Linear Algebra 1

Zambelli Lorenzo
BSc Applied Mathematics

November 2020- January 2021

Matrices and Systems of equations

Systems of linear equations

Definition 1 (a linear equation in n unknowns) *A linear equation in n unknowns $x_1, x_2, ..., x_n$ is an equation of the form*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, ..., a_n$ and b are real numbers.

Definition 2 (a linear system of m equations in n unknowns) *A linear system of m equations in n unknowns $x_1, x_2, ..., x_n$ is a system of equations of the form*

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

where a_{ij}'s and bi's are real numbers

Terminology 1 ($m \times n$ system) *By an $m \times n$ system, we mean a linear system of m equations in n unknowns*

Terminology 2 *Here three definitions:*

- *By a* <u>*solution*</u> *of an $m \times n$ system, we mean n numbers $x_1, x_2, ..., x_n$ that satisfies all m equations.*
- *if a linear system has no solurion, we say that it is* <u>*inconsistent*</u>
- *if it has at least one solution, we say that is* <u>*consistent*</u>
- *The set of all solutions is called the* <u>*solution set*</u>

Definition 3 (Equivalent) *Two systems of equations involving the same variables are said to be equivalent if they have the same solution set*

Homogeneous systems

Definition 4 (Homogeneous systems) *An $m \times n$ linear system of the form*

$$a_{11}x_1 + a_{12}x_2 + \cdots + x_{1n}x_n = 0$$

Remark! Every $m \times n$ homogeneous system has the trivial solution $x_1 = x_2 = \cdots = x_n = 0$

Theorem 1 *An $m \times n$ homogeneous system has a nontrivial solution if $n > m$*

Matrices

Definition 5 (square matrix) *A matrix is said to be square if it has the same number of rows and columns.*

Definition 6 (coefficient matrix, augmented matrix) *Consider a linear system of m equations in n unknowns $x_1, x_2, ..., x_n$*

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

we associate two matrices to such a system:

Coefficient matrix: $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ ***Augmented matrix:*** $\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$

Definition 7 (Elementary row operations) *The following row operations on a matrix are called elementary row operations:*

1. *Interchange two rows*
2. *multiply a row by a nonzero real number*
3. *Replace a row by its sum with a multiple of another row*

Remark!: Elementary row operations on the augmented matrix would not change the solution set and hence create equivalent system.

Row Echelon form

Definition 8 (Row Echelon form) *A matrix is said to in row echelon form if it satisfies the following:*

- *the first nonzero entry in each nonzero row is 1*

- *if row k does not consist entirely of zeros, then the number of leading zero entries in row $k+1$ is greater than the number of leading zero entries in row k,*
- *