathbf{1}$. This entry is called a pivot.
- If row k does not consists entirely of zeros, then the number of leading zero entries in row $k+1$ is greater then the number of leading zeros in row k.
- Zero rows appear below the rows having nonzero entries.

The process of using ERO to transform a linear system into one whose augmented matrix is in row echelon form is called Gaussian Elimination.

Corollary. *The given system is inconsistent if and only if the* REF *of its augmented matrix contains a row of the form*:

$$[\mathbf{0}\ \mathbf{0}\ \ \ldots \mathbf{0}|\ \mathbf{1}] \tag{6.1}$$

Constructive proof of existence of REF $A = [a_{ij}]_{i=j=1}^{m,n}$

0. If $A = 0$ (Zero matrix) A in REF done! Assume $A \neq 0$.

1. If $a_{11} \neq 0$:

a. divide the first row by a_{11}: $\frac{1}{a_{11}} R_1 \rightarrow R_1$ to obtain $A_1 = [a_{ij}^{(1)}]$.

Note: $a_{11}^1 = 1$.

b. Subtract $a_{i1}^{(1)}$ times row 1 from row $i \geq 2$: $-a_{i1} R_1 + R_i \rightarrow R_i$ for $i = 2, \ldots, m$

c. Put all zero rows to be the last rows

d. GO TO

2. If $a_{11} = \ldots = a_{(i-1)0} = 0$ and $a_{i1} \neq 0$ for some $1 < i \leq m$: $R_1 \leftrightarrow R_i$

GO TO 1.

3. Suppose that the first $k - 1$ columns of A are zero, but not $k - th$ row.

So $A = [0_{m \times k-1}\ B$, where B obtained from A by removing first $k - 1$ zero rows

REF of $\boldsymbol{A}$ is $\boldsymbol{C} = [\mathbf{0}_{m\times k-1}\ \boldsymbol{C}']$, $\boldsymbol{C}'$ REF of $\boldsymbol{B}$.

Replace $\boldsymbol{A}$ by $\boldsymbol{B}$ and GO TO $\mathbf{1}$.

Examples of REF

$$\begin{pmatrix} 1 & a & b & c \\ 0 & 1 & d & e \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & a & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Five possible REF of $(a\ b\ c\ d)$ (1×4 matrix):

$$\begin{array}{ll} (1\ u\ v\ w) & \text{if } a \neq 0, \\ (0\ 1\ p\ q) & \text{if } a = 0,\ b \neq 0, \\ (0\ 0\ 1\ r) & \text{if } a = b = 0,\ c \neq 0, \\ (0\ 0\ 0\ 1) & \text{if } a = b = c = 0,\ d \neq 0, \\ (0\ 0\ 0\ 0) & \text{if } a = b = c = d = 0. \end{array}$$

Overdetermined System m (number of equations) $> n$ (number of unknowns):

if there are more equations then unknowns.

Usually (but not always) overdetermined system are inconsistent.

Underdetermined System $m < n$:

if there are less equations then unknowns.

Usually (but not always) underdetermined system are solvable with many solutions.

7 Lecture 3

The general solution of the system in REF.

Assume that REF does not contain a row of the form (6.1):

$$[0\ 0\ \ldots 0|\ 1].$$

The variables associated with pivots are called lead variables. The rest of the variables are called free variables. The solution of the system is given by expressing each lead variable as a linear (affine) function of free variables.

Examples

$$\begin{pmatrix} 1 & -2 & 3 & -1 & | & 0 \\ 0 & 1 & 3 & 1 & | & 4 \\ 0 & 0 & 0 & 1 & | & 5 \end{pmatrix}$$

x_1, x_2, x_4 are lead variables, x_3 is a free variable.

$x_4 = 5,\ x_2+3x_3+x_4 = 4 \Rightarrow x_2 = -3x_3-x_4+4$

$x_2 = -3x_3 - 1,\ x_1 - 2x_2 + 3x_3 + -x_4 = 0 \Rightarrow$

$x_1 = 2x_2-3x_3+x_4 = 2(-3x_3-1)-3x_3+5 \Rightarrow$

$$x_1 = -9x_3 + 3$$

The simplest way (but not the fastest) to find the general solution of the system is to find its RREF.

Reduce Row Echelon Form (RREF):

- The matrix is in REF.
- If 1 is a pivot on row k and column p then all other elements on the column p are zero.

Examples

$$\begin{pmatrix} 1 & 0 & b & 0 \\ 0 & 1 & d & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Bringing a matrix to RREF is called Gauss-Jordan reduction.

It is easy to find from RREF the solution of the system:

$$\begin{pmatrix} 1 & 0 & b & 0 & | & u \\ 0 & 1 & d & 0 & | & v \\ 0 & 0 & 0 & 1 & | & w \end{pmatrix}$$

x_1, x_2, x_4 lead variables x_3 free variable

$$x_1 + bx_3 = u \Rightarrow x_1 = -bx_3 + u,$$

$$x_2 + dx_3 = v \Rightarrow x_2 = -dx_3 + v,$$

$$x_4 = w.$$

8 Vector and Matrix Notations

Vectors: Row Vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ is $1 \times n$ matrix

Column Vector $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}$ is $m \times 1$ matrix. For convenience of notation we denote column vector $\mathbf{u}$ as $\mathbf{u} = (u_1, u_2, \ldots, u_m)^\top$

Vectors with two coordinates represent vectors in the plane $\mathbf{x} = (x_1, x_2)$ represents a vector joining the origin with $P = (x_1, x_2)$.

$a\mathbf{x} = a(x_1, x_2) := (ax_1, ax_2)$ stretch of $\mathbf{x}$ by factor a.

$\mathbf{x}+\mathbf{y} = (x_1, x_2)+(y_1, y_2) := (x_1+y_1, x_2+y_2)$ represents vector obtained by the parallelepiped law.

Draw the two dimensional picture.

The coordinates of a vector and real numbers are called scalars

Caution!: In Leon's book scalars are often denoted by Greek letters: $\alpha, \beta, \gamma....$ In these notes scalars are denoted by small Latin letters, while vector are in a different font: $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{v}, \mathbf{w}$ are vectors, while $a, b, c, d, x, y, z, u, v, w$ are scalars.

The rules for multiplications of vector by scalars and additions of vectors are:

$$a\mathbf{x} = a(x_1, ..., x_n) := (ax_1, ..., ax_n),$$
$$\mathbf{x} + \mathbf{y} = (x_1, ..., x_n) + (y_1, ..., y_n) :=$$
$$(x_1 + y_1, ..., x_n + y_n),$$

the set of all vectors with n coordinates is denoted by $\mathbb{R}^n$.

$$a\mathbf{u} = a \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix},$$

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} :=$$

$$\begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_m + v_m \end{pmatrix}$$

The zero vector $\mathbf{0}$ has all its coordinate $\mathbf{0}$.

$-\mathbf{x} := (-1)\mathbf{x} := (-x_1, ..., -x_n)$

$\mathbf{x} + (-\mathbf{x}) = \mathbf{x} - \mathbf{x} = \mathbf{0}$.

9 Lecture

Homogeneous Systems of Equations

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = 0 \end{array}$$

Augmented Matrix $(A|0)$.

HSE is always solvable:

$$x_1 = x_2 = \ldots = x_n = 0.$$

Trivial Solution

The number of pivots does not exceed m.

If $n > m$ there is at least $n - m$ free variables.

If $n > m$ HSE has infinite number of nontrivial solutions

10 Products Matrix with vector

scalar product: $(u_1, u_2, u_3) \cdot (x_1, x_2, x_3) = u_1x_1 + u_2x_2 + u_3x_3$.

Product of row vector with column vector with the same number of coordinates:

$$\mathbf{u}\mathbf{x} = (u_1 \ u_2 ... u_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = u_1x_1 + u_2x_2 + ... + u_nx_n$$

product of $m \times n$ A and column vector $\mathbf{x} \in \mathbb{R}^n$:

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{nn}x_n \end{pmatrix} \in \mathbb{R}^m$$

The system of m equations in n unknowns

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n = b_m \end{array}$$

can be compactly written as

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

A is an $m \times n$ coefficient matrix, $\mathbf{x} \in \mathbb{R}^n$ is the columns vector of unknowns and $\mathbf{b} \in \mathbb{R}^m$ is the given column vector.

Clearly $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$, where A is $m \times n$ matrix, and $\mathbf{x}, \mathbf{y}$ are two column vectors with n coordinates

11 General solution of systems of LE

A vector $\mathbf{u} = (u_1, \ldots, u_n)^\top$ satisfying $A\mathbf{u} = \mathbf{b}$ is called a particular solution to system $A\mathbf{x} = \mathbf{b}$.

Thm 1. The general solution of the system

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

of m equations in n unknowns is of the form

$$\mathbf{x} = \mathbf{u} + \mathbf{y},$$

where $\mathbf{u}$ is a particular solution of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{y}$ is the general solution of the homogeneous system

$$A\mathbf{y} = \mathbf{0}$$

Proof. Write $\mathbf{x} = \mathbf{u} + \mathbf{y}$. Then $A(\mathbf{u} + \mathbf{y}) = A\mathbf{u} + A\mathbf{y} = \mathbf{b} + A\mathbf{y}$. Hence $\mathbf{x}$ is a solution to $A\mathbf{x} = \mathbf{b}$ if and only if $A\mathbf{y} = \mathbf{0}$.

12 Existence of REF and RREF

Thm 2. Let A be a given $m \times n$ matrix. Then

1. It is always possible to bring A to a row echelon form C (REF), where C is $m \times n$ matrix, by using elementary row operations
   (a) C usually is not unique
   (b) If we consider the homogeneous system $A\mathbf{x} = \mathbf{0}$, then the lead and the free variables are uniquely determined, i.e. they do not depend on a particular form of C.
2. It is always possible to bring A to a reduced row echelon form F (RREF), by using elementary row operations, and F is unique.

13 Proof of Theorem 2

1. We show the existence of REF of A by induction on m.

I. Assume $m = 1$. So $A = (a_{11}\ a_{12}\ \ldots\ a_{1n})$.

i. If $A = 0$ then A is already in row echelon form.

ii. $A \neq 0$. Let a_{1k} be the first nonzero element of the row matrix A. Then $a_{1k}^{-1}A$ is the row echelon form of A.

II. Assume the induction hypothesis that any positive integer M any $M \times n$ matrix A can be brought to a REF using elementary row operations, i.e. we assume the induction hypothesis for $m = M$.

III. Assume that $A = (a_{ij})$ $(M+1) \times n$ matrix, i.e. $m = M + 1$.

i. Suppose first that $a_{11} \neq 0$. Let $a_{11}\ R_1 \to R_1$ to obtain A_1. Then $R_i \to R_i - a_{i1} \times R_1$ for $i = 2, \ldots, M+1$ to obtain the matrix A_2 with a pivot on the entry $(1,1)$ and all other entries of the first column of A_2 are zero.

$$A_2 = \begin{pmatrix} 1 & a_{12,2} & \cdots & a_{1n,2} \\ 0 & a_{22,2} & \cdots & a_{2n,2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & a_{m2,2} & \cdots & a_{mn,2} \end{pmatrix}$$

If $n = 1$ A_2 is the row echelon form of A.

Assume $n > 1$. Let B_2 be the following $M \times (n-1)$ matrix:

$$B_2 = \begin{pmatrix} a_{22,2} & a_{23,2} & \cdots & a_{2n,2} \\ a_{32,2} & a_{33,2} & \cdots & a_{3n,2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m2,2} & a_{m3,2} & \cdots & a_{mn,2} \end{pmatrix}$$

Use the induction hypothesis to deduce the existence of ERO to bring B_2 to REF D_2. Apply the same row operation on the last M rows of A_2 to bring A_2 to a REF

$$C_2 = \begin{pmatrix} 1 & * \\ 0 & D_2 \end{pmatrix}$$ (block matrix form)

ii. Suppose $a_{11} = 0$ but the first column of A is not zero column. Let $a_{i1} \neq 0$, (you may choose i to be the smallest number $i > 1$ to satisfy this assumption.) Switch rows 1 and i, i.e. perform $R_1 \leftrightarrow R_i$, to obtain A_1. Now use the previous case i. to bring A_1 to REF C, which is a REF of A.

iii. Suppose $A = 0$. Then A in REF.

iv. Suppose $A \neq 0$ and the first k columns of A are zero ($1 \leq k < n$). Let B be $(M + 1) \times (n - k)$ matrix obtained from A by deleting the first k columns. Now use cases (i-ii) to bring B to a REF D.

Use the same row operations on A to bring it to REF:
$C := (0_{m\times k}\ D)$,

where $0_{m\times k}$ denotes the zero matrix of order $m \times k$.

14 Non-uniqueness of REF

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

Do $R_2 - R_1 \to R_2$ to obtain REF

$$A_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Now switch the two rows of A, i.e. $R_1 \leftrightarrow R_2$ to obtain

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

Let $R_2 - R_1 \to R_2$ to obtain $B_1 = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$

Let $-R_2 \to R_2$ to obtain REF $B_2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$

Note $A_1 \neq B_2$

15 Lead & free variables uniqueness

Assume that A $m \times n$. Assume that B and C are two $m \times n$ matrices which are two REF of A. Then $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ are equivalent systems to $A\mathbf{x} = \mathbf{0}$.

We show by induction on n that the systems $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ have the same set of lead and free variables.

I. $n = 1$.

i. $A = 0$. Then $B = C = 0$. x_1 is free variable.

ii. $A \neq 0$. Then $B = C = (1\ 0 \ldots 0)^\top$. x_1 is lead variable

II. Assume that the statement holds for $n = N$.

III. Let $n = N + 1$, i.e. A, B, C are $m \times (N + 1)$. Let A_1, B_1, C_1 are $n \times N$ matrices obtained from A, B, C by deleting their last columns respectively. Note that B_1 and C_1 are REF of A_1. The homogeneous systems $A_1\mathbf{x}_1 = \mathbf{0}, B_1\mathbf{x}_1 = \mathbf{0}, C_1\mathbf{x}_1 = \mathbf{0}$ obtained from the equivalent systems $A\mathbf{x} = 0, B\mathbf{x} = 0, C\mathbf{x} = 0$ by letting $x_n = 0$.

The induction hypothesis claims that among $x_1, ..., x_{n-1}$ the lead and free variables in the systems $B_1\mathbf{x}_1 = \mathbf{0}, C_1\mathbf{x}_1 = \mathbf{0}$ are the same. This is equivalent to to the statement that among $x_1, ..., x_{n-1}$ the lead and free variables in the systems $B\mathbf{x} = \mathbf{0}, C\mathbf{x} = \mathbf{0}$ are the same. It is left to show that x_n in the both system is either lead or free. Set all free variables in $x_1, ..., x_{n-1}$ to be zero.

i. Assume that x_n is a lead variable in $B\mathbf{x} = \mathbf{0}$ we deduce that $x_n = 0$. Hence in the equivalent system $C\mathbf{x} = \mathbf{0}$ $x_n = 0$. Therefore x_n is a lead variable too in $C\mathbf{x} = \mathbf{0}$.

ii. Assume that x_n is free variable in $B\mathbf{x} = \mathbf{0}$. So the value of x_n can be anything. Hence in the equivalent system $C\mathbf{x} = \mathbf{0}$ x_n can have any value. Therefore x_n is a free variable too in $C\mathbf{x} = \mathbf{0}$.

16 Uniqueness of RREF

Assume that A $m \times n$. Assume that B and C are two $m \times n$ matrices which are two RREF of A. Then $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ are equivalent systems to $A\mathbf{x} = \mathbf{0}$.

We claim that $B = C$.

From previous part we now that the equivalent systems $B\mathbf{x} = \mathbf{0}, C\mathbf{x} = \mathbf{0}$ have the same lead and free variables. By transferring the free variables of in these equivalent systems to right hand-side we obtain each lead variable as the same linear function of free variables. Hence $B = C$.

$\square$

17 Row equivalence of matrices

Definition

a. Denote by $\mathbb{R}^{m\times n}$ the set of $m \times n$ matrices with real entries

b. Let $A, B \in \mathbb{R}^{m\times n}$. B Is called row equivalent to A, denoted by $B \sim A$, if B can be obtained from A using ERO

Thm 3. Let $A, B \in \mathbb{R}^{m\times n}$. Then

a. $B \sim A \iff A \sim B$

b. $B \sim A$ if and only if A and B have the same REF C.

Remark. Assume that $B \sim A$ and B has the row echelon form. Thm 2 yields these facts independent of choice of B

1. The number of nonzero rows of B is called rank of A, and is denoted by $\mathrm{rank}\, A$.

2. The pivots of A are the first nonzero elements in each nonzero row of B, which are equal to 1. Their location: $(1, j_1), \ldots, (r, j_r),\ 1 \le i_1 < \ldots < i_r \le n$, where $r = \mathrm{rank}\, A$. So $x_{j_1}, \ldots, x_{j_r}$ free variables

18 Vector Spaces-

A set $\mathbf{V}$ is called a vector space if:

I. For each $\mathbf{x}, \mathbf{y} \in \mathbf{V}$, $\mathbf{x} + \mathbf{y}$ is an element of $\mathbf{V}$.

(Addition)

II. For each $\mathbf{x} \in \mathbf{V}$ and $\boldsymbol{a} \in \mathbb{R}$, $\boldsymbol{a}\mathbf{x}$ is an element of $\mathbf{V}$.

(Multiplication by scalar)

The two operations satisfy the following laws:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$, commutative law
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$, associative law
3. $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for each $\mathbf{x}$, neutral element $\mathbf{0}$
4. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$, unique anti element
5. $\boldsymbol{a}(\mathbf{x} + \mathbf{y}) = \boldsymbol{a}\mathbf{x} + \boldsymbol{a}\mathbf{y}$ for each $\mathbf{x}, \mathbf{y}$, distributive law
6. $(\boldsymbol{a} + \boldsymbol{b})\mathbf{x} = \boldsymbol{a}\mathbf{x} + \boldsymbol{b}\mathbf{x}$, distributive law
7. $(\boldsymbol{ab})\mathbf{x} = \boldsymbol{a}(\boldsymbol{b}\mathbf{x})$, distributive law
8. $\mathbf{1x} = \mathbf{x}$.

corollary: $\mathbf{0x} = \mathbf{0}$ neutral element:

$\mathbf{0x} = (\mathbf{0} + \mathbf{0})\mathbf{x} = \mathbf{0x} + \mathbf{0x} \Rightarrow$

$\mathbf{0} = \mathbf{0x} - \mathbf{0x} = (\mathbf{0x} + \mathbf{0x}) - \mathbf{0x} = \mathbf{0x}.$

Examples:

1. $\mathbb{R}$ - Real Line
2. $\mathbb{R}^{\boldsymbol{2}}$ = Plane
3. $\mathbb{R}^{\boldsymbol{3}}$ - Three dimensional space
4. $\mathbb{R}^{\boldsymbol{n}}$ - $\boldsymbol{n}$-dimensional space
5. $\mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$ - Space of $\boldsymbol{m} \times \boldsymbol{n}$ matrices

$m \times n$ matrices

$A = (a_{ij}),\ i = 1, ..., m, j = 1, ...n$

denoted by $\mathbb{R}^{m\times n}$.

(We can identify $\mathbb{R}^{m\times n}$ with vectors $\mathbb{R}^{mn}$ in some cases)

We can multiply matrices by a scalar

$sA = s(a_{ij}) = (sa_{ij})$ and add two matrices of the same dimension:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} +$$

$$\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{pmatrix}$$

The zero matrix $\mathbf{0}$ is an $m \times n$ whose al entries are equal to 0:

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$-A = -(a_{ij}) := (-a_{ij}) = (-1)A$ and

$A + (-A) = A - A = \mathbf{0}$,

$A - B = A + (-B)$

19 Transpose of a matrix $\boldsymbol{A}^\top$

Let $$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Then $$\boldsymbol{A}^\top = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

$$(\boldsymbol{A} + \boldsymbol{B})^{\mathbf{T}} = \boldsymbol{A}^{\mathbf{T}} + \boldsymbol{B}^{\mathbf{T}}$$

$$(a\boldsymbol{A})^\top = a\boldsymbol{A}^\top$$

6a. Diagonal matrices (denoted by $\mathcal{D}_n \subset \mathbb{R}^{n\times n}$): Those are square matrices whose all off-diagonal entries are 0:

$$\mathrm{diag}(d_1, d_2, ..., d_n) = \begin{pmatrix} d_1 & 0 & \ldots & 0 & 0 \\ 0 & d_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & d_n \end{pmatrix}$$

$$\textit{Example}: \ \mathrm{diag}(3, -2, 7) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

Claim: The sum and the product of two diagonal matrices is a diagonal matrix:

$\mathrm{diag}(d_1, ..., d_n) + \mathrm{diag}(q_1, ..., q_n) =$
$\mathrm{diag}(d_1 + q_1, ..., d_n + q_n)$,
$\mathrm{diag}(d_1, ..., d_n)\,\mathrm{diag}(q_1, ..., q_n) =$
$\mathrm{diag}(d_1 q_1, ..., d_n q_n)$,

$\mathrm{diag}(d_1, ..., d_n)$ is invertible $\iff$ $d_1 ... d_n \neq 0$ and
$\mathrm{diag}(d_1, ..., d_n)^{-1} = \mathrm{diag}(d_1^{-1}, ..., d_n^{-1})$

6b. Upper Triangular Matrices (denoted by $\mathcal{UT}_n \subset \mathbb{R}^{n\times n}$): Those are square matrices where all elements below the main diagonal entries are 0:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

$$\textit{Example}: \begin{pmatrix} 3 & 0.1 & -8 \\ 0 & -2 & 6.1 \\ 0 & 0 & 7 \end{pmatrix}$$

Claim: The sum and the product of two upper triangular matrices is an upper triangular matrix.

Claim: An upper triangular matrix is invertible $\iff$ its all diagonal entries are nonzero. The inverse of an upper triangular matrix is upper triangular.

6c. Lower Triangular Matrices (denoted by $\mathcal{LT}_n \subset \mathbb{R}^{n\times n}$): Those are square matrices where all elements above the main diagonal entries are $\mathbf{0}$:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 & 0 \\ a_{21} & a_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n(n-1)} & a_{nn} \end{pmatrix}$$

$$\textit{Example}: \begin{pmatrix} 3 & 0 & 0 \\ 0.1 & -2 & 0 \\ -8 & 6.1 & 7 \end{pmatrix}$$

Claim: The sum and the product of two lower triangular matrices is a lower triangular matrix.

Claim: A lower triangular matrix is invertible $\iff$ its all diagonal entries are nonzero. The inverse of a lower triangular matrix is lower triangular.

Claim: A matrix is lower triangular $\iff$ its transpose is upper triangular.

7. $\mathcal{P}_n$ - Space of polynomials of degree at most n: $\mathcal{P}_n := \{p(x) = a_n x^n + a_{n-2} x^{n-2} + ... + a_1 x + a_0\}$.

8. $C[a, b]$ - Space of continuous functions on the interval $[a, b]$.

Note. The examples 1 - 7 are finite dimensional vector spaces. 8 - is infinite dimensional vector space.

Note. In this course all vector spaces are finite dimensional and isomorphic to $\mathbb{R}^n$ (or $\mathbb{C}^n$ as in Chapter 6).

20 Subspaces

Let $\mathbf{V}$ be a vector space. A subset $\mathbf{W}$ of $\mathbf{V}$ is called a subspace of $\mathbf{V}$ if the following two conditions hold:

a. for any $\mathbf{x}, \mathbf{y} \in \mathbf{W} \Rightarrow \mathbf{x} + \mathbf{y} \in \mathbf{W}$,

b. for any $\mathbf{x} \in \mathbf{W},\ \boldsymbol{a} \in \mathbb{R} \Rightarrow \boldsymbol{a}\mathbf{x} \in \mathbf{W}$.

Note: The zero vector $\mathbf{0} \in \mathbf{W}$ since by the condition a. for any $\mathbf{x} \in \mathbf{W}$ one has $\mathbf{0} = \mathbf{0x} \in \mathbf{W}$.

Equivalently: $\mathbf{W} \subseteq \mathbf{V}$ is a subspace $\iff$ $\mathbf{W}$ is a vector space with respect to the addition and the multiplication by a scalar defined in $\mathbf{V}$.

Claim The conditions a. and b. above are equivalent to one condition

If $\mathbf{x}, \mathbf{y} \in \mathbf{U}$ then $\boldsymbol{a}\mathbf{x} + \boldsymbol{b}\mathbf{y} \in \mathbf{U}$ for any scalars $\boldsymbol{a}, \boldsymbol{b}$

Every vector space $\mathbf{V}$ has the following two subspaces:

1. $\mathbf{V}$.

2. The trivial subspace consisting of the zero element: $\mathbf{W} = \{\mathbf{0}\}$.

Examples of subspaces

1. $\mathbb{R}^2$ - Plane: the whole space, lines through the origin, the trivial subspace.

2. $\mathbb{R}^3$ $\mathbf{3}$-dimensional space: the whole space, planes through the origin, lines through the origin, the trivial subspace.

3. For $A \in \mathbb{R}^{m \times n}$ the null space of A, denoted by $N(A)$, is a subspace of $\mathbb{R}^n$ consisting of all vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

Note: $N(A)$ is also called the kernel of A, and denoted by $\ker A$. (See below the explanation for this term.)

4. For $A \in \mathbb{R}^{m \times n}$ the range of A, denoted by $R(A)$, is a subspace of $\mathbb{R}^m$ consisting of all vectors $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. Equivalently $R(A) = A\mathbb{R}^n$.

In 3. and 4. A is viewed as a transformation $A : \mathbb{R}^n \to \mathbb{R}^m$: The vector $\mathbf{x} \in \mathbb{R}^n$ is mapped to the vector $A\mathbf{x} \in \mathbb{R}^m$ ($\mathbf{x} \mapsto A\mathbf{x}$.) So $R(A)$ is the range of the transformation induced by A and $N(A)$ the set of vectors mapped to zero vector in R^m.

21 Linear combination & span

For $\mathbf{v_1}, \ldots, \mathbf{v_k} \in \mathbf{V}$ and $a_1, \ldots, a_k \in \mathbb{R}$ the vector

$$a_1\mathbf{v_1} + a_2\mathbf{v_2} + \ldots + a_k\mathbf{v_k}$$

is called a linear combination of $\mathbf{v_1}, \ldots, \mathbf{v_k}$.

The set of all linear combinations of $\mathbf{v_1}, \ldots, \mathbf{v_k}$ is called the span of $\mathbf{v_1}, \ldots, \mathbf{v_k}$ and denoted by $\mathbf{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$.

Claim: $\mathbf{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is a linear subspace of $\mathbf{V}$.

Fact: All subspaces in a finite dimensional vector spaces are always given as $\mathbf{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ for some corresponding vectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$.

Examples:

1. Any line through the origin in $1, 2, 3$ dimensional space is spanned by any nonzero vector on the line.

2. Any plane through the origin in $2, 3$ dimensional space is spanned by any two nonzero vectors not lying on a line, i.e. non collinear vectors.

3. $\mathbb{R}^3$ spanned by any 3 non planar vectors.

In the following examples $A \in \mathbb{R}^{m\times n}$.

4. Consider the null space $\mathbf{N}(A)$. Let $B \in \mathbb{R}^{m\times n}$ be the RREF of A. B has p pivots and $k := n - p$ free variables. Let $\mathbf{v}_i \in \mathbb{R}^n$ be the following solution of $A\mathbf{x} = \mathbf{0}$. Let the $i - th$ free variable be equal to 1 while all other free variables are equal to 0. Then $\mathbf{N}(A) = \mathbf{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$.

5. Consider the range $\mathbf{R}(A)$, which is a subspace of $\mathbb{R}^m$. View $A = [\mathbf{c}_1 ... \mathbf{c}_n]$ as a matrix composed of n columns $\mathbf{c}_1, ..., \mathbf{c}_n \in \mathbb{R}^m$. Then $\mathbf{R}(A) = \mathbf{span}(\mathbf{c}_1, ..., \mathbf{c}_n)$.

Proof. Observe that for $\mathbf{x} = (x_1, ..., x_n)^\mathrm{T}$ one has $A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + ... + x_n\mathbf{c}_n$.

Corollary. The system $A\mathbf{x} = \mathbf{b}$ is solvable $\iff$ $\mathbf{b}$ is a linear combination of the columns of A.

Problem. Let $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbb{R}^n$. When $\mathbf{b} \in \mathbb{R}^n$ is a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_k$?

Answer. Let $C := [\mathbf{v}_1 \ \mathbf{v}_2 ... \ \mathbf{v}_k] \in \mathbb{R}^{n\times k}$. Then $\mathbf{b} \in \mathbf{span}(\mathbf{v}_1, ..., \mathbf{v}_k) \iff$ the system $A\mathbf{y} = \mathbf{b}$ is solvable.

Example. $\mathbf{v_1} = (1,1,0)^T, \mathbf{v_2} = (2,3,-1)^T, \mathbf{v_3} = (3,1,2)^T, \mathbf{x} = (2,1,1)^T, \mathbf{y} = (2,1,0)^T \in \mathbb{R}^3$.
Show $\mathbf{x} \in \mathbf{W} := \operatorname{span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}), \mathbf{y} \notin \mathbf{W}$.

Spanning set of a vector space

$\mathbf{v_1}, ..., \mathbf{v_k}$ is called a spanning set of $\mathbf{V} \iff \mathbf{V} = \operatorname{span}(\mathbf{v_1}, ..., \mathbf{v_k})$

Example: Let $\mathbf{V}_{even}, \mathbf{V}_{odd} \subset \mathcal{P}_4$ be the subspaces of even and odd polynomials of degree 4 at most. Then $\mathbf{V}_{even} = \operatorname{span}(1, x^2, x^4)$, $\mathbf{V}_{odd} = \operatorname{span}(x, x^3)$.

Example: which of these sets is a spanning set of R^3?

a. $[(1,1,0)^T, (1,0,1)^T]$,

b. $[(1,1,0)^T, (1,0,1)^T, (0,1,-1)^T]$,

c. $[(1,1,0)^T, (1,0,1)^T, (0,1,-1)^T, (0,1,0)^T]$.

Theorem. $\mathbf{v_1}, ..., \mathbf{v_k} \in \mathbf{V}$ is a spanning set of $R^n \iff k \geq n$ and REF of $A = [\mathbf{v_1}\ \mathbf{v_2} ... \mathbf{v_k}] \in \mathbb{R}^{n \times k}$ has n pivots.

2.5.07 Lemma: Let $\mathbf{v_1}, ..., \mathbf{v_k} \in \mathbf{V}$ and assume $\mathbf{v_i} \in W := \operatorname{span}(\mathbf{v_1}, ..., \mathbf{v_{i-1}}, \mathbf{v_{i+1}}, ..., \mathbf{v_k})$.
Then $\operatorname{span}(\mathbf{v_1}, ..., \mathbf{v_k}) = W$.

Corollary. Let $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$. Form $A = [\mathbf{v}_1 \ \mathbf{v}_2 ... \mathbf{v}_n] \in \mathbb{R}^{m \times n}$. Let $B \in \mathbb{R}^{m \times n}$ be REF of A. Then $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_n)$ is spanned by $\mathbf{v}_{j_1}, ..., \mathbf{v}_{j_r}$ corresponding to the columns of B at which the pivots are located.

Proof Assume that x_i is a free variable. Set $x_i = 1$ and all other free variables are zero. We obtain a nontrivial solution $\mathbf{a} = (a_1, \dots, a_n)^\top$ such that $a_i = 1$ and $a_k = 0$ if x_k is another free variable. $A\mathbf{a} = \mathbf{0}$ implies that $\mathbf{v}_i \in \mathrm{span}(\mathbf{v}_{j_1}, \dots, \mathbf{v}_{j_r})$. □

Work out example in the class

Corollary. Let $A \in \mathbb{R}^{m \times n}$ and assume that $B \in \mathbb{R}^{m \times n}$ be REF of A. Then $\mathbf{R}(A)$-the column space of A is spanned by the columns of A corresponding to the columns of B at which the pivots are located.

Corollary. Let $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^n$. Then $\mathbf{v}_1, ..., \mathbf{v}_n$ span $\mathbb{R}^n \iff$ REF of $A := [\mathbf{v}_1 \ \mathbf{v}_2 ... \mathbf{v}_n]$ has n pivots.

Definition: A square matrix $A \in \mathbb{R}^{n \times n}$ is called nonsingular if REF of A has n pivots

22 Linear Independence

$\mathbf{v_1, ..., v_n} \in \mathbf{V}$ are linearly independent $\iff$ the equality $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_n\mathbf{v}_n = \mathbf{0}$ implies that $a_1 = a_2 = ... = a_n = 0$.

Equivalently $\mathbf{v_1, ..., v_n} \in \mathbf{V}$ are linearly independent $\iff$ every vector in $\mathrm{span}(\mathbf{v_1, ..., v_n})$ can be written as a linear combination of $\mathbf{v_1, ..., v_n}$ in a unique (one) way. (Explain!)

$\mathbf{v_1, ..., v_n} \in \mathbf{V}$ are linearly dependent $\iff$ $\mathbf{v_1, ..., v_n} \in \mathbf{V}$ are not linearly independent.

Equivalently $\mathbf{v_1, ..., v_n} \in \mathbf{V}$ are linearly dependent $\iff$ there exists a nontrivial linear combination of $\mathbf{v_1, ..., v_n}$ which equals to zero vector: $a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n = \mathbf{0}$ and $|a_1| + ... + |a_n| > 0$.

Lemma The following statements are equivalent

(i) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ l.d.

(ii) $\mathbf{v}_i \in W := \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, ..., \mathbf{v}_n)$ for some i.

Proof (i)⇒(ii). $a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n = \mathbf{0}$ for some $(a_1, \ldots, a_n)^\top \neq \mathbf{0}$. Hence $a_i \neq 0$ for some i. So $\mathbf{v}_i = \frac{-1}{a_i}(a_1\mathbf{v}_1 + ... + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \ldots + a_n\mathbf{v}_n)$.

(ii)⇒(i) $\mathbf{v}_i = a_1\mathbf{v}_1 + ... + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \ldots + a_n\mathbf{v}_n$. So $a_1\mathbf{v}_1 + ... + a_{i-1}\mathbf{v}_{i-1} + (-1)\mathbf{v}_i + a_{i+1}\mathbf{v}_{i+1} + \ldots + a_n\mathbf{v}_n = \mathbf{0}$ □

Claim Let $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$. Form $A = [\mathbf{v}_1 ... \mathbf{v}_n] \in \mathbb{R}^{m\times n}$. Then $\mathbf{v}_1, ..., \mathbf{v}_n$ are linearly independent $\iff$ $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. $\iff$ (REF of A has n pivots).

