Outline of Lectures
in Applied Linear Algebra
Fall 2006

Shmuel Friedland
University of Illinois at Chicago

http://www.math.uic.edu/ ∼friedlan

Notes of Fall 2004

updated during the course

1 Main Topics of the Course

- MATRICES AND SYSTEMS OF EQUATIONS
- DETERMINANTS
- VECTOR SPACES
- LINEAR TRANSFORMATIONS
- ORTHOGONALITY
- EIGENVALUES

Text: Steven J. Leon, *Linear Algebra with Applications*, MacMillan, 7th edition, 2005.

Software: MatLab,

equivalent: Maple, Matematica.

2 Many 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 -8-28-06

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

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 2 -8-30-06

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 $(A|\mathrm{b}),\ (A|B)$

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

c. Elementary Row Operations (ERO)

- Interchange two rows

$$R_i \longleftrightarrow R_j, \quad (R_2 \longleftrightarrow R_4), \quad i \neq j.$$

- 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

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

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

d. Pivotal Row

e. Back(ward) Substitution for triangular form

Row Echelon Form of a matrix.

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

$$[\mathbf{0}\ \mathbf{0}\ \ \ldots \mathbf{0}|\ \mathbf{1}] \tag{4.1}$$

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):

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

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.

5 Lecture 3 -9-1-06

The general solution of the system in REF.

Assume that REF does not contain a row of the form (4.1):

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

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

$$\left(\begin{array}{cccc|c} 1 & -2 & 3 & -1 & 0 \\ 0 & 1 & 3 & 1 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{array}\right)$$

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:

$$\left(\begin{array}{cccc|c} 1 & 0 & b & 0 & u \\ 0 & 1 & d & 0 & v \\ 0 & 0 & 0 & 1 & w \end{array}\right)$$

x_1, x_2, x_4 lead variables x_3 free variable

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

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

$$x_4 = w.$$

Homogeneous Systems of Equations

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

Augmented Matrix $(A|0)$.

HSE is always solvable:

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

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

6 Applications

1. Traffic Flow

At each intersection the number of cars entering must be the same as the number leaving.

2. Markov chains

The sum of the total weights from each node is equal to 1.

3. Electrical Networks

Kirchhoff's Laws

1. At every node the sum of incoming ($+$ sign) and outgoing ($-$ sign) currents is equal to 0.

2. Around every closed loop the algebraic sum of voltage must equal to the algebraic sum of the voltage drops.

3. Ohm's law the voltage drops is current times resistance: $E = iR$.

7 Lecture 6 -9-3-04

§3. Matrix Notation

Vectors: Row Vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ 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.

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 $\mathbf{0}$.

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

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

$m \times n$ 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

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

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

8 Laws for $A \pm B$ and aA

1. $A + B = B + A$, commutative law
2. $(A + B) + C = A + (B + C)$, associative law
3. $A + 0 = A$ for each A, neutral element
4. $A + (-A) = 0$, unique anti element
5. $a(A + B) = aA + aB$ for each A, B, distributive law
6. $(a + b)A = aA + bA$, distributive law
7. $(ab)A = a(bA)$, distributive law
8. $1A = A$.

corollary: $0A = 0$ zero matrix:

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

$0 = 0A - 0A = (0A + 0A) - 0A = 0A.$

9 Products of Matrices

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:

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

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

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

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

The system of m equations in n unknowns

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

can be compactly written as

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

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

Multiplication 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.

$$A = (a_{ij})_{i=j=1}^{i=m,j=n}, \quad B = (b_{jk})_{j=k=1}^{j=n,k=p},$$

$$C = (b_{ik})_{i=k=1}^{i=m,k=p},$$

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + ... + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

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.

Finished here on 9.8.06

10 Lecture 9-11-06

Transpose of a matrix A^{T}.

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

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

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

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

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 $A^T = A$.

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

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

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

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

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

Indeed $(AA^T)^T = (A^T)^T A^T = AA^T$

$(A^T A)^T = A^T (A^T)^T = A^T A$

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:

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

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 $A^{k+q} = A^kA^q = A^qA^k$.

$A^0 := I_m$.

A invertible if there exists A^{-1} such that

$AA^{-1} = A^{-1}A = I_m$

$(\iff A^1A^{-1} = A^{-1}A^1 = A^0)$

If A invertible then $A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$ (condition NC):

$A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I\mathbf{x} = \mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$

We will show that later that if NC holds then A 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

11 Elementary Matrices – 9.13.06

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:

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

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

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

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

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

  $R_i + a \times R_j \longrightarrow R_i$

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

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

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 $\boldsymbol{E}$: $\boldsymbol{A} \longrightarrow \boldsymbol{E}\boldsymbol{A}$.

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

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

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

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

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

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

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

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

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

Elementary Row Operations in terms of Elementary Matrices

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

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

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

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

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

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

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:
$A^{-1} = M = E_k \ldots E_1$.