23 Basis and dimension

Definition: $\mathbf{v}_1, ..., \mathbf{v}_n$ form a basis in $\mathbf{V}$ if $\mathbf{v}_1, ..., \mathbf{v}_n$ are linearly independent and span $\mathbf{V}$.

Equivalently: Any vector in $\mathbf{V}$ can be expressed as a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_n$ in a unique way.

Theorem 3: Assume that $\mathbf{v}_1, ..., \mathbf{v}_n$ spans $\mathbf{V}$. Then any collection of m vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbf{V}$, such that $m > n$ is linearly dependent.

Proof Let
$\mathbf{u}_j = a_{1j}\mathbf{v}_1 + \ldots + a_{nj}\mathbf{v}_n, j = 1, \ldots, m$ Let $A = (a_{ij}) \in \mathbb{R}^{n \times m}$. Homogeneous system $A\mathbf{x} = \mathbf{0}$ has more variables than equations. It has a free variable, hence a nontrivial solution $\mathbf{x} = (x_1, \ldots, x_m)^\top \neq \mathbf{0}$. It follows $x_1\mathbf{u}_1 + ... + x_m\mathbf{u}_m = \mathbf{0}$. $\square$

Corollary If $[\mathbf{v}_1, ..., \mathbf{v}_n]$ and $[\mathbf{u}_1, ..., \mathbf{u}_m]$ are bases in $\mathbf{V}$ then $m = n$.

Definition: $\mathbf{V}$ is called n-dimensional, if $\mathbf{V}$ has a basis consisting of n -elements. The dimension of $\mathbf{V}$ is n, which is denoted by $\dim \mathbf{V}$.

The dimension of the trivial space $\{0\}$ is 0.

Theorem 4. Let $\dim \mathbf{V} = n$ Then

(i) Any set of n linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis in $\mathbf{V}$.

(ii) Any set of n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ that span $\mathbf{V}$ is a basis in $\mathbf{V}$.

Proof (i). Let $\mathbf{v} \in \mathbf{V}$. Thm 4 implies $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}$ l.d.: $a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n + a\mathbf{v} = \mathbf{0}, (a_1, \ldots, a_n, a)^\top \neq \mathbf{0}$. If $a = 0$ it follows $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are l.d. contradiction! So $\mathbf{v} = \frac{-1}{a}(a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n)$.

(ii). Need to show $\mathbf{v}_1, \ldots, \mathbf{v}_n$ l.i. If not Lemmas p'45, p'42 and Thm 3 contradict that V has n l.i. vectors. $\square$

Theorem 5. Let $\dim \mathbf{V} = n$. Then:

a. No set of less than n vectors can span $\mathbf{V}$.

b. Any spanning set of more than n vectors can be paired down to form a basis for $\mathbf{V}$.

c. Any subset of less than n linearly independent vectors can be extended to basis of $\mathbf{V}$.

Proof a. If less than n vectors span $\mathbf{V}$, $\mathbf{V}$ can not have n l.i. vectors.

b. See Pruning Lemma.

c. See Completion Lemma.

24 Pruning Lemma

Prunning Lemma. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be vectors in a vector space $\mathbf{V}$. Let $\mathbf{W} = \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ and $k = \dim \mathbf{W}$. Then $0 \leq k \leq m$.

- $k = 0$ if and only if $\mathbf{v}_i = \mathbf{0}$ for $i = 1, \ldots, m$.
- Assume that $k > 0$. Then $\mathbf{W}$ has a basis $\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_k}$, where $1 \leq i_1 < \ldots < i_k \leq m$.

Proof. By Thm 3 (p'46) $k \leq m$.

$k = 0$ if and only if $\mathbf{W} = \{0\}$, which is equivalent to the assumption that each $\mathbf{v}_i = \mathbf{0}$.

Assume that $k > 0$. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent. Then by definition $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a basis.

Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent. by Lemma p'45
$\mathbf{v}_j \in \mathbf{U} := \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_m)$.
Hence $\mathbf{U} = \mathbf{W}$. Continue this process to conclude the lemma

25 Completion lemma

Lemma. Let $\mathbf{V}$ be a vector space of dimension n. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbf{V}$ be m linearly independent vectors. (Hence $m \leq n$.) Then there exist $n - m$ vectors $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis in $\mathbf{V}$.

Proof. If $m = n$ then by Thm 4 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis.

Assume that $m < n$. Hence by Thm 4 $\mathbf{W} := \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_m) \neq \mathbf{V}$. Let $\mathbf{v}_{m+1} \in \mathbf{V}$ and $\mathbf{v}_{m+1} \notin \mathbf{W}$. We claim that $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1}$ are linearly independent. Suppose that $a_1\mathbf{v}_1 + \ldots + a_{m+1}\mathbf{v}_{m+1} = \mathbf{0}$. If $a_{m+1} \neq 0$ then $\mathbf{v}_{m+1} = -\frac{1}{a_{m+1}}(a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m) \in \mathbf{W}$, which contradicts our assumption. So $a_{m+1} = 0$. Hence $a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = \mathbf{0}$. As $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent $a_1 = \ldots = a_m = 0$. So $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1}$ are l.i.

Continue in this manner to deduce the lemma.

26 Row & column spaces of matrix

Def. Let $A \in \mathbb{R}^{m\times n}$.

(a) Let $\mathbf{r}_1, \ldots, \mathbf{r}_m \in \mathbb{R}^{1\times n}$ be the m rows of A. Then the row space of A is $\mathrm{span}(\mathbf{r}_1, \ldots, \mathbf{r}_m)$, which is a subspace of $\mathbb{R}^{1\times n}$.

(b) Let $\mathbf{c}_1, \ldots, \mathbf{c}_n \in \mathbb{R}^{m}$ be the n columns of A. Then the column space of A is $\mathrm{span}(\mathbf{c}_1, \ldots, \mathbf{c}_m)$, which is a subspace of $\mathbb{R}^m = \mathbb{R}^{m\times 1}$.

Claim Let $A, B \in \mathbb{R}^{m\times n}$ and assume that $A \sim B$. Then A and B have the same row spaces

Recall that the column space of A can be identified with the range of A, denoted by $R(A)$. The row space of A can be identified with $R(A^\top)$.

Proof We can obtain B from A

$$A \overset{ERO_1}{\longrightarrow} A_1 \overset{ERO_2}{\longrightarrow} A_2 \overset{ERO_3}{\longrightarrow} \ldots A_{k-1} \overset{ERO_k}{\longrightarrow} B$$

using a sequence of ERO.

Need to show that $A \overset{ERO_1}{\to} A_1$

ERO I: $R_i \longleftrightarrow R_j$. For example $R_1 \longleftrightarrow R_2$. Clearly $\mathrm{span}(\mathrm{r}_1, \mathrm{r}_2, \mathrm{r}_3, \ldots, \mathrm{r}_m) =$ $\mathrm{span}(\mathrm{r}_2, \mathrm{r}_1, \mathrm{r}_3, \ldots, \mathrm{r}_m)$

Hence the row spaces of A and A_1 are the same.

ERO II: $aR_i \longleftrightarrow R_i$, where $a \neq 0$. For example $aR_1 \longleftrightarrow R_1$. Clearly $\mathrm{span}(\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_m) =$ $\mathrm{span}(a\mathrm{r}_1, \mathrm{r}_2, \mathrm{r}_3, \ldots, \mathrm{r}_m)$ since $x_1\mathrm{r}_1 = y_1(ar_1)$ by letting $x_1 = ay_1$ or $y_1 = \frac{x_1}{a}$. Hence the row spaces of A and A_1 are the same.

ERO III: $R_i \longleftrightarrow R_i + aR_j$, where $i \neq j$. For example $R_1 \longleftrightarrow R_1 + aR_2$. Straightforward argument yields $\mathrm{span}(\mathrm{r}_1, \mathrm{r}_2) = \mathrm{span}(\mathrm{r}_1 + a\mathrm{r}_2, \mathrm{r}_2)$. Hence the row spaces of A and A_1 are the same.

$\square$

27 Dimension and basis for row, column and null space

Let $A \in \mathbb{R}^{m\times n}$ and let B be its REF.

Rank of A, denoted by $\mathbf{rank}\, A$ is the number of pivots in B, which is the number of nonzero rows in B.

a. A basis of the row space of A, which is a basis for $\mathrm{R}(A^{\mathrm{T}})$, consists of nonzero rows in B.
$\dim \mathrm{R}(A^{\mathrm{T}}) = \mathbf{rank}\, A$. (number of lead variables.)

Reason: Two row equivalent matrices A and C have the same row space. (But not the same column space!)

b. A basis of column space of A consists of the columns of A in which the pivots of B located.
$\dim \mathrm{R}(A) = \mathbf{rank}\, A$.

c. A basis of the null space of A obtained by letting each free variable to be equal 1 and all the other free variable equal to 0 and then finding the corresponding solution of $A\mathrm{x} = \mathbf{0}$. The dimension of $\mathrm{N}(A)$ called the nullity of A is the number of free variables:
$\mathbf{nul}\, A := \dim \mathrm{N}(A) = n - \mathbf{rank}\, A$.

28 A basis of $N(A)$: Example

Consider the homogeneous system $A\mathbf{x} = \mathbf{0}$ and assume that the RREF of A is given by

$$B = \begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -5 \end{pmatrix}$$

$B\mathbf{x} = \mathbf{0}$ is the system

$$\begin{aligned} x_1 + 2x_2 + \qquad\quad 3x_4 &= 0 \\ x_3 - 5x_4 &= 0 \end{aligned}$$

Note that x_1, x_3 are lead variables and x_2, x_3 are free variables. Express lead variables as functions of free variables: $x_1 = -2x_2 - 3x_4, \; x_3 = 5x_4$

First set $x_2 = 1, x_4 = 0$ to obtain $x_1 = -2, x_3 = 0$. So the whole solution is $\mathbf{u} = (-2, 1, 0, 0)^\top$

Second set $x_2 = 0, x_4 = 1$ to obtain $x_1 = -3, x_4 = 5$. So the whole solution is $\mathbf{v} = (-3, 0, 5, 1)^\top$

$\mathbf{u}, \mathbf{v}$ is a basis in $N(A)$.

29 Useful facts

a. The column and the row space of A have the same dimension. Hence $\mathbf{rank}\, A^{\mathrm{T}} = \mathbf{rank}\, A$.

b. Standard basis in $\mathbb{R}^n$ are given by the n columns of $n \times n$ identity matrix I_n.

$\mathbf{e}_1 = (1, 0)^{\mathrm{T}}, \mathbf{e}_2 = (0, 1)^{\mathrm{T}}$ is a standard basis in $\mathbb{R}^2$.
$\mathbf{e}_1 = (1, 0, 0)^{\mathrm{T}}, \mathbf{e}_2 = (0, 1, 0)^{\mathrm{T}}, \mathbf{e}_3 = (0, 0, 1)^{\mathrm{T}}$ is a standard basis in $\mathbb{R}^3$.

c. $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in \mathbb{R}^n$ form a basis in $\mathbb{R}^n \iff A := [\mathbf{v}_1\ \mathbf{v}_2 ... \mathbf{v}_n]$ has n pivots.

d. $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbb{R}^n$.

Question: Find the dimension and a basis of $\mathbf{V} := \mathbf{span}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k)$.

Answer: Form a matrix $A = [\mathbf{v}_1\ \mathbf{v}_2 ... \mathbf{v}_k] \in \mathbb{R}^{n \times k}$. Then $\mathbf{dim}\, \mathbf{V} = \mathbf{rank}\, A$ Let B be REF of A. Then the vectors $\mathbf{v}_j$ corresponding to the columns of B where the pivots are located form a basis in $\mathbf{V}$.

30 The space $\mathcal{P}_n$

To find the dimension and a basis of a subspace in $\mathcal{P}_n$ One corresponds to each polynomial $p(x) = a_0 + a_1x + ... + a_nx^n$ the vector $(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1}$ and treats these problems as corresponding problems in $\mathbb{R}^{n+1}$

31 Sum of two subspaces

Definition For any two subspaces $\mathbf{U}, \mathbf{W} \subseteq \mathbf{V}$ Denote
$\mathbf{U} + \mathbf{W} := \{\mathbf{v} := \mathbf{u} + \mathbf{w},\ \mathbf{u} \in \mathbf{U}, \mathbf{w} \in \mathbf{W}\}$,
where we take all possible vectors $\mathbf{u} \in \mathbf{U}, \mathbf{w} \in \mathbf{W}\}$.

Thm 6: Let $\mathbf{V}$ be a vector space and $\mathbf{U}, \mathbf{W}$ be subspaces in $\mathbf{V}$. Then

(a) $\mathbf{U} + \mathbf{W}$ and $\mathbf{U} \cap \mathbf{W}$ are subspace of $\mathbf{V}$.

(b) Assume that $\mathbf{V}$ is finite dimensional. Then

1. $\mathbf{U}, \mathbf{W}, \mathbf{U} \cap \mathbf{W}$ are finite dimensional Let $l = \dim \mathbf{U} \cap \mathbf{W} \geq 0, p = \dim \mathbf{U} \geq 0, q = \dim \mathbf{W} \geq 0$ (So $l \leq p, l \leq q$.)
2. There exists a basis in $\mathbf{v}_1, \ldots, \mathbf{v}_m$ in $\mathbf{U} + \mathbf{W}$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_l$ is a basis in $\mathbf{U} \cap \mathbf{W}$, $\mathbf{v}_1, \ldots, \mathbf{v}_p$ a basis in $\mathbf{U}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_l, \mathbf{v}_{p+1}, \ldots, \mathbf{v}_{p+q-l}$ is a basis in $\mathbf{W}$.
3. $\dim(\mathbf{U}+\mathbf{W}) = \dim \mathbf{U} + \dim \mathbf{W} - \dim \mathbf{U} \cap \mathbf{W}$

Identity $\#(A \cup B) = \#A + \#B - \#(A \cap B)$

for finite sets A, B is analogous to 3.

32 Proofs

(a) 1. Let $\mathbf{u}, \mathbf{w} \in \mathbf{U} \cap \mathbf{W}$. Since $\mathbf{u}, \mathbf{w} \in \mathbf{U}$ it follows $a\mathbf{u} + b\mathbf{w} \in \mathbf{U}$. Similarly $a\mathbf{u} + b\mathbf{w} \in \mathbf{W}$. Hence $a\mathbf{u} + b\mathbf{w} \in \mathbf{U} \cap \mathbf{W}$ and $\mathbf{U} \cap \mathbf{W}$ is a subspace. (See Claim on p' 38.)

(a) 2. Assume that $\mathbf{u_1}, \mathbf{u_2} \in \mathbf{U}, \mathbf{w_1}, \mathbf{w_2} \in \mathbf{W}$. Then
$a(\mathbf{u_1} + \mathbf{w_1}) + b(\mathbf{u_2} + \mathbf{w_2}) =$
$(a\mathbf{u_1} + b\mathbf{u_2}) + (a\mathbf{w_1} + b\mathbf{w_2}) \in \mathbf{U} + \mathbf{W}$ Hence $\mathbf{U} + \mathbf{W}$ is a subspace.

(b) 1. Any subspace of an m an dimensional space has dimension m at most.

(b) 2. Let $\mathbf{v_1}, \ldots, \mathbf{v_l}$ be a basis in $\mathbf{U} \cap \mathbf{W}$. Complete this linearly independent set in $\mathbf{U}$ and $\mathbf{W}$ to a basis
$\mathbf{v_1}, \ldots, \mathbf{v_p}$ in $\mathbf{U}$ and a basis
$\mathbf{v_1}, \ldots, \mathbf{v_l}, \mathbf{v_{p+1}}, \ldots, \mathbf{v_{p+q-l}}$ in $\mathbf{W}$ Hence any for any $\mathbf{u} \in \mathbf{U}, \mathbf{w} \in \mathbf{W}$
$\mathbf{u} + \mathbf{w} \in \mathrm{span}(\mathbf{v_1}, \ldots, \mathbf{v_{p+q-l}})$. Hence
$\mathbf{U} + \mathbf{W} = \mathrm{span}(\mathbf{v_1}, \ldots, \mathbf{v_{p+q-l}})$.

We show that $\mathbf{v}_1, \ldots, \mathbf{v}_{p+q-l}$ l.i. Suppose that

$a_1\mathbf{v}_1 + \ldots + a_{p+q-l}\mathbf{v}_{p+q-l} = \mathbf{0}$ So

$\mathbf{u} := a_1\mathbf{v}_1 + \ldots a_p\mathbf{v}_p =$

$-a_{p+1}\mathbf{v}_{p+1} + \ldots - a_{p+q-l}\mathbf{v}_{p+q-l} := \mathbf{w}$

Note $\mathbf{u} \in \mathbf{U}, \mathbf{w} \in \mathbf{W}$. So $\mathbf{w} \in \mathbf{U} \cap \mathbf{W}$. Hence

$\mathbf{w} = b_1\mathbf{v}_1 + \ldots + \mathbf{b}_l\mathbf{v}_l$. Since

$\mathbf{v}_1, \ldots, \mathbf{v}_l, \mathbf{v}_{p+1}, \ldots, \mathbf{v}_{p+q-l}$ l.i.

$a_{p+1} = \ldots a_{p+q-l} = b_1 = \ldots = b_l = \mathbf{0}$. So

$\mathbf{w} = \mathbf{0} = \mathbf{u}$. Since $\mathbf{v}_1, \ldots, \mathbf{v}_p$ l.i.

$a_1 = \ldots = a_p = \mathbf{0}$. Hence $\mathbf{v}_1, \ldots, \mathbf{v}_{p+q-l}$ l.i.

(b) 3. Note from (b) 2 $\mathbf{dim}(\mathbf{U} + \mathbf{W}) = p + q - l$. $\square$

Note $\mathbf{U} + \mathbf{W} = \mathbf{W} + \mathbf{U}$.

Definition: The subspace $\mathbf{X} := \mathbf{U} + \mathbf{W}$ is called a direct sum of $\mathbf{U}$ and $\mathbf{W}$, if any vector $\mathbf{v} \in \mathbf{U} + \mathbf{W}$ has a unique representation of the form $\mathbf{v} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u} \in \mathbf{U}, \mathbf{w} \in \mathbf{W}$. Equivalently, if $\mathbf{u}_1 + \mathbf{w}_1 = \mathbf{u}_2 + \mathbf{w}_2$, where $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}, \mathbf{w}_1, \mathbf{w}_2 \in \mathbf{W}$, then $\mathbf{u}_1 = \mathbf{u}_2, \mathbf{v}_1 = \mathbf{v}_2$.

A direct sum of $\mathbf{U}$ and $\mathbf{W}$ is denoted by $\mathbf{U} \oplus \mathbf{W}$

Claim For two finite dimensional vectors subspaces $\mathbf{U}, \mathbf{W} \subseteq \mathbf{V}$ TFAE (the following are equivalent):

(a) $\mathbf{U} + \mathbf{W} = \mathbf{U} \oplus \mathbf{W}$

(b) $\mathbf{U} \cap \mathbf{W} = \{\mathbf{0}\}$

(c) $\dim \mathbf{U} \cap \mathbf{W} = \mathbf{0}$

(d) $\dim(\mathbf{U} + \mathbf{W}) = \dim \mathbf{U} + \dim \mathbf{W}$

(e) For any bases $\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{w}_1, \ldots, \mathbf{w}_q$ in $\mathbf{U}, \mathbf{W}$ respectively $\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{w}_1, \ldots, \mathbf{w}_q$ is a basis in $\mathbf{U} + \mathbf{W}$.

Proof Straightforward

Example 1. Let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{l \times n}$. Then

$$N(A) \cap N(B) = N(\begin{pmatrix} A \\ B \end{pmatrix})$$

Note $\mathbf{x} \in N(A) \cap N(B) \iff A\mathbf{x} = \mathbf{0} = B\mathbf{x}$

Example 2. Let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times l}$. Then $R(A) + R(B) = R((A\ B))$.

Note $R(A) + R(B)$ is the span of the columns of A and B

33 Sums of many subspaces

Defn Let $\mathbf{U}_1, \ldots, \mathbf{U}_k$ be k subspaces of $\mathbf{V}$. Then $\mathbf{X} := \mathbf{U}_1 + \ldots + \mathbf{U}_k$ is the subspace consisting all vectors of the form $\mathbf{u}_1 + \mathbf{u}_2 + \ldots + \mathbf{u}_k$, where $\mathbf{u}_i \in \mathbf{U}_i, i = 1, \ldots, k$. $\mathbf{U}_1 + \ldots + \mathbf{U}_k$ is called a direct sum of $\mathbf{U}_1 \ldots, \mathbf{U}_k$, and denoted by $\oplus_{i=1}^k \mathbf{U}_i := \mathbf{U}_1 \oplus \ldots \oplus \mathbf{U}_k$ if any vector in $\mathbf{X}$ can be represented in a unique way as $\mathbf{u}_1 + \mathbf{u}_2 + \ldots + \mathbf{u}_k$, where $\mathbf{u}_i \in \mathbf{U}_i, i = 1, \ldots, k$.

Claim For finite dimensional vectors subspaces $\mathbf{U}_i \subseteq \mathbf{V}, i = 1, \ldots, k$ TFAE (the following are equivalent):

(a) $\mathbf{U}_1 + \ldots + \mathbf{U}_k = \oplus_{i=1}^k \mathbf{U}_i$,

(b) $\dim(\mathbf{U}_1 + \ldots + \mathbf{U}_k) = \sum_{i=1}^k \dim \mathbf{U}_i$

(e) For any bases $\mathbf{u}_{1,i}, \ldots, \mathbf{u}_{p_i,i}$ in $\mathbf{U}_i, i = 1, \ldots, k$ the vectors $\mathbf{u}_{j,i}, j = 1, \ldots, p_i, i = 1, \ldots, k$ form a basis in $\mathbf{U}_1 + \ldots + \mathbf{U}_k$.

34 Fields

Defn: A set $\mathbb{F}$ is called a field if for any two elements $a, b \in \mathbb{F}$ one has two operations $a + b,\ ab$, such that $a + b, ab \in \mathbb{F}$ and these two operations satisfy the following properties:

A. The addition operation has the same properties as the addition operation of vector spaces (page 30):

1. $a + b = b + a$, commutative law
2. $(a + b) + c = a + (b + c)$, associative law
3. There exists unique neutral element 0 such that

$a + 0 = a$ for each a,

4. For each a there exists a unique anti element $a + (-a) = 0$,

B. The multiplication operation has similar properties as the addition operation

5. $ab = ba$, commutative law
6. $(ab)c = a(bc)$, associative law
7. There exists unique identity element 1 such that

$a1 = a$ for each a,

8. For each $a \neq 0$ there exists a unique inverse $aa^{-1} = 1$,

C. The distributive law:

9. a(b+c)=ab+ac

Note The commutativity implies $(b+c)a = ba + ca$.

$0a = a0 = 0$ for all $a \in \mathbb{F}$:

$$0a = (0+0)a = 0a + 0a \Rightarrow 0a = 0$$

Examples of Fields

1. Real numbers $\mathbb{R}$
2. Rational numbers $\mathbb{Q}$
3. Complex numbers $\mathbb{C}$

35 Finite Fields

Defn Denote by $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$ the set of positive integers and the set of whole integers respectively Let $m \in \mathbb{N}$. $i, j \in \mathbb{Z}$ are called equivalent modulo m, denoted as $i \equiv j \mod m$, if $i - j$ is divisible by m. mod m is an equivalence relation (easy to show). Denote by $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ the set of equivalence classes, usually identified with $\{0, \ldots, m-1\}$.

(Any integer $i \in \mathbb{Z}$ induces a unique element $a \in \{0, \ldots, m-1\}$ such that $i - a$ is divisible by m.)

In $\mathbb{Z}_m$ define $a + b, ab$ by taking representatives $i, j \in \mathbb{Z}$.

Claim For any $m \in \mathbb{N}$, $\mathbb{Z}_m$ satisfies all the properties on p'62-62, except 8 for some m.

Property 8 holds, i.e. $\mathbb{Z}_m$ is a field, if and only if m is a prime number.

($p \in \mathbb{N}$ is a prime number if p is divisible by 1 and p only)

Proof. Note that $\mathbb{Z}$ satisfies all the properties on p'62-62, except 8. ($0, 1$ are the zero and the identity element of $\mathbb{Z}$.) Hence $\mathbb{Z}_m$ satisfies all the properties on p'62-62, except 8.

Suppose m is composite $m = ln, l, n \in \mathbb{N}, l, n > 1$. Then $l, n \in 2, \ldots, m-2$ and ln is zero element in $\mathbb{Z}_m$. So l and n can not have inverses.

Suppose $m = p$ prime. Take $i \in \{1, \ldots, m-1\}$. Look at $S := \{i, 2i, \ldots, (m-1)i\} \subset \mathbb{Z}_m$. Consider $ki - ji = (k-j)i$ for $1 \le j < k \le m-1$. So $(k-j)i$ is not divisible by p. Hence $S = \{1, \ldots, m-1\}$ as a subset of $\mathbb{Z}_m$. So there is exactly one integer $j \in [1, m-1]$ such that $ji = 1$. i.e. j is the inverse of $i \in \mathbb{Z}_m$.

Thm 7. The number of elements in a finite field $\mathbb{F}$ is p^k, where p is prime and $k \in \mathbb{N}$. For each prime $p > 1$ and $k \in \mathbb{N}$ there exists a finite field $\mathbb{F}$ with p^k elements. Such $\mathbb{F}$ is unique up to an isomorphism.

36 Vector spaces over fields

Defn. Let $\mathbb{F}$ be a field. Then $\mathbf{V}$ is called vector field over $\mathbb{F}$ if $\mathbf{V}$ satisfies all the properties stated on p'30, where the scalars are the elements of $\mathbb{F}$.

Example For any $n \in \mathbb{N}$

$$\mathbb{F}^n := \{\mathbf{x} = (x_1, \ldots, x_n)^\top : \ x_1, \ldots, x_n \in \mathbb{F}\}$$

is a vector space over $\mathbb{F}$.

We can repeat all the notions that we developed for vector spaces over $\mathbb{R}$ for a general field $\mathbb{F}$.

For example $\dim \mathbb{F}^n = n$

If $\mathbb{F}$ is a finite field with $\#\mathbb{F}$ elements, then $\mathbb{F}^n$ is a finite vector space with $(\#\mathbb{F})^n$ elements.

Finite vector spaces are very useful in coding theory.

37 One-to-one and onto maps

Defn. T is called a transformation or map from the source space $\mathbf{V}$ to the target space $\mathbf{W}$, if to each element $\mathbf{v} \in \mathbf{V}$ the transformation T corresponds an element $\mathbf{w} \in \mathbf{W}$. We denote $\mathbf{w} = T(\mathbf{v})$, and $T : \mathbf{V} \to \mathbf{W}$. (In other books T is called a map.)

Example 1: A function $f(x)$ on the real line $\mathbb{R}$ can be regarded as a transformation $f : \mathbb{R} \to \mathbb{R}$.

Example 2: A function $f(x, y)$ on the plane $\mathbb{R}^2$ can be regarded as a transformation $f : \mathbb{R}^2 \to \mathbb{R}$.

Example 3: A transformation $f : \mathbf{V} \to \mathbb{R}$ is called a real valued function on $\mathbf{V}$.

Example 4: Let $\mathbf{V}$ be a map of USA, where at each point we plot the vector of the wind blowing at this point. Then we get a transformation $T : \mathbf{V} \to \mathbb{R}^2$.

T is called one-to-one, or injective, denoted by $1-1$, if for any $x, y \in \mathbf{V}$ one has $Tx \neq Ty$, i.e. the image of two different elements of $\mathbf{V}$ by T are different.

T is called onto, or surjective if
$T\mathbf{V} = \mathbf{W} \iff \text{Range } (T) = \mathbf{W}$, i.e, for each $y \in Y$ there exists $x \in X$ so that $Tx = y$.

Example 1. $\mathbf{V} = \mathbb{N}, T : \mathbb{N} \to \mathbb{N}$ given by $T(i) = 2i$. T is $1-1$ but not onto. However $T : \mathbb{N} \to \text{Range } T$ is one-to-one and onto.

Example 2. $Id : \mathbf{V} \to \mathbf{V}$ defined as $Id(x) = x$ for all $x \in \mathbf{V}$ is one-to-one and onto map of $\mathbf{V}$ onto itself

Claim. Let X, Y be two sets. Assume that $F : X \to Y$ is one-to-one and onto. Then there exists a one-to-one and onto map $G : Y \to X$ such that
$F \circ G = Id_Y, G \circ F = Id_X$. G is the inverse of F denoted by F^{-1}. Note $(F^{-1})^{-1} = F$.

38 Isomorphism of vector spaces

Defn. Two vector spaces $\mathbf{U}, \mathbf{V}$ over field $\mathbb{F}(= \mathbb{R})$ are called isomorphic if there exists one-to-one and onto map $L : \mathbf{U} \to \mathbf{V}$, which preserves the linear structure on $\mathbf{U}, \mathbf{V}$:

1. $L(\mathbf{u_1} + u_2) = L(\mathbf{u_1}) + L(\mathbf{u_2})$ for all $\mathbf{u_1}, \mathbf{u_2} \in \mathbf{U}$. (Note that the first addition is in $\mathbf{U}$, and the second addition is in $\mathbf{V}$.)

2. $L(a\mathbf{u}) = aL(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}, a \in \mathbb{F}(= \mathbb{R})$.

Note that the above two conditions are equivalent to one condition

3. $L(a_1\mathbf{u_1} + a_2\mathbf{u_2}) = a_1L(\mathbf{u_1}) + a_2L(\mathbf{u_2})$ for all $\mathbf{u_1}, \mathbf{u_2} \in \mathbf{U}, a_1, a_2 \in \mathbb{F}(= R)$.

Intuitively $\mathbf{U}$ and $\mathbf{V}$ are isomorphic if they are the same spaces modulo renaming, where L is the renaming function

If $L : \mathbf{U} \to \mathbf{V}$ is an isomorphism then $L(\mathbf{0_U}) = \mathbf{0_V}$:

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

Claim. The inverse of isomorphism is an isomorphism

39 Iso. of fin. dim. vector spaces

Thm 8. Two finite dimensional vector spaces $\mathbf{U}, \mathbf{V}$ over $F (= \mathbb{R})$ are isomorphic if and only if they have the same dimension.

Proof. (a) $\dim \mathbf{U} = \dim \mathbf{V} = n$. So $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}, \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are bases in $\mathbf{U}, \mathbf{V}$ respectively. Define $T : \mathbf{U} \to \mathbf{V}$ by $T(a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n) = a_1\mathbf{v}_1 + \ldots a_n\mathbf{v}_n$. Since any $\mathbf{u} \in \mathbf{U}$ is of the form $\mathbf{u} = a_1\mathbf{u}_1 + \ldots a_n\mathbf{u}_n$ T is a mapping from $\mathbf{U}$ to $\mathbf{V}$. It is straightforward to check that T is linear. As $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis in $\mathbf{V}$, it follows that T is onto. Furthermore $T\mathbf{u} = \mathbf{0}$ implies $a_1, \ldots, a_n = 0$. Hence $\mathbf{u} = \mathbf{0}$, i.e. $T^{-1}\mathbf{0} = \mathbf{0}$. Suppose that $T(\mathbf{x}) = T(\mathbf{y})$. Hence $\mathbf{0}_{\mathbf{V}} = T(x) - T(y) = T(x - y)$. Since $T^{-1}\mathbf{0}_{\mathbf{V}} = \mathbf{0}_{\mathbf{U}} \Rightarrow \mathbf{x} - \mathbf{y} = \mathbf{0}$, i.e. T is 1-1.

(b) Assume $T : \mathbf{U} \rightarrow \mathbf{V}$ is an isomorphism. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis in $\mathbf{U}$. Denote $T(\mathbf{u}_i) = \mathbf{v}_i, i = 1, \ldots, n$. The linearity of T yields $T(a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n) = a_1\mathbf{v}_1 + \ldots a_n\mathbf{v}_n$. Assume that $a_1\mathbf{v}_1 + \ldots a_n\mathbf{v}_n = \mathbf{0}$. Then $a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n = \mathbf{0}$. Since $\mathbf{u}_1, \ldots, \mathbf{u}_n$ l.i. $a_1 = \ldots = a_n = 0$, i.e. $\mathbf{v}_1, \ldots, \mathbf{v}_n$ l.i.. For an $\mathbf{v} \in \mathbf{V}$, there exists $\mathbf{u} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n \in \mathbf{U}$ s.t. $\mathbf{v} = T\mathbf{u} = T(a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n) = a_1\mathbf{v}_1 + \ldots a_n\mathbf{v}_n$. So $\mathbf{V} = \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis. So $\dim \mathbf{U} = \dim \mathbf{V} = n$. $\square$

Corollary. Any finite dimensional vector space is isomorphic to $\mathbb{R}^n$ $(\mathbb{F}^n)$.

Example. $\mathcal{P}_n$- the set of polynomials of degree n at most isomorphic to $\mathbb{R}^{n+1}$:
$T((a_0, \ldots, a_n)^\top) = a_0 + a_1x + \ldots + a_nx^n$.

40 Isomorphisms of $\mathbb{R}^n$

Defn. $A \in \mathbb{R}^{n\times n}$ is nonsingular if any REF of A has n pivots, i.e. RREF of A is I_n, the $n \times n$ diagonal matrix which has all $1's$ on the main diagonal.

Note that the columns of I_n: $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form a standard basis of $\mathbb{R}^n$.

Thm 9. $T : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism if and only if there exists a nonsingular matrix $A \in \mathbb{R}^{n\times n}$ such that $T(\mathbf{x}) = A\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^n$.

Proof. (a) Suppose $A \in \mathbb{R}^{n\times n}$ is nonsingular. Let $T(\mathbf{x}) = A\mathbf{x}$. Clearly T linear. Since any system $A\mathbf{x} = \mathbf{b}$ has a unique solution T is onto and $1-1$.

(b) Assume $T : \mathbb{R}^n \to \mathbb{R}^n$ isomorphism. Let $T\mathbf{e}_i = \mathbf{c}_i, i = 1, \ldots, n$. Proof of Thm 8, p' 69, $\mathbf{c}_1, \ldots, \mathbf{c}_n$ are linearly independent. Let $A = [\mathbf{c}_1\ \mathbf{c}_2\ \ \ldots \mathbf{c}_n]$. $\textbf{rank}\, A = n$ so A is nonsingular. Note $T((a_1, \ldots, a_n)^\top) = T(\sum_{i=1}^n a_i\mathbf{e}_i) = \sum_{i=1} a_iT(\mathbf{e}_i) = \sum_{i=1}^n a_i\mathbf{c}_i = A(a_1, \ldots, a_n)^\top$

41 Examples

Defn The matrix A corresponding to the isomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ in Thm 9 is called the representation matrix of T.