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: $R_2 + 2R_1 \longrightarrow R_2,\ R_3 - 3R_1 \longrightarrow R_3$:

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

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

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

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

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

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

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

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

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

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

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 $\iff$

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

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

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

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

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

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
$\mathrm{diag}(d_1, ..., d_n)^{-1} = \mathrm{diag}(d_1^{-1}, ..., d_n^{-1})$

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 $\mathbf{0}$:

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

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

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.

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 0:

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

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

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.

12 Lecture 9-13-04

DETERMINANTS

For a square matrix $A \in \mathbb{R}^{n \times n}$ determinant of A denoted by $\det \mathbf{A}$, (or $\det (\mathbf{A})$ as in Leon's book), is a real number such that $\det \mathbf{A} \neq 0 \iff A$ is invertible.

$$(a) \quad \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \mathbf{ad} - \mathbf{bc}.$$

$$(b) \quad \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} =$$

$$\mathbf{aei} + \mathbf{bfg} + \mathbf{cdh} - \mathbf{ceg} - \mathbf{afh} - \mathbf{bdi}$$

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) $\mathbf{det\ A} \neq \mathbf{0} \iff$ Row Echelon Form of $\boldsymbol{A}$ has the maximal number of possible pivots $\iff$ Reduced Row Echelon Form of $\boldsymbol{A}$ is the identity matrix.

$\boldsymbol{A}$ Is called singular if $\mathbf{det\ A} = \mathbf{0}$.

(e) The determinant of a matrix having at least one zero row or column is $\mathbf{0}$.

(f) $\mathbf{det\ A} = \mathbf{det\ A^T}$: The determinant of $\boldsymbol{A}$ is equal to the determinant of $\boldsymbol{A}^{\mathbf{T}}$.

(g) $\mathbf{det\ AB} = \mathbf{det\ A\ det\ B}$: The determinant of the product of matrices is equal to the product of determinants.$\Rightarrow$

(h) If $\boldsymbol{A}$ is invertible then $\mathbf{det\ A^{-1}} = \frac{\mathbf{1}}{\mathbf{det\ A}}$.

$\boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A} \Rightarrow$

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

Determinants of Elementary Matrices

(i) $\mathbf{\det E_I = -1}$ where E_I corresponds to interchanging two rows: $R_i \leftrightarrow R_j$.

$$\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 0 \cdot 0 - 1 \cdot 1 = -1.$$

(j) $\mathbf{\det E_{II} = a}$ where E_{II} corresponds to multiplying a row by a: $R_i \to aR_i$. (Note that E_{II} is diagonal.)

$$\det \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} = \mathrm{a}.$$

(k) $\mathbf{\det E_{III} = 1}$ 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)

$$\det \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} = 1. \quad (\mathrm{R_2 + aR_1 \to R_2})$$

Computing Determinants using Elementary Matrices

Let $A \in \mathbb{R}^{n\times n}$ and perform k ERO:

$$A \xrightarrow{ERO_1} A_1 \xrightarrow{ERO_2} A_2 \xrightarrow{ERO_3} \dots A_{k-1} \xrightarrow{ERO_k} A_k$$

where A_k is a Row Echelon Form of A.

$$A_1 = E_1 A, \ A_2 = E_2 A_1 = E_2 E_1 A, \dots$$

$$A_k = E_k \dots E_1 A \Rightarrow$$

$$A_k = MB, \ M = E_k E_{k-1} ... E_2 E_1$$

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

$A = M^{-1} A_k \Rightarrow \det \mathbf{A} = \det \mathbf{M}^{-1} \det \mathbf{A_k} = \frac{\det \mathbf{A_k}}{\det \mathbf{E_1} \cdot \det \mathbf{E_2} \cdot \dots \cdot \det \mathbf{E_k}}$.

A_k is upper triangular. Hence the determinant of A is equal to the product of diagonal elements of A_k divided by the product of the determinants of the elementary matrices appearing in the above decomposition of A.

If A has two identical rows or columns then $\det \mathbf{A} = 0$.

Proof. Assume that rows i and j are equal. Subtract row i from row j to obtain $A_1 = E_{III} A$, which has a zero $j-th$ row. Hence $0 = \det \mathbf{A_1} = \det \mathbf{A}$.

As $\det \mathbf{A^T} = \det \mathbf{A}$ the second case follows too.

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 $(-1)^{i+j}\det \mathrm{M_{ij}}$.

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

$$M_{32} = \begin{pmatrix} a & c \\ d & f \end{pmatrix},$$

$$A_{32} = -af + cd.$$

Expansion of the determinant by row i:

$$\mathbf{det\ A = a_{i1}A_{i1} + a_{i2}A_{i2} + ... + a_{in}A_{in}}$$

$$= \sum_{j=1}^{n} a_{ij} A_{ij}$$

Expansion of the determinant by column p:

$$\mathbf{det\ A = a_{1p}A_{1p} + a_{2p}A_{2p} + ... + a_{np}A_{np}}$$

$$= \sum_{j=1}^{n} a_{jp} A_{jp}$$

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.