Examples: (a) The identity isomorphism $Id : \mathbb{R}^n \to \mathbb{R}^n$, i.e. $Id(\mathbf{x}) = \mathbf{x}$, is represented I_n, as $I_n\mathbf{x} = \mathbf{x}$. Hence I_n is called the identity matrix.

(b) The dilatation isomorphism $T(\mathbf{x}) = a\mathbf{x}, a \neq 0$ is represented by aI_n.

(c) The reflection of $\mathbb{R}^2$: $R((a, b)^\top) = (a, -b)^\top$ is represented by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

(d) A rotation by an angle θ in $\mathbb{R}^2$:
$(a, b)^\top \mapsto (\cos\theta a + \sin\theta b, -\sin\theta a + \cos\theta b)^\top$
represented by $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

42 Linear Transformations (Homomorphisms)

T is called a transformation or map from the source space $\mathbf{V}$ to the target space $\mathbf{W}$, if to each element $\mathbf{v} \in \mathbf{V}$ the transformation T corresponds an element $\mathbf{w} \in \mathbf{W}$. We denote $\mathbf{w} = T(\mathbf{v})$, and $T : \mathbf{V} \to \mathbf{W}$. (In other books T is called a map.)

Definition: Let $\mathbf{V}$ and $\mathbf{W}$ be two vector spaces. A transformation $T : \mathbf{V} \to \mathbf{W}$ is called linear if

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.
2. $T(a\mathbf{v}) = aT(\mathbf{v})$ for any scalar $a \in \mathbb{R}$.

Equivalently: $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ and $a, b \in \mathbb{R}$.

Corollary: If $T : \mathbf{V} \to \mathbf{W}$ is linear then $T(\mathbf{0}_\mathbf{V}) = \mathbf{0}_\mathbf{W}$.

Proof $\mathbf{0}_\mathbf{W} = 0T(\mathbf{v}) = T(0\mathbf{v}) = T(\mathbf{0}_\mathbf{V})$.

Linear transformation is also called linear operator

Example: Let $A \in \mathbb{R}^{m \times n}$ and define $T : \mathbb{R}^n \to \mathbb{R}^m$ as $T(\mathbf{v}) = A\mathbf{v}$. Then T is a linear transformation.

$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$,

$A(a\mathbf{v}) = a(A\mathbf{v})$.

$\mathrm{R}(T)$ Range of T. $\mathrm{R}(T)$ Is a subspace of $\mathbf{W}$.
$\dim \mathrm{R}(T) = \mathrm{rank}\, T$ is called the rank of T.

$\ker T$ kernel of T, null space of T, all vectors in $\mathbf{V}$ mapped by T to a zero vector in $\mathbf{W}$. $\ker T$ is a subspace of $\mathbf{V}$. $\dim \ker T = \mathrm{nul}\, T$ is called the nullity of T.

Proof. $aT(\mathbf{u}) + bT(\mathbf{v}) = T(a\mathbf{u} + b\mathbf{v})$.

$T(\mathbf{u}) = T(\mathbf{v}) = \mathbf{0} \Rightarrow T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v}) = a\mathbf{0} + b\mathbf{0} = \mathbf{0}$.

Thm. 10 : Any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by some $A \in \mathbb{R}^{m\times n}$: $T\mathbf{x} = A\mathbf{x}$ for each $\mathbf{x} \in \mathbb{R}^n$.

Prf. Let $T(\mathbf{e}_i) = \mathbf{c}_i \in \mathbb{R}^m, i = 1, \ldots, n$. Then $A = [\mathbf{c}_1 \ \ldots \ c_n]$.

Examples: (a) $C^k(a, b)$ all continuous functions on he interval (a, b) with k continuous derivatives. $C^0(a, b) = C(a, b)$ the set of continuous functions in (a, b). Let $p(x), q(x) \in C(a, b)$. Then $L : C^2(a, b) \to C(a, b)$ given by $L(f)(x) = f''(x) + p(x)f'(x) + q(x)f(x)$ is a linear operator. $\ker L$ is the subspace of all functions f satisfying the second order linear differential equation: $y''(x) + p(x)y'(x) + q(x)y(x) = 0$.

It is known that the above ODE has a unique solution satisfies the initial conditions, IC: $y(x_0) = a_1, y'(x_0) = a_2$ for any fixed $x_0 \in (a, b)$. Hence $\dim\ker L = 2$. Using the theory of ODE one can show that $\mathrm{R}(L) = C(a, b)$.

(b) $L : \mathcal{P}_n \to \mathcal{P}_{n-2}$ given by $L(f) = f''$ is a linear operator. L is onto and $\dim\ker L = 2$ if $n \geq 2$.

43 Rank-nullity theorem

Thm 11. For linear $T : \mathbf{V} \to \mathbf{W}$

$\mathbf{rank}\, T + \mathbf{nul}\, T = \mathbf{dim}\, \mathbf{V}$.

Remark. If $\mathbf{V} = \mathbb{R}^n, \mathbf{W} = \mathbb{R}^m$ by Thm 10 $T\mathbf{x} = A\mathbf{x}$. for some $A \in \mathbb{R}^{m\times n}$. $\mathbf{rank}\, T = \mathbf{rank}\, A = \#$ of lead variables, $\mathbf{nul}\, T = \mathbf{nul}\, A = \mathbf{dim}\, N(A) = \#$ number of free variables, so the total number of variables is $n = \mathbf{dim}\, \mathbb{R}^n$.

Proof. (a) Suppose that $\mathbf{nul}\, T = \mathbf{0}$. Then T is $\mathbf{1} - \mathbf{1}$. So $T : \mathbf{V} \to \mathbf{R}(T)$ isomorphism. $\mathbf{dim}\, \mathbf{V} = \mathbf{rank}\, T$.

(b) If $\mathbf{ker}\, T = \mathbf{V}$ then $\mathbf{R}(T) = \{\mathbf{0}\}$ so $\mathbf{nul}\, T = \mathbf{dim}\, \mathbf{V}, \mathbf{rank}\, T = \mathbf{0}$.

(c) $\mathbf{0} < m := \mathbf{nul}\, T < n := \mathbf{dim}\, \mathbf{V}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be a basis in $\mathbf{ker}\, T$. Complete these set of l.i. vectors to a basis of $\mathbf{V}$: $\mathbf{v}_1, \ldots, v_n$. Show that $T(\mathbf{v}_{m+1}), \ldots, T(\mathbf{v}_n)$ is a basis in $\mathbf{R}(T)$. Hence $n - m = \mathbf{rank}\, T$. So

$\mathbf{rank}\, T + \mathbf{nul}\, T = m + (n - m) = \mathbf{dim}\, \mathbf{V}$. $\square$

44 Matrix representations of linear transformations

Let $\mathbf{V}$ and $\mathbf{W}$ be finite dimensional vector spaces with the bases $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$. Let $T : \mathbf{V} \to \mathbf{W}$ be a linear transformation. Then T induces the representation matrix $A \in \mathbb{R}^{m\times n}$ as follows. The column j of A is the coordinate vector of $T(\mathbf{v}_j)$ in the basis $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$.

The definition of A can be formally stated as $[T(\mathbf{v}_1)\ T(\mathbf{v}_2) \ldots T(\mathbf{v}_n)] = [\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]A$.

A is called the representation matrix of T in the bases $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$.

Thm 12. Assume the above assumptions. Assume that $\mathbf{a} \in \mathbb{R}^n$ is the coordinate vector of $\mathbf{v} \in \mathbf{V}$ in the basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $\mathbf{b} \in \mathbb{R}^m$ is the coordinate vector of $T(\mathbf{v}) \in \mathbf{W}$ in the basis $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$. Then $\mathbf{b} = A\mathbf{a}$.

45 Composition of maps

Definition: Let $\mathbf{U}, \mathbf{V}, \mathbf{W}$ be three sets. Assume that we have two maps $S : \mathbf{U} \to \mathbf{V}, T : \mathbf{V} \to \mathbf{W}$.
$T \circ S : \mathbf{U} \to \mathbf{W}$ defined by $T \circ S(\mathbf{u}) = T(S(\mathbf{u}))$ is called the composition map, and denoted TS.

Example 1: $f : \mathbb{R} \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}$. Then $(f \circ g)(x) = f(g(x)), (g \circ f)(x) = g(f(x))$.

Example 2: $f : \mathbb{R}^2 \to \mathbb{R}$, i.e. $f = f(x, y)$, $g : \mathbb{R} \to \mathbb{R}$ then $(g \circ f)(x, y) = g(f(x, y))$, while $f \circ g$ is not defined

Claim. Let $\mathbf{U}, \mathbf{V}, \mathbf{W}$ be vector spaces. Assume that the maps $S : \mathbf{U} \to \mathbf{V}, T : \mathbf{V} \to \mathbf{W}$ are linear. Then $T \circ S : \mathbf{U} \to \mathbf{W}$ is linear.

Proof.

$$T(S(a\mathbf{u}_1 + b\mathbf{u}_2)) = T(aS(\mathbf{u}_1) + bS(\mathbf{u}_2)) =$$
$$aT(S(\mathbf{u}_1)) + bT(S(\mathbf{u}_2)) =$$
$$a(T \circ S)(\mathbf{u}_1) + b(T \circ S)(\mathbf{u}_2). \qquad \square$$

46 Product of matrices

We can multiply A times B if the number of columns in the matrix A is equal to the number of columns in B.

Equivalently A is $m \times n$ matrix and B is $n \times p$ matrix. The resulting matrix $C = AB$ is $m \times p$ matrix. The (i, k) entry of AB is obtained by multiplying $i - th$ row of A and $k - th$ column of B.

$$A = (a_{ij})_{i=j=1}^{i=m,j=n}, \quad B = (b_{jk})_{j=k=1}^{j=n,k=p},$$
$$C = (b_{ik})_{i=k=1}^{i=m,k=p},$$
$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + ... + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

So A, B can be viewed as linear transformations
$B : \mathbb{R}^p \to \mathbb{R}^n, B(\mathrm{u}) = B\mathrm{u}$,
$A : \mathbb{R}^n \to \mathbb{R}^m, A(\mathrm{v}) = B\mathrm{v}$

So AB represents the composition map
$AB : \mathbb{R}^p \to \mathbb{R}^m$.

Example

$$\begin{pmatrix} 1 & -2 \\ -3 & 4 \\ 0 & 2 \\ -7 & -1 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} =$$

$$\begin{pmatrix} a-2d & b-2e & c-2f \\ -3a+4d & -3b+4e & -3c+4f \\ 2d & 2e & 2f \\ -7a-d & -7b-e & -7c-f \end{pmatrix}$$

Note in general $AB \neq BA$ for several reasons

1. AB may be defined but not BA, (as in the above example), or the other way around.

2. AB and BA defined $\iff$ $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times m} \Rightarrow$ $AB \in R^{m\times m}, BA \in \mathbb{R}^{n\times n}$

3. $A, B \in \mathbb{R}^{n\times n}$ usually for $n > 1$ $AB \neq BA$,

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

Rules involving products and additions of matrices

Note: whenever we write additions and products of matrices we assume that they are all defined, i.e. the dimensions of corresponding matrices match.

1. $(AB)C = A(BC)$, associative law.
2. $A(B + C) = AB + AC$, distributive law.
3. $(A + B)C = AC + BC$, distributive law.
4. $a(AB) = (aA)B = A(aB)$, algebra rule.

47 Transpose of a matrix A^{T}

.

Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$

Then $A^{\mathrm{T}} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$

$(A + B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$

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

Examples

$$\begin{pmatrix} -1 & 2 \\ a & b \\ e^{10} & \pi \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} -1 & a & e^{10} \\ 2 & b & \pi \end{pmatrix}.$$

$$\left(\begin{pmatrix} 2 & 3 & -4 \\ 5 & -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 3 & -4 \\ 10 & 1 \end{pmatrix} \right)^{\mathrm{T}} =$$

$$\begin{pmatrix} -33 & -12 \\ -8 & 14 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} -33 & -8 \\ -12 & 14 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 2 \\ 3 & -4 \\ 10 & 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 & 3 & -4 \\ 5 & -1 & 0 \end{pmatrix}^{\mathrm{T}} =$$

$$\begin{pmatrix} -1 & 3 & 10 \\ 2 & -4 & 1 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 3 & -1 \\ -4 & 0 \end{pmatrix} = \begin{pmatrix} -33 & -8 \\ -12 & 14 \end{pmatrix}$$

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

Then $\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{n \times m}$ and $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$.

$$\left(\begin{pmatrix} -1 & 2 \\ a & b \\ e^{10} & \pi \end{pmatrix}^{\mathrm{T}} \right)^{\mathrm{T}} = \begin{pmatrix} -1 & a & e^{10} \\ 2 & b & \pi \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} -1 & 2 \\ a & b \\ e^{10} & \pi \end{pmatrix}.$$

Symmetric Matrices

$A \in \mathbb{R}^{m\times m}$ is called symmetric if $A^{\mathrm{T}} = A$.

The $i-th$ row of a symmetric matrix is equal to its $i-th$ column for $i = 1, ..., m$.

Equivalently: $A = (a_{ij})_{i,j=1}^{m}$ symmetric $\iff$ $a_{ij} = a_{ji}$ for all $i, j = 1, ..., m$.

Examples of 2×2 and 3×3 symmetric matrices:

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

Note symmetricity with respect to the main diagonal

$A \in \mathbb{R}^{m\times n} \Rightarrow$

$A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$ and $AA^{\mathrm{T}} \in \mathbb{R}^{m\times m}$ are symmetric.

Indeed $(AA^{\mathrm{T}})^{\mathrm{T}} = (A^{\mathrm{T}})^{\mathrm{T}}A^{\mathrm{T}} = AA^{\mathrm{T}}$

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

Identity Matrix

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

I_n is in RREF with no zero rows.

I_n is a symmetric matrix.

Example $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Property of the identity matrix:

$I_m A = A I_n = A,$ `for all` $A \in \mathbb{R}^{m \times n}$

Example: $I_2 A$, where $A \in \mathbb{R}^{2 \times 3}$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

Square matrices: $A \in \mathbb{R}^{m \times m}$.

I. Positive Powers of Square Matrices

$$A^2 := AA$$

$$A^3 := A(AA) = (AA)A = A^2A = AA^2$$

is equal to the product of A three times AAA

If k positive integer $A^k := A...A$ - product of A k times

If k, q positive integers $A^{k+q} = A^k A^q = A^q A^k$.

$$A^0 := I_m.$$

A invertible if there exists A^{-1} such that
$$AA^{-1} = A^{-1}A = I_m$$

Thm 12. Let $A \in \mathbb{R}^{m \times m}$. View $A : \mathbb{R}^m \to \mathbb{R}^m$ as a linear transformation. TFAE

a. A 1-1.

b. A onto.

c. $A : \mathbb{R}^m \to \mathbb{R}^m$ is isomorphism.

d. A is invertible.

Applications of matrix powers for Markov chains

In one town people catch cold and recover every day at the following rate: 90% of healthy stay in the morning healthy the next morning; 60% of sick in the morning recover the next morning.

Find the transition matrix of this phenomenon after one day, two days, and after many days.

$a_{HH} = 0.9, a_{SH} = 0.1, a_{HS} = 0.6, a_{SS} = 0.4$

$$A = \begin{pmatrix} 0.9 & 0.6 \\ 0.1 & 0.4 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_H \\ x_S \end{pmatrix}.$$

Note that if $\mathbf{x}^{\mathrm{T}} = (x_H, x_S)$ represents the number of healthy and sick in a given day, then the situation in the next day is given by

$(0.9x_H + 0.6x_S, 0.1x_H + 0.4x_S)^{\mathrm{T}} = A\mathbf{x}$ Hence the number of healthy and sick after two days are given by $A(A\mathbf{x}) = A^2\mathbf{x}$, i.e. the transition matrix given by A^2:

$$\begin{pmatrix} 0.9 & 0.6 \\ 0.1 & 0.4 \end{pmatrix} \begin{pmatrix} 0.9 & 0.6 \\ 0.1 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.87 & 0.78 \\ 0.13 & 0.22 \end{pmatrix}$$

The transition matrix after k days is given by A^k. It can be shown that

$$\lim_{k\to\infty} A^k = \begin{pmatrix} \frac{6}{7} & \frac{6}{7} \\ \frac{1}{7} & \frac{1}{7} \end{pmatrix} \sim \begin{pmatrix} 0.857 & 0.857 \\ 0.143 & 0.143 \end{pmatrix}.$$

The reason for these numbers is the equilibrium state for which we have the equations $A\mathrm{x} = \mathrm{x} = I_2\mathrm{x} \Rightarrow$
$(A - I_2)\mathrm{x} = 0 \Rightarrow 0.1x_H = 0.6x_S \Rightarrow$
$x_H = 6x_S$. If
$x_H + x_S = 1 \Rightarrow x_H = \frac{6}{7}, x_S = \frac{1}{7}$.

In the equilibrium stage $\frac{6}{7}$ of all population: $x_H + x_S$ are healthy and $\frac{1}{7}$ of all population is sick.

Inverse matrices

Suppose $A \in \mathbb{R}^{n\times n}$ is invertible. Then the system $A\mathrm{x} = \mathrm{b}$ where $\mathrm{x} = (x_1\ x_2 ... x_n)^\mathrm{T},\ \mathrm{b} = (b_1\ b_2 ... b_n)^\mathrm{T} \in \mathbb{R}^n$, i.e. the system of n equations and n unknowns has a unique solution: $\mathrm{x} = A^{-1}\mathrm{b}$.

Indeed multiply the above system by A^{-1} to obtain $A^{-1}(A\mathrm{x}) = (A^{-1}A)\mathrm{x} = I_n\mathrm{x} = \mathrm{x} = A^{-1}\mathrm{b}$.

Inverse of 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

if $ad - bc \neq 0$.

If $ad - bc = 0$ then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d \\ -c \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} -b \\ a \end{pmatrix} = 0$$

So A is not invertible.

If $A_1, ..., A_k \in \mathbb{R}^{n\times n}$ invertible then $A_1 ... A_k$ are invertible and $(A_1 ... A_k)^{-1} = A_k^{-1} ... A_1^{-1}$. 9.11.06

48 Elementary Matrices

Elementary Matrix is a square matrix of order m which is obtained by applying one of the three Elementary Row Operations to the identity matrix I_m.

- Interchange two rows $R_i \longleftrightarrow R_j$.

  Example: Apply $R_1 \longleftrightarrow R_3$ to I_3:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow E_I = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

- Multiply i-th row by $a \neq 0$: $aR_i \longrightarrow R_i$

  Example: Apply $aR_2 \longrightarrow R_2$ to I_3:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow E_{II} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Replace a row by its sum with a multiple of another row

  $R_i + a \times R_j \longrightarrow R_i$

  Example: Apply $R_1 + a \times R_3 \longrightarrow R_1$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to E_{III} = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

All elementary matrices are invertible.

The inverse of an elementary matrix is given by another elementary matrix of the same kind corresponding to reversing the first elementary operation:

- The inverse of E_I is E_I: $E_I E_I = E_I^2 = I_m$.

  Example:

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

- The inverse of E_{II} corresponding to $aR_i \longrightarrow R_i$ is E_{II}^{-1} corresponding to $\frac{1}{a}R_i \longrightarrow R_i$

  Example:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{a} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- The inverse of E_{III} corresponding to $R_i + aR_j \longrightarrow R_i$ is E_{III}^{-1} corresponding to $R_i - aR_j \longrightarrow R_i$ Example:

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

Let $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$. Then performing an elementary row operation on $\boldsymbol{A}$ is equivalent to multiplying $\boldsymbol{A}$ by the corresponding elementary matrix $\boldsymbol{E}$: $\boldsymbol{A} \to \boldsymbol{E}\boldsymbol{A}$.

Example I: Apply $\boldsymbol{R_1} \leftrightarrow \boldsymbol{R_3}$ to $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{3} \times \boldsymbol{2}}$:

$$\begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix} \to \begin{pmatrix} y & z \\ w & x \\ u & v \end{pmatrix} =$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix}$$

Example II: Apply $\boldsymbol{aR_2 \rightarrow R_2}$ to $\boldsymbol{A \in \mathbb{R}^{3\times 2}}$:

$$\begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix} \rightarrow \begin{pmatrix} u & v \\ aw & ax \\ y & z \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix}$$

Example III: Apply $\boldsymbol{R_1 + a \times R_3 \longrightarrow R_1}$: to $\boldsymbol{A \in \mathbb{R}^{3\times 2}}$:

$$\begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix} \rightarrow \begin{pmatrix} u + ay & v + az \\ w & x \\ y & z \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u & v \\ w & x \\ y & z \end{pmatrix}$$

Elementary Row Operations in terms of Elementary Matrices

Let $B \in \mathbb{R}^{m\times p}$ and perform k ERO:

$$B \overset{ERO_1}{\rightarrow} B_1 \overset{ERO_2}{\rightarrow} B_2 \overset{ERO_3}{\rightarrow} \dots B_{k-1} \overset{ERO_k}{\rightarrow} B_k$$

$$B_1 = E_1 B,\ B_2 = E_2 B_1 = E_2 E_1 B, \dots$$

$$B_k = E_k \dots E_1 B \Rightarrow$$

$$B_k = MB,\ M = E_k E_{k-1} \dots E_2 E_1$$

M is invertible matrix since $M^{-1} = E_1^{-1} E_2^{-1} \dots E_k^{-1}$.

The system $A\mathrm{x} = \mathrm{b}$, represented by the augmented matrix $B := (A|\mathrm{b})$, after k ERO is given by $B_k = (A_k|\mathrm{b}_k) = MB = M(A|\mathrm{b}) = (MA, M\mathrm{b})$

and represents the system $MA\mathrm{x} = M\mathrm{b}$. As M invertible

$$M^{-1}(MA\mathrm{x}) = A\mathrm{x} = M^{-1}(M\mathrm{b}) = \mathrm{b}.$$

Thus performing elementary row operations on a system results in equivalent system, i.e. the original and the new system of equations have the same solutions.

The inverse of a matrix as products of elementary matrices

Let A_k be the reduced row echelon form of A. Then $A_k = MA$.

Assume that $A \in \mathbb{R}^{n\times n}$. As M invertible A invertible $\iff$ A_k invertible:
$A = M^{-1}A_k \Rightarrow A^{-1} = A_k^{-1}M$.

If A invertible $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, hence A_k has n pivots (no free variables). Thus $A_k = I_n$ and $A^{-1} = M$!

Summary $A \in \mathbb{R}^{n\times n}$ is invertible $\iff$ its reduced row echelon form is the identity matrix. If A is invertible its inverse is given by the product of the elementary matrices:
$A^{-1} = M = E_k \dots E_1$.

Gauss-Jordan algorithm to compute the inverse of A:

- form the matrix $B = (A|I_n)$.
- Perform the ERO to obtain RREF of B: $C = (D|F)$.
- A is invertible $\iff$ $D = I_n$.
- If $D = I_n$ then $A^{-1} = F$.

Numerical Example

$$A = \begin{pmatrix} 1 & 2 & -1 \\ -2 & -5 & 5 \\ 3 & 7 & -5 \end{pmatrix}.$$

Write $B = (A|I_3)$ and observe the $(1,1)$ entry in B is a pivot: $B = \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ -2 & -5 & 5 & 0 & 1 & 0 \\ 3 & 7 & -5 & 0 & 0 & 1 \end{array}\right)$

Perform ERO: $R_2 + 2R_1 \to R_2,\ R_3 - 3R_1 \to R_3$:

$$B_1 = \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -1 & 3 & 2 & 1 & 0 \\ 0 & 1 & -2 & -3 & 0 & 1 \end{array}\right).$$

To make $(2,2)$ entry pivot do: $-R_2 \to R_2$:

$$B_2 = \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & -1 & 0 \\ 0 & 1 & -2 & -3 & 0 & 1 \end{array}\right).$$

To eliminate $(1,2), (1,3)$ entries do

$R_1 - 2R_2 \to R_1,\ R_3 - R_2 \to R_3$

$$B_3 = \left(\begin{array}{ccc|ccc} 1 & 0 & 5 & 5 & 2 & 0 \\ 0 & 1 & -3 & -2 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right).$$

$(3,3)$ is a pivot. To eliminate $(1,3), (2,3)$ entries do:

$R_1 - 5R_3 \to R_1,\ R_2 + 3R_3 \to R_2$

$$B_4 = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 10 & -3 & -5 \\ 0 & 1 & 0 & -5 & 2 & 3 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right).$$

So $B_4 = (I_3|F)$ is RREF of B. Thus A has inverse:

$$A^{-1} = \begin{pmatrix} 10 & -3 & -5 \\ -5 & 2 & 3 \\ -1 & 1 & 1 \end{pmatrix}.$$

Why Gauss-Jordan algorithm works

(For Gauss see later. Wilhelm Jordan (1842-1899), German geodesist)

Perform ERO operations on $B = (A|I_n)$ to obtain RREF of B, which is given by $B_k =$ $MB = M(A|I_n) = (MA|MI_n) = (MA|M)$. $M \in \mathbb{R}^{n \times n}$ is an invertible matrix, which is a product of elementary matrices.

A is invertible $\iff$ RREF of A is I_n $\iff$

The first n columns of B have n pivots $\iff$

$MA = I_n \iff M = A^{-1} \iff$

$B_k = (I_n|A^{-1})$.

Claim. $A \in \mathbb{R}^{n \times n}$ is invertible if and only if $A^\top$ is invertible. Furthermore $(A^\top)^{-1} = (A^{-1})^\top$.

Proof. The first part of Claim follows from $\mathbf{rank}\, A = \mathbf{rank}\, A^\top$. (Recall A invertible iff $\mathbf{rank}\, A = n$.)

The second part follows from the identity $I_n = I_n^\top = (AA^{-1})^\top = (A^{-1})^\top A^\top$.

49 Change of basis

Assume that $\mathbf{V}$ is an n-dimensional vector space. Let $\mathbf{v} = \mathbf{v}_1, ..., \mathbf{v}_n$ be a basis in $\mathbf{V}$. Notation: $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$. Then any vector $\mathbf{x} \in \mathbf{V}$ can be uniquely presented as $\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$.

There is one to one correspondence between $\mathbf{x} \in \mathbf{V}$ and the coordinate vector of $\mathbf{x}$ in the basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$:
$\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}} \in \mathbb{R}^n$. Thus if
$\mathbf{y} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \ldots b_n\mathbf{v}_n$, so
$\mathbf{y} \leftrightarrow \mathbf{b} = (b_1, b_2, ..., b_n)^{\mathrm{T}} \in \mathbb{R}^n$ then

$r\mathbf{x} \leftrightarrow r\mathbf{a}$ and $\mathbf{x} + \mathbf{y} \leftrightarrow \mathbf{a} + \mathbf{b}$.

Thus $\mathbf{V}$ is isomorphic $\mathbb{R}^n$.

Denote $\mathbf{x} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n] \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$

Let $\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n$ be n vectors in $\mathbf{V}$. Write

$\mathbf{u}_j = u_{1j}\mathbf{v}_1 + u_{2j}\mathbf{v}_2 + \ldots + u_{nj}\mathbf{v}_j, j = 1, ..., n$.

Define $U = \begin{pmatrix} u_{11} & u_{12} & \ldots & u_{1n} \\ u_{21} & u_{22} & \ldots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ u_{n1} & u_{n2} & \ldots & u_{nn} \end{pmatrix}$.

Claim: $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ is a basis in $\mathbf{V} \iff U$ is invertible.

Let $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ is a basis in $\mathbf{V}$. Then

$$[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n] = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]U. \qquad (49.1)$$

U is called the transition matrix from basis $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ to basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$. Denoted as

$[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n] \xrightarrow{U} [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$

Claim: U^{-1} is the transition matrix from basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ to basis $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$:

$[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n] \xleftarrow{U^{-1}} [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$.

Proof Multiply (49.1) by U^{-1} to obtain

$[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]U^{-1} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$.

Let $\mathbf{x} = [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n](b_1, b_2, ..., b_n)^\mathrm{T} \iff$ $\mathbf{x} = b_1\mathbf{u}_1 + \ldots b_n\mathbf{u}_n$, i.e. the vector coordinates of $\mathbf{x}$ in the basis $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ is $\mathbf{b} := (b_1, b_2, ..., b_n)^\mathrm{T}$.

Then the coordinate vector of $\mathbf{x}$ in the basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ is $\mathbf{a} = U\mathbf{b}$.

Proof: $\mathbf{x} = [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]\mathbf{b} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]U\mathbf{a}$.

If $\mathbf{a} \in \mathbb{R}^n$ is the coordinate vector of $\mathbf{x}$ in the basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ then $U^{-1}\mathbf{a}$ is the coordinate vector of $\mathbf{x}$ in the basis $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$.

Theorem 12: Let $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n] \xrightarrow{U} [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n] \xrightarrow{W} [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$. Then $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n] \xrightarrow{U^{-1}W} [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$.

Proof. $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n] = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]W = ([\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]U^{-1})W$.

Note To obtain $U^{-1}W$ take $A := [U\ W] \in R^{n\times(2n)}$ and bring it to RREF $B = [I\ C]$. Then $C = U^{-1}W$.

50 An example

Let

$$\mathbf{u} = [\begin{pmatrix} \mathbf{1} \\ \mathbf{2} \end{pmatrix}, \begin{pmatrix} \mathbf{1} \\ \mathbf{3} \end{pmatrix}], \mathbf{v} = [\begin{pmatrix} \mathbf{3} \\ \mathbf{4} \end{pmatrix}, \begin{pmatrix} \mathbf{4} \\ \mathbf{5} \end{pmatrix}]$$

Find the transition matrix from the basis $\mathbf{w}$ to basis $\mathbf{u}$.

Solution. Introduce the standard basis $\mathbf{v} = [\mathbf{e_1}, \mathbf{e_2}]$ in $\mathbb{R}^2$.

So

$$\mathbf{u} = [\mathbf{e_1}, \mathbf{e_2}] \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} \end{pmatrix}, \mathbf{w} = [\mathbf{e_1}, \mathbf{e_2}] \begin{pmatrix} \mathbf{3} & \mathbf{4} \\ \mathbf{4} & \mathbf{5} \end{pmatrix}$$

Hence the transition matrix is

$$\begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{3} & \mathbf{4} \\ \mathbf{4} & \mathbf{5} \end{pmatrix}.$$

To find this matrix get the RREF of

$$\begin{pmatrix} \mathbf{1} & \mathbf{1} & | & \mathbf{3} & \mathbf{4} \\ \mathbf{2} & \mathbf{3} & | & \mathbf{4} & \mathbf{5} \end{pmatrix}$$

which is

$$\begin{pmatrix} \mathbf{1} & \mathbf{0} & | & \mathbf{5} & \mathbf{7} \\ \mathbf{0} & \mathbf{1} & | & -\mathbf{2} & -\mathbf{3} \end{pmatrix}$$

Answer

$$\begin{pmatrix} \mathbf{5} & \mathbf{7} \\ -\mathbf{2} & -\mathbf{3} \end{pmatrix}$$

51 Change of the representation matrix under the change of bases

$T : \mathbf{V} \to \mathbf{W}$ linear trans. T represented by A in $\mathbf{v}, \mathbf{w}$ bases: $[T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)] = [\mathbf{w}_1, \ldots, \mathbf{w}_m]A$.

Change basis in $\mathbf{W}$

$[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m] \xrightarrow{P} [\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_m]$. Then the representation matrix of T in bases $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $[\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_m]$ is given by the matrix PA, P invertible.

Proof. $[T(\mathbf{v}_1)\ T(\mathbf{v}_2) \ldots T(\mathbf{v}_n)] =$
$[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]A = [\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_m]PA$.

we change basis in $\mathbf{V}$

$[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n] \xrightarrow{Q} [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$. Then the representation matrix of T in bases $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ and $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$ is given by the matrix AQ^{-1}. Proof:
$[T(\mathbf{v}_1)\ T(\mathbf{v}_2) \ldots T(\mathbf{v}_n)] =$
$[T(\mathbf{u}_1)\ T(\mathbf{u}_2) \ldots T(\mathbf{u}_n)]Q = [\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]A$
Hence $[T(\mathbf{u}_1)\ T(\mathbf{u}_2) \ldots T(\mathbf{u}_n)] =$
$[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]AQ^{-1}$. Corollary: The representation matrix of T in bases $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ and $[\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_m]$ is given by the matrix PAQ^{-1}.

52 Example

$D : \mathcal{P}_2 \to \mathcal{P}_1, D(p) = p'$. Choose bases $[1, x, x^2], [1, x]$ in $\mathcal{P}_2, \mathcal{P}_1$ respectively.
$D(1) = 0 = 0 \cdot 1 + 0 \cdot x, D(x) = 1 = 1 \cdot 1 + 0 \cdot x, D(x^2) = 2x = 0 \cdot 1 + 2 \cdot x$.

Representation matrix of T in this basis is $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

Change the basis to $[1 + 2x, x - x^2, 1 - x + x^2]$ in $\mathcal{P}_2$. One can find the new representation matrix A_1 in 2 ways. First

$$D(1 + 2x) = 2, D(x - x^2) = 1 - 2x,$$

$$D(1 - x + x^2) = -1 + 2x$$

Hence

$$A_1 = \begin{pmatrix} 2 & 1 & -1 \\ 0 & -2 & 2 \end{pmatrix}$$

Second

$$[1+2x, x-x^2, 1-x+x^2] =$$

$$[1, x, x^2]\begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

So

$$A_1 = \begin{pmatrix} 2 & 1 & -1 \\ 0 & -2 & 2 \end{pmatrix} =$$

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

Now choose a new basis in $\mathcal{P}_1$: $[1+x, 2+3x]$. Then

$$[1+x, 2+3x] = [1, x]\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

Hence the representation matrix of D in bases

$[1+2x, x-x^2, 1-x+x^2]$ and $[1+x, 2+3x]$

is $$A_2 = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^{-1}\begin{pmatrix} 2 & 1 & -1 \\ 0 & -2 & 2 \end{pmatrix} =$$

$$\frac{1}{1\cdot 3 - 1\cdot 2}\begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & -1 \\ 0 & -2 & 2 \end{pmatrix} =$$

$$\begin{pmatrix} 6 & 7 & -7 \\ -2 & -3 & 3 \end{pmatrix}$$

So $D(1+2x) = 2 = 6(1+x) - 2(2+3x)$,

$D(x-x^2) = 1-2x = 7(1+x) - 3(2+3x)$,

$D(1-x+x^2) = -1+2x =$
$-7(1+x) + 3(2+3x)$

53 Equivalence of matrices

Definition. $A, B \in \mathbb{R}^{m\times n}$ are called equivalent if there exist two invertible matrices $P \in \mathbb{R}^{m\times m}, R \in \mathbb{R}^{n\times n}$ such that $B = PAR$.

Claim 1. Equivalence of matrices is an equivalence relation.

Thm 13. $A, B \in \mathbb{R}^{m\times n}$ are called equivalent if and only if they have the same rank.

Proof. Let $E_{k,m,n} = (e_{ij})_{i,j=1}^{m,n} \in \mathbb{R}^{m\times n}$ be a matrix such that $e_{11} = e_{22} = \ldots = e_{kk} = 1$ and all other entries of $E_{k,m,n}$ are equal to zero. We claim that A is equivalent to $E_{k,m,n}$, where $\operatorname{rank} A = k$.

Let $SA = C$, where C is RREF of A and S invertible. Then RREF of $C^\top$ is $E_{k,n,m}$! (Prove). So $UC^\top = E_{k,n,m} \Rightarrow CU^\top = E_{k,m,n} = SAU^\top$, where U is invertible.

Claim. $A, B \in \mathbb{R}^{m\times n}$ are equivalent iff they represent the same linear transformation $T : \mathbf{V} \to \mathbf{W}, \dim \mathbf{V} = n, \dim \mathbf{W} = m$ in different bases.

54 Scalar Product in $\mathbb{R}^n$

In $\mathbb{R}^2$ scalar or dot product is defined for

$\mathbf{x} = (x_1, x_2)^\mathbf{T}, \mathbf{y} = (y_1, y_2)^\mathbf{T} \in \mathbb{R}^2$:

$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 = \mathbf{y}^\mathbf{T}\mathbf{x}$.

In $\mathbb{R}^3$ scalar or dot product is defined for

$\mathbf{x} = (x_1, x_2, x_3)^\mathbf{T}, \mathbf{y} = (y_1, y_2, y_3)^\mathbf{T} \in \mathbb{R}^3$:

$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3 = \mathbf{y}^\mathbf{T}\mathbf{x}$.

In $\mathbb{R}^n$ scalar or dot product is defined for

$\mathbf{x} = (x_1, \ldots, x_n)^\mathbf{T}, \mathbf{y} = (y_1, \ldots, y_n)^\mathbf{T} \in \mathbb{R}^n$:

$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \ldots + x_n y_n = \mathbf{y}^\mathbf{T}\mathbf{x}$.

The length of $\mathbf{x} = (x_1, \ldots, x_n)^\mathbf{T} \in \mathbb{R}^n$ is

$||\mathbf{x}|| := \sqrt{\mathbf{x}^\mathbf{T}\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$.

$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are called orthogonal if $\mathbf{y}^\mathbf{T}\mathbf{x} = \mathbf{x}^\mathbf{T}\mathbf{y} = \mathbf{0}$.

55 Cauchy-Schwarz inequality

$|\mathbf{x}^\mathbf{T}\mathbf{y}| \le ||\mathbf{x}||\ ||\mathbf{y}||$ CSI

Equality holds iff $\mathbf{x}, \mathbf{y}$ are linearly dependent, equivalently if $\mathbf{y} \neq \mathbf{0}$ then $\mathbf{x} = a\mathbf{y}$ for some $a \in \mathbb{R}$.

Proof If either $\mathbf{x}$ or $\mathbf{y}$ are zero vectors then equality holds in CSI. Suppose $\mathbf{y} \neq \mathbf{0}$. Then for $t \in \mathbb{R}$
$f(t) := (\mathbf{x} - t\mathbf{y})^\mathbf{T}(\mathbf{x} - t\mathbf{y}) =$
$||\mathbf{y}||^2 t^2 - 2(\mathbf{x}^\mathbf{T}\mathbf{y})t + ||\mathbf{x}||^2 \ge 0$ The equation $f(t) = 0$ is either unsolvable, in the case $f(t)$ is always positive, or has one solution. Hence CSI holds. Equality holds if $\mathbf{x} - a\mathbf{y} = \mathbf{0}$.

The cosine of the angle between two nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is $\cos\theta = \frac{\mathbf{y}^\mathbf{T}\mathbf{x}}{||\mathbf{x}||\ ||y||}$: (Cosine Law)
$||\mathbf{y} - \mathbf{x}||^2 = ||\mathbf{y}||^2 + ||\mathbf{x}||^2 - 2||\mathbf{y}||\ ||\mathbf{x}|| \cos\theta$
Use $||\mathbf{z}||^2 = \mathbf{z}^\mathbf{T}\mathbf{z}$ to deduce the formula for $\cos\theta$.

So if $\mathbf{x} \perp \mathbf{y}$ Pithagoras theorem holds:

$$||\mathbf{x} - \mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 = ||\mathbf{x} + \mathbf{y}||^2.$$

56 Cauchy

Augustin Louis Cauchy Born: 21 Aug 1789 in Paris, France
Died: 23 May 1857 in Sceaux (near Paris), France

his achievement is summed as follows:- ... Cauchy's creative genius found broad expression not only in his work on the foundations of real and complex analysis, areas to which his name is inextricably linked, but also in many other fields. Specifically, in this connection, we should mention his major contributions to the development of mathematical physics and to theoretical mechanics... we mention ... his two theories of elasticity and his investigations on the theory of light, research which required that he develop whole new mathematical techniques such as Fourier transforms, diagonalisation of matrices, and the calculus of residues.

Cauchy was first to state the Cauchy-Schwarz inequality, and stated it for sums.

http://www-history.mcs.st-and.ac.uk/Biographies/Cauchy.html

57 Schwarz

Hermann Amandus Schwarz Born: 25 Jan 1843 in Hermsdorf, Silesia (now Poland) Died: 30 Nov 1921 in Berlin, Germany

His most important work is a Festschrift for Weierstrass's 70th birthday. @articleSchwarz1885, author = "H. A. Schwarz", title = "Ueber ein die Flächen kleinsten Flächeninhalts betreffendes Problem der Variationsrechnung", journal = "Acta societatis scientiarum Fennicae", volume = "XV", year = 1885, pages = "315–362" Schwarz answered the question of whether a given minimal surface really yields a minimal area. An idea from this work, in which he constructed a function using successive approximations, led Emile Picard to his existence proof for solutions of differential equations. It also contains the inequality for integrals now known as the 'Schwarz inequality'.

Schwarz was the third person to state the Cauchy-Schwarz inequality, stated it for integrals over surfaces

58 Bunyakovsky

Viktor Yakovlevich Bunyakovsky Born: 16 Dec 1804 in Bar, Podolskaya gubernia (now Vinnitsa oblast), Ukraine Died: 12 Dec 1889 in St. Petersburg, Russia

Bunyakovskii was first educated at home and then went abroad, obtaining a doctorate from Paris in 1825 after working under Cauchy.

Bunyakovskii published over 150 works on mathematics and mechanics. He is best known (in Russia) for his discovery of the Cauchy-Schwarz inequality, published in a monograph in 1859 on inequalities between integrals. This is twenty-five years before Schwarz's work. In the monograph Bunyakovskii gave some results on the functional form of the inequality.

@articleBunyakovskii1859, author = " V. Bunyakovski**ui**", title = "Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies", journal = "Mém. Acad. St. Petersbourg", year = 1859, volume = 1

59 Scalar and vector projection

The scalar projection of $\mathbf{x} \in \mathbb{R}^n$ on nonzero $\mathbf{y} \in \mathbb{R}^n$ is given by $\frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{y}||} = \cos\theta ||\mathbf{x}||$.

The vector projection of $\mathbf{x} \in \mathbb{R}^n$ on nonzero $\mathbf{y} \in \mathbb{R}^n$ is given by $\frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{y}||^2}\mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$.

Example. Let

$\mathbf{x} = (2, 1, 3, 4)^T, \mathbf{y} = (1, -1, -1, 1)^T$.

a. Find the cosine of angle between $\mathbf{x}, \mathbf{y}$.

b. Find the scalar and vector projection of $\mathbf{x}$ on $\mathbf{y}$.

Solution

$$||\mathbf{y}|| = \sqrt{1^2 + (-1)^2 + (-1)^2 + 1^2} = \sqrt{4} = 2$$

$$||\mathbf{x}|| = \sqrt{2^2 + 1^2 + 3^2 + 4^2} = \sqrt{30},$$

$$\mathbf{x}^T\mathbf{y} = 2 - 1 - 3 + 4 = 2, \cos\theta = \frac{2}{2\sqrt{30}} = \frac{1}{\sqrt{30}}$$

Scalar projection $\frac{2}{2} = 1$,

Vector projection $\frac{2}{4}\mathbf{y} = (.5, -.5, -.5, .5)^T$.

60 Orthogonal subspaces

Definitions: Two subspaces $\mathbf{U}$ and $\mathbf{V}$ in $\mathbb{R}^n$ are called orthogonal if any $\mathbf{u} \in \mathbf{U}$ is orthogonal to any $\mathbf{v} \in \mathbf{V}$: $\mathbf{v}^{\mathrm{T}}\mathbf{u} = \mathbf{0}$. This is denoted by $\mathbf{U} \perp \mathbf{V}$.

Example in $\mathbb{R}^3$: $\mathbf{U}$ is an orthogonal line to the plane $\mathbf{V}$, which intersect at the origin.

For a subspace $\mathbf{U}$ of $\mathbb{R}^n$ $\mathbf{U}^{\perp}$ denotes all vectors in $\mathbb{R}^n$ orthogonal to $\mathbf{U}$.

Claim 1: Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ span $\mathbf{U} \subseteq \mathbb{R}^n$. Form a matrix $A = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_k) \in \mathbb{R}^{n \times k}$. Then

(a): $N(A^{\mathrm{T}}) = \mathrm{U}^{\perp}$,

(b): $\dim \mathrm{U}^{\perp} = n - \dim \mathrm{U}$,

(c): $(\mathrm{U}^{\perp})^{\perp} = \mathrm{U}$.

Note: (b-c) Holds for any subspace $\mathbf{U} \subseteq \mathbb{R}^n$

Proof (a) follows from definition.

(b) follows from $\dim \mathbf{U} = \mathrm{rank}\, A$,

$\mathrm{nul}\, A^{\mathrm{T}} = n - \mathrm{rank}\, A^{\mathrm{T}} = n - \mathrm{rank}\, A$.

(c) follows from the observations $(\mathrm{U}^{\perp})^{\perp} \supseteq \mathbf{U}$,

$\dim(\mathbf{U}^{\perp})^{\perp} = n - \dim \mathbf{U}^{\perp} =$
$n - (n - \dim \mathbf{U}) = \dim \mathbf{U}$

Corollary: $\mathbb{R}^n = \mathbf{U} \oplus \mathbf{U}^{\perp}$.

Proof Observe that if $\mathbf{x} \in \mathbf{U} \cap \mathbf{U}^{\perp}$ then
$\mathbf{x}^{\mathbf{T}}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0} \Rightarrow \mathbf{U} \cap \mathbf{U}^{\perp} = \{\mathbf{0}\}$.

(b) of Claim 1 yields $\dim \mathbf{U} + \dim \mathbf{U}^{\perp} = n$.

Claim 2: For $A \in \mathbb{R}^{n \times m}$:

(a): $\mathbf{N}(A^{\mathbf{T}}) = \mathbf{R}(A)^{\perp}$

(b): $\mathbf{N}(A^{\mathbf{T}})^{\perp} = \mathbf{R}(A)$.

Proof. Any vector in $N(A^{\mathbf{T}})$ satisfies
$A^{\mathbf{T}}\mathbf{y} = \mathbf{0} \iff \mathbf{y}^{\mathbf{T}}A = \mathbf{0}$. Any vector $\mathbf{z} \in \mathbf{R}(A)$ is of the form $\mathbf{z} = A\mathbf{x}$. So
$\mathbf{y}^{\mathbf{T}}\mathbf{z} = \mathbf{y}^{\mathbf{T}}A\mathbf{x} = (\mathbf{y}^{\mathbf{T}}A)\mathbf{x} = \mathbf{0}^{\mathbf{T}}\mathbf{x} = \mathbf{0}$. So
$\mathbf{N}(A^{\mathbf{T}}) \subseteq \mathbf{R}(A)^{\perp}$. Recall
$\dim \mathbf{N}(A^{\mathbf{T}}) = n - \operatorname{rank} A^{\mathbf{T}} = n - \operatorname{rank} A$
Claim 1 yields
$\dim \mathbf{R}(A)^{\perp} = n - \dim \mathbf{R}(A) = n - \operatorname{rank} A$.
Hence (a) follows. Apply $\perp$ operation to (a) and use (c) of Claim 1 to deduce (b).

61 Example

Let $\mathbf{u} = (1,2,3,4)^{\mathrm{T}}, \mathbf{v} = (2,4,5,2)^{\mathrm{T}}, \mathbf{w} = (3,6,8,6)^{\mathrm{T}}$. Find a basis in $\mathbf{span(u,v,w)}^{\perp}$.

Solution: Set $A = [\mathbf{u}\ \mathbf{v}\ \mathbf{w}]$. Then

$$A^{\mathrm{T}} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 2 \\ 3 & 6 & 8 & 6 \end{pmatrix}$$. RREF of A^{T} is:

$$B = \begin{pmatrix} 1 & 2 & 0 & -14 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence a basis in $N(A^{\mathrm{T}}) = N(B)$ is:
$(-2,1,0,0,)^{\mathrm{T}}, (14,0,-6,1)^{\mathrm{T}}$.

Note that a basis of the row space of A^{T} is given by the nonzero rows of B. Hence a basis of $\mathbf{span(u,v,w)}$ is given by $(1,2,0,-14)^{\mathrm{T}}, (0,0,1,6)^{\mathrm{T}}$.

Fredholm alternative: Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then either $A\mathbf{x} = \mathbf{b}$ is solvable or there exists $\mathbf{y} \in \mathbf{N}(A^\mathrm{T})$ such that $\mathbf{y}^\mathrm{T}\mathbf{b} \neq \mathbf{0}$.

Proof. $A\mathbf{x} = \mathbf{b}$ solvable iff $\mathbf{b} \in \mathbf{R}(A)$. (a) of Claim 2 yields $\mathbf{R}(A)^\perp = \mathbf{N}(A^\mathrm{T})$. So $A\mathbf{x} = \mathbf{b}$ not solvable iff $\mathbf{N}(A^\mathrm{T})$ is not orthogonal to $\mathbf{b}$.

62 Projection on a subspace

Let $\mathbf{U}$ be a subspace of $\mathbb{R}^n$. Let $\mathbb{R}^m = \mathbf{U} \oplus \mathbf{U}^\perp$ and $\mathbf{b} \in \mathbb{R}^m$. Express $\mathbf{b} = \mathbf{u} + \mathbf{v}$ where $\mathbf{u} \in \mathbf{U}, \mathbf{v} \in \mathbf{U}^\perp$. Then $\mathbf{u}$ is called the projection of $\mathbf{b}$ on $\mathbf{U}$ and denoted by $P_\mathrm{U}(\mathrm{b})$: $(\mathrm{b} - P_\mathrm{U}(\mathrm{b})) \perp \mathrm{U}$.

Claim 1: $P_\mathrm{U} : \mathbb{R}^n \to \mathbf{U}$ is a linear transformation.

Claim 2: $P_\mathrm{U}(\mathrm{b})$ is the unique solution of the minimal problem: $\min_{\mathrm{x}\in\mathrm{U}} ||\mathrm{b} - \mathrm{x}|| = ||\mathrm{b} - P_\mathrm{U}||$.

Least Square Theorem: Let $A \in \mathbb{R}^{m\times n},\ \mathbf{b} \in \mathbb{R}^m$. Then the system $A^\mathrm{T}A\mathbf{x} = A^\mathrm{T}\mathbf{b}$ is always solvable. Any solution $\mathbf{z}$ to this system is called the least square solution of $A\mathbf{x} = \mathrm{b}$. Furthermore $P_{\mathrm{R}(A)}(\mathrm{b}) = A\mathrm{z}$.

Proofs

Claim 1: $\boldsymbol{\alpha}\mathbf{b} = \boldsymbol{\alpha}\mathbf{u} + \boldsymbol{\alpha}\mathbf{v}$. As $\boldsymbol{\alpha}\mathbf{u} \in \mathbf{U}$ and $\boldsymbol{\alpha}\mathbf{v} \in \mathbf{U}^{\perp}$ it follows $P_{\mathbf{U}}(\boldsymbol{\alpha}\mathbf{b}) = \boldsymbol{\alpha}\mathbf{u} = \boldsymbol{\alpha} P_{\mathbf{U}}(\mathbf{b})$. Let $\mathbf{c} = \mathbf{x} + \mathbf{y}, \mathbf{x} \in \mathbf{U}, \mathbf{y} \in \mathbf{U}^{\perp}$. Then $\mathbf{b} + \mathbf{c} = (\mathbf{u} + \mathbf{x}) + (\mathbf{v} + \mathbf{y})$ and $\mathbf{u} + \mathbf{x} \in \mathbf{U}, \mathbf{v} + \mathbf{y} \in \mathbf{U}^{\perp}$. Hence $P_{\mathbf{U}}(\mathbf{b} + \mathbf{c}) = (\mathbf{u} + \mathbf{x}) = P_{\mathbf{U}}(\mathbf{b}) + P_{\mathbf{U}}(\mathbf{c})$.

Claim 2: As $\mathbf{b} - P_{\mathbf{U}}(\mathbf{b}) \perp \mathbf{U}$ for any $\mathbf{x} \in \mathbf{U}$: $||\mathbf{b} - \mathbf{x}||^2 = ||(\mathbf{b} - P_{\mathbf{U}}(\mathbf{b})) + (P_{\mathbf{U}}(\mathbf{b}) - \mathbf{x})||^2 = ||\mathbf{b} - P_{\mathbf{U}}(\mathbf{b})||^2 + ||P_{\mathbf{U}}(\mathbf{b}) - \mathbf{x}||^2 \geq ||\mathbf{b} - P_{\mathbf{U}}(\mathbf{b})||^2$.

LST: $A^{\mathbf{T}} A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^{\mathbf{T}} A^{\mathbf{T}} A\mathbf{x} = \mathbf{0} \iff ||A\mathbf{x}||^2 = \mathbf{0} \Rightarrow \mathbf{x} \in \mathrm{N}(A) \Rightarrow \mathbf{x} \in \mathrm{N}(A^{\mathbf{T}} A)$. Let $B := A^{\mathbf{T}} A$ and $B^{\mathbf{T}} = B$. If $\mathbf{y} \in \mathrm{N}(B^{\mathbf{T}})$ then $A\mathbf{y} = \mathbf{0} \Rightarrow \mathbf{y}^{\mathbf{T}} A^{\mathbf{T}} = \mathbf{0} \Rightarrow \mathbf{y}^{\mathbf{T}} A^{\mathbf{T}} \mathbf{b} = \mathbf{0}$. Fredholm alternative yields that $A^{\mathbf{T}} A\mathbf{x} = A^{\mathbf{T}} \mathbf{b}$ is solvable.

Note $A^{\mathbf{T}} A\mathbf{z} = A^{\mathbf{T}} \mathbf{b} \iff A^{\mathbf{T}} \mathbf{b} - A^{\mathbf{T}} A\mathbf{z} = \mathbf{0} \iff A^{\mathbf{T}}(\mathbf{b} - A\mathbf{z}) \iff (\mathbf{b} - A\mathbf{z}) \perp \mathrm{R}(A)$. As $A\mathbf{z} \in \mathrm{R}(A)$ we deduce that $P_{\mathrm{R}(A)}(\mathbf{b}) = A\mathbf{z}$.

63 Johann Carl Friedrich Gauss

Born: 30 April 1777 in Brunswick, Duchy of Brunswick (now Germany)

Died: 23 Feb 1855 in Göttingen, Hanover (now Germany)

The method of least squares, established independently by two great mathematicians, Adrien Marie Legendre (1752 1833) of Paris and Carl Friedrich Gauss

In June 1801, Zach, an astronomer whom Gauss had come to know two or three years previously, published the orbital positions of Ceres, a new ”small planet” which was discovered by G Piazzi, an Italian astronomer on 1 January, 1801. Unfortunately, Piazzi had only been able to observe 9 degrees of its orbit before it disappeared behind the Sun. Zach published several predictions of its position, including one by Gauss which differed greatly from the others. When Ceres was rediscovered by Zach on 7 December 1801 it was almost exactly where Gauss had predicted. Although he did not disclose his methods at the time, Gauss had used his least squares approximation method.

http://www-history.mcs.st-and.ac.uk/Biographies/Gauss.html

64 Example

Consider the system of three equations in two variables:

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 3 \\ -2x_1 & + & 3x_2 & = & 1 \\ 2x_1 & - & x_2 & = & 2 \end{array} \Rightarrow A\mathrm{x} = \mathrm{b}$$

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix}, \mathrm{b} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix},$$

$$\mathrm{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \hat{A} = \begin{pmatrix} 1 & 1 & | & 3 \\ -2 & 3 & | & 1 \\ 2 & -1 & | & 2 \end{pmatrix}$$

RREF of $\hat{A}$: $B = \begin{pmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \end{pmatrix}$

Hence the original system is unsolvable!

The least square system

$A^TAx = A^Tb \iff Cx = c$:

$C = A^TA =$

$$\begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} =$$

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix}, \ c = A^Tb =$$

$$\begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

Since C invertible the solution of the LSP is:

$$x = C^{-1}c = \frac{1}{9 \cdot 11 - (-7)^2} \begin{pmatrix} 11 & 7 \\ 7 & 9 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} =$$

$\begin{pmatrix} 1.66 \\ 1.42 \end{pmatrix}$ Hence $Ax = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.90 \end{pmatrix}$ Is the projection

of b on the column space of A.

65 Finding the projection on span

Claim To find the projection of $\mathbf{b} \in \mathbb{R}^m$ on the subspace $\mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n) \subseteq \mathbb{R}^m$:

a. Form the matrix $A = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_n] \in \mathbb{R}^{m \times n}$.

b. Solve the system $A^{\mathrm{T}} A\mathbf{x} = A^{\mathrm{T}}\mathbf{b}$.

c. For any solution $\mathbf{x}$ of b. $A\mathbf{x}$ is the projection.

Claim: Let $A \in \mathbb{R}^{m \times n}$. Then $\mathrm{rank}\, A = n \iff A^{\mathrm{T}} A$ is invertible. In that case $\mathbf{z} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}}\mathbf{b}$ is the least square solution of $A\mathbf{x} = \mathbf{b}$. Also $A(A^{\mathrm{T}} A)^{-1}\mathbf{b}$ is the projection of $\mathbf{b}$ on the column space of A.

Proof. $A\mathbf{x} = \mathbf{0} \iff ||A\mathbf{x}|| = 0 \iff \mathbf{x}^{\mathrm{T}} A^{\mathrm{T}} A\mathbf{x} = 0 \iff A^{\mathrm{T}} A\mathbf{x} = \mathbf{0}$

So $N(A) = N(A^{\mathrm{T}} A)$.

$\mathrm{rank}\, A = n \iff N(A) = \{\mathbf{0}\} = N(A^{\mathrm{T}} A) \iff A^{\mathrm{T}} A$ invertible.

66 The best fit line

Fitting a straight line $y = a + bx$ in the $X - Y$ plane through m given points

$(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$.

Solution: The line should satisfy m conditions:

$$\begin{array}{ccccc} 1 \cdot a & + & x_1 \cdot b & = & y_1 \\ 1 \cdot a & + & x_2 \cdot b & = & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 \cdot a & + & x_m \cdot b & = & y_m \end{array} \Rightarrow$$

$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

$$A \qquad \mathbf{z} = \mathbf{c}.$$

The least squares system $A^T A\mathbf{z} = A^T\mathbf{c}$:

$$\begin{pmatrix} m & x_1 + x_2 + \ldots + x_m \\ x_1 + x_2 + \ldots + x_m & x_1^2 + x_2^2 + \ldots + x_m^2 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + \ldots + y_m \\ x_1 y_1 + x_2 y_2 + \ldots + x_m y_m \end{pmatrix}.$$

$\det \mathbf{A}^T\mathbf{A} =$
$\mathbf{m}(\mathbf{x}_1^2 + \mathbf{x}_2^2 + \ldots + \mathbf{x}_m^2) - (\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_m)^2$.

$\det \mathbf{A}^T\mathbf{A} = 0 \iff \mathbf{x}_1 = \mathbf{x}_2 = \ldots = \mathbf{x}_m$.

If $\det \mathbf{A}^T\mathbf{A} \neq 0$ then

$$a = \frac{(\sum_{i=1}^m x_i^2)(\sum_{i=1}^m y_i) - (\sum_{i=1}^m x_i)(\sum_{i=1}^m x_i y_i)}{\det \mathbf{A}^T\mathbf{A}}$$

$$b = \frac{-(\sum_{i=1}^m x_i)(\sum_{i=1}^m y_i) + m(\sum_{i=1}^m x_i y_i)}{\det \mathbf{A}^T\mathbf{A}}$$

67 Example

Given three points in $\mathbb{R}^2$: $(0,1), (3,4), (6,5)$. Find the best least square fit by a linear function $y = a + bx$ to these three points.

Solution.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}, \mathbf{z} = \begin{pmatrix} a \\ b \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix}$$

$$\mathbf{z} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{c} =$$

$$\begin{pmatrix} 3 & 9 \\ 9 & 45 \end{pmatrix}^{-1} \begin{pmatrix} 10 \\ 42 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

The best least square fit by a linear function is $y = \frac{4}{3} + \frac{2}{3}x$.

68 Orthonormal sets

$\mathbf{v_1}, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ is called an orthogonal set (OS) if $\mathbf{v}_i^\top \mathbf{v}_j = 0$ if $\mathbf{i} \neq j$, i.e. any two vectors in this set is an orthogonal pair.

Example 1: The standard basis $\mathbf{e_1}, \ldots, \mathbf{e}_m \in \mathbb{R}^m$ is an orthogonal set

Example 2: The vectors $\mathbf{v_1} = (3, 4, 1, 0)^\top, \mathbf{v_2} = (4, -3, 0, 2)^\top, \mathbf{v_3} = (0, 0, 0, 0)^\top$ are three orthogonal vectors in $\mathbb{R}^4$.

Theorem. An orthogonal set of nonzero vectors is linearly independent.

Proof. Suppose that $a_1\mathbf{v_1} + a_2\mathbf{v_2} + \ldots + a_n\mathbf{v}_n = \mathbf{0}$. Multiply by $\mathbf{v_1}^\top$: $0 = \mathbf{v_1}^\top \mathbf{0} = \mathbf{v_1}^\top(a_1\mathbf{v_1} + a_2\mathbf{v_2} + \ldots + a_n\mathbf{v}_n) = a_1\mathbf{v_1}^\top\mathbf{v_1} + a_2\mathbf{v_1}^\top\mathbf{v_2} + \ldots + a_n\mathbf{v_1}^\top\mathbf{v}_n$ Since $\mathbf{v_1}^\top\mathbf{v}_i = 0$ for $i > 1$ we obtain: $0 = a_1(\mathbf{v_1}^\top\mathbf{v_1}) = a_1||\mathbf{v_1}||^2$. Since $||\mathbf{v_1}|| > 0$ we deduce $a_1 = 0$. Continue in the same manner to deduce that all $a_i = 0$. $\square$

$\mathbf{v_1, ..., v_n} \in \mathbf{V}$ is called an orthonormal set (ONS) if $\mathbf{v_1, ..., v_n}$ is an orthogonal set and each $\mathbf{v}_i$ has length 1.

In Example 1 $\mathbf{e}_1, \ldots, \mathbf{e}_m$ is an ONS.

In Example 2 the set $\{\frac{1}{\sqrt{26}}\mathbf{v}_1, \frac{1}{\sqrt{29}}\mathbf{v}_2\}$ is an ONS.

Notation: Let $I_n \in \mathbb{R}^{n\times n}$ be an identity matrix. Let $\delta_{ij},\ i,j = 1,\ldots,n$ be the entries of I_n. So $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$ for $i = 1,\ldots,n$.

Remark δ_{ij} are called the Kronecker's delta in honor of Leopold Kronecker (1823-1891)

http://www-history.mcs.st-and.ac.uk/Biographies/Kronecker.html

$\mathbf{v}_1,\ldots,\mathbf{v}_n$ ONS $\iff$ $\mathbf{v}_i^\top \mathbf{v}_j = \delta_{ij}$ for $i,j = 1,...,n$.

Normalization: A nonzero OS $\mathbf{u}_1, ..., \mathbf{u}_n$ can be normalized to an ONS by $\mathbf{v}_i := \frac{\mathbf{u}_i}{||\mathbf{u}_i||}$ for $i = 1, \ldots, n$.

Theorem. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be ONS in $\mathbb{R}^m$. Denote $\mathbf{U} := \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Then

- Any vector $\mathbf{u} \in \mathbf{U}$ can be written as a unique linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$: $\mathbf{u} = \sum_{i=1}^n (\mathbf{v}_i^\top \mathbf{u})\mathbf{v}_i$.
- For any $\mathbf{v} \in \mathbb{R}^m$ the orthogonal projection $P_\mathbf{U}(\mathbf{v})$ on the subspace $\mathbf{U}$ is given by $P_\mathbf{U}(\mathbf{v}) = \sum_{i=1}^n (\mathbf{v}_i^\top \mathbf{v})\mathbf{v}_i$. In particular $||\mathbf{v}||^2 = \mathbf{v}^\top \mathbf{v} \geq \sum_{i=1}^n |\mathbf{v}_i^\top \mathbf{v}|^2$

  (Bessel's inequality: http://www-history.mcs.st-andrews.ac.uk/Biographies/Bessel.html)

  and equality holds $\iff$ $\mathbf{v} \in \mathbf{U}$.
- If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal basis (OB) in $\mathbf{V}$ then for any vector $\mathbf{v} \in \mathbf{V}$ one has: $\mathbf{v} = \sum_{i=1}^n (\mathbf{v}_i^\top \mathbf{v})\mathbf{v}_i$ and $||\mathbf{v}||^2 = \sum_{i=1}^n |\mathbf{v}_i^\top \mathbf{v}|^2$.

  (Parseval's formula: http://www-history.mcs.st-andrews.ac.uk/Biographies/Parseval.html)

Example 1. Let $\mathbf{U} = \mathbf{span}(\mathbf{e_1}, \mathbf{e_2}) \subset \mathbb{R}^4$.

Any vector in $\mathbf{U}$ is

$\mathbf{u} = (u_1, u_2, 0, 0)^\top = u_1\mathbf{e_1} + u_2\mathbf{e_2}$.

Note that $u_1 = \mathbf{e}_1^\top\mathbf{u}, u_2 = \mathbf{e}_2^\top\mathbf{u}$.

So $\mathbf{u} = (\mathbf{e}_1^\top\mathbf{u})\mathbf{e_1} + (\mathbf{e}_2^\top\mathbf{u})\mathbf{e_2}$.

Note $\mathbf{U}^\perp = \mathbf{span}(\mathbf{e_3}, \mathbf{e_4})$.

For any vector $\mathbf{v} = (v_1, v_2, v_3, v_4)^\top$.

$P_{\mathrm{U}}(v) = \mathbf{w} := (v_1, v_2, 0, 0)^\top$ since $\mathbf{w} \in \mathbf{U}$ and

$\mathbf{v} - \mathbf{w} = (0, 0, v_3, v_4) \in \mathbf{U}^\perp$. Clearly

$\mathbf{w} = (\mathbf{e}_1^\top v)\mathbf{e_1} + (\mathbf{e}_2^\top v)\mathbf{e_2}$.

$\mathbf{v}^\top\mathbf{v} = v_1^2 + v_2^2 + v_3^2 + v_4^2 \geq \mathbf{w}^\top\mathbf{w} = v_1^2 + v_2^2$.

Equality holds iff $v_3 = v_4 = 0$, i.e. $\mathbf{v} \in \mathbf{U}$.

Example 2: Let $\mathbf{v_1} = \frac{1}{2}(1,1,1,1)^\top$, $\mathbf{v_2} = \frac{1}{2}(1,-1,1,-1)^\top$, $\mathbf{v_3} = \frac{1}{2}(1,-1,-1,1)^\top$, $\mathbf{v_4} = \frac{1}{2}(-1,-1,1,1)^\top$

Check that $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}$ is an OB in $\mathbb{R}^4$. Let $\mathbf{U} = \mathrm{span}(\mathbf{v_1}, \mathbf{v_2})$. Show

a. $\mathbf{u} \in \mathbf{U} \iff \mathbf{u} = (a, b, a, b)^\top$.

b. $\mathbf{U}^\perp = \mathrm{span}(\mathbf{v_3}, \mathbf{v_4})^\top$.

c. $\mathbf{v} \in \mathbf{U}^\perp \iff \mathbf{v} = (c, d, -c, -d)^\top$.

d. $P_{\mathbf{U}}((x_1, x_2, x_3, x_4)^\top) = \frac{x_1+x_2+x_3+x_4}{2}\mathbf{v_1} + \frac{x_1-x_2+x_3-x_4}{2}\mathbf{v_2} = \frac{1}{2}(x_1+x_3, x_2+x_4, x_1+x_3, x_2+x_4)^\top$

e.

$$x_1^2+x_2^2+x_3^2+x_4^2 \geq \frac{1}{2}((x_1+x_3)^2+(x_2+x_4)^2)$$

Equality holds if and only if $x_1 = x_3, x_2 = x_4$, i.e. $(x_1, x_2, x_3, x_4)^\top \in \mathbf{U}$.

Orthogonal Matrices

$Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix if $Q^{\mathrm{T}}Q = I$.

Equivalently, the columns of Q form an OB in $\mathbb{R}^n$

Equivalently $Q^{-1} = Q^{\mathrm{T}}$. Hence $QQ^{\mathrm{T}} = I$.

Equivalently $(Q\mathbf{y})^{\mathrm{T}}(Q\mathbf{x}) = \mathbf{y}^{\mathrm{T}}\mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
I.e. $Q : \mathbb{R}^n \to \mathbb{R}^n$ preserves angles & lengths of vectors.