13 Lecture 9-15-04

Adjoint Matrix and Cramer's Rule

For $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$

the adjoint matrix is defined as

$$\mathbf{adj\ A} = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}$$

where A_{ij} is the (i, j) cofactor of A.

Note that the i-th row of $\mathbf{adj\ A}$ is $(A_{1i}\ A_{2i} \dots A_{ni})$.

Examples:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \mathbf{adj\ A} =$$

$$\begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\mathbf{adj\ A} = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}$$

$$A_{33} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \mathrm{a}_{11}\mathrm{a}_{22} - \mathrm{a}_{12}\mathrm{a}_{21}$$

$$A_{12} = -\det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} =$$

$$-\mathrm{a}_{21}\mathrm{a}_{33} + \mathrm{a}_{23}\mathrm{a}_{31}.$$

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:

$$A\ \mathbf{adj\ A} = \mathbf{(adjA)A} = \mathbf{(det\ A)I},$$

where I is the identity matrix of the corresponding size.

Proof. Consider the (i, j) element of the product $A\ \mathbf{adj\ A}$: $a_{i1}A_{j1} + a_{i2}A_{j2} + \ldots + a_{in}A_{jn}$. 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 $\mathbf{det\ A}$, which is the (i, i) entry of the diagonal matrix $\mathbf{(det\ A)I}$.

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 $\mathbf{det\ B_j} = \mathbf{0}$. This shows $A\ Adj A = \mathbf{(det\ A)I}$. Similarly $\mathbf{(adj\ A)A} = \mathbf{(det\ A)I}$.

Corollary: $\mathbf{det\ A} \neq \mathbf{0} \Rightarrow \mathbf{A^{-1}} = \frac{\mathbf{1}}{\mathbf{det\ A}}\mathbf{adj\ A}$.

Cramer's Rule

Consider the linear system of n equations with n unknowns:

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \dots & + & a_{nn}x_n = b_n \end{array}$$

Let $A \in \mathbb{R}^{n\times n}, \mathrm{b} = (b_1, \dots, b_n)^{\mathrm{T}}$ be the coefficient matrix and the column vector corresponding to the right-hand side of these system. That is the above system is $A\mathrm{x} = b, \ \mathrm{x} = (x_1, \dots, x_n)^{\mathrm{T}}$. 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 $j = 1, \dots, n$.

Proof of Cramer's Rule:

Since $\det \mathbf{A} \neq 0$, $A^{-1} = \frac{1}{\det \mathbf{A}} \mathrm{adj}\ \mathbf{A}$. Hence the solution to the system $A\mathbf{x} = \mathbf{b}$ is given by:

$A^{-1}\mathbf{x} = \frac{1}{\det \mathbf{A}} \mathrm{adj}\ \mathbf{A}\mathbf{b}$. Writing down the formula for the matrix $\mathrm{adj}\ \mathbf{A}$ 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. $\square$

14 Vector Spaces- 9-22-04

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 $a \in \mathbb{R}$, $a\mathbf{x}$ 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. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ for each $\mathbf{x}, \mathbf{y}$, distributive law
6. $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$, distributive law
7. $(ab)\mathbf{x} = a(b\mathbf{x})$, distributive law
8. $1\mathbf{x} = \mathbf{x}$.

corollary: $0\mathbf{x} = \mathbf{0}$ neutral element:
$0\mathbf{x} = (0 + 0)\mathbf{x} = 0\mathbf{x} + 0\mathbf{x} \Rightarrow$
$\mathbf{0} = 0\mathbf{x} - 0\mathbf{x} = (0\mathbf{x} + 0\mathbf{x}) - 0\mathbf{x} = 0\mathbf{x}$.

Examples:

1. $\mathbb{R}$ - Real Line

2. $\mathbb{R}^2$ = Plane

3. $\mathbb{R}^3$ - Three dimensional space

4. $\mathbb{R}^n$ - n-dimensional space

5. $\mathbb{R}^{m\times n}$ - Space of $m \times n$ matrices

6. Spaces of upper triangular, lower triangular and diagonal matrices

7. $\mathcal{P}_n$ - Space of polynomials of degree less than n:
$\mathcal{P}_n := \{p(x) =$
$a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_1 x + a_0\}.$

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).

15 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 $\mathbf{x} \in \mathbf{W}, \ a \in \mathbb{R} \Rightarrow a\mathbf{x} \in \mathbf{W}$.

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

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

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

1. $\mathbf{V}$.

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

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 $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$: 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.

16 Linear combination & span 9-27-4

For $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbf{V}$ and $a_1, ..., a_k \in \mathbb{R}$ the vector

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_k\mathbf{v}_k$$

is called a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_k$.

The set of all linear combinations of $\mathbf{v}_1, ..., \mathbf{v}_k$ is called the span of $\mathbf{v}_1, ..., \mathbf{v}_k$ and denoted by $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$.

Claim: $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ is a linear subspace of $\mathbf{V}$.

Fact: All subspaces in a finite dimensional vector spaces are always given as $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ for some corresponding vectors $\mathbf{v}_1, ..., \mathbf{v}_k$.