Equivalently $||Q\mathbf{x}||^2 = ||\mathbf{x}||^2$ for all $\mathbf{x} \in \mathbb{R}^n$. I.e.
$Q : \mathbb{R}^n \to \mathbb{R}^n$ preserves length

Example 1: I_n is an orthogonal matrix since
$I_n I_n^{\mathrm{T}} = I_n I_n = I_n$. (Note $I_n = [\mathbf{e}_1\ \mathbf{e}_2\ \ldots\ \mathbf{e}_n]$)

Example 2: $Q = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$.

(Note $Q = [\mathbf{e}_3\ \mathbf{e}_1\ \mathbf{e}_2]$)

Example 3: $Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Example 4: Any 2×2 orthogonal matrix is either of the form $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ rotation

or $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ reflection

$P \in \mathbb{R}^{n\times n}$ is called a permutation matrix if in each row and column of P there is one nonzero entry which equals to 1.

A permutation matrix is orthogonal.

If P is a permutation matrix and $(y_1, \ldots, y_n)^{\mathrm{T}} = P(x_1, \ldots, x_n)^{\mathrm{T}}$ then the coordinates of $\mathbf{y}$ a permutation of the coordinates of $\mathbf{x}$, which does not depend on the coordinates of $\mathbf{x}$.

n columns of $A \in \mathbb{R}^{m\times n}$ form an OB in the columns space $\mathrm{R}(A)$ of $A \iff A^{\mathrm{T}}A = I_n$. In that case the LSS of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{z} = A^{\mathrm{T}}\mathbf{b}$, which is the projection of $\mathbf{b}$ the column space of A.

69 Gram-Schmidt orthogon. process

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be linearly independent vectors in $\mathbb{R}^m$. Then the Gram-Schmidt (orthogonalization) process gives a recursive way to generate ONS $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ from $\mathbf{x}_1, \ldots, \mathbf{x}_n$, such that $\mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k)$ for $k = 1, \ldots, n$. If $m = n$, i.e. $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a basis of $\mathbb{R}^n$ then $\mathbf{q}_1, \ldots, \mathbf{q}_n$ is an ONB of $\mathbb{R}^n$.

GS-algorithm:

$r_{11} := ||\mathbf{x}_1||, \mathbf{q}_1 := \frac{1}{r_{11}}\mathbf{x}_1$

$r_{12} := \mathbf{q}_1^\top \mathbf{x}_2, \mathbf{p}_1 := r_{12}\mathbf{q}_1, r_{22} := ||\mathbf{x}_2 - \mathbf{p}_1||, \mathbf{q}_2 := \frac{1}{r_{22}}(\mathbf{x}_2 - \mathbf{p}_1)$.

$r_{13} := \mathbf{q}_1^\top \mathbf{x}_3, r_{23} := \mathbf{q}_2^\top \mathbf{x}_3, \mathbf{p}_2 := r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2, r_{33} := ||\mathbf{x}_3 - \mathbf{p}_2||, \mathbf{q}_3 := \frac{1}{r_{33}}(\mathbf{x}_3 - \mathbf{p}_2)$.

Assume that $\mathbf{q}_1, \ldots, \mathbf{q}_k$ were computed. Then $r_{1(k+1)} := \mathbf{q}_1^\top \mathbf{x}_{k+1}, \ldots, r_{1(k+1)} := \mathbf{q}_k^\top \mathbf{x}_{k+1}, \mathbf{p}_k := r_{1(k+1)}\mathbf{q}_1 + \ldots r_{k(k+1)}\mathbf{q}_k, r_{(k+1)(k+1)} := ||\mathbf{x}_{k+1} - \mathbf{p}_k||$ and $\mathbf{q}_{k+1} := \frac{1}{r_{(k+1)(k+1)}}(\mathbf{x}_{k+1} - \mathbf{p}_k)$.

70 Explanation of G-S process

$r_{i(k+1)} := \mathbf{q}_i^\top \mathbf{x}_{k+1}$

is the scalar projection of $\mathbf{x}_{k+1}$ on $\mathbf{q}_i$.

p_k is the projection of $\mathbf{x}_{k+1}$ on $\mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k) = \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$.

Hence $\mathbf{x}_{k+1} - \mathbf{p}_k \perp \mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k)$.
$\mathbf{r}_{(k+1)(k+1)} = ||\mathbf{x}_{k+1} - \mathbf{p}_k||$ is the distance of $\mathbf{x}_{k+1}$ to $\mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k) = \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$.

The assumption that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent yields that $r_{(k+1)(k+1)} > 0$.

Hence $q_{k+1} = r_{(k+1)(k+1)}^{-1}(\mathbf{x}_{k+1} - \mathbf{p}_k)$

is a vector of unit length orthogonal to $\mathbf{q}_1, \ldots, \mathbf{q}_k$

71 Example

Let $\mathbf{x_1} = (1,1,1,1)^\top$, $\mathbf{x_2} = (-1,4,4,-1)^\top$,
$\mathbf{x_3} = (4,-2,2,0)^\top$

$r_{11} = ||\mathbf{x_1}|| = 2$, $\mathbf{q_1} = \frac{1}{r_{11}}\mathbf{x_1} = (\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})^\top$
$r_{12} = \mathbf{q_1^\top x_2} = 3$,

$\mathbf{p_1} = r_{12}\mathbf{q_1} = 3\mathbf{q_1} = (\frac{3}{2},\frac{3}{2},\frac{3}{2},\frac{3}{2})^\top$ $\mathbf{x_2} - \mathbf{p_1} = (-\frac{5}{2},\frac{5}{2},\frac{5}{2},-\frac{5}{2})^\top$, $\mathbf{r_{22}} = ||\mathbf{x_2} - \mathbf{p_1}|| = 5$

$\mathbf{q_2} = \frac{1}{r_{22}}(\mathbf{x_2} - \mathbf{p_1}) = (-\frac{1}{2},\frac{1}{2},\frac{1}{2},-\frac{1}{2})^\top$
$r_{13} = \mathbf{q_1^\top x_3} = 2$, $\mathbf{r_{23}} = \mathbf{q_2^\top x_3} = -2$,
$\mathbf{p_2} = r_{13}\mathbf{q_1} + r_{23}\mathbf{q_2} = (2,0,0,2)^\top$,
$\mathbf{x_3} - \mathbf{p_2} = (2,-2,2,-2)^\top$,

$r_{33} = ||\mathbf{x_3} - \mathbf{p_2}|| = 4$,

$\mathbf{q_3} = \frac{1}{r_{33}}(\mathbf{x_3} - \mathbf{p_2}) = (\frac{1}{2},-\frac{1}{2},\frac{1}{2},-\frac{1}{2})^\top$

72 QR Factorization

Let $A = [\mathbf{a}_1\ \mathbf{a}_2 \ldots \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ matrix and assume that $\operatorname{rank} A = n \iff$ the columns of A are linearly independent. Perform G-S process with the book keeping as above:

- $r_{11} := ||\mathbf{a}_1||,\ \mathbf{q}_1 := \frac{1}{r_{11}}\mathbf{a}_1$.
- Assume that $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$ were computed. Then
  $r_{ik} := \mathbf{q}_i^{\mathrm{T}}\mathbf{a}_k$ for $i = 1, \ldots, k-1$.
  $\mathbf{p}_{k-1} := r_{1k}\mathbf{q}_1 + r_{2k}\mathbf{q}_2 + \ldots r_{(k-1)k}\mathbf{q}_{k-1}$
  and
  $r_{kk} := ||\mathbf{a}_k - \mathbf{p}_{k-1}||,\ \mathbf{q}_k := \frac{1}{r_{kk}}(\mathbf{a}_k - \mathbf{p}_{k-1})$
  for $k = 2, \ldots, n$.

Let $Q = [\mathbf{q}_1\ \mathbf{q}_2 \ldots \mathbf{q}_n] \in \mathbb{R}^{m\times n}$ and

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{23} & \ldots & r_{2n} \\ 0 & 0 & r_{33} & \ldots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & r_{nn} \end{pmatrix}$$

Then $A = QR$, $Q^{\mathrm{T}}Q = I_n$ and $A^{\mathrm{T}}A = R^{\mathrm{T}}R$.

The Least Squares Solution of $A\mathrm{x} = \mathrm{b}$ is given by the upper triangular system $R\hat{\mathrm{x}} = Q^{\mathrm{T}}\mathrm{b}$ which can be solved by back substitution.

Formally $\hat{\mathrm{x}} = R^{-1}Q^{\mathrm{T}}\mathrm{b}$.

Proof $A^{\mathrm{T}}A\mathrm{x} = R^{\mathrm{T}}Q^{\mathrm{T}}QR\mathrm{x} = R^{\mathrm{T}}R\mathrm{x} = A^{\mathrm{T}}\mathrm{b} = R^{\mathrm{T}}Q^{\mathrm{T}}\mathrm{b}$. Multiply from left by $(R^{\mathrm{T}})^{-1}$ to get
$R\hat{\mathrm{x}} = Q^{\mathrm{T}}\mathrm{b}$

Note: $QQ^{\mathrm{T}}\mathrm{b}$ is the projection of b on the columns space of A.

The matrix $P := QQ^{\mathrm{T}}$ is called an orthogonal projection. It is symmetric and $P^2 = P$, as $(QQ^{\mathrm{T}})(QQ^{\mathrm{T}}) = Q(Q^{\mathrm{T}}Q)Q^{\mathrm{T}} = Q(I)Q^{\mathrm{T}} = QQ^{\mathrm{T}}$.

Note $QQ^{\mathrm{T}} : \mathbb{R}^m \to \mathbb{R}^m$ is the orthogonal projection

Equivalently: The assumption that $\operatorname{rank} A = n$ is equivalent to the assumption that $A^{\mathrm{T}}A$ is invertible. So the LSS $A^{\mathrm{T}}A\hat{x} = A^{\mathrm{T}}\mathrm{b}$ has unique solution $\hat{\mathrm{x}} = (A^{\mathrm{T}}A)^{-1}\mathrm{b}$. Hence the projection of b on the column space of A is $P\mathrm{b} = A\hat{\mathrm{x}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathrm{b}$. Hence $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$.

73 An example of QR algorithm

Let $A = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3] = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{pmatrix}$ be the matrix corresponding to the Example of G-S Pr. above.

Then

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix} = \begin{pmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{pmatrix}$$

$$Q = [\mathbf{q}_1\ \mathbf{q}_2\ \mathbf{q}_3] = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Explain why in this example $A = QR$!

Note $QQ^T : \mathbb{R}^4 \to \mathbb{R}^4$ is the projection on $\mathrm{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$

74 Gram and Schmidt

Jorgen Pedersen Gram: Born: 27 June 1850 in Nustrup (18 km W of Haderslev), Denmark Died: 29 April 1916 in Copenhagen, Denmark Gram is best remembered for the Gram-Schmidt orthogonalisation process which constructs an orthogonal set of from an independent one. The process seems to be a result of Laplace and it was essentially used by Cauchy in 1836.

http://www-history.mcs.st-and.ac.uk/Biographies/Gram.html

Erhard Schmidt Born: 13 Jan 1876 in Dorpat, Germany (now Tartu, Estonia) Died: 6 Dec 1959 in Berlin, Germany Schmidt published a two part paper on integral equations in 1907 in which he reproved Hilbert's results in a simpler fashion, and also with less restrictions. In this paper he gave what is now called the Gram-Schmidt orthonormalisation process for constructing an orthonormal set of functions from a linearly independent set.

http://www-history.mcs.st-and.ac.uk/Biographies/Schmidt.html

75 Inner Product Spaces

Let $\mathbf{V}$ be a vector space. Then the function $\langle\cdot,\cdot\rangle : \mathbf{V}\times\mathbf{V}\to\mathbb{R}$ is called an inner product on $\mathbf{V}$ if the following conditions hold:

- For each pair $\mathbf{x},\mathbf{y}\in\mathbf{V}$ $\langle\mathbf{x},\mathbf{y}\rangle$ is a real number.
- $\langle\mathbf{x},\mathbf{y}\rangle=\langle\mathbf{y},\mathbf{x}\rangle$. (symmetricity.)
- $\langle\mathbf{x}+\mathbf{z},\mathbf{y}\rangle=\langle\mathbf{x},\mathbf{y}\rangle+\langle\mathbf{z},\mathbf{y}\rangle$. (linearity)
- $\langle\alpha\mathbf{x},\mathbf{y}\rangle=\alpha\langle\mathbf{x},\mathbf{y}\rangle$ for any scalar $\alpha\in\mathbb{R}$. (linearity)
- For any $\mathbf{0}\neq\mathbf{x}\in\mathbf{V}$ $\langle\mathbf{x},\mathbf{x}\rangle>\mathbf{0}$. (positivity)

Note:

- The two linearity conditions can be put in one condition: $\langle\alpha\mathbf{x}+\beta\mathbf{z},\mathbf{y}\rangle=\alpha\langle\mathbf{x},\mathbf{y}\rangle+\beta\langle\mathbf{z},\mathbf{y}\rangle$.
- The symmetricity condition yields linearity in the second variable: $\langle\mathbf{x},\alpha\mathbf{y}+\beta\mathbf{z}\rangle=\alpha\langle\mathbf{x},\mathbf{y}\rangle+\beta\langle\mathbf{x},\mathbf{z}\rangle$.
- Each linearity condition implies $\langle\mathbf{0},\mathbf{y}\rangle=\mathbf{0}\Rightarrow\langle\mathbf{0},\mathbf{0}\rangle=\mathbf{0}$.
- $\langle\mathbf{x},\mathbf{x}\rangle\geq\mathbf{0}$ For any $\mathbf{x}\in\mathbf{V}$.

Examples:

- $\mathbf{V} = \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x}$.
- $\mathbf{V} = \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T D \mathbf{x}$,

  $D = \text{diag}(d_1, ..., d_n)$ is a diagonal matrix with positive diagonal entries. Then

  $\mathbf{y}^T D \mathbf{x} = d_1 x_1 y_1 + \ldots + d_n x_n y_n$.
- $\mathbf{V} = \mathbb{R}^{m \times n}$, $\langle A, B \rangle = \sum_{i,j=1}^{m,n} a_{ij} b_{ij}$.
- $\mathbf{V} = \mathrm{C}[a, b]$, $\langle f, g \rangle = \int_a^b f(x) g(x) dx$.
- $\mathbf{V} = \mathrm{C}[a, b]$, $\langle f, g \rangle = \int_a^b f(x) g(x) p(x) dx$, where $p(x) \in \mathrm{C}[a, b]$, $p(x) \geq 0$ and $p(x) = 0$ at most at a finite number of points.
- $\mathbf{V} = P_n$: all polynomials of degree $n - 1$ at most. Let $t_1 < t_2 < \ldots < t_n$ be any n real numbers.

  $\langle p, q \rangle := \sum_{i=1}^{n} p(t_i) q(t_i)$

  $= p(t_1) q(t_1) + \ldots + p(t_n) q(t_n)$

The norm (length) of the vector $\mathbf{x}$ is $||\mathbf{x}|| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

Cauchy-Schwarz inequality: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq ||\mathbf{x}|| \; ||\mathbf{y}||$.

The cosine of the angle between $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$:
$\cos\theta := \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}|| \; ||\mathbf{y}||}$.

$\mathbf{x}$ and $\mathbf{y}$ are orthogonal if: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{0}$.

Two subspace $\mathbf{X}, \mathbf{Y}$ of $\mathbf{V}$ are orthogonal if any $\mathbf{x} \in \mathbf{X}$ is orthogonal to any $\mathbf{y} \in \mathbf{Y}$.

The Parallelogram Law;

$$||\mathbf{u} + \mathbf{v}||^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle =$$
$$||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle.$$

The Pythagorean Law:
$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0} \Rightarrow ||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2.$$

Scalar projection of $\mathbf{u}$ on $\mathbf{v} \neq \mathbf{0}$: $\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{v}||}$.

Vector projection of $\mathbf{u}$ on $\mathbf{v} \neq \mathbf{0}$: $\frac{\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}}{\langle \mathbf{v}, \mathbf{v} \rangle}$.

The distance between $\mathbf{u}$ and $\mathbf{v}$ is defined by $||\mathbf{u} - \mathbf{v}||$.

76 Orthonormal sets

Let $\mathbf{V}$ Inner Product Space (IPS). $\mathbf{v_1}, \ldots, \mathbf{v}_n \in \mathbf{V}$ is called an orthogonal set (OS) if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ if $\mathbf{i} \neq j$, i.e. any two vectors in this set is an orthogonal pair.

Theorem. An orthogonal set of nonzero vectors is linearly independent.

$\mathbf{v_1}, \ldots, \mathbf{v}_n \in \mathbf{V}$ is called an orthonormal set (ONS) if $\mathbf{v_1}, \ldots, \mathbf{v}_n$ is an orthogonal set and each $\mathbf{v}_i$ has length 1, i.e. $\mathbf{v_1}, \ldots, \mathbf{v}_n$ ONS $\iff$ $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$ for $i, j = 1, \ldots, n$.

Example: In $\mathbf{C}[-\pi, \pi]$ with $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx$ the set $1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots, \cos nx, \sin nx$ is a nonzero ONS.

An orthonormal basis in $\mathbf{C}[-\pi, \pi]$ is $\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin nx}{\sqrt{\pi}}, \ldots$

77 Fourier series

Every $f(x) \in \mathrm{C}[-\pi, \pi]$ can be expanded in Fourier series

$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos(nx) + b_n \sin(nx))$

$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,dx$

are the even Fourier coefficients of f, and

$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,dx$

are the odd Fourier coefficients of f.

Parseval equality is

$\frac{a_0^2}{2} + \sum_{n=1}^{\infty}\left(a_n^2 + b_n^2\right) = \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|^2\,dx$.

Dirichlet's theorem: If $f \in \mathrm{C}^1((-\infty, \infty))$ and $f(x + 2\pi) = f(x)$, i.e. f is differentiable and periodic, then the Fourier series converge absolutely for each $x \in \mathbb{R}$ to $f(x)$.

This is an infinite version of the identity on p'131 $\mathbf{u} = \sum_{i=1}^{\infty}\langle \mathbf{u}, \mathbf{v}_i\rangle \mathbf{v}_i$ where $\mathbf{v}_1, \ldots, \mathbf{v}_n, \ldots$ is an orthonormal basis in a complete IPS,

Such a complete infinite dimensional IPS is called a Hilbert space.

78 Jean Baptiste Joseph Fourier

Born: 21 March 1768 in Auxerre, Bourgogne, France

Died: 16 May 1830 in Paris, France

It was during his time in Grenoble that Fourier did his important mathematical work on the theory of heat. His work on the topic began around 1804 and by 1807 he had completed his important memoir On the Propagation of Heat in Solid Bodies. The memoir was read to the Paris Institute on 21 December 1807 and a committee consisting of Lagrange, Laplace, Monge and Lacroix was set up to report on the work. Now this memoir is very highly regarded but at the time it caused controversy.

There were two reasons for the committee to feel unhappy with the work. The first objection, made by Lagrange and Laplace in 1808, was to Fourier's expansions of functions as trigonometrical series, what we now call Fourier series. Further clarification by Fourier still failed to convince them.

http://www-history.mcs.st-and.ac.uk/Biographies/Fourier.html

79 J. Peter Gustav Lejeune Dirichlet

Born: 13 Feb 1805 in Dren, French Empire (now Germany)

Died: 5 May 1859 in Göttingen, Hanover (now Germany)

Dirichlet is also well known for his papers on conditions for the convergence of trigonometric series and the use of the series to represent arbitrary functions. These series had been used previously by Fourier in solving differential equations. Dirichlet's work is published in Crelle's Journal in 1828. Earlier work by Poisson on the convergence of Fourier series was shown to be non-rigorous by Cauchy. Cauchy's work itself was shown to be in error by Dirichlet who wrote of Cauchy's paper:-

The author of this work himself admits that his proof is defective for certain functions for which the convergence is, however, incontestable.

Because of this work Dirichlet is considered the founder of the theory of Fourier series.

http://www-history.mcs.st-and.ac.uk/Biographies/Dirichlet.html

80 David Hilbert

Born: 23 Jan 1862 in Königsberg, Prussia (now Kaliningrad, Russia) Died: 14 Feb 1943 in Göttingen, Germany

Today Hilbert's name is often best remembered through the concept of Hilbert space. Irving Kaplansky, writing in [2], explains Hilbert's work which led to this concept:-

Hilbert's work in integral equations in about 1909 led directly to 20th-century research in functional analysis (the branch of mathematics in which functions are studied collectively). This work also established the basis for his work on infinite-dimensional space, later called Hilbert space, a concept that is useful in mathematical analysis and quantum mechanics. Making use of his results on integral equations, Hilbert contributed to the development of mathematical physics by his important memoirs on kinetic gas theory and the theory of radiations.

http://www-history.mcs.st-and.ac.uk/Biographies/Hilbert.html

81 Lecture 4-2-07

DETERMINANTS

For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ determinant of $\boldsymbol{A}$ denoted by $\det \mathbf{A}$, (in Hefferon book $|\boldsymbol{A}| := \det \mathbf{A}$), is a real number such that $\det \mathbf{A} \neq 0 \iff \boldsymbol{A}$ is invertible.

$$(a)\quad \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \mathrm{ad} - \mathrm{bc}.$$

$$(b)\quad \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} =$$

$$\mathrm{aei} + \mathrm{bfg} + \mathrm{cdh} - \mathrm{ceg} - \mathrm{afh} - \mathrm{bdi}$$

A way to remember this formula:

$$\begin{pmatrix} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{pmatrix}$$

The product of diagonals starting from a, b, c, going south west have positive signs, the products of diagonals starting from c, a, b and going south east have negative signs.

(c) The determinant of diagonal matrix, upper triangular matrix and lower triangular is equal to the product of the diagonal entries.

(d) $\det \mathbf{A} \neq \mathbf{0} \iff$ Row Echelon Form of A has the maximal number of possible pivots $\iff$ Reduced Row Echelon Form of A is the identity matrix.

A Is called singular if $\det \mathbf{A} = \mathbf{0}$.

(e) The determinant of a matrix having at least one zero row or column is $\mathbf{0}$.

(f) $\det \mathbf{A} = \det \mathbf{A}^{\mathbf{T}}$: The determinant of A is equal to the determinant of $A^{\mathbf{T}}$.

(g) $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$: The determinant of the product of matrices is equal to the product of determinants.$\Rightarrow$

(h) If A is invertible then $\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}$.

$$I = A^{-1}A \Rightarrow$$

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

We will demonstrate some of these properties later

82 Determinant as multilinear functn

Prop 1:View $A \in \mathbb{R}^{n\times n}$ as composed of n-columns $A = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n]$. Then $\det \mathbf{A}$ is a multilinear function in each column separately. Fix all columns except the column $\mathbf{c}_i$. Let $\mathbf{c}_i = a\mathbf{x} + b\mathbf{y}$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$. Then
$\det [\mathbf{c}_1, \ldots, \mathbf{c}_{i-1}, \mathbf{a}\mathbf{x} + \mathbf{b}\mathbf{y}, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_n] =$
$\mathbf{a} \det [\mathbf{c}_1, \ldots, \mathbf{c}_{i-1}, \mathbf{x}, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_n] +$
$\mathbf{b} \det [\mathbf{c}_1, \ldots, \mathbf{c}_{i-1}, \mathbf{y}, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_n]$ for each $i = 1, \ldots, n$.

Prop 2: $\det \mathbf{A}$ is skew-symmetric, (anti-symmetric): The exchange of any two columns of A changes the sign of determinant. For example:
$\det [\mathbf{c}_2, \mathbf{c}_1, \ldots, \mathbf{c}_n] = -\det [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n]$.
(The skew symmetricity yields that the determinant of A is zero if A has two identical columns)

Prop 3: $\det \mathbf{I}_n = 1$.

Claim: These three properties determine uniquely the determinant function

Remark: Above claims hold for rows as in Hefferon.

83 Examples

$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \mathbf{ad - bc.}$ It is linear in the columns $\mathbf{c_1} = (a, c)^\top, \mathbf{c_2} = (b, d)^\top$ and in the rows $(a, b), (c, d)$.

Clearly $\det \begin{pmatrix} b & a \\ d & c \end{pmatrix} = \det \begin{pmatrix} c & d \\ a & b \end{pmatrix} = \mathbf{bc - ad} = -\det \mathbf{A}$

84 Permutations

Defn: A bijection, i.e. $1-1$ and onto map, $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, is called a permutation of the set $\{1, 2, \ldots, n\}$. The set of all permutations of $\{1, 2, \ldots, n\}$ is called the symmetric group on n-elements, an is denoted by S_n.

$\sigma(i)$ is the image of the number i for $i = 1, \ldots, n$. (Note that $1 \le \sigma(i) \le n$ for $i = 1, \ldots, n$.

$\iota \in \mathrm{S}_n$ is called the identity element, (or map), if $\iota(i) = i$ for $i = 1, \ldots, n$.

Claim The number of elements in S_n is $n! = 1 \cdot 2 \cdots n$.

Proof. $\sigma(1)$ can have n choices: $1, \ldots, n$. $\sigma(2)$ can have all choices: $1, \ldots, n$ except $\sigma(1)$, i.e. $n-1$ choices. $\sigma(3)$ can have all choices except $\sigma(1), \sigma(2)$, i.e. $\sigma(3)$ has $n-3$ choices. Hence total number of σ-s is $n(n-1)\ldots 1 = n!$.

Defn $\tau \in \mathrm{S}_n$ is transposition, if there exists $1 \le i < j \le n$ so that $\tau(i) = j, \tau(j) = i$, and $\tau(k) = k$ for all $k \ne i, j$.

Since $\sigma, \omega \in \mathbf{S}_n$ is bijections, we can compose them $\sigma \circ \omega$ which is an element in $\mathbf{S}_n$, $((\sigma \circ \omega)(i) = \sigma(\omega(i)))$. we denote this composition by $\sigma\omega$ and view this composition as a product in $\mathbf{S}_n$.

Thm. Any $\sigma \in \mathbf{S}_n$ is a product of transpositions. There are many different products of transpositions to obtain σ. All these products of transpositions have the same parity of elements. (Either all products have even number of elements only, or have odd numbers of elements only.

Defn For $\sigma \in \mathbf{S}_n$,

$\mathrm{sgn}(\sigma) = 1$ if σ is a product of even number of transpositions

$\mathrm{sgn}(\sigma) = -1$ if σ is a product of odd number of transpositions

Claim $\mathrm{sgn}(\sigma\omega) = \mathrm{sgn}(\sigma)\mathrm{sgn}(\omega)$.

Prf Express σ and ω as a product of transpositions. Then $\sigma\omega$ is also a product of transpositions. Now count the parity.

85 $\mathbf{S_2}$

$\mathbf{S_2}$ consists of two elements:

(a) the identity ι: $\iota(1) = 1, \iota(2) = 2$

(b) the transposition τ: $\tau(1) = 2, \tau(2) = 1$.

Note $\tau^2 = \tau\tau = \iota$ since
$\tau(\tau(1)) = \tau(2) = 1, \tau(\tau(2)) = \tau(1) = 2$.

So ι is a product of any any even number of τ, i.e.
$\iota = \tau^{2m}$, while $\tau = \tau^{2m+1}$ for $m = 0, 1, \ldots$.

Note that this is true for any transposition $\tau \in \mathbf{S}_n, n \geq 2$.

Thus $\operatorname{sgn}(\iota) = 1, \operatorname{sgn}(\tau) = -1$ for any $n \geq 2$.

86 $\mathbf{S_3}$

$\mathbf{S_3}$ consists of $\mathbf{6}$ elements. Identity: ι.

There are three transpositions in S_3:

$\tau_1(1) = 1, \tau_1(2) = 3, \tau_1(3) = 2$,

$\tau_2(1) = 3, \tau_2(2) = 2, \tau_2(3) = 1$

$\tau_3(1) = 2, \tau_3(2) = 1, \tau_3(3) = 3$.

(τ_j fixes j.)

There are two cyclic permutations

$\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$

$\omega(1) = 3, \omega(2) = 1, \omega(3) = 2$

Note $\omega\sigma = \sigma\omega = \iota$, i.e. $\sigma^{-1} = \omega$.

Show $\sigma = \tau_1\tau_2 = \tau_2\tau_3, \omega = \tau_2\tau_1 = \tau_3\tau_2$.

So $\operatorname{sgn}(\iota) = \operatorname{sgn}(\sigma) = \operatorname{sgn}(\omega) = 1$

$\operatorname{sgn}(\tau_1) = \operatorname{sgn}(\tau_2) = \operatorname{sgn}(\tau_3) = -1$.

87 Rigorous definition of determinant

For

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{m2} & \cdots & a_{nn} \end{pmatrix} \in \mathbb{R}^{n\times n}$$

$$\det \mathbf{A} = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

Note that $\det \mathbf{A}$ has $n!$ summands in the above sum.

88 Cases $n = 2, 3$

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} =$$

$$a_{1\iota(1)}a_{2\iota(2)} - a_{1\tau(1)}a_{2\tau(2)} = a_{11}a_{22} - a_{12}a_{21}$$

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} =$$

$$a_{1\iota(1)}a_{2\iota(2)}a_{3\iota(3)} + a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)} + a_{1\omega(1)}a_{2\omega(2)}a_{3\omega(3)}$$

$$-a_{1\tau_1(1)}a_{2\tau_1(2)}a_{3\tau_1(3)} - a_{1\tau_2(1)}a_{2\tau_2(2)}a_{3\tau_2(3)} - a_{1\tau_3(1)}a_{2\tau_3(2)}a_{3\tau_3(3)} =$$

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$-a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}$$

89 Determinant of $\mathcal{U}T_n$ and $\mathcal{L}T_n$

Thm: The determinant of upper or lower triangular matrix is equal to the product of diagonal entries.

Prf. Let $A = (a_{ij})_{i,j=1}^n$ and assume that A is upper triangular $A \in \mathcal{U}T_n$. So $a_{ij} = 0$ for $i > j$. Let $\sigma \in \mathrm{S}_n$ be a permutation. mcIf $i > \sigma(i)$ then $a_{i\sigma(i)} = 0$. Hence
$f(\sigma) := a_{1\sigma(1)}a_{2\sigma(2)} \dots a_{n\sigma(n)} = 0$ if $i > \sigma(i)$ for some i. So $f(\sigma)$ may not be equal to zero if $i \leq \sigma(i)$ for $i = 1, \dots, n$. This statement is true if only $\sigma = \iota$. Thus

$$\det \mathbf{A} = \mathrm{sgn}(\iota)\mathbf{a}_{1\iota(1)}\mathbf{a}_{2\iota(2)} \dots \mathbf{a}_{n\iota(n)} = \mathbf{a}_{11}\mathbf{a}_{22} \dots \mathbf{a}_{nn}$$

Determinants of Elementary Matrices

(i) $\det E_I = -1$ where E_I corresponds to interchanging two rows: $R_i \leftrightarrow R_j$.

$$\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 0 \cdot 0 - 1 \cdot 1 = -1.$$

(Follows from $\operatorname{sgn}(\tau\sigma) = -\operatorname{sgn}(\sigma)$.)

(j) $\det E_{II} = a$ where E_{II} corresponds to multiplying a row by a: $R_i \to aR_i$. (Note that E_{II} is diagonal.)

$$\det \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} = a.$$ (Follows from multilinearity.)

(k) $\det E_{III} = 1$ where E_{III} corresponds to adding to one row a multiple of another row: $R_i + aR_j \to R_i$. (E_{III} is either upper triangular or lower triangular)

$$\det \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} = 1. \quad (R_2 + aR_1 \to R_2)$$

(Follows from multilinearity and the fact that the determinant of a matrix with two identical rows is equal to zero)

Observe that for any elementary matrix E

$$\det E^{-1} = (\det E)^{-1}$$

Computing Determinants using Elementary Matrices

Let $A \in \mathbb{R}^{n\times n}$ and perform k ERO:

$$A \stackrel{ERO_1}{\longrightarrow} A_1 \stackrel{ERO_2}{\longrightarrow} A_2 \stackrel{ERO_3}{\longrightarrow} \dots A_{k-1} \stackrel{ERO_k}{\longrightarrow} A_k$$

where A_k is a Row Echelon Form of A.

(More general A_k is an upper triangular matrix if we do not force pivots to be equal 1.)

$$A_1 = E_1 A,\ A_2 = E_2 A_1 = E_2 E_1 A, \dots$$

$$A_k = E_k \dots E_1 A \Rightarrow$$

$$A_k = MB,\ M = E_k E_{k-1} \dots E_2 E_1$$

M is invertible matrix since $M^{-1} = E_1^{-1} E_2^{-1} \dots E_k^{-1}$.

$$A = M^{-1} A_k = E_1^{-1} E_2^{-1} \dots E_k^{-1} A_k$$

Since each E_i^{-1} is elementary matrix

$$\det \mathrm{E_i^{-1}} (\mathrm{E_{i+1}^{-1}} \dots \mathrm{E_k^{-1}} \mathrm{A_k}) =$$
$$\det \mathrm{E_i^{-1}} \det \mathrm{E_{i+1}^{-1}} \dots \mathrm{E_k^{-1}} \mathrm{A_k} =$$
$$(\det \mathrm{E_i})^{-1} \det \mathrm{E_{i+1}^{-1}} \dots \mathrm{E_k^{-1}} \mathrm{A_k}$$

Hence

$$\det \mathrm{A} =$$
$$(\det \mathrm{E_1})^{-1} (\det \mathrm{E_2})^{-1} \dots (\det \mathrm{E_k})^{-1} \det \mathrm{A_k} =$$
$$\frac{\det \mathrm{A_k}}{\det \mathrm{E_1} \cdot \det \mathrm{E_2} \cdot \ldots \cdot \det \mathrm{E_k}}$$

90 Example

Find the determinant of $A = \begin{pmatrix} -2 & -1 & -3 \\ 4 & 2 & 1 \\ -6 & 3 & -4 \end{pmatrix}$

Perform the following ERO

$R_2 + 2R_1 \to R_2,\ R_3 - 3R_1 \to R_3$:

$$A_2 = \begin{pmatrix} -2 & -1 & -3 \\ 0 & 0 & -5 \\ 0 & 6 & 5 \end{pmatrix}$$

Perform $R_3 \leftrightarrow R_2$

$$A_3 = \begin{pmatrix} -2 & -1 & -3 \\ 0 & 6 & 5 \\ 0 & 0 & -5 \end{pmatrix}$$

So $\det \mathbf{A}_3 = (-2)(6)(-5) = 60$.

Note that all the elementary matrices corresponding to the above ERO have determinant 1 except $R_3 \leftrightarrow R_2$, with derterminant -1. Hence $\det \mathbf{A} = -60$.

Minors and Cofactors

For $A \in \mathbb{R}^{n\times n}$ the matrix $M_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ denotes the submatrix of A obtained from A by deleting row i and column j. The determinant of M_{ij} is called (i,j)-minor of A. The cofactor A_{ij} is defined to be $(-1)^{i+j}\det \mathbf{M_{ij}}$.

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

$$M_{32} = \begin{pmatrix} a & c \\ d & f \end{pmatrix},$$

$$A_{32} = -af + cd.$$

Expansion of the determinant by row i:

$$\det \mathbf{A} = \mathbf{a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}}$$
$$= \sum_{j=1}^{n} a_{ij}A_{ij}$$

Expansion of the determinant by column p:

$$\det \mathbf{A} = \mathbf{a_{1p}A_{1p} + a_{2p}A_{2p} + \ldots + a_{np}A_{np}}$$
$$= \sum_{j=1}^{n} a_{jp}A_{jp}$$

One can compute also the determinant of A using repeatedly the row or column expansions.

Warning: Computationally the method of using row/column expansion is very inefficient.

Expansion of determinant by row/column is used primarily for theoretical computations.

91 Examples

Expand the determinant of $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ by the second row:

$$\det A = dA_{21} + eA_{22} + fA_{23} =$$

$$d(-1)\det \begin{pmatrix} b & c \\ h & i \end{pmatrix} + e \det \begin{pmatrix} a & c \\ g & i \end{pmatrix} +$$

$$f(-1)\det \begin{pmatrix} a & b \\ g & h \end{pmatrix} =$$

$$(-d)(bi - hc) + e(ai - cg) + (-f)(ah - bg) =$$

$$aei + bfg + cdh - ceg - afh - bdi$$

Find $\det \begin{pmatrix} -1 & 1 & -1 & 3 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ -1 & -1 & -1 & 2 \end{pmatrix}$

Expand by the row or column which has the maximal number of zeros. We expand by the first column:

$\det A = a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31} + a_{41}A_{41} = a_{11}A_{11} + a_{41}A_{41}$ since $a_{21} = a_{31} = 0$ Observe that $(-1)^{1+1} = 1, (-1)^{1+4} = -1$. Hence

$$\det A = (-1)\det \begin{pmatrix} 3 & 1 & 1 \\ 0 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix} +$$

$$(-1)(-1)\det \begin{pmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \\ 0 & 2 & 2 \end{pmatrix}$$ Expand the first determinant by the second row and the second determinant by the third row

$$\mathbf{det\ A} = (-1)\Big(3\ \det\begin{pmatrix} 2 & 2 \\ -1 & 2 \end{pmatrix} +$$

$$(-1)\det\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}\Big) + \Big((-2)\det\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} +$$

$$2\ \det\begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}\Big) = -18 + 16 + 8 = 6$$

Another way to find $\mathbf{det\ A}$,

$$A = \begin{pmatrix} -1 & 1 & -1 & 3 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ -1 & -1 & -1 & 2 \end{pmatrix}$$

Perform ERO: $R_4 - R_1 \to R_4$ to obtain

$$B = \begin{pmatrix} -1 & 1 & -1 & 3 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & -2 & 0 & -1 \end{pmatrix}.$$ So

$\det \mathbf{A} = \det \mathbf{B}$. Expand $\det \mathbf{B}$ by the first column to obtain $\det \mathbf{B} = -\det \mathbf{C}$, $C = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 2 & 2 \\ -2 & 0 & -1 \end{pmatrix}$.

Perform the ERO $R_1 - 0.5R_2 \to R_1$ to obtain

$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 2 \\ -2 & 0 & -1 \end{pmatrix}$$

Expand $\det \mathbf{D}$ by the first row to get

$\det \mathbf{D} = (3)(2 \cdot (-1) - 2 \cdot 0) = -6$.

Hence $\det \mathbf{A} = 6$.

92 Lecture

Adjoint Matrix and Cramer's Rule

For $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$

the adjoint matrix is defined as

$$\text{adj } \mathbf{A} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

where A_{ij} is the (i, j) cofactor of A.

Note that the i-th row of $\text{adj } \mathbf{A}$ is $(A_{1i} \; A_{2i} \ldots A_{ni})$.

Examples:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \text{adj } \mathbf{A} =$$

$$\begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\text{adj } \mathbf{A} = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}$$

$$A_{33} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \mathrm{a}_{11}\mathrm{a}_{22} - \mathrm{a}_{12}\mathrm{a}_{21}$$

$$A_{12} = -\det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} =$$

$$-\mathrm{a}_{21}\mathrm{a}_{33} + \mathrm{a}_{23}\mathrm{a}_{31}.$$

A way to remember to get the adjoint matrix correctly:

$$\mathbf{adj\ A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}^{\mathbf{T}} =$$

$$\begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}$$

The properties of the adjoint matrix:

$$A \text{ adj } \mathbf{A} = (\text{adj}\mathbf{A})\mathbf{A} = (\det \mathbf{A})\mathbf{I},$$

where I is the identity matrix of the corresponding size.

Proof. Consider the (i, j) element of the product A adj $\mathbf{A}$: $a_{i1}A_{j1} + a_{i2}A_{j2} + \ldots + a_{in}A_{jn}$. Assume first that $i = j$. Then this sum is the expansion of the determinant of A by $i - th$ row. Hence it is equal to $\det \mathbf{A}$, which is the (i, i) entry of the diagonal matrix $(\det \mathbf{A})\mathbf{I}$.

Assume now that $i \neq j$. Then the above sum is the expansion of the determinant of a matrix C obtained from A by replacing row j in A by row i of A. Since C has two identical row, hence $\det \mathbf{C} = \mathbf{0}$. This shows $A \text{ adj } \mathbf{A} = (\det \mathbf{A})\mathbf{I}$. Similarly $(\text{adj } \mathbf{A})\mathbf{A} = (\det \mathbf{A})\mathbf{I}$.

Corollary: $\det \mathbf{A} \neq 0 \Rightarrow \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \text{adj } \mathbf{A}$.

93 Example

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$ Find $\mathbf{adj}\ \mathbf{A}$ and A^{-1}.

$A_{11} = 24$, $A_{12} = -0$, $A_{13} = 0$, $A_{21} = -12$, $A_{22} = 6$, $A_{23} = -0$, $A_{31} = 10 - 12 = -2$, $A_{32} = -5$, $A_{33} = 4$, $\mathbf{adj}\ \mathbf{A} =:$

$$\begin{pmatrix} 24 & 0 & 0 \\ -12 & 6 & 0 \\ -2 & -5 & 4 \end{pmatrix}^{\top} = \begin{pmatrix} 24 & -12 & -2 \\ 0 & 6 & -5 \\ 0 & 0 & 4 \end{pmatrix}$$

Since A is upper triangular $\det\ \mathbf{A} = 1 \cdot 4 \cdot 6 = 24$

$$A^{-1} = \frac{1}{24} \begin{pmatrix} 24 & -12 & -2 \\ 0 & 6 & -5 \\ 0 & 0 & 4 \end{pmatrix}$$

Cramer's Rule

Consider the linear system of n equations with n unknowns:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \dots & + & a_{nn}x_n = b_n \end{array}$$

Let $A \in \mathbb{R}^{n\times n}, \mathrm{b} = (b_1, ..., b_n)^{\mathrm{T}}$ be the coefficient matrix and the column vector corresponding to the right-hand side of these system. That is the above system is $A\mathrm{x} = b,\ \mathrm{x} = (x_1, ..., x_n)^{\mathrm{T}}$. Denote by $B_j \in \mathbb{R}^{n\times n}$ the matrix obtained from A by replacing the $j-th$ column in A by: $B_j =$

$$\begin{pmatrix} a_{11} & \dots & a_{1(j-1)} & b_1 & a_{1(j+1)} & \cdots & a_{1n} \\ a_{21} & \dots & a_{2(j-1)} & b_2 & a_{2(j+1)} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n(j-1)} & b_n & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix}$$

Then $x_j = \frac{\det \mathrm{B_j}}{\det \mathrm{A}}$ for $j = 1, ..., n$.

Proof of Cramer's Rule:

Since $\det \mathbf{A} \neq 0$, $A^{-1} = \frac{1}{\det \mathbf{A}} \operatorname{adj} \mathbf{A}$. Hence the solution to the system $A\mathbf{x} = \mathbf{b}$ is given by:
$A^{-1}\mathbf{x} = \frac{1}{\det \mathbf{A}} \operatorname{adj} \mathbf{A}\ \mathbf{b}$. Writing down the formula for the matrix $\operatorname{adj} \mathbf{A}$ we get:
$x_j = \frac{A_{1j}b_1 + A_{2j}b_2 + \ldots + A_{nj}b_n}{\det \mathbf{A}}$.

The numerator of this quotient is the expansion of $\det \mathbf{B_j}$ by the column j. □

Example: Find the value of x_2 in the system

$$\begin{aligned} x_1 &+ 2x_2 - x_3 = 0 \\ -2x_1 &- 5x_2 + 5x_3 = 3 \\ 3x_1 &+ 7x_2 - 5x_3 = 0 \end{aligned}$$

$$x_2 = \frac{\det \begin{pmatrix} 1 & 0 & -1 \\ -2 & 3 & 5 \\ 3 & 0 & -5 \end{pmatrix}}{\det \begin{pmatrix} 1 & 2 & -1 \\ -2 & -5 & 5 \\ 3 & 7 & -5 \end{pmatrix}}$$

Expand the determinant of the denominator by the second column to obtain $\det \begin{pmatrix} 1 & 0 & -1 \\ -2 & 3 & 5 \\ 3 & 0 & -5 \end{pmatrix} =$

$$3 \det \begin{pmatrix} 1 & -1 \\ 3 & -5 \end{pmatrix} = 3(-5+3) = -6$$

on the coefficient matrix $A = \begin{pmatrix} 1 & 2 & -1 \\ -2 & -5 & 5 \\ 3 & 7 & -5 \end{pmatrix}$

Perform the ERO

$R_1 + 3R_2 \to R_2, R_2 - 3R_1 \to R_3$ to obtain

$A_2 = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 3 \\ 0 & 1 & -2 \end{pmatrix}$. Expand $\det \mathbf{A_2}$ by the first column to obtain

$\det \mathbf{A} = \det \mathbf{A_2} = 1(2-3) = -1$. So $x_2 = 6$.

(Note that A^{-1} was computed on p' 100. Check the answer by comparing it to $A^{-1}(0,3,0)^\top = (-9,6,3)^\top$.)

94 History of determinants

Historically, determinants were considered before matrices. Originally, a determinant was defined as a property of a system of linear equations. The determinant ”determines” whether the system has a unique solution (which occurs precisely if the determinant is non-zero). In this sense, two-by-two determinants were considered by Cardano at the end of the 16th century and larger ones by Leibniz about 100 years later. Following him Cramer (1750) added to the theory, treating the subject in relation to sets of equations.

It was Vandermonde (1771) who first recognized determinants as independent functions. Laplace (1772) gave the general method of expanding a determinant in terms of its complementary minors: Vandermonde had already given a special case. Immediately following, Lagrange (1773) treated determinants of the second and third order. Lagrange was the first to apply determinants to questions outside elimination theory; he proved many special cases of general identities.

Gauss (1801) made the next advance. Like Lagrange, he made much use of determinants in the theory of numbers. He introduced the word determinants (Laplace had used resultant), though not in the present signification, but rather as applied to the discriminant of a quantic. Gauss also arrived at the notion of reciprocal (inverse) determinants, and came very near the multiplication theorem.

The next contributor of importance is Binet (1811, 1812), who formally stated the theorem relating to the product of two matrices of m columns and n rows, which for the special case of m = n reduces to the multiplication theorem. On the same day (Nov. 30, 1812) that Binet presented his paper to the Academy, Cauchy also presented one on the subject. (See Cauchy-Binet formula.) In this he used the word determinant in its present sense, summarized and simplified what was then known on the subject, improved the notation, and gave the multiplication theorem with a proof more satisfactory than Binet's. With him begins the theory in its generality.

Source:

http://en.wikipedia.org/wiki/Determinant

(See section History)

95 Eigenvalues and Eigenvectors

Let $\mathbb{C}$ be the field of complex numbers. Let $A \in \mathbb{C}^{n \times n}$. $\mathbf{x} \in \mathbb{C}^n$ is called an eigenvector (characteristic vector) if $\mathbf{x} \neq \mathbf{0}$ and there exists $\lambda \in \mathbb{C}$ such that $A\mathbf{x} = \lambda \mathbf{x}$. λ is called an eigenvalue (characteristic value of A.

Claim: λ is an eigenvalue of A if and only if $\det\,(\mathbf{A} - \lambda \mathbf{I}) = 0$.

The polynomial $p(\lambda) := \det\,(\mathbf{A} - \lambda \mathbf{I})$ is called a characteristic polynomial of A.

$$p(\lambda) = (-1)^n(\lambda^n - \sigma_1\lambda^{n-1} + \sigma_2\lambda^{n-2} + \ldots + (-1)^n\sigma_n)$$

is a polynomial of degree n. The fundamental theorem of algebra states that $p(\lambda)$ has n roots (eigenvalues) $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $p(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$.

Given an eigenvalue λ then a basis to the null space $\mathrm{N}(A - \lambda I)$ is a basis for the eigenspace of eigenvectors of A corresponding to λ.

96 Example 1

Consider the Markov chain given by

$$A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}$$

($\%70$ of Healthy remain Healthy and $\%20$ of Sick recover.)

$$A - \lambda I = \begin{pmatrix} 0.7 - \lambda & 0.2 \\ 0.3 & 0.8 - \lambda \end{pmatrix}$$

$\det\,(\mathbf{A} - \lambda\mathbf{I}) = (0.7 - \lambda)(0.8 - \lambda) - 0.2 \cdot 0.3 = \lambda^2 - 1.5\lambda + 0.5$ is the characteristic polynomial of A.

$\det\,(\mathbf{A} - \lambda\mathbf{I}) = (\lambda - 1)(\lambda - 0.5)$.

Eigenvalues of A are the zeros of the characteristic polynomial, i.e. solutions of $\det\,(\mathbf{A} - \lambda\mathbf{I}) = 0$:

$\lambda_1 = 1,\ \lambda_2 = 0.5$.

To find a basis for the null space of $A - \lambda_1 I = A - I$ denoted by $N(A - \lambda_1 I)$ we need to bring the matrix $A - I$ to RREF:

$$A - I = \begin{pmatrix} -0.3 & 0.2 \\ 0.3 & -0.2 \end{pmatrix}$$ So

$$B = \begin{pmatrix} 1 & -\frac{2}{3} \\ 0 & 0 \end{pmatrix}$$ is RREF of $A - I$.

$N(B)$ corresponds to the system $x_1 - \frac{2}{3}x_2 = 0$. Since x_1 is a lead variable and x_2 is free $x_1 = \frac{2x_2}{3}$. By choosing $x_2 = 1$ we get the eigenvector $\mathbf{x}_1 = (\frac{2}{3}, 1)^\top$ which corresponds to the eigenvalue 1.

Note that the steady state of the Markov chain corresponds to the coordinates of $\mathbf{x}_1$. More precisely the ratio of Heathy to Sick is $\frac{x_1}{x_2} = \frac{2}{3}$.

To find a basis for the null space of $A - \lambda_2 I = A - 0.5I$ denoted by $N(A - \lambda_2 I)$ we need to bring the matrix $A - 0.5I$ to RREF:

$$A - 0.5I = \begin{pmatrix} 0.2 & 0.2 \\ 0.3 & 0.3 \end{pmatrix} \text{ So } C = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

is RREF of $A - 0.5I$.

$N(C)$ corresponds to the system $x_1 + x_2 = 0$. Since x_1 is a lead variable and x_2 is free $x_1 = -x_2$. By choosing $x_2 = 1$ we get the eigenvector $\mathbf{x}_2 = (-1, 1)^\top$ which corresponds to the eigenvalue 0.5.

97 Example 2

Let $A = \begin{pmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix}$.

So $A - \lambda I = \begin{pmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{pmatrix}$

Expand $\det\,(\mathbf{A} - \lambda\mathbf{I})$ by the first row:

$$(2-\lambda)\big((-2-\lambda)(2-\lambda)+3\big)+$$
$$(-1)(-3)\big(1(2-\lambda)-1\big)+1\big(-3+(2+\lambda)\big)=$$
$$(2-\lambda)(\lambda^2-1)+3(1-\lambda)+(\lambda-1)=$$
$$(\lambda-1)\big((2-\lambda)(\lambda+1)-3+1\big)=$$
$$(\lambda-1)(-\lambda^2+\lambda)=-\lambda(\lambda-1)^2$$

$\lambda_1 = 0$ is a simple root and $\lambda_2 = 1$ is a double root.

$$\boldsymbol{A} - \boldsymbol{\lambda_1 I} = \boldsymbol{A} = \begin{pmatrix} \mathbf{2} & \mathbf{-3} & \mathbf{1} \\ \mathbf{1} & \mathbf{-2} & \mathbf{1} \\ \mathbf{1} & \mathbf{-3} & \mathbf{2} \end{pmatrix}$$ RREF of $\boldsymbol{A}$ is

$$\boldsymbol{B} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{1} & \mathbf{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$ The null space $\boldsymbol{N(B)}$ given by

$\boldsymbol{x_1 = x_3, x_2 = x_3}$, where $\boldsymbol{x_3}$ is the free variable. Set $\boldsymbol{x_3 = 1}$ to obtain that $\mathbf{x_1} = \mathbf{(1, 1, 1)}^\top$ is an eigenvector corresponding to $\boldsymbol{\lambda_1 = 0}$.

$$A - \lambda_2 I = \begin{pmatrix} 1 & -3 & 1 \\ 1 & -3 & 1 \\ 1 & -3 & 1 \end{pmatrix}$$ RREF of $A - \lambda_2 I$ is

$$B = \begin{pmatrix} 1 & -3 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$ The null space $N(B)$ given

by $x_1 = 3x_2 - x_3$, where x_2, x_3 are the free variable.

Set $x_2 = 1, x_3 = 0$ to obtain that $\mathbf{x_2} = (3, 1, 0)^\top$.

Set $x_2 = 0, x_3 = 1$ to obtain that $\mathbf{x_3} = (-1, 0, 1)^\top$.

so $\mathbf{x_2}, \mathbf{x_3}$ are two (linearly independent) eigenvectors corresponding to the double zero $\lambda_2 = 1$.

98 Similarity

Definition. Let $\mathbf{V}$ be a vector space with a basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$. Let $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation. Then the representation matrix $A = [\mathbf{a}_1\ \mathbf{a}_2 \ldots \mathbf{a}_n] \in \mathbb{R}^{n \times n}$ of T in the basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ is given as follows: The column j of A, denoted by $\mathbf{a}_j \in \mathbb{R}^n$, is the coordinate vector of $T(\mathbf{v}_j)$. That is $T(\mathbf{v}_j) = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]\mathbf{a}_j$ for $j = 1, \ldots, n$.

Change a basis in $\mathbf{V}$:
$[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n] \xrightarrow{Q} [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$. Then the representation matrix of T in the bases $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ is given by the matrix QAQ^{-1}.

Definition. $A, B \in \mathbb{R}^{n \times n}$ are called similar if $B = QAQ^{-1}$ for some invertible matrix $Q \in \mathbb{R}^{n \times n}$.

Definition. For $A \in \mathbb{R}^{n \times n}$ trace of A is the sum of the diagonal elements of A.

Claim. Two similar matrices A and B have the same trace and the same determinant.

Expressing $\det\,(\mathbf{A}-\lambda\mathbf{I})$ as sum of $n!$ product of elements of $A-\lambda I$ (p' 160) we get $\det\,(\mathbf{A}-\lambda\mathbf{I}) = (-1)^{\mathbf{n}}\lambda^{\mathbf{n}} + (-1)^{\mathbf{n}-1}\,\mathrm{tr}\,\mathbf{A}\,\lambda^{\mathbf{n}-1} + \ldots + \det\,\mathbf{A}$.

Hence

$$\mathrm{tr}\,A := a_{11} + a_{22} + \ldots + a_{nn} = \lambda_1 + \lambda_2 + \ldots + \lambda_n.$$

$$\det\,\mathbf{A} = \lambda_1\lambda_2\ldots\lambda_{\mathbf{n}}.$$

Two matrices A, B in $\mathbb{C}^{n\times n}$ similar if $B = QAQ^{-1}$ for some invertible $Q \in \mathbb{C}^{n\times n}$.

Claim: Similar matrices have the same characteristic polynomial.

$$B = QAQ^{-1} \Rightarrow B-\lambda I = Q(A-\lambda I)Q^{-1} \Rightarrow$$
$$\det\,(\mathbf{B}-\lambda\mathbf{I}) =$$
$$\det\,\mathbf{Q}\,\det\,(\mathbf{A}-\lambda\mathbf{I})\,\det\,\mathbf{Q}^{-1} = \det\,(\mathbf{A}-\lambda\mathbf{I}).$$

Hence two similar matrices have the same trace and determinant.

Claim: Suppose that $A, B \in \mathrm{M}_n(\mathbb{C})$ have the same characteristic polynomial $p(\lambda)$. If $p(\lambda)$ has n distinct roots then A and B are similar.

99 Examples

Suppose that A is upper triangular. Hence $A - \lambda I$ is also upper triangular. Thus

$$\det\,(\mathbf{A} - \lambda\mathbf{I}) = (\mathbf{a}_{11} - \lambda)(\mathbf{a}_{22} - \lambda)\ldots(\mathbf{a}_{nn} - \lambda)$$

(See p' 162). The eigenvalues of upper or lower triangular matrix are given by its diagonal entries, (counted with multiplicities!)

Example: $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$, $A - \lambda I =$

$$\begin{pmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{pmatrix}$$

$$\det\,(\mathbf{A} - \lambda\mathbf{I}) = (\mathbf{a}_{11} - \lambda)(\mathbf{a}_{22} - \lambda)(\mathbf{a}_{33} - \lambda)$$

In particular, the eigenvalues of the diagonal matrices are given by its diagonal entries, (counted with multiplicities!)

Let $A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}$ (p' 183) Recall

$\det (\mathbf{A} - \lambda \mathbf{I}) = (1 - \lambda)(0.5 - \lambda)$. Let

$D = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$. So

$\det (\mathbf{A} - \lambda \mathbf{I}) = \det (\mathbf{D} - \lambda \mathbf{I})$

We show that A and D are similar. Recall that

$A\mathbf{x}_1 = \mathbf{x}_1, A\mathbf{x}_2 = 0.5\mathbf{x}_2$. Let

$X = (\mathbf{x}_1 \ \mathbf{x}_2) = \begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix}$

So $AX = XD$ (Check it!. This is equivalent to the fact that $\mathbf{x}_1, \mathbf{x}_2$ are the corresponding eigenvectors) As $\det \mathbf{X} = \frac{5}{3} \neq 0$ X is invertible and $A = XDX^{-1}$. So A and X are similar.

This demonstrate the claim on the bottom of page 190.

100 Matrices nonsimilar to diagonal m.

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

Both matrices are upper triangular so
$\det (A - \lambda I) = \det (B - \lambda I) = \lambda^2$.

Since $TAT^{-1} = 0 = A \neq B$, A and B are not similar.

Claim: B is not similar to a diagonal matrix

Proof Suppose B similar to $D = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. As
$\det (B - \lambda I) = \lambda^2 = \det (D - \lambda I) =$
$(a - \lambda)(b - \lambda)$ we must have $a = b = 0$, i.e.
$D = A$. We showed above that A and B are not similar.

101 Defective matrices

Defn λ_0 is called a defective eigenvalue of $B \in \mathbb{C}^{n\times n}$ if the multiplicity of λ_0 in $\det\,(\mathrm{B}-\lambda\mathrm{I})\ (=0)$ is strictly greater than $\dim N(B-\lambda_0 I)$.

$B \in \mathbb{C}^{n\times n}$ is called defective if it has at least one defective eigenvalue.

Note that $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is defective since the only eigenvalue $\lambda_0 = 0$ is defective: $\operatorname{rank}(B-0I) = \operatorname{rank} B = 1, \dim N(B) = 2 - \operatorname{rank} B = 1$, since the multiplicity of $\lambda_0 = 0$ in $\det\,(\mathrm{B}-\lambda\mathrm{I}) = \lambda^2$ is 2.

Definition $A \in \mathbb{C}^{n\times n}$ is called diagonable if A is similar to a diagonal matrix $D \in \mathbb{C}^{n\times n}$. (The diagonal entries of D are the eigenvalues of A counted with multiplicities.)

Diagonability Thm: $B \in \mathbb{C}^{n\times n}$ is diagonable matrix if and only if B is not defective.

Note that $A \in \mathbb{R}^{3\times 3}$ given on p' 186 is not defective, hence according to the above Theorem A is diagonable.

102 Proof of Diagonability Thm

1. Let $D = \mathrm{diag}(d_1, \ldots, d_n) =$
$\mathrm{diag}(d_1, d_2, ..., d_n) =$

$$\begin{pmatrix} d_1 & 0 & \ldots & 0 & 0 \\ 0 & d_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & d_n \end{pmatrix}$$ Then

$$\det\,(\mathrm{D} - \lambda \mathrm{I}) = (\mathrm{d}_1 - \lambda)(\mathrm{d}_2 - \lambda)\ldots(\mathrm{d}_\mathrm{n} - \lambda)$$

The eigenvalues of D are the diagonal entries. The multiplicity m of the eigenvalue λ_0 is the number of times it appears on the diagonal entry.

The matrix $D - \lambda_0 I$ has exactly m zero elements on the diagonal. Each nonzero diagonal entry can be made to a pivot in RREF of B. Hence $\mathrm{rank}\, B = n - m$ and $\dim N(B) = \mathrm{nul}\, B = n - \mathrm{rank}\, B = m$. So λ_0 is non-defective.

Thus each eigenvalue of a diagonal matrix is non-defective

103 Example

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \det\,(\mathbf{D} - \lambda\mathbf{I}) =$$

$-\lambda(1-\lambda)^2$.

$\lambda_0 = 0$ is a simple eigenvalue of D. $\operatorname{rank} D = 2$ since RREF of A is $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\operatorname{nul} A = 1$.

$\lambda_1 = 1$ is a double eigenvalue of D.

$\operatorname{rank}(D - I) = 1$ since RREF of $A - I$ is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. So $\operatorname{nul}(A - I) = 2$.

2. Claim: If $A, B \in \mathbb{C}^{n\times n}$ are similar then for each $\lambda \in \mathbb{C}$ $\mathbf{nul}(A - \lambda I) = \mathbf{nul}(B - \lambda I)$.

Proof. Assume that $B = TAT^{-1}$ so $B - \lambda I = T(A - \lambda I)T^{-1}$. Hence $T^{-1}N(B - \lambda I) = N(A - \lambda I)$. (Check this claim by computation using the fact that T is invertible.)

Since similar matrices have the same characteristic polynomial we deduce:

Corollary: Each eigenvalue of a diagonable matrix is non defective. Hence a diagonable matrix is not defective.

3. Lemma: Let $\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_p}$ be p eigenvectors of A corresponding to p distinct eigenvalues. Then $\mathbf{y_1}, \ldots, \mathbf{y_p}$ are linearly independent.

Proof: By induction on p.

$p = 1$: By definition an eigenvector $\mathbf{y}_1 \neq \mathbf{0}$. Hence $\mathbf{y}_1$ l.i.

$p = k$: Assume that Lemma holds.

$p = k + 1$. Assume that
$A\mathbf{y}_i = \lambda_i \mathbf{y}_i, \mathbf{y}_i \neq \mathbf{0}, i = 1, \ldots, k+1$ and
$\lambda_i \neq \lambda_j$ for $i \neq j$. Suppose that
$a_1\mathbf{y}_1 + \ldots + a_k\mathbf{y}_k + a_{k+1}\mathbf{y}_{k+1} = \mathbf{0}$. (*)

So

$$A\mathbf{0} = \mathbf{0} = A(a_1\mathbf{y}_1 + \ldots + a_k\mathbf{y}_k + a_{k+1}\mathbf{y}_{k+1}) =$$
$$a_1A\mathbf{y}_1 + \ldots + a_kA\mathbf{y}_k + a_{k+1}A\mathbf{y}_{k+1} =$$
$$a_1\lambda_1\mathbf{y}_1 + \ldots + a_k\lambda_k\mathbf{y}_k + a_{k+1}\lambda_{k+1}\mathbf{y}_{k+1}$$

Mulitply (*) by λ_{k+1} and subtract it from the last equality above to get

$$a_1(\lambda_1 - \lambda_{k+1})\mathbf{y}_1 + \ldots + a_k(\lambda_k - \lambda_{k+1})\mathbf{y}_k = \mathbf{0}$$

The induction hypothesis implies that
$a_i(\lambda_i - \lambda_{k+1}) = 0$ for $i = 1, \ldots, k$. Since
$\lambda_i - \lambda_{k+1} \neq 0$ for $i < k+1$ we get
$a_i = 0, i = 1, \ldots, k$. Use these equalities in (*) to obtain $a_{k+1}\mathbf{y}_{k+1} = \mathbf{0} \Rightarrow a_{k+1} = 0$. So
$\mathbf{y}_1, \ldots, \mathbf{y}_{k+1}$ are linearly independent. □

104 Diagonalization Thm

4. Theorem Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det\,(\mathbf{A}-\lambda\mathbf{I}) = (\lambda_1-\lambda)^{m_1}(\lambda_2-\lambda)^{m_2}\ldots(\lambda_k-\lambda)^{m_k}$, where $\lambda_i \neq \lambda_j$ for $i \neq j$ and $1 \leq m_i$ (the multiplicity of λ_i). Assume that $\dim \mathrm{N}(A-\lambda_i I) = m_i$ and $\mathrm{N}(A-\lambda_i I) = \mathrm{span}(\mathrm{x}_{i1},\ldots,\mathrm{x}_{im_i})$ for $i = 1,\ldots,k$.

(This is equivalent to the assumption that A is not defective.)

Form the matrix whose columns are the vectors which span the null spaces $X = (\mathrm{x}_{11}\ldots\mathrm{x}_{1m_1}\ \mathrm{x}_{21}\ldots\mathrm{x}_{2m_2}\ldots\mathrm{x}_{km_k}) \in \mathbb{C}^{n\times n}$ and the diagonal matrix whose entries are the eigenvalues of A: $D = \mathrm{diag}(\lambda_1\ \ldots\lambda_k)$, where the diagonal entry λ_i repeats m_i times for $i = 1, ..., k$.

Then X is an invertible matrix and $A = XDX^{-1}$, i.e. A is similar to D.

105 Proof

We claim that columns of X are l.i. Assume to the contrary that $\sum_{i=1}^{k}\sum_{j=1}^{m_i} b_{ij}\mathbf{x}_{ij} = 0$ (*)

and not all $b_{ij} = 0$. Let $1 \le i_1 < \ldots < i_p \le k$ be the set of all i such that $\mathbf{y}_i := \sum_{j=1}^{m_i} b_{ij}\mathbf{x}_{ij} \neq 0$. Since $\mathbf{x}_{i1}, \ldots, x_{im_i}$ are l.i. this assumption equivalent to the assumption that the equality $b_{i1} = \ldots = b_{im_i} = 0$ does not hold. (So $p \ge 1$.) Hence (*) is equivalent to $\mathbf{y}_{i_1} + \mathbf{y}_{i_1} + \ldots + \mathbf{y}_{i_p} = 0$. Note that our assumptions imply that $\mathbf{y}_{i_l}$ is an eigenvector corresponding to λ_{i_l}. Since $\lambda_{i_l} \neq \lambda_{i_m}$ for $l \neq m$ we ge a contradiction to Lemma on p' 197. Hence all the columns of X are l.i.. So X is invertible.

A straightforward calculation shows $AX = XD$. As X is invertible $A = XDX^{-1}$. □

106 Corollaries

Corollary 1: Let $A \in \mathbb{C}^{n\times n}$ and assume that the characteristic polynomial of A has only simple roots. Then A is diagonable

Proof. So
$\det(\mathbf{A} - \lambda \mathbf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_n - \lambda)$,
where $\lambda_i \neq \lambda_j$ for $i \neq j$. Since $\det(\mathbf{A} - \lambda_i \mathbf{I}) = 0$ let $A\mathbf{y}_i = \lambda_i \mathbf{y}_i, \mathbf{y}_i \neq \mathbf{0}$. Lemma on p' 197 yields that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ l.i.. Let $X = (\mathbf{y}_1\ \mathbf{y}_2 \ldots \mathbf{y}_n) \in \mathbb{C}^{n\times n}$. So X is invertible. As above $AX = XD$ where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. So $A = XDX^{-1}$. □

Corollary Assume that $A, B \in \mathbb{C}^{n\times n}$ and suppose that $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{B} - \lambda\mathbf{I})$, (i.e. A and B have the same characteristic polynomial. If $p(\lambda)$ have simple roots then A and B are similar.

Proof Let D be the diagonal matrix as in Corollary 1. So
$A = XDX^{-1}, B = YDY^{-1} \Rightarrow$
$A = (XY^{-1})B(XY^{-1})^{-1}$.

107 Examples

1. See example on page 192

2. Let $\boldsymbol{A} = \begin{pmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix}$. (p' 186)

$\mathbf{det}\ (\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}) = -\boldsymbol{\lambda}(\boldsymbol{\lambda} - \mathbf{1})^{\mathbf{2}}$.

$\boldsymbol{X} = (\mathbf{x_1}\ \mathbf{x_2}\ \mathbf{x_3}) = \begin{pmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1} = \begin{pmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix}$$

Powers of diagonable matrices

$A = XDX^{-1} \Rightarrow A^m = XD^mX^{-1}$,

$D^m = \text{diag}(\lambda_1^m \dots \lambda_n^m),\ m = 1, \dots$

Iteration process:

$\mathbf{x}_m = A\mathbf{x}_{m-1},\ m = 1, \dots \Rightarrow \mathbf{x}_m = A^m\mathbf{x}_0$.

Under what conditions $\mathbf{x}_m$ converges to $\mathbf{x} := \mathbf{x}(\mathbf{x}_0)$?

If A is diagonable then $\mathbf{x}_m$ converges to $\mathbf{x}$ for all $\mathbf{x}_0$ if and only if each eigenvalue of A either $|\lambda| < 1$ or $\lambda = 1$.