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 $\mathbf{N}(A) = \mathbf{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$.

9-29-04

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 $\mathbf{x} \in \mathbf{W} := \mathrm{span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}), \mathbf{y} \notin \mathbf{W}$.

Spanning set of a vector space

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

Example: Let $\mathbf{V}_{even}, \mathbf{V}_{odd} \subset \mathcal{P}_5$ be the subspaces of even and odd polynomials of degree 4 at most. Then $\mathbf{V}_{even} = \mathrm{span}(1, x^2, x^4)$, $\mathbf{V}_{odd} = \mathrm{span}(x, x^3)$.

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.

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

10-1-04

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

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 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span $\mathbb{R}^n \iff \det\, [\mathbf{v}_1 \ \mathbf{v}_2 \ldots \mathbf{v}_n] \neq \mathbf{0}$.

17 Linear Independence

$\mathbf{v}_1, ..., \mathbf{v}_n \in V$ 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 $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ are linearly independent $\iff$ every vector in $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_n)$ can be written as a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_n$ in a unique (one) way. (Explain!)

$\mathbf{v}_1, ..., \mathbf{v}_n \in V$ are linearly dependent $\iff$ $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ are not linearly independent.

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

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).

18 Wronskian

Let $f_1, ..., f_n \in C^{n-1}(a, b)$. Then the Wronskian of $f_1, ..., f_n$ is defined as the following determinant:

$W[f_1, ..., f_n](x) :=$

$$\begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \vdots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

Claim If $W[f_1, f_2, ..., f_n](c) \neq 0$ for some $c \in (a, b)$ then $f_1, ..., f_n$ are linearly independent functions in $C(a, b)$.

19 Basis and dimension 10-9-04

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

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 $\dim \mathbf{V}$.

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

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

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

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

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

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

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

20 Dimension and basis for row, column and null space-(§3.6 Leon)

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 $\mathbf{0}$ 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:
$\mathbf{nul}\, A := \dim \mathrm{N}(A) = n - \mathbf{rank}\, A$.

21 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 $\mathbb{R}^n \iff \det\ [\mathbf{v}_1\ \mathbf{v}_2 ... \mathbf{v}_n] \neq 0$.

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 $\dim \mathbf{V} = \mathbf{rank}\ A$ 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 $\mathbf{V}$.

22 Change of basis 10-13-04

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 $\mathbf{u} \in \mathbf{V}$ 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

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

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: $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is a basis in

$\mathbf{V} \iff \det \mathbf{U} \neq \mathbf{0}.$

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

$$[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n] = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]U. \qquad (22.1)$$

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 (22.1) 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 $\mathbf{x} = [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n](a_1, a_1, ..., a_n)^\mathrm{T} \iff$
$\mathbf{x} = a_1\mathbf{u}_1 + \ldots a_n\mathbf{u}_n$, i.e. the vector coordinates of $\mathbf{x}$ in the basis $[\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n]$ is $\mathbf{a} := (a_1, a_2, ..., a_n)^\mathrm{T}$.

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

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

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

Theorem: 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 $C = U^{-1}W$.

23 Linear Transformations 10-18-04

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 $T : \mathbf{V} \longrightarrow \mathbf{W}$. (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 $f : \mathbb{R} \longrightarrow \mathbb{R}$.

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

Example 3: A transformation $f : \mathbf{V} \longrightarrow \mathbb{R}$ 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 $T : \mathbf{V} \longrightarrow \mathbb{R}^2$.

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

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

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

Proof $\mathbf{0} = 0T(\mathrm{v}) = T(0\mathrm{v}) = T(\mathbf{0})$.

Linear transformation is also called linear operator

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

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

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

- $\mathrm{R}(T)$ is the image of T. $\mathrm{R}(T)$ Is a subspace of $\mathbf{W}$.
- $\ker T$ is the set of all vectors in $\mathbf{V}$ which mapped by T to a zero vector in $\mathbf{W}$. $\ker T$ is a subspace of $\mathbf{V}$.

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

$T(\mathrm{u}) = T(\mathrm{v}) = 0 \Rightarrow T(a\mathrm{u} + b\mathrm{v}) = aT(\mathrm{u}) + bT(\mathrm{v}) = a0 + b0 = 0$.

24 Matrix representations of linear transformations 10-18-04

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 $[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_m]$.

Theorem. 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 $\mathbf{b} = A\mathbf{a}$.

25 Change of the representation matrix under the change of bases

First we change a 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.

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

Second we 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 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}.

26 Scalar Product in $\mathbb{R}^n$ 10-23-06

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 $\mathbf{y}^{\mathbf{T}}\mathbf{x} = \mathbf{x}^{\mathbf{T}}\mathbf{y} = 0$.
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 $\cos\theta$.

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 $\frac{\mathbf{x}^\mathrm{T}\mathbf{y}}{||\mathbf{y}||} = \cos\theta ||\mathbf{x}||$.

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

27 Orthogonal subspaces 10-25-04

Two subspaces $\mathbf{U}$ and $\mathbf{V}$ 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: $\mathbf{N}(A) \perp \mathbf{R}(A^\mathrm{T})$.

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

Claim: Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be a basis in $\mathbf{U}$. Form a matrix $A = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_k) \in \mathbb{R}^{n\times k}$. Then $N(A^\mathrm{T}) = \mathbf{U}^\perp$ and $\dim \mathbf{U}^\perp = n - \dim \mathbf{U}$.

Claim: $\mathbb{R}(A^\mathrm{T})^\perp = \mathbf{N}(A)$ and $\mathbf{N}(A)^\perp = \mathbf{R}(A^\mathrm{T})$.

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

We write $\mathbb{R}^n = \mathbf{U} \oplus \mathbf{V}$ if $\mathbf{U}$ and $\mathbf{V}$ are subspaces of $\mathbb{R}^n$ and any vector $\mathbf{x} \in \mathbb{R}^n$ can be represented $\mathbf{x} = \mathbf{u} + \mathbf{v}$ for some unique vectors $\mathbf{u} \in \mathbf{U},\ \mathbf{v} \in \mathbf{V}$.

Equivalently if $\mathbf{u}_1, \ldots, \mathbf{u}_p$ is a basis of $\mathbf{U}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_q$ is a a basis in $\mathbf{V}$ then $p + q = n$ and $\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q$ is a basis of $\mathbb{R}^n$.

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 $\mathbf{y}^\mathrm{T}\mathbf{b} \neq \mathbf{0}$.

10-27-06 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 $P_\mathrm{U}(\mathbf{b})$: $(\mathbf{b} - P_\mathrm{U}(\mathbf{b})) \perp \mathbf{U}$.

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

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

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 $A\mathbf{x} = \mathbf{b}$. Furthermore $P_{\mathrm{R}(A)}(\mathbf{b}) = A\mathbf{z}$.

10-30-06 – Proofs

Claim 1: $\alpha \mathbf{b} = \alpha \mathbf{u} + \alpha \mathbf{v}$. As $\alpha \mathbf{u} \in \mathbf{U}$ and $\alpha \mathbf{v} \in \mathbf{U}^{\perp}$ it follows $P_{\mathbf{U}}(\alpha \mathbf{b}) = \alpha \mathbf{u} = \alpha P_{\mathbf{U}}(\mathbf{b})$. 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: $A^{\mathbf{T}}A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^{\mathbf{T}}A^{\mathbf{T}}A\mathbf{x} = \mathbf{0} \iff ||A\mathbf{x}||^2 = \mathbf{0} \Rightarrow \mathbf{x} \in \mathbf{N}(A) \Rightarrow \mathbf{x} \in \mathbf{N}(A^{\mathbf{T}}A)$. Let $B := A^{\mathbf{T}}A$ and $B^{\mathbf{T}} = B$. If $\mathbf{y} \in \mathbf{N}(B^{\mathbf{T}})$ 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.

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

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

Example: 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:

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

$$\underbrace{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\mathrm{z}} \underset{=}{=} \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}}_{\mathrm{b}.}.$$

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

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

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

$\det \mathrm{A^T A} =$
$\mathrm{m(x_1^2 + x_2^2 + \ldots + x_m^2) - (x_1 + x_2 + \ldots + x_m)^2}$.

$\det \mathrm{A^T A} = 0 \iff \mathrm{x_1 = x_2 = \ldots = x_m}$.

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

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

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

28 Inner Product Spaces 11-1-06

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)
- $\langle \boldsymbol{\alpha}\mathbf{x}, \mathbf{y} \rangle = \boldsymbol{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle$ for any scalar $\boldsymbol{\alpha} \in \mathbb{R}$. (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: $\langle \boldsymbol{\alpha}\mathbf{x} + \boldsymbol{\beta}\mathbf{z}, \mathbf{y} \rangle = \boldsymbol{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle + \boldsymbol{\beta}\langle \mathbf{z}, \mathbf{y} \rangle$.
- The symmetricity condition yields linearity in the second variable: $\langle \mathbf{x}, \boldsymbol{\alpha}\mathbf{y} + \boldsymbol{\beta}\mathbf{z} \rangle = \boldsymbol{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle + \boldsymbol{\beta}\langle \mathbf{x}, \mathbf{z} \rangle$.
- 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:

- $\mathbf{V} = \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x}$.
- $\mathbf{V} = \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T D \mathbf{x}$,
  $D = \mathrm{diag}(d_1, ..., d_n)$ is a diagonal matrix with positive diagonal entries. Then
  $\mathbf{y}^T D \mathbf{x} = d_1 x_1 y_1 + \ldots + d_n x_n y_n$.
- $\mathbf{V} = \mathbb{R}^{m \times n}$, $\langle A, B \rangle = \mathrm{tr}\, B^T A$.
- $\mathbf{V} = \mathrm{C}[a, b]$, $\langle f, g \rangle = \int_a^b f(x) g(x) dx$.
- $\mathbf{V} = \mathrm{C}[a, b]$, $\langle f, g \rangle = \int_a^b f(x) g(x) p(x) dx$, 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 $||\mathbf{u} - \mathbf{v}||$.