Markov Chains: $A \in \mathbb{R}^{n\times n}$ is called column (row) stochastic if all entries of A are nonnegative and the sum of each column (row) is 1. That is $A^\mathrm{T}\mathbf{e} = \mathbf{e}$, ($A\mathbf{e} = \mathbf{e}$), where $\mathbf{e} = (1, 1, \dots, 1)^\mathrm{T}$. Under mild assumptions, e.g. all entries of A are positive $\lim_{m\to\infty} A^m\mathbf{x}_0 = \mathbf{x}$. If A is column stochastic and $\mathbf{e}^\mathrm{T}\mathbf{x}_0 = 1$ then the limit vector is a unique probability eigenvector of A:

$A\mathbf{x} = \mathbf{x}, \quad \mathbf{x} = (x_1, \dots, x_n)^\mathrm{T}$,

$0 < x_1, \dots, x_n, \quad x_1 + x_2 + \dots x_n = 1$.

108 Examples

1. See example on page 192: $A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} =$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix}$$

$$A^k = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}^k =$$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1^k & 0 \\ 0 & (0.5)^k \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix}$$

$$\lim_{k \to \infty} A^k = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix} =$$

$\begin{pmatrix} \frac{2}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{3}{5} \end{pmatrix}$ (columns give proportions of healthy and sick)

2. From page $\mathbf{202}$ $\boldsymbol{A^k = A}$ since $\mathbf{diag(0,1,1)^k} =$ $\mathbf{diag(0^k, 1^k, 1^k) = diag(0,1,1)}$.

(This follows also from the straightforward computation $\boldsymbol{A^2 = A}$.

$\boldsymbol{A}$ is called projection, or involution if $\boldsymbol{A^2 = A}$.

For projection $\boldsymbol{\lim_{k\to\infty} A^k = A}$.

Systems of linear ordinary differential equations (SOLODE)

$$\begin{array}{ccccccccc} y_1' & = & a_{11}y_1 & + & a_{12}y_2 & +\dots+ & a_{1n}y_n \\ y_2' & = & a_{21}y_1 & + & a_{22}y_2 & +\dots+ & a_{2n}y_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\vdots & \\ y_n' & = & a_{n1}y_1 & + & a_{n2}y_2 & +\dots+ & a_{1n}y_n \end{array}$$

In matrix terms we write: $\mathbf{y}' = A\mathbf{y}$, where $\mathbf{y} = \mathbf{y}(t) = (y_1(t), y_2(t), ..., y_n(t))^\mathbf{T}$ and $A \in \mathbb{C}^{n\times n}$ a constant matrix.

We guess a solution of the form $\mathbf{y}(t) = e^{\lambda t}\mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{C}^n$ is a constant vector. we assume that $\mathbf{x} \neq \mathbf{0}$, otherwise we have a constant non-interesting solution $\mathbf{x} = \mathbf{0}$. Then $\mathbf{y}' = (e^{\lambda t})'\mathbf{x} = \lambda e^{\lambda t}\mathbf{x}$. The system $\mathbf{y}' = A\mathbf{y}$ is equivalent to $\lambda e^{\lambda t}\mathbf{x} = A(e^{\lambda t}\mathbf{x})$. Since $e^{\lambda t} =\neq \mathbf{0}$ divide by $e^{\lambda t}$ to get $A\mathbf{x} = \lambda\mathbf{x}$.

Corollary: If $\mathbf{x}(\neq \mathbf{0})$ is an eigenvector of A corresponding to the eigenvalue λ then $\mathbf{y}(t) = e^{\lambda t}\mathbf{x}$ is a nontrivial solution of the given SOLODE.

Theorem Assume that $A \in \mathbb{C}^{n\times n}$ is diagonable:

$\det(\mathbf{A} - \lambda\mathbf{I}) =$
$(\lambda_1 - \lambda)^{m_1}(\lambda_2 - \lambda)^{m_2} \ldots (\lambda_k - \lambda)^{m_k}$, where
$\lambda_i \neq \lambda_j$ for $i \neq j$, $1 \leq m_i$ (the multiplicity of λ_i), and
$\dim \mathrm{N}(A - \lambda_i I) = m_i$,
$\mathrm{N}(A - \lambda_i) = \mathrm{span}(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i})$ for
$i = 1, \ldots, k$. Then the general solution of SOLODE is:
$\mathbf{y}(t) = \sum_{i=1,j=1}^{k,m_i} C_{ij} e^{\lambda_i (t-t_0)} \mathbf{x}_{ij}$.

$\mathbf{y}(t)$ is determined by the initial condition $\mathbf{y}(t_0) = \mathbf{c}$.

109 Examples

1.

$$\begin{aligned} y_1' &= 0.7y_1 + 0.2y_2 \\ y_2' &= 0.3y_1 + 0.8y_2 \end{aligned}$$

The right-hand side is given by $A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}$

which was studies on p' 183.

$\det(\mathbf{A} - \lambda \mathbf{I}) = (\lambda - 1)(\lambda - 0.5)$.

$A\mathbf{x}_1 = \mathbf{x}_1, A\mathbf{x}_2 = 0.5\mathbf{x}_2,$

$\mathbf{x}_1 = (\frac{2}{3}, 1)^\top, \mathbf{x}_2 = (-1, 1)^\top$.

The general solution of the system

$\mathbf{y}(t) = c_1 e^t \mathbf{x}_1 + c_2 e^{0.5t}\mathbf{x}_2$:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = c_1 e^t \begin{pmatrix} \frac{2}{3} \\ 1 \end{pmatrix} + c_2 e^{0.5t} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$y_1(t) = \frac{2c_1 e^t}{3} - c_2 e^{0.5t}$

$y_2(t) = c_1 e^t + c_2 e^{0.5t}$

2.

$$\begin{aligned} y_1' &= 2y_1 - 3y_2 + y_3 \\ y_2' &= y_1 - 2y_2 + y_3 \\ y_3' &= y_1 - 3y_2 + 2y_3 \end{aligned}$$

$$A = \begin{pmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix}$$ as on p' 202

$$\det\ (\mathbf{A} - \lambda\mathbf{I}) = -\lambda(\lambda - 1)^2$$

$$\lambda_1 = 0, \lambda_2 = \lambda_3 = 1$$

$$X = (\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3) = \begin{pmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

General solution $\mathbf{y}(t) = c_1 e^0 \mathbf{x}_1 + c_2 e^t \mathbf{x}_2 + c_3 e^t \mathbf{x}_3$:

$$y_1(t) = c_1 + 3c_2 e^t - c_3 e^t$$

$$y_2(t) = c_1 + c_2 e^t$$

$$y_3(t) = c_1 \qquad\qquad + c_3 e^t$$

110 Initial conditions

$\mathbf{y}(0) = \mathbf{y}_0^\top$ are equivalent always to $X\mathbf{c} = \mathbf{y}_0$.

Solve this system either by Gauss elimination or

$\mathbf{c} = X^{-1}\mathbf{y}_0$.

Example 1: In the system of ODE on page 208 find the solution satisfying IC $\mathbf{y}(0) = (1, 2)^\top$.

Solution This condition is equivalent to

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{9}{5} \\ \frac{1}{5} \end{pmatrix}$$

(The inverse is taken from page 204)

Now substitute these values of c_1, c_2 in the formulas on p' 208.

Complex eigenvalues of real matrices

Claim: Let $A \in \mathbb{R}^{n \times n}$ and assume $\lambda := \alpha + i\beta,\ \alpha, \beta \in \mathbb{R}$ is non-real eigenvalue ($\beta \neq 0$). Then the corresponding eigenvector $\mathbf{x} = \mathbf{u} + i\mathbf{v},\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ($A\mathbf{u} = \lambda\mathbf{u}$) is non-real ($\mathbf{v} \neq \mathbf{0}$). Furthermore $\overline{\lambda} = \alpha - i\beta \neq \lambda$ is another eigenvalue of A with the corresponding eigenvector $\overline{\mathbf{x}} = \mathbf{u} - i\mathbf{v}$.

The corresponding contributions of the above two complex eigenvectors to the solution of $\mathbf{y}' = A\mathbf{y}$ is

$$e^{\alpha t} C_1 (\cos(\beta t)\mathbf{u} - \sin(\beta t)\mathbf{v}) +$$
$$e^{\alpha t} C_2 (\sin(\beta t)\mathbf{u} + \cos(\beta t)\mathbf{v}).$$

These two solutions can be obtained by considering the real linear combination of the real and the imaginary part of the complex solution $e^{\lambda t}\mathbf{x}$.

Recall the Euler's formula for e^z where $z = a + ib,\ a, b \in \mathbb{R}$:

$$e^z = e^{a+ib} = e^a e^{ib} = e^a(\cos b + i \sin b).$$

Second Order Linear Differential Systems

$\mathbf{y}'' = A_1\mathbf{y} + A_2\mathbf{y}'$,

$$A_1, A_2 \in \mathbb{C}^{n\times n}, \ \mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}.$$

Let $\mathbf{z} = (y_1, \ldots, y_n, y_1', \ldots, y_n')^{\mathrm{T}}$. Then

$$\mathbf{z}' = A\mathbf{z}, \text{ where } A = \begin{pmatrix} 0_n & I_n \\ A_1 & A_2 \end{pmatrix} \in \mathbb{C}^{2n\times 2n}.$$

Here 0_n is $n \times n$ zero matrix and I_n is $n \times n$ identity matrix.

The initial conditions are

$\mathbf{y}(t_0) = \mathbf{a} \in \mathbb{C}^n, \ \mathbf{y}'(t_0) = \mathbf{b} \in \mathbb{C}^n$ which are equivalent to the initial conditions $\mathbf{z}(t_0) = \mathbf{c} \in \mathbb{C}^{2n}$.

The solution of the second order differential system with n unknown functions can be solved by converting this system to the first order system with $2n$ unknown functions.

111 Exponential of a Matrix

For $A \in \mathbb{C}^{n\times n}$ let
$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$

If $D = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ then
$e^D = \mathrm{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n})$.

If A is diagonable, i.e. $A = XDX^{-1}$ then
$e^A = Xe^D X^{-1}$.

$e^{tA} = I + tA + \frac{1}{2!}t^2A^2 + \frac{1}{3!}t^3A^3 + \dots$
$(e^{tA})' = 0 + A + \frac{1}{2!}2tA^2 + \frac{1}{3!}3t^2A^3 + \dots = Ae^{At}$

If A is diagonable $A = XDX^{-1}$ then
$tA = X(tD)X^{-1} \Rightarrow$
$e^{At} = X\,\mathrm{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t})X^{-1}$.

The matrix $Y(t) := e^{(t-t_0)A}$ satisfies the matrix differential equation $Y'(t) = AY(t) = Y(t)A$ with the initial condition $Y(t_0) = I$.

(As in the scalar case, i.e. A is 1×1 matrix.)

The solution of $\mathbf{y}' = A\mathbf{y}$ with the initial condition $\mathbf{y}(t_0) = \mathbf{a}$ is given by $\mathbf{y}(t) = e^{(t-t_0)A}\mathbf{a}$.

112 Examples

1.

$$A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix}$$

$$e^{A} =$$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^{1} & 0 \\ 0 & e^{0.5} \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix} =$$

$$\begin{pmatrix} \frac{2e-3e^{0.5}}{5} & \frac{2e-2e^{0.5}}{5} \\ \frac{3e-3e^{0.5}}{5} & \frac{3e+2e^{0.5}}{5} \end{pmatrix}$$

$$e^{tA} =$$

$$\begin{pmatrix} \frac{2}{3} & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^t & 0 \\ 0 & e^{0.5t} \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix} =$$

$$\begin{pmatrix} \frac{2e^t - 3e^{0.5t}}{5} & \frac{2e^t - 2e^{0.5t}}{5} \\ \frac{3e^t - 3e^{0.5t}}{5} & \frac{3e^t + 2e^{0.5t}}{5} \end{pmatrix}$$

In the system of ODE on page 208 the solution satisfying IC $\mathbf{y}(0) = (1, 2)^\top$ is given as.

$$\mathbf{y}(t) = e^{At}\mathbf{y}(0) =$$

$$\begin{pmatrix} \frac{2e^t - 3e^{0.5t}}{5} & \frac{2e^t - 2e^{0.5t}}{5} \\ \frac{3e^t - 3e^{0.5t}}{5} & \frac{3e^t + 2e^{0.5t}}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{6e^t - 7e^{0.5t}}{5} \\ \frac{9e^t + e^{0.5t}}{5} \end{pmatrix}$$

Compare this solution with the solution given on page 211

2.

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$ is defective.

Compute e^B, e^{tB} using power series (p' 213). Note $B^2 = 0$. Hence $B^k = 0$ for $k \geq 2$. So

$$e^B = I + B + \frac{1}{2!}B^2 + \frac{1}{3!}B^3 + \ldots = I + B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$e^{tB} = I + tB + \frac{1}{2!}t^2B^2 + \frac{1}{3!}t^3B^3 + \ldots = I + tB = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

Hence the system of ODLE $\begin{array}{rcl} y_1' & = & y_2 \\ y_2' & = & 0 \end{array}$ Has the general solution

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = e^{tB} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 t \\ c_2 \end{pmatrix}$$

113 Spectral Theory of Real Symmetric Matrices

Theorem Let $A = A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. Then all eigenvalues of A are real. A is orthogonally similar to a real diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$:
$A = QDQ^{-1} = QDQ^{T}$, where Q is an orthogonal matrix $Q^{\mathrm{T}} = Q^{-1}$. The columns of Q is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Procedure: Find the characteristic polynomial of A and compute its eigenvalues: $\det\,(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)^{m_1}(\lambda_2 - \lambda)^{m_2}\ldots(\lambda_k - \lambda)^{m_k}$, where $\lambda_i \neq \lambda_j$ for $i \neq j$ and $1 \leq m_i$ (the multiplicity of λ_i). Then $\dim \mathrm{N}(A - \lambda_i I) = m_i$ and $\mathrm{N}(A - \lambda_i I) = \mathrm{span}(\mathrm{x}_{i1}, \ldots, \mathrm{x}_{im_i})$. (This is done by solving the homogeneous system $(A - \lambda_i)\mathrm{x} = \mathbf{0}$ which has m_i free variables.) Perform Gram-Schmidt process on $\mathrm{x}_{i1}, \ldots, \mathrm{x}_{im_i}$ to obtain $\mathrm{y}_{i1}, \ldots, \mathrm{y}_{im_i}$ for $i = 1, \ldots, k$. Form the orthogonal matrix $Q = (\mathrm{y}_{11} \ldots \mathrm{y}_{1m_1}\ \mathrm{y}_{21} \ldots \mathrm{y}_{2m_2} \ldots \mathrm{y}_{km_k}) \in \mathbb{R}^{n\times n}$.

114 Examples

1.

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} \end{pmatrix}, \det(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}) =$$

$$\det \begin{pmatrix} \mathbf{2} - \boldsymbol{\lambda} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} - \boldsymbol{\lambda} \end{pmatrix} = (\mathbf{2} - \boldsymbol{\lambda})^{\mathbf{2}} - \mathbf{1}^{\mathbf{2}} =$$

$(\mathbf{1} - \boldsymbol{\lambda})(\mathbf{3} - \boldsymbol{\lambda}),\ \boldsymbol{\lambda}_{\mathbf{1}} = \mathbf{3}, \boldsymbol{\lambda}_{\mathbf{2}} = \mathbf{1}$.

RREF of $\boldsymbol{A} - \boldsymbol{\lambda}_{\mathbf{1}}\boldsymbol{I} = \boldsymbol{A} - \mathbf{3}\boldsymbol{I} = \begin{pmatrix} -\mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{pmatrix}$ is

$\begin{pmatrix} \mathbf{1} & -\mathbf{1} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ A basis of $\boldsymbol{N}(\boldsymbol{A} - \mathbf{3}\boldsymbol{I})$ is

$\mathbf{x_1} = (\mathbf{1}, \mathbf{1})^\top$. Perform Gram-Schmidt on $\mathbf{x_1}$:

$\boldsymbol{r}_{\mathbf{11}} = ||\mathbf{x_1}|| = \sqrt{\mathbf{1^2 + 1^2}} = \sqrt{\mathbf{2}}, \mathbf{q_1} =$

$\frac{\mathbf{1}}{||\mathbf{x_1}||}\mathbf{x_1} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}}(\mathbf{1}, \mathbf{1})^\top = (\frac{\mathbf{1}}{\sqrt{\mathbf{2}}}, \frac{\mathbf{1}}{\sqrt{\mathbf{2}}})^\top$

RREF of $A - \lambda_2 I = A - I = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is

$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ A basis of $N(A - I)$ is $\mathbf{x_2} = (-1, 1)^\top$.

Note that

$$\mathbf{x_1} \perp \mathbf{x_2} \iff \mathbf{x_1}^\top \mathbf{x_2} = (1)(-1) + (1)(1) = 0.$$

Perform Gram-Schmidt on $\mathbf{x_2}$:

$r_{11} = ||\mathbf{x_2}|| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}, \mathbf{q_2} = \frac{1}{||\mathbf{x_2}||}\mathbf{x_2} = \frac{1}{\sqrt{2}}(-1, 1)^\top = (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$.

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

$$A = QDQ^{-1} = QDQ^\top = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

2. $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$, $A - \lambda I =$

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

Expand $\det\ (\mathbf{A} - \lambda\mathbf{I})$ by the first row and use the formula for 2×2 determinant to obtain:

$$(1-\lambda)\big((4-\lambda)(1-\lambda)-2^2\big)+(-2)\big(2(1-\lambda)-2\big)+1\big(2^2-1(4-\lambda)\big) = (1-\lambda)(\lambda^2-5\lambda)+4\lambda+\lambda = \lambda\big((1-\lambda)(\lambda-5)+5\big) = \lambda^2(6-\lambda)$$

$\lambda_1 = 6,\ \lambda_2 = \lambda_3 = 0$.

RREF of

$$\boldsymbol{A-\lambda_1 I = A - 6I = \begin{pmatrix} -5 & 2 & 1 \\ 2 & -2 & 2 \\ 1 & 2 & -5 \end{pmatrix}}$$ is

$$\boldsymbol{B = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}}$$ A basis of $\boldsymbol{N(B)}$ is

$\mathbf{x_1 = (1, 2, 1)}^\top$ (Set the free variable $\boldsymbol{x_3 = 1}$.)

Perform GS on $\mathbf{x_1}$:

$\boldsymbol{r_{11}} = ||\mathbf{x_1}|| = \sqrt{\mathbf{1^2 + 2^2 + 1^2}} = \sqrt{\mathbf{6}}$,

$\mathbf{q_1} = \frac{\mathbf{1}}{||\mathbf{x_1}||}\mathbf{x_1} = (\frac{\mathbf{1}}{\sqrt{\mathbf{6}}}, \frac{\mathbf{2}}{\sqrt{\mathbf{6}}}, \frac{\mathbf{1}}{\sqrt{\mathbf{6}}})^\top$.

RREF of $A - \lambda_2 I = A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$ is

$C = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (Note x_2, x_3 are free variables) A basis of $N(C)$ are $\mathbf{x_2}, \mathbf{x_3}$ where

$\mathbf{x_2} = (-2, 1, 0)^\top$ ($x_2 = 1, x_3 = 0$)

$\mathbf{x_3} = (-1, 0, 1)^\top$ ($x_2 = 0, x_3 = 1$)

So $\mathbf{x_2}, \mathbf{x_3}$ are two linearly independent eigenvectors corresponding to a double eigenvalue $\lambda_2 = \lambda_3 = 0$.

Note that $\mathbf{x_1} \perp \mathrm{span}(\mathbf{x_2}, \mathbf{x_3})$ as $\mathbf{x_1}^\top \mathbf{x_2} = 1(-2) + 2(1) + 1(0) = 0 = \mathbf{x_1}^\top \mathbf{x_3} = 1(-1) + 2(0) + 1(1).$

Since $\lambda_1 = 6 \neq \lambda_2 = \lambda_3 = 0$

So $\mathbf{x_1}$ is orthogonal to any eigenvector corresponding to $\lambda = 0$.

Gram-Schmidt process on

$\mathbf{x}_2 = (-2, 1, 0)^\top, \mathbf{x}_3 = (-1, 0, 1)^\top$:

$\mathbf{r}_{11} = ||\mathbf{x}_2|| = \sqrt{(-2)^2 + 1^2 + 0^2} = \sqrt{5}, \mathbf{q}_2 = \frac{1}{||\mathbf{x}_2||}\mathbf{x}_2 = (-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0)^\top.$

$r_{12} = \mathbf{q}_2^\top \mathbf{x}_3 = \frac{2}{\sqrt{5}},$

$\mathbf{p}_1 = r_{12}\mathbf{q}_2 = (-\frac{4}{5}, \frac{2}{5}, 0)^\top$

$\mathbf{x}_3 - \mathbf{p}_1 = (-\frac{1}{5}, -\frac{2}{5}, 1)^\top$

$r_{22} = ||\mathbf{x}_3 - \mathbf{p}_1|| = \frac{\sqrt{30}}{5},$

$\mathbf{q}_3 = \frac{1}{r_{22}}(\mathbf{x}_3 - \mathbf{p}_1) = (-\frac{1}{\sqrt{30}}, -\frac{2}{\sqrt{30}}, \frac{5}{\sqrt{30}})^\top$

$$Q = \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{5}{\sqrt{30}} \end{pmatrix}, A = QDQ^\top =$$

$$\begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{5}{\sqrt{30}} \end{pmatrix} \begin{pmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ -\frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{30}} & \frac{5}{\sqrt{30}} \end{pmatrix}$$

115 Prf spectral theorem for sym. mat

1. Assume λ is a complex eigenvalue of a real symmetric A with the corresponding eigenvalue $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$: $A\mathbf{x} = \lambda\mathbf{x}$. Let $\mathbf{x}^H := \overline{\mathbf{x}}^{\mathrm{T}} = (\overline{x}_1, \ldots, \overline{x}_n)$. Then $\mathbf{x}^H\mathbf{x} = |x_1|^2 + \ldots + |x_n|^2 > 0$. Thus $\mathbf{x}^H A\mathbf{x} = \lambda\mathbf{x}^H\mathbf{x}$. So $\overline{\lambda}\mathbf{x}^H\mathbf{x} = \overline{\mathbf{x}^H A\mathbf{x}} = \mathbf{x}^{\mathrm{T}} A\overline{\mathbf{x}} = (\mathbf{x}^{\mathrm{T}} A\overline{\mathbf{x}})^{\mathrm{T}} = \mathbf{x}^H A^{\mathrm{T}}\mathbf{x} = \mathbf{x}^H A\mathbf{x} = \lambda\mathbf{x}^H\mathbf{x} \Rightarrow \overline{\lambda} = \lambda$. Thus λ is a real number.

Every eigenvalue of A is real

2. We show by induction that A can be diagonalized by an orthogonal matrix, i.e. $A = QDQ^{-1} = QDA^{\top}, Q^{\top}Q = I, D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

a. $n = 1$. Then $a = 1a1^{-1}$. Any 1×1 matrix is symmetric and diagonal $Q = 1$ is 1×1 is an orthogonal matrix.

b. $n = m$ Assume that any $m \times m$ real symmetric matrix is orthogonally similar to a diagonal matrix.

c. $n = m + 1$. Since λ is real the eigenvector $\mathbf{x}$ corresponding to λ can be chosen real and $||\mathbf{x}|| = 1$. Choose an orthonormal basis $\mathbf{y}_1, \ldots, \mathbf{y}_{n-1}$ in the orthogonal complement of $\mathrm{span}(\mathbf{x}) \subset \mathbb{R}^n$. Then $O = (\mathbf{y}_1 \ldots \mathbf{y}_{n-1}\ \mathbf{x}) \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. Now $B = O^{\mathrm{T}} A O$ is symmetric $B^{\mathrm{T}} = (O^{\mathrm{T}} A^{\mathrm{T}} O)^{\mathrm{T}} = O^{\mathrm{T}} A^{\mathrm{T}} O = B$ and

$$B = \begin{pmatrix} c_{11} & \cdots & c_{1(n-1)} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{(n-1)1} & \cdots & c_{(n-1)(n-1)} & 0 \\ 0 & \cdots & 0 & \lambda \end{pmatrix}$$

Note that

$B = O^{\mathrm{T}} A O = O^{\mathrm{T}} (A\mathbf{y}_1 \ \ldots A\mathbf{y}_{n-1} \ A\mathbf{x}) = O^{\mathrm{T}} (A\mathbf{y}_1 \ \ldots A\mathbf{y}_{n-1} \ \lambda\mathbf{x})$, which explains the $n-1$ zeros on the last column of B. Since B is symmetric B also have $n-1$ zeros on the last row. Also the matrix $C = (c_{ij})_1^{n-1} \in \mathbb{R}^{(n-1)\times(n-1)}$ is symmetric. Use the induction) to deduce that

$Q_1^\top C Q_1 = D_1, Q_1 Q_1^\top = I_{n-1}$. Define

$$Q_2 = \begin{pmatrix} Q_1 & 0_{n-1} \\ 0_{n-1}^\top & 1 \end{pmatrix}, 0_{n-1}^\top = (\underbrace{0, \ldots, 0}_{n-1}$$

Then Q_2 is orthogonal and

$$D = \begin{pmatrix} D_1 & 0_{n-1} \\ 0_{n-1}^\top & \lambda \end{pmatrix} =$$

$$\begin{pmatrix} Q_1^\top C Q_1 & 0_{n-1} \\ 0_{n-1}^\top & \lambda \end{pmatrix} =$$

$$Q_2^\top \begin{pmatrix} C & 0_{n-1} \\ 0_{n-1}^\top & \lambda \end{pmatrix} Q_2 = Q_2^\top O^\top A Q_2 O =$$

$$(OQ_2)^\top A (OQ_2)$$

Since the product of two orthogonal matrices $OQ_2(OQ_2)^\top = OQ_2Q_2^\top O^\top = OO^\top = I$ we obtain that $A = OQ_2D(OQ_2)^\top$, i.e. A is orthogonally similar to a diagonal matrix $\square$

3. Claim: Let A be real symmetric and $\mathbf{x}, \mathbf{y}$ be two eigenvectors corresponding to two different eigenvalues λ, μ. Then $\mathbf{x}$ is orthogonal to $\mathbf{y}$.

Proof: $\mathbf{y}^{\mathbf{T}}A\mathbf{x} = (\mathbf{y}^{\mathbf{T}}A\mathbf{x})^{\mathbf{T}} = \mathbf{x}^{\mathbf{T}}A\mathbf{y} \Rightarrow \lambda\mathbf{y}^{\mathbf{T}}\mathbf{x} = \mu\mathbf{x}^{\mathbf{T}}\mathbf{y} \Rightarrow (\lambda - \mu)\mathbf{y}^{\mathbf{T}}\mathbf{x} = 0 \Rightarrow \mathbf{y}^{\mathbf{T}}\mathbf{x} = 0$.

Hence in the procedure for finding the orthonormal matrix Q it is enough to perform the Gram-Schmidt process on a basis of each null space of $A - \lambda_i I$.

116 Quadratic forms

For $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$
$Q(\mathbf{x}) = a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + \ldots + 2a_{(n-1)n}x_{n-1}x_n = \sum_{i=1}^n a_{ii}x_i^2 + 2\sum_{1\le i<j\le n} a_{ij}x_ix_j$

is called the quadratic form in n variables.

Example 1: $Q(x_1, x_2) = 2x_1^2 + 2x_2^2 + 2x_1x_2$

Observe

$$(x_1, x_2)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix} = x_1(2x_1 + x_2) + x_2(x_1 + 2x_2) = Q(x_1, x_2) = Q(\mathbf{x})$$

Example 2: $Q(x_1,x_2,x_3) =$
$x_1^2 + 4x_2^2 + x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3$

Observe $(x_1,x_2,x_3)\begin{pmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} =$

$$(x_1,x_2,x_3)\begin{pmatrix}x_1+2x_2+x_3\\ 2x_1+4x_2+2x_3\\ x_1+2x_2+x_3\end{pmatrix} =$$

$x_1(x_1+2x_2+x_3) + x_2(2x_1+4x_2+2x_3) +$
$x_3(x_1+2x_2+x_3) = Q(x_1,x_2,x_3)$

Claim: To each quadratic form $Q(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ given on previous page corresponds a unique symmetric matrix
$A = (a_{ij})_{i,j=1}^n R^{n\times n} =$

$$\begin{pmatrix}a_{11} & a_{12} & \dots & a_{1n}\\ a_{12} & a_{22} & \dots & a_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ a_{1n} & a_{2n} & \dots & a_{nn}\end{pmatrix}$$ such that

$Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$. (Proof straightforward!)

Note that $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \sum_{i,j=1}^{n} a_{ij}x_ix_j$. Indeed the term $a_{ii}x_i^2$ comes from $i = j$. The term $2a_{ij}x_ix_j, i < j$ in the above sum comes from $a_{ij}x_ix_j$ and $a_{ji}x_jx_i$ (Recall $a_{ij} = a_{ji}$!)

Note that if $D = \mathrm{diag}(d_1, \ldots, d_n)$ is a diagonal matrix then

$$\mathbf{x}^\top D\mathbf{x} = d_1x_1^2 + d_2x_2^2 + \ldots + d_nx_n^2 = \sum_{i=1}^{n} d_1x_i^2.$$

117 Rayleigh quotient

Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then $\frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\mathrm{T}\mathbf{x}}$ for $0 \neq \mathbf{x} \in \mathbb{R}^n$ is called the Rayleigh quotient. Equivalently consider the quadratic form $\mathbf{x}^\mathrm{T} A\mathbf{x}$ with the normalization $||\mathbf{x}|| = 1 (= \mathbf{x}^\mathrm{T}\mathbf{x})$.

Arrange eigenvalues of A in a decreasing order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, where each eigenvalue is repeated with its multiplicities. Then
$\lambda_1 = \max_{0\neq \mathbf{x}\in\mathbb{R}^n} \frac{\mathbf{x}^\mathrm{T} A\mathbf{x}}{\mathbf{x}^\mathrm{T}\mathbf{x}} = \max_{||\mathbf{x}||=1} \mathbf{x}^\mathrm{T} A\mathbf{x}$.
Equality achieved only for eigenvector of A corresponding to λ_1.

$\lambda_n = \min_{0\neq \mathbf{x}\in\mathbb{R}^n} \frac{\mathbf{x}^\mathrm{T} A\mathbf{x}}{\mathbf{x}^\mathrm{T}\mathbf{x}} = \min_{||\mathbf{x}||=1} \mathbf{x}^\mathrm{T} A\mathbf{x}$.
Equality achieved only for eigenvector of A corresponding to λ_n.

Proof. $A = QDQ^\mathrm{T},\ D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Let
$\mathbf{y} := Q^\mathrm{T}\mathbf{x} \Rightarrow \mathbf{x}^\mathrm{T}\mathbf{x} = \mathbf{y}^\mathrm{T}\mathbf{y} = \mathbf{y}_1^2 + \ldots + \mathbf{y}_n^2$,
$\mathbf{x}^\mathrm{T} A\mathbf{x} = \mathbf{y}^\mathrm{T} D\mathbf{y} = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 \leq \lambda_1^2 \mathbf{y}^\mathrm{T}\mathbf{y}$.
This implies the maximal characterization. Similarly:
$\mathbf{y}^\mathrm{T} D\mathbf{y} \geq \lambda_n \mathbf{y}^\mathrm{T}\mathbf{y}$ which implies the minimal characterization.

118 Examples

1.

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \det\,(\mathbf{A} - \lambda \mathbf{I}) =$$

$(1-\lambda)(3-\lambda),\ \lambda_1 = 3, \lambda_2 = 1$. (See page 218).

So

$$1 \leq \frac{\mathbf{x}^\top A \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \frac{2x_1^2 + 2x_1x_2 + 2x_2^2}{x_1^2 + x_2^2} \leq 3$$

The maximum achieved if and only if

$\mathbf{x} = a\mathbf{x}_1 = (a, a)^\top$.

The minimum is achieved if and only if

$\mathbf{x} = b\mathbf{x}_2 = (-b, b)^\top$

119 LU factorization

Def: A square matrix $A \in \mathbb{R}^{n\times n}$ has an LU factorization if $A = LU$ where $L \in \mathbb{R}^{n\times n}$ is a lower triangular matrix with all diagonal entries equal to 1 and $U \in \mathbb{R}^{n\times n}$ is an upper triangular matrix with nonzero diagonal entries:

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ l_{21} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{n(n-1)} & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1(n-1)} & u_{1n} \\ 0 & u_{22} & \dots & u_{2(n-1)} & u_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & u_{nn} \end{pmatrix}$$

Thm: $A = (a_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n\times n}$ has LU factorization if and if all the n leading principal minors of A are nonzero:

$$a_{11} \neq 0, \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \neq 0, \ldots$$

$$\det \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \neq 0.$$

Moreover U is obtained by Gauss elimination without making the pivots equal to 1 and no permutation of rows. Further L^{-1} is the product of elementary matrices corresponding to Gauss eliminations.

In particular the LU factorization is unique.

120 Example

$$A = \begin{pmatrix} 2 & 4 & 2 \\ 1 & 5 & 2 \\ 4 & -1 & 9 \end{pmatrix}$$ Perform the elementary row operations $R_2 - 0.5R_1 \to R_2, R_3 - 2R_1 \to R_3$ to obtain $B_1 = \begin{pmatrix} 2 & 4 & 2 \\ 0 & 3 & 1 \\ 0 & -9 & 5 \end{pmatrix} = L_1 A$ where

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$ Perform the elementary row operation $R_3 + 3R_2 \to R_3$ to obtain

$$U = \begin{pmatrix} 2 & 4 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 8 \end{pmatrix} = L_2 B_1$$ where

$$\mathbf{L_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}$$

So $\mathbf{U = (L_2 L_1) A \Rightarrow A = LU, L = L_1^{-1} L_2^{-1}} =$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 2 & -3 & 1 \end{pmatrix}$$

121 Proof of LU factorization Thm

Let $A_k = (a_{ij})_{i,j=1}^k \mathbb{R}^{k\times k}, k = 1, \ldots, n$ Suppose first that A has LU factorization. Let $L_k = (l_{ij})_{i,j=1}^k, U_k = (u_{ij})_{i,j=1}^k \in \mathbb{R}^{k\times k}$ be the k leading submatrices of order k of A, L, U respectively. A straightforward calculation shows:

$$A_k = L_k U_k \quad (*)$$

Hence $\det\ \mathrm{A_k} = \det\ \mathrm{L_k} \det\ \mathrm{U_k}$. Since L_k is upper triangular with one on the diagonal $\det\ \mathrm{L_k} = 1$. Since U_k is upper triangular $\det\ \mathrm{U_k} = \mathrm{u_{11}} \ldots \mathrm{u_{kk}}$. So $\det\ \mathrm{A_k} = \det\ \mathrm{U_k}$. Since all diagonal elements of U different from zero $\det\ \mathrm{A_k} \neq 0$ for $k = 1, \ldots, n$. In particular

$$a_{11} = u_{11}, u_{ii} = \frac{\det\ \mathrm{A_i}}{\det\ \mathrm{A_{i-1}}}, i = 2, \ldots, n$$

Assume now that $\det A_i \neq 0, i = 1, \ldots, n$. We prove that we can do Gauss elimination without making the pivots equal to 1 and no permutation of rows. Since $\det A_1 = a_1 1$ we perform the elementary row operations of the third kind: $R_i - \frac{a_{i1}}{a_{11}} R_1, i = 2, \ldots, n$ to obtain $B_1 = (b_{ij,1})_{i,j=1}^n = L_1 A$, where B_1 has the first column $(a_{11}, 0, \ldots, 0)^\top$, and L_1 is a product of lower triangular elementary matrices with 1 on the main diagonal. So $A = M_1 B_1, M_1 = (m_{ij,1})_{i,j=1}^n = L_1^{-1}$. M_1 is a lower triangular matrix with ones on the main diagonal. Let $M_{1,2} = (m_{ij,1})_{i,j=1}^2, B1,2 = (b_{ij,1})_{i,j=1}^2$. Then $M_{1,2}$ is lower triangular with ones on the diagonal, $B_{1,2}$ is upper triangular and $A_2 = M_{1,2} B_{1,2}$. Thus $0 \neq \det A_2 = \det M_{1,2} \det B_{1,2} = 1(b_{11,1} b_{22,1})$. So $b_{22,1} \neq 0$. Apply the elementary row operations $R_i - \frac{b_{i2,1}}{b_{22,1}} R_2, i = 3, \ldots, n$ To obtain the matrix $B_2 = (b_{ij,2})_{i,j=1}^n$ whose first and the second columns are $(a_{11}, 0, \ldots, 0)^\top, (b_{12,1}, b_{22,1}, 0, \ldots, 0)^\top$.

So $B_2 = L_2 B_1$, L_2 is a product of lower triangular elementary matrices with 1 on the main diagonal. So $A = M_1 B_1 = M_1 M_2 B_2, M_2 = (m_{ij,2})_{i,j=1}^n = L_2^{-1}$. M_2 is a lower triangular matrix with ones on the main diagonal. We proceed as above to show that the condition $\det\ \mathbf{A}_3 \neq 0$ implies $b_{33,1} \neq 0$. Continue in this manner to obtain $U = B_{n-1} = L_{n-1} L_{n-2} \dots L_1 A$, is upper triangular with nonzero entries on the diagonal, and $L_1, \dots, L_{n-1}$ are lower triangular with ones on the diagonal. Then $M_i = L_i^{-1}$ is lower triangular with ones on the diagonal for $i = 1, \dots, n-1$. Then $A = M_1 \dots M_{n-1} U$. So $L = M_1 M_2 \dots M_{n-1}$ is a lower diagonal matrix with one on the diagonal.

□

122 $LDL^\top$ factorization

Definition. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ has $LDL^\top$ factorization if $A = LDL^\top$, where L is a lower triangular matrix with one on the diagonal and D is a diagonal matrix with nonzero diagonal entries.

Thm. A symmetric matrix has $LDL^\top$ factorization if and only if A has LU factorization, i.e. all leading minors of A are different from 0.

The $LDL^\top$ factorization obtained from the LU factorization by letting the diagonal entries of D to be the diagonal entries of U.

After finding U determine D from the diagonal entries of U. Then $L^\top = D^{-1}U, L = (L^\top)^\top$.

123 Example

$$A = \begin{pmatrix} 2 & -2 & 4 \\ -2 & -1 & 5 \\ 4 & 5 & -18 \end{pmatrix}$$ Perform the following ERO

$R_2 + R_1 \to R_2, R_3 - 2R_1 \to R_3$ to obtain

$$B_1 = \begin{pmatrix} 2 & -2 & 4 \\ 0 & -3 & 9 \\ 0 & 9 & -26 \end{pmatrix}$$ Perform the ERO

$R_3 + 3R_2 \to R_3$ on B_1 to obtain

$$B_2 = \begin{pmatrix} 2 & -2 & 4 \\ 0 & -3 & 9 \\ 0 & 0 & 1 \end{pmatrix}$$ So $U = B_2$,

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$D^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, L^\top = D^{-1}U =$$

$$\begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}, L = (L^\top)^\top =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -3 & 1 \end{pmatrix}$$

Check that $A = LDL^\top$!

124 Proof of $LDL^\top$ factorization

Suppose first that a symmetric $A = LDL^\top$, where L is lower triangular with ones on the diagonal and D is a diagonal matrix with nonzero entries. Observe that $U = DL^\top$ is an upper diagonal matrix whose diagonal entries are the diagonal entries of D, which are different from zero. Hence A has an LU factorization.

Suppose a symmetric A has LU factorization. Let the diagonal entries of D to be the diagonal entries of U. Then $M = D^{-1}U$ is an upper triangular matrix with ones on the main diagonal. So $A = LDM$. Since A symmetric $A = A^\top = (LDM)^\top = M^\top(D^\top)L^\top = M^\top(DL^\top)$. Since $V = DL^\top$ is upper triangular with the diagonal entries equal to the diagonal entries of D it follows that $A = M^\top V$ is another LU decomposition of A. Since the LU decomposition is unique it follows that $L = M^\top \Rightarrow M = L^\top \Rightarrow A = LDM = LDL^\top$.

□

125 Positive Definite Matrices

$A = A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ is called positive definite, denoted by $A \succ 0$, if $\mathbf{x}^{\mathrm{T}} A \mathbf{x} > 0$ for any $0 \neq \mathbf{x} \in \mathbb{R}^n$.

From the minimal characterization of the smallest eigenvalues of A it follows $A \succ 0$ if and only if all the eigenvalues of A are positive: $\lambda_i > 0, i = 1, \ldots, n$.

Thm $A = (a_{ij})_{i,j=1}^n = A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ is positive definite if and only if the n leading principal minors of A are positive: $a_{11} > 0, \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \ldots$

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} > 0.$$

(such an $i \times i$ determinant is called the $i-th$ principal minor of A.)

Proof: Assume that $A = (a_{ij})_{i,j=1}^{n} \succ 0$. Then A has positive eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n > 0$. Hence $\det\ \mathrm{A} = \lambda_1 \ldots \lambda_n > 0$. Let $A_k = (a_{ij})_{i,j}^{k} \in \mathbb{R}^{k \times k}$. Let $\mathbf{x} = (x_1, \ldots, x_k, 0, \ldots, 0)^\top, \mathbf{x}_k = (x_1, \ldots, x_k)^\top$.
Note that $\mathbf{x}^\top A \mathbf{x} = \mathbf{x}_k^\top A_k \mathbf{x}_k$. Since $A \succ 0$ we get that $\mathbf{x}^\top A \mathbf{x} = \mathbf{x}_k^\top A_k \mathbf{x}_k > 0$ if $\mathbf{x}_k \neq 0$. Hence $A_k \succ 0 \Rightarrow \det\ \mathrm{A_k} > 0, \mathrm{k} = 1, \ldots, \mathrm{n}$.

Assume now that all the leading principle minors of A are positive. Then $A = LU = LDL^\top$ where L is lower diagonal $D = \mathrm{diag}(u_{11}, \ldots, u_{nn}$, and $u_{11}, \ldots, u_{nn}$ are the diagonal entries of the upper triangular matrix U. Recall the formulas from page 237

$$a_{11} = u_{11}, u_{ii} = \frac{\det\ \mathrm{A_i}}{\det\ \mathrm{A_{i-1}}}, i = 2, \ldots, n$$

So $u_{ii} > 0, i = 1, \ldots, n \Rightarrow D \succ 0$. Observe $\mathbf{x}^\top A \mathbf{x} = \mathbf{x}^\top LDL^\top \mathbf{x} = \mathbf{y}^\top D \mathbf{y}$, where $\mathbf{y} = L^\top \mathbf{x}$. So $\mathbf{y}^\top D \mathbf{y} > 0$ if $\mathbf{y} \neq 0$. Since $\det\ \mathrm{L}^\top = 1$ $\mathbf{y} = 0 \iff \mathbf{x} = 0$, hence $A \succ 0$. $\square$

126 Cholesky decomposition

Thm: $A \succ 0$ if and only if $A = MM^\top$, where M is a lower triangular with positive entries on the diagonal

Proof 1. Assume $A \succ 0$. So $A = LDL^\top$ decomposition, where $D = \text{diag}(d_1, \ldots, d_n) \succ 0$. Define $D_1 = \text{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n})$. Then $A = MM^\top$, where $M = LD_1$.

2. Suppose that $A = MM^\top$, where M is a lower triangular with positive entries on the diagonal. So M is invertible. Note $\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top MM^\top\mathbf{x} = (M^\top\mathbf{x})^\top(M^\top\mathbf{x}) = ||M^\top\mathbf{x}||^2 \geq \mathbf{0}$. Since $M^\top\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$. Hence $A \succ \mathbf{0}$. $\square$

127 Andre-Louis Cholesky

Born: 15 Oct 1875 in Montguyon, Charentes Maritime, France Died: 31 Aug 1918 in North France.

Cholesky entered l'cole Polytechnique on 15 October 1895 He then joined the army becoming a second lieutenant, and went to study at the school d'Application de l'Artillerie et du Gnie starting in October 1897. He completed his course in 1899 and he maintained his steady improvement for now he was placed 5th out of 86 students who qualified in that year.

Cholesky died from wounds received on the battle field on 31 August 1918 at 5 o'clock in the morning in the North of France. After his death one of his fellow officers, Commandant Benoit, published Cholesky's method of computing solutions to the normal equations for some least squares data fitting problems in Note sur une methode de resolution des equation normales provenant de l'application de la methode des moindres carrs a un systeme d'equations lineaires en nombre inferieure a celui des inconnues. Application de la methode a la resolution d'un systeme defini d'equations lineaires (Procede du Commandant Cholesky), published in the Bulletin geodesique in 1924.

The Cholesky Factorization (or Cholesky Decomposition) takes a symmetric positive definite matrix A and writes it as A = LL’ where L is a lower triangular matrix with positive diagonal entries (sometimes called the Cholesky triangle), and L’ is the transpose of L. To solve Ax = b one now needs to solve LL’x = b so put y = L’x which gives Ly = b which is solved for y, then y = L’x is solved for x to obtain the solution. The beauty of the method is that it is trivial to solve equations of the type Mx = b when M is a triangular matrix.

The method received little attention after its publication in 1924 but Jack Todd included it in his analysis courses in King’s College, London, during World War II. In 1948 the method was analysed in a paper by Fox, Huskey and Wilkinson while in the same year Turing published a paper on the stability of the method.

Def.: A symmetric $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is called

1. Nonnegative definite , denoted by $\boldsymbol{A} \succeq \mathbf{0}$ if $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} \geq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$.

2. Negative definite , denoted by $\boldsymbol{A} \prec \mathbf{0}$ if $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} < \mathbf{0}$ for any $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.

3. Nonpositive definite , denoted by $\boldsymbol{A} \preceq \mathbf{0}$ if $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} \leq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$.

4. Indefinite if $\boldsymbol{A}$ has at least one positive and one negative eigenvalue.

Clearly

a. $\boldsymbol{A} \succ \mathbf{0} \iff -\boldsymbol{A} \prec \mathbf{0}$

b. $\boldsymbol{A} \succeq \mathbf{0} \iff -\boldsymbol{A} \preceq \mathbf{0}$.

The maximum and minimum characterization of the eigenvalues of $\boldsymbol{A}$ yield

Cor. $\boldsymbol{A} \succeq \mathbf{0} \iff \boldsymbol{\lambda}_1 \geq \ldots \geq \boldsymbol{\lambda}_n \geq \mathbf{0}$,

$\boldsymbol{A} \preceq \mathbf{0} \iff \mathbf{0} \geq \boldsymbol{\lambda}_1 \geq \ldots \geq \boldsymbol{\lambda}_n$.

Corollary: A symmetric $A \in \mathbb{R}^{n\times n}$ is negative definite if the leading prinipal minors have alternating sums:

$$a_{11} < 0, \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \ldots$$

$$(-1)^n \det \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} > 0$$

It is more difficult to characterize nonnegative definite or nonpositive definite symmetric matrices in terms of principal minors

Def. A principle minor of a square matrix A is obtained by erasing the same rows and columns of A and taking the determinant of the remaining square matrix

Thm. A symmetric A is nonnegative definite if and only if its all principle minors are nonnegative

Note that $A \succeq 0 \Rightarrow \det \mathbf{A} = \lambda_1 \ldots \lambda_n \geq 0$. It is not difficult to show that $A \succeq 0$ implies the nonnegativity of all principle minors. The sufficiency is more involved.

128 Examples

1. $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$.

a. $A \succ 0 \iff a > 0, ac - b^2 > 0$

b. $A \succeq 0 \iff a \geq 0, c \geq 0, ac - b^2 \geq 0$

c. $A \prec 0 \iff a < 0, ac - b^2 > 0$

d. $A \preceq 0 \iff a \leq 0, c \leq 0, ac - b^2 \geq 0$

f. A is indefinite iff $ac - b^2 < 0$, since $ac - b^2 = \det \mathbf{A} = \lambda_1 \lambda_2$.

2. $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$, page 220.

$\lambda_1 = 6, \lambda_2 = \lambda_3 = 0$ So $A \succeq 0$. Principal minors of order one are the diagonal elements $1, 4, 1$. All other principle minors are equal to zero.

3. Let $A = \begin{pmatrix} 2 & -4 & 6 \\ -4 & 12 & -4 \\ 6 & -4 & 35 \end{pmatrix}$.

a. Find $LDL^\top$ factorization of A:

Perform the following row operations on A:
$R_2 + 2R_1 \to R_2, R_3 - 3R_1 \to R_1$ to obtain
$B_1 = \begin{pmatrix} 2 & -4 & 6 \\ 0 & 4 & 8 \\ 0 & 8 & 17 \end{pmatrix}$ Perform the following row operation on B_1: $R_3 - 2R_2 \to R_3$ to obtain
$B_2 = \begin{pmatrix} 2 & -4 & 6 \\ 0 & 4 & 8 \\ 0 & 0 & 1 \end{pmatrix}$ So

$$U = B_2, D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$L^\top = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -4 & 6 \\ 0 & 4 & 8 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$, $LDL^\top$ factorization of A is $A =$

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

b. Show that A is positive definite.

Since A has $LDL^\top$ factorization and all diagonal entries of D are positive $A \succ 0$.

c. Find the Cholesky factorization of A:

Let $D_1 = \sqrt{D} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{4} & 0 \\ 0 & 0 & \sqrt{1} \end{pmatrix} =$

$\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ Then $M = LD_1 =$

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

$$\begin{pmatrix} \sqrt{2} & 0 & 0 \\ -2\sqrt{2} & 2 & 0 \\ 3\sqrt{2} & 4 & 1 \end{pmatrix}$$

and $A = MM^\top$.

129 $LDL^\top$ for negative definite

Since $A = A^\top$ is negative definite if and only if $-A$ is positive definite we deduce

Thm: A symmetric matrix is negative definite if and only if it has $LDL^\top$ factorization, where all diagonal entries of D are negative.

130 Classification of critical points

1. One variable: Let $f(t)$ be a continuous function with a continuous derivative on the open interval $a < t < b$. $c \in (a, b)$ is called critical if $f'(c) = 0$. Recall the well known fact that if $f(c),\ c \in (a, b)$ is a local minimum or maximum then c is a critical point.

Problem: Given a critical point $c \in (a, b)$ of f when c is a local minimum or maximum?

Second order criteria for critical points Let $f \in C^2(a, b)$, i.e. f has two continuous derivatives in (a, b). Assume that $f'(c) = 0, c \in (a, b)$. Then

(a) If $f''(c) > 0$ then c is a local minimum. More precisely, there exists $\varepsilon > 0$ so that $f(c) < f(t)$ for any t such that $0 < |t - c| < \varepsilon$.

(b) If $f''(c) < 0$ then c is a local maximum. More precisely, there exists $\varepsilon > 0$ so that $f(c) > f(t)$ for any t such that $0 < |t - c| < \varepsilon$.

(c) If $f(c)$ is a local minimum $f''(c) \geq 0$.

(d) If $f(c)$ is a local maximum $f''(c) \leq 0$.

Proof. Recall the Taylor formula with the remainder
$f(t) = f(c)+f'(c)(t-a)+\frac{1}{2}f''(s(t))(t-a)^2 = f(c) + \frac{1}{2}f''(s(t))(t-a)^2, |s(t)-c| \leq |t-c|.$

Now use continuity of the second derivative at c to deduce the conditions (a) and (b).

Suppose that $f(c)$ is a local minimum. Then the condition (b) Can not hold. Hence (c) holds.

Similarly, if $f(c)$ is a local maximum then (d) holds.

1. Many variables: Let $D \subset \mathbb{R}^n, n \geq 2$ be an open set and $f : D \to \mathbb{R}$ be a function. Recall that $C^k(D)$ is the set of all function with continuous partial derivatives up to order k. Assume that $f \in C^1(D)$, i.e. f is continuous and it has continuous first order partial derivatives. Then $\mathbf{c} \in D$ is called a critical point if

$$\nabla f(\mathbf{c}) := \left(\frac{\partial f}{\partial x_1}(\mathbf{c}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{c})\right) = \mathbf{0}.$$

Definition Assume that $f \in C^2(D)$. For $\mathbf{x} \in D$ define the symmetric matrix $H(f)(\mathbf{x}) :=$

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \ldots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix}$$

If $c \in D$ is a critical point of f then $H(f)(\mathbf{c})$ is called the Hessian matrix of f at $\mathbf{c}$.

Second order criteria for critical points:

Let $f \in C^2(D)$. Assume that $\nabla f(\mathbf{c}) = \mathbf{0}, \mathbf{c} \in D$. Then

(a) If $H(f)(\mathbf{c}) \succ 0$ then $\mathbf{c}$ is a local minimum. More precisely, there exists $\varepsilon > 0$ so that $f(\mathbf{c}) < f(\mathbf{x})$ for any $\mathbf{x}$ such that $0 < ||\mathbf{x} - \mathbf{c}|| < \varepsilon$.

(b) If $H(f)(\mathbf{c}) \prec 0$ then $\mathbf{c}$ is a local maximum. More precisely, there exists $\varepsilon > 0$ so that $f(\mathbf{c}) > f(\mathbf{x})$ for any $\mathbf{x}$ such that $0 < ||\mathbf{x} - \mathbf{c}|| < \varepsilon$.

(c) If $f(\mathbf{c})$ is a local minimum $H(f)(\mathbf{c}) \succeq 0$.

(d) If $f(\mathbf{c})$ is a local maximum $H(f)(\mathbf{c}) \preceq 0$.

Proof The proof follows from the following formula. Fix the direction $\mathbf{y} \in \mathbb{R}^n, ||\mathbf{y}|| = 1$ and let $g(t, \mathbf{y}) = f(\mathbf{c} + t\mathbf{y})$. Then $g(t, \mathbf{y}) \in C^2(-\varepsilon, \varepsilon)$ and $g'(0, \mathbf{y}) = \nabla f(c)\mathbf{y} = 0$. The Taylor formula with remainder using chain rule is $f(\mathbf{c} + t\mathbf{y}) = g(t, \mathbf{y}) = f(\mathbf{c}) + \mathbf{y}^\top H(f)(\mathbf{c} + s(t, \mathbf{y})\mathbf{y})\mathbf{y}$ Use continuity of the second derivatives of f at $\mathbf{c}$ to obtain the conditions (a) and (b). (c), (d) obtained similarly to one variable case.

3. Indefinite case

Assume that we are in the several variable case, $\nabla f(\mathbf{c}) = \mathbf{0}$ and $H(f)(\mathbf{c})$ an indefinite symmetric matrix. By linear change of coordinates $\mathbf{x} = \mathbf{c} + Q\mathbf{z}$, where Q is orthogonal matrix, we may assume that $\mathbf{c} = \mathbf{0}$ and $H(f)(\mathbf{0}) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Assume that $f \in C^3(D)$. Then the Taylor expansion of f is $f(\mathbf{x}) = f(\mathbf{0}) + \mathbf{x}^\top D\mathbf{x} +$ higher order term.

Since $H(f)(\mathbf{0})$ has at least one positive and one negative eigenvalue the quadratic form

$$\mathbf{x}^\top D\mathbf{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2$$

is indefinite, i.e. it takes positive and negative values.

Hence $\mathbf{c}$ is a saddle point.

Recall that $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is indefinite iff $\det \mathbf{A} = \mathbf{ac} - \mathbf{b}^2 < \mathbf{0}$.

131 Singular Value Decomposition

Let $A \in \mathbb{R}^{m \times n}$. Then there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ and generalized diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\min(m,n)}) \in \mathbb{R}^{m \times n}$, with the diagonal entries

$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$, such that
$A = U \Sigma V^{\mathrm{T}}$. (SVD)

If $m = n$ then $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix.

If $m > n$ then
$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & \sigma_n \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

If $n > m$ then Σ^{T} is as above.

$\sigma_1, \ldots, \sigma_n$ are called the singular values of A.

The number of positive singular values of A is equal to $\mathrm{rank}\, A$.

Finding SVD

Assume that $m \geq n$. Form the symmetric matrix $B = A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$. Then B is nonnegative definite: $0 \leq \mathbf{x}^{\mathrm{T}}B\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^n$ since $\mathbf{x}^{\mathrm{T}}B\mathbf{x} = ||A\mathbf{x}||^2$. Hence all the eigenvalues of B are nonnegative. As $B\mathbf{x} = 0 \iff A\mathbf{x} = 0$ it follows $\mathrm{rank}\, B = \mathrm{rank}\, A = r$. Then the eigenvalues of B are $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ arranged in a decreasing order with the corresponding multiplicities. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ be an orthonormal set of eigenvectors of B: $B\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ for $i = 1, \ldots, n$. Form the orthogonal matrix $V := (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$. Then $B = V\,\mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)V^{\mathrm{T}}$. The vectors $\mathbf{u}_i := \frac{1}{\sigma_i}A\mathbf{v}_i \in \mathbb{R}^m$ is an orthonormal set of vectors for $i = 1, \ldots, r$. Let $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ be an orthonormal basis for $\mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_r)^{\perp}$. Then $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $U^{\mathrm{T}}U = I_m$. Thus $A = U\,\mathrm{diag}(\sigma_1, \ldots, \sigma_n)V^{\mathrm{T}}$.

If $m < n$ form the symmetric nonnegative definite matrix $C = AA^{\mathrm{T}} \in \mathbb{R}^{m\times m}$ and $\mathrm{rank}\, A = \mathrm{rank}\, A^{\mathrm{T}} = \mathrm{rank}\, C = r$. Then the eigenvalues of C are $\sigma_1^2, \ldots, \sigma_m^2$ arranged in a decreasing order with their multiplicities. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^m$ be an orthonormal set of eigenvectors of C: $C\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i$ for $i = 1, \ldots, n$. Form the orthogonal matrix $U := (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$. Then $C = U\, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2) U^{\mathrm{T}}$. The vectors $\mathbf{v}_i := \frac{1}{\sigma_i} A^{\mathrm{T}} \mathbf{u}_i \in \mathbb{R}^n$ is an orthonormal set of vectors for $i = 1, \ldots, r$. Let $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ be an orthonormal basis for $\mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)^{\perp}$. Then $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$ and $V^{\mathrm{T}} V = I_n$. Thus $A = U\, \mathrm{diag}(\sigma_1, \ldots, \sigma_n) V^{\mathrm{T}}$.

132 Example

Let $A = \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix}$ Since $m = 3 > n = 2$ it is advisable to compute $B = A^\top A =$

$$\begin{pmatrix} 6 & -3 & 0 \\ -2 & 5 & -4 \end{pmatrix} \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix} =$$

$$\begin{pmatrix} 45 & -27 \\ -27 & 45 \end{pmatrix}$$

$$\det\,(\mathrm{B} - \lambda \mathrm{I}) = \det \begin{pmatrix} 45 - \lambda & -27 \\ -27 & 45 - \lambda \end{pmatrix} =$$

$$(45 - \lambda)^2 - (-27)^2 = (45 - \lambda + 27)((45 - \lambda - 27) = (72 - \lambda)(27 - \lambda), \lambda_1 = 72, \lambda_2 = 18$$

The two positive singular values of A are

$$\sigma_1 = \sqrt{72} = 6\sqrt{2}, \quad \sigma_2 = \sqrt{18} = 3\sqrt{2}$$

To find the orthogonal matrix $V = (\mathbf{v_1}\ \mathbf{v_2} \ldots \mathbf{v}_n)$ in SVD decomposition of $A = U\Sigma V^\top$, we need to diagonalize the matrix

$B = A^\top A = VDV^\top, D = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots)$.

The RREF of

$$B - \lambda_1 I = \begin{pmatrix} -27 & -27 \\ -27 & -27 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} x_2$$

is a free variable. Set $x_2 = 1$ to see that the eigenvector $\mathbf{x_1} = (-1, 1)^\top$ is a basis in $N(B - \lambda_1 I)$. The Gram-Schmidt process on $\mathbf{x_1}$ gives

$\mathbf{v_1} = \frac{1}{||\mathbf{x_1}||}\mathbf{x_1} = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$

The RREF of

$$B - \lambda_2 I = \begin{pmatrix} 27 & -27 \\ -27 & 27 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

x_2 is a free variable. Set $x_2 = 1$ to see that the eigenvector $\mathbf{x_2} = (1, 1)^\top$ is a basis in $N(B - \lambda_2 I)$. The Gram-Schmidt process on $\mathbf{x_1}$ gives

$\mathbf{v_2} = \frac{1}{||\mathbf{x_2}||}\mathbf{x_2} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$

Hence $V = \begin{pmatrix} \frac{1}{-\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

Recall that $U = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_m)$ is an orthogonal matrix. The first r-columns of U, where $r = \text{rank}\, A$, which is also the number of positive singular values of A is determined by the formula $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i, i = 1, \ldots, r$:

$$\mathbf{u}_1 = \frac{1}{6\sqrt{2}} \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} =$$

$$\frac{1}{6\cdot 2} \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$$

$$\mathbf{u_2} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} =$$

$$\frac{1}{3 \cdot 2} \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

Note that $\mathbf{u_1}, \mathbf{u_2}$ is an orthonormal set of two vectors

To find $\mathbf{u_3}$ we observe that $\mathbf{u_1^\top u_3} = 0, \mathbf{u_2^\top u_3} = 0$, which is equivalent to the fact that $\mathbf{u_3}$ is in the null space of

$$C = (\mathbf{u_1} \ \mathbf{u_2})^\top = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{pmatrix}$$ The RREF of

C is $\begin{pmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 1 & -1 \end{pmatrix}$. $\mathbf{w} = (\frac{1}{2}, 1, 1)^\top$ is a basis in $N(C)$. Perform GS process on $\mathbf{w}$ to obtain

$\mathbf{u_3} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})^\top$.

So $\boldsymbol{U} = (\mathbf{u_1}\ \mathbf{u_2}\ \mathbf{u_3}) = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{pmatrix}$ Hence

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{-\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

(Note that in this example we have a very special case $\boldsymbol{V}^\top = \boldsymbol{V}$. One has to pay attention to the formula $\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top$

Let $\boldsymbol{U_2} = (\mathbf{u_1}\ \mathbf{u_2}) = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{pmatrix}, \boldsymbol{\Sigma_2} =$

$\begin{pmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{pmatrix}$ Since the last row of $\boldsymbol{\Sigma}$ is zero row we deduce

$$\boldsymbol{U\Sigma} = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \\ 0 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{pmatrix} = \boldsymbol{U_2 \Sigma_2}$$

The reduced SVD of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{U_2 \Sigma_2 V_2^\top} =$

$$\begin{pmatrix} -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{1}{-\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

(where $\boldsymbol{V_2} = \boldsymbol{V}$)

133 Basic properties of $A^\top A$, $AA^\top$

Lemma: For any $A \in \mathbb{R}^{m\times n}$
$\mathbf{rank}\, A = \mathbf{rank}\, A^\top A = \mathbf{rank}\, AA^\top$.

Proof. From page 122 $A\mathbf{x} = \mathbf{0} \iff A^\top A\mathbf{x} = \mathbf{0}$ So $\mathbf{nul}\, A = \mathbf{nul}\, A^\top A \Rightarrow \mathbf{rank}\, A = n - \mathbf{nul}\, A = n - \mathbf{nul}\, A^\top A = \mathbf{rank}\, A^\top A$

Hence $\mathbf{rank}\, A = \mathbf{rank}\, A^\top = \mathbf{rank}(A^\top)^\top A^\top = \mathbf{rank}\, AA^\top$

Lemma: Let $A = A^\top \in \mathbb{R}^{n\times n}$. Then $\mathbf{nul}\, A$ is the number of zero eigenvalues of A, and $\mathbf{rank}\, A$ is the number of nonzero eigenvalues of A

Proof. $N(A)$ is the subspace of all $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$. This subspace is nontrivial, i.e. $\mathbf{nul}\, A = \dim N(A) > 0$. $\mathbf{nul}\, A = 0$ iff and only if A has no zero eigenvalues. So $\mathbf{rank}\, A = n$ iff al the eigenvalues of A are nonzero.

Assume that $\mathbf{nul}\, A > 0$. Since A is diagonable $\dim N(A)$ is equal to the number of zero eigenvalues of A.

Lemma: $A^\top A$ and $A^\top A$ are nonnegative definite, and the number of positive eigenvalues of $A^\top A$ and $A^\top A$ is equal to the rank of A.

Proof. $\mathbf{x}^\top A^\top A\mathbf{x} = (A\mathbf{x})^\top (A\mathbf{x}) = ||A\mathbf{x}||^2 \geq 0$ hence $A^\top A \succeq \mathbf{0}$. Similarly $AA^\top \succeq \mathbf{0}$.

134 The Reduced SVD of A

Let $A \in \mathbb{R}^{m\times n}$ and $A = U\Sigma V^\top$

$U = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_m) \in \mathbb{R}^{m\times m}, V =$ $(\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n)$ are orthogonal matrices. $\Sigma \in \mathbb{R}^{m\times n}$ is "diagonal" matrix with the singular values on the diagonal, see p'261. Moreover $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$, while other singular values equal to zero. (This follows from the fact that $\sigma_1^2 \geq \sigma_2^{\geq} \sigma_n^2 \geq 0$ are the eigenvalues of $A^\top A$, which have exactly $r = \mathbf{rank}\, A$ positive eigenvalues.

Recall that $A^\top A\mathbf{v}_i = \sigma_i \mathbf{v}_i,\ i =$ $1, \ldots, n,\ A^\top \mathbf{u}_j = \sigma_j \mathbf{u}_j,\ j = 1, \ldots, m$

(See pages 261-263) $\mathbf{v}_i, \mathbf{u}_j$ are called the right and the left singular vectors of A

For $p \le r = \operatorname{rank} A$ let $U_p := (\mathbf{u}_1, \ldots, \mathbf{u}_p) \in \mathbb{R}^{m \times p}$, $V_p := (\mathbf{v}_1, \ldots, \mathbf{v}_p) \in \mathbb{R}^{n \times p}$ be the matrices obtained from U, V by retaining their first p columns respectively. Let $\Sigma_p = \operatorname{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p \times p}$ and $r = \operatorname{rank} A$.

Claim $A = U_r \Sigma_r V_r^{\mathrm{T}} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\top + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top \ (*)$

(Reduced Singular Value Decomposition (RSVD)).

Proof. Let $\Sigma_{r,1} \in \mathbb{R}^{r \times n}$ be the matrix obtained from $\Sigma \in \mathbb{R}^{m \times n}$ by deleting the last $m - r$ zero rows of Σ. As in the example on p' 269 $U\Sigma = U_r \Sigma_{r,1}$. So $A = U\Sigma V = U_r \Sigma_{r,1} V^\top$. Since Σ_r is obtained by deleting the last $n - r$ columns it follows that

$$\Sigma_{r,1} V^\top = (V \Sigma_{r,1}^\top)^\top = (V_r \Sigma_r)^\top = \Sigma_r^\top V_r^\top = \Sigma_r V_r^\top.$$

Hence $A = U_r \Sigma_{r,1} V^\top = U_r \Sigma_r V_r^\top$

The last equality in (*) is obtained by straightforward computation

Advantages of RSVD: First, the computation of U_r, V_r are faster than the computation of U, V. Second the storage memory for$U_r,\ V_r,\ \Sigma_r$ is $r(m+n+1)$ may be may be much less than the storage memory for U, V, Σ, which is m^2+n^2+r if $r << \min(m,n)$.

For $p < r$ let

$$A_p := U_p\Sigma_p V_p^{\mathrm{T}} = \sigma_1 \mathrm{u}_1 \mathrm{v}_1^\top + \ldots + \sigma_p \mathrm{u}_p \mathrm{v}_p^\top = U \operatorname{diag}(\sigma_1, \ldots, \sigma_p, 0, \ldots, 0) V^{\mathrm{T}}.$$

Then $\operatorname{rank} A_p = p$ and A_p is the best l_2 approximation among all matrices $E \in \mathbb{R}^{m\times n}, \operatorname{rank} E \le p$:

$$||A-E||_F^2 \ge ||A-A_p||_F^2 = \sigma_{p+1}^2 + \ldots + \sigma_r^2.$$

Note that the storage memory for A_p is $p(m+n+1)$

135 Example

Find the best rank one approximation to

$$A = \begin{pmatrix} 6 & -2 \\ -3 & 5 \\ 0 & -4 \end{pmatrix}$$

Answer: The best rank one approximation is $A_1 = U_1 \Sigma_1 V_1^\top = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$. Using the results from the example on p 264 we obtain

$$A_1 = 6\sqrt{2} \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) =$$

$$2 \begin{pmatrix} -2 \\ 2 \\ -1 \end{pmatrix} (-1, 1) = \begin{pmatrix} 4 & -4 \\ -4 & 4 \\ 2 & -2 \end{pmatrix}$$

Applications to Digital Image Processing

In digital image processing a big matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is generated by recording a_{ij}: the information on the nature of the light at the place (i, j) on the grid. There are two major problems.

1. There are errors in some entries a_{ij} that should be corrected to improve the picture.

2. Can one condense the information stored in A such that it storage will be much smaller than mn?

Usually any picture has a lot of redundant information. That is the effective rank of A: the number eigenvalues that are not equal to zero numerically, denoted by p is relatively small. By considering A_p one filters a lot of noise and decreases the storage memory.