29 Orthonormal sets 11-3-06

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.

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 $i = 1, \ldots, n$.

$\mathbf{v}_1, \ldots, \mathbf{v}_n$ 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.

Normalization: A nonzero OS $\mathbf{u_1, ..., u_n}$ 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 $\mathbf{V}$. Denote $\mathbf{U} := \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Then

- Any vector $\mathbf{u} \in \mathbf{U}$ can be written as a unique linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$: $\mathbf{u} = \sum_{i=1}^{n} \langle \mathbf{u}, \mathbf{v}_i \rangle \mathbf{v}_i$.

- For any $\mathbf{v} \in \mathbf{V}$ the orthogonal projection $P_{\mathbf{U}}(\mathbf{v})$ on the subspace $\mathbf{U}$ is given by $P_{\mathbf{U}}(\mathbf{v}) = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i$. In particular $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle \geq \sum_{i=1}^{n} |\langle \mathbf{v}, \mathbf{v}_i \rangle|^2$ (Bessel's inequality) 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: $\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i$ (Parseval's formula) and $||\mathbf{v}||^2 = \sum_{i=1}^{n} |\langle \mathbf{v}, \mathbf{v}_i \rangle|^2$.

Orthogonal Matrices 11-8-06

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

Equivalently, the columns of Q form an OB in $\mathbb{R}^n$ with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^{\mathrm{T}} \mathbf{x}$.

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

Equivalently $(Q\mathbf{y})^{\mathrm{T}}(Q\mathbf{x}) = \mathbf{y}^{\mathrm{T}}\mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Equivalently $||Q\mathbf{x}||^2 = ||\mathbf{x}||^2$ for all $\mathbf{x} \in \mathbb{R}^n$.

11-9-04. $P \in \mathbb{R}^{n \times n}$ 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 $(y_1, ..., y_n)^{\mathrm{T}} = P(x_1, ..., x_n)^{\mathrm{T}}$ 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.

30 Gram-Schmidt process 11-10-06

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be linearly independent vectors in IPS $\mathbf{V}$. Then the Gram-Schmidt (orthogonalization) process gives a recursive way to generate ONS $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbf{V}$ from $\mathbf{x}_1, \ldots, \mathbf{x}_n$, such that $\mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ for $k = 1, \ldots, n$. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a basis of $\mathbf{V}$ then $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an ONB of $\mathbf{V}$.

GS-algorithm:

- $\mathbf{u}_1 := \frac{1}{||\mathbf{x}_1||}\mathbf{x}_1$.
- $\mathbf{p}_1 := \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \mathbf{u}_1$, $\mathbf{u}_2 := \frac{1}{||\mathbf{x}_2 - \mathbf{p}_1||}(\mathbf{x}_2 - \mathbf{p}_1)$.
- $\mathbf{p}_2 := \langle \mathbf{x}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_3, \mathbf{u}_2 \rangle \mathbf{u}_2$.
  $\mathbf{u}_3 := \frac{1}{||\mathbf{x}_3 - \mathbf{p}_2||}(\mathbf{x}_3 - \mathbf{p}_2)$.
- Assume that $\mathbf{u}_1, \ldots, \mathbf{u}_k$ were computed. Then $\mathbf{p}_k := \langle \mathbf{x}_{k+1}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_{k+1}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \ldots \langle \mathbf{x}_{k+1}, \mathbf{u}_k \rangle \mathbf{u}_k$ and
  $\mathbf{u}_{k+1} := \frac{1}{||\mathbf{x}_{k+1} - \mathbf{p}_k||}(\mathbf{x}_{k+1} - \mathbf{p}_k)$.

11-13-06 QR Factorization

Let $A = [\mathbf{a}_1\ \mathbf{a}_2 \ldots \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ matrix and assume that $\mathrm{rank}\, A = n \iff$ the columns of A are linearly independent. Perform G-S process with the following book keeping:

- $r_{11} := ||\mathbf{a}_1||,\ \mathbf{q}_1 := \frac{1}{r_{11}}\mathbf{a}_1.$
- 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
  $r_{kk} := ||\mathbf{a}_k - \mathbf{p}_{k-1}||,\ \mathbf{q}_k := \frac{1}{r_{kk}}(\mathbf{q}_k - \mathbf{p}_{k-1})$
  for $k = 2, ..., n$.

Let $Q = [\mathbf{q}_1\ \mathbf{q}_2 \ldots \mathbf{q}_n] \in \mathbb{R}^{m\times n}$ and

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{13} & \ldots & r_{2n} \\ 0 & 0 & r_{33} & \ldots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & r_{nn} \end{pmatrix}$$

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 $\hat{\mathrm{x}} = R^{-1}Q^{\mathrm{T}}\mathrm{b}$.

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 $(QQ^{\mathrm{T}})(QQ^{\mathrm{T}}) = Q(Q^{\mathrm{T}}Q)Q^{\mathrm{T}} = Q(I)Q^{\mathrm{T}} = QQ^{\mathrm{T}}$.

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

$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}.$$

31 Eigenvalues and Eigenvectors

11-13-06. 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 $\det\,(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$.

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.

$\mathbf{tr}\, A := a_{11} + a_{22} + \ldots + a_{nn} = \lambda_1 + \lambda_2 + \ldots + \lambda_n.$

$\det\ \mathbf{A} = \lambda_1 \lambda_2 \ldots \lambda_\mathbf{n}.$

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.

32 Similarity 11-15-06

Definition. Let $\mathbf{V}$ be a vector space with a basis $[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$. Let $T : \mathbf{V} \longrightarrow \mathbf{V}$ 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.

Systems of linear ordinary differential equations (SOLODE)

$$\begin{array}{ccccccc} y_1' & = & a_{11}y_1 & + & a_{12}y_2 & +\ldots+ & a_{1n}y_n \\ y_2' & = & a_{21}y_1 & + & a_{22}y_2 & +\ldots+ & a_{2n}y_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\vdots & \\ y_n' & = & a_{n1}y_1 & + & a_{n2}y_2 & +\ldots+ & a_{1n}y_n \end{array}$$

In matrix terms we write: $\mathbf{y}' = A\mathbf{y}$, where $\mathbf{y} = \mathbf{y}(t) = (y_1(t), y_2(t), ..., y_n(t))^{\mathrm{T}}$ and $A \in \mathbb{C}^{n\times n}$ a constant matrix.

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 Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det\ (\mathrm{A} - \lambda\mathrm{I}) = (\lambda_1 - \lambda)^{\mathrm{m}_1}(\lambda_2 - \lambda)^{\mathrm{m}_2}\ldots(\lambda_{\mathrm{k}} - \lambda)^{\mathrm{m}_{\mathrm{k}}}$, 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 $\mathrm{N}(A - \lambda_i) = \mathrm{span}(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i})$ 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 $\mathbf{y}(t_0) = \mathbf{c}$.

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

$\mathbf{z}' = A\mathbf{z}$, 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.

11-15-04 Diagonalization

Theorem Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det\,(\mathbf{A}-\lambda\mathbf{I}) = (\lambda_1-\lambda)^{\mathrm{m}_1}(\lambda_2-\lambda)^{\mathrm{m}_2}\ldots(\lambda_{\mathrm{k}}-\lambda)^{\mathrm{m}_{\mathrm{k}}}$, 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 $\mathrm{N}(A-\lambda_i I) = \mathrm{span}(\mathbf{x}_{i1},\ldots,\mathbf{x}_{im_i})$ for $i = 1,\ldots,k$. Form the matrix whose columns are the vectors which span the null spaces $X = (\mathbf{x}_{11}\ldots\mathbf{x}_{1m_1}\;\mathbf{x}_{21}\ldots\mathbf{x}_{2m_2}\ldots\mathbf{x}_{km_k}) \in \mathbb{C}^{n\times n}$ 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 $i = 1, \ldots, k$.

Then X is an invertible matrix and $A = XDX^{-1}$, i.e. A is similar to D.

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.

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.)

Powers of diagonable matrices

$A = XDX^{-1} \Rightarrow A^m = XD^mX^{-1}$,

$D^m = \mathrm{diag}(\lambda_1^m \ldots \lambda_n^m), \ m = 1, \ldots$

Iteration process:

$\mathbf{x}_m = A\mathbf{x}_{m-1}, \ m = 1, \ldots \Rightarrow \mathbf{x}_m = A^m \mathbf{x}_0$.

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 $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}}$. 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:

$A\mathbf{x} = \mathbf{x}, \quad \mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$,

$0 < x_1, \ldots, x_n, \quad x_1 + x_2 + \ldots x_n = 1$.

Exponential of a Matrix

For $A \in \mathbb{C}^{n\times n}$ let
$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \ldots$

If $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ then
$e^D = \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n})$.

If A is diagonable, i.e. $A = XDX^{-1}$ then
$e^A = Xe^D X^{-1}$.

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 $Y(t_0) = I$.

The solution of $\mathbf{y}' = A\mathbf{y}$ with the initial condition $\mathbf{y}(t_0) = \mathbf{a}$ is given by $\mathbf{y}(t) = e^{(t-t_0)A}\mathbf{a}$.

Defective matrices: $A \in \mathbb{C}^{n\times n}$ is called defective if it is not diagonable. Equivalently, there exists an eigenvalue λ_i of A of multiplicity $m_i > 1$ such that
$\dim \mathrm{N}(A - \lambda_i I) < m_i$. Example:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \ \det\,(\mathrm{A} - \lambda \mathrm{I}) = \lambda^2 =$$

$(\lambda - 0)^2, \ \dim \mathrm{N}(\mathrm{A}) = 1$.

11-29-06 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: $\det\,(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)^{\mathrm{m_1}}(\lambda_2 - \lambda)^{\mathrm{m_2}} \ldots (\lambda_{\mathrm{k}} - \lambda)^{\mathrm{m_k}}$, 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 $\mathrm{N}(A - \lambda_i I) = \mathrm{span}(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i})$. (This is done by solving the homogeneous system $(A - \lambda_i)\mathbf{x} = \mathbf{0}$ which has m_i free variables.) Perform Gram-Schmidt process on $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i}$ to obtain $\mathbf{y}_{i1}, \ldots, \mathbf{y}_{im_i}$ for $i = 1, \ldots, k$. Form the orthogonal matrix $Q = (\mathbf{y}_{11} \ldots \mathbf{y}_{1m_1}\ \mathbf{y}_{21} \ldots \mathbf{y}_{2m_2} \ldots \mathbf{y}_{km_k}) \in \mathbb{R}^{n\times n}$.

Main ideas of the proof of the theorem

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.

2. 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 $B = O^{\mathrm{T}}AO$ is symmetric $B^{\mathrm{T}} = (O^{\mathrm{T}}A^{\mathrm{T}}O)^{\mathrm{T}} = O^{\mathrm{T}}A^{\mathrm{T}}O = B$ and

$$B = \begin{pmatrix} c_{11} & \ldots & c_{1(n-1)} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{(n-1)1} & \ldots & c_{(n-1)(n-1)} & 0 \\ 0 & \ldots & 0 & \lambda \end{pmatrix}$$

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 this process (or induction) to deduce that C is diagonalizable by an orthogonal matrix. Hence A is diagonalizable by an orthogonal matrix.

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: $\mathbf{y}^{\mathrm{T}} A\mathbf{x} = (\mathbf{y}^{\mathrm{T}} A\mathbf{x})^{\mathrm{T}} = \mathbf{x}^{\mathrm{T}} A\mathbf{y} \Rightarrow \lambda\mathbf{y}^{\mathrm{T}}\mathbf{x} = \mu\mathbf{x}^{\mathrm{T}}\mathbf{y} \Rightarrow (\lambda-\mu)\mathbf{y}^{\mathrm{T}}\mathbf{x} = 0 \Rightarrow \mathbf{y}^{\mathrm{T}}\mathbf{x} = 0$.

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 $A - \lambda_i I$.

Singular Value Decomposition 11-30-04

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 $A = U\Sigma V^T$. (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 $B\mathbf{x} = \mathbf{0} \iff A\mathbf{x} = \mathbf{0}$ 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 $\operatorname{rank} A = \operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} C = r$. 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 $C = U \operatorname{diag}(\sigma_1^2, \ldots, \sigma_m^2) U^{\mathrm{T}}$. 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 $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)^{\perp}$. Then $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$ and $V^{\mathrm{T}} V = I_n$. Thus $A = U \operatorname{diag}(\sigma_1, \ldots, \sigma_n) V^{\mathrm{T}}$.

The Reduced SVD and Best Approximations for A.

For $p \leq \min(m,n)$ 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 $\Sigma_p = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p\times p}$ and $r = \mathrm{rank}\, A$. Then $A = U_r \Sigma_r V_r^{\mathrm{T}}$ (RSVD).

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

$$A_p := U_p \Sigma_p V_p^{\mathrm{T}} =$$
$$U\, \mathrm{diag}(\sigma_1, \ldots, \sigma_p, 0, \ldots, 0) V^{\mathrm{T}}.$$

Then $\mathrm{rank}\, A_p = p$ and A_p is the best l_2 approximation among all matrices $E \in \mathbb{R}^{m\times n}, \mathrm{rank}\, E \leq p$:
$||A - E||_F^2 \geq ||A - A_p||_F^2 = \sigma_{p+1}^2 + \ldots + \sigma_r^2$.

Note that the storage memory for A_p is $p(m+n+1)$

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.

Rayleigh quotient and mini-max characterizations

Let $A = A^{\mathrm{T}} \in \mathbb{R}^{n \times n}$. Then $\frac{\mathbf{x}^{\mathrm{T}} A \mathbf{x}}{\mathbf{x}^{\mathrm{T}} \mathbf{x}}$ for $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ 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
$\mathbf{y} := Q^{\mathrm{T}} \mathbf{x} \Rightarrow \mathbf{x}^{\mathrm{T}} \mathbf{x} = \mathbf{y}^{\mathrm{T}} \mathbf{y} = \mathbf{y}_1^2 + \ldots + \mathbf{y}_n^2$,
$\mathbf{x}^{\mathrm{T}} A \mathbf{x} = \mathbf{y}^{\mathrm{T}} D \mathbf{y} = \lambda_1 y_1^2 + \ldots + y_n^2 \leq \lambda_1^2 \mathbf{y}^{\mathrm{T}} \mathbf{y}$.
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.

Positive Definite Matrices

$A = A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ is called positive definite, denoted by $A > 0$, 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 $A > 0$ 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 following n determinants are positive: $a_{11} > 0, \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \ldots$

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} > 0.$$

Proof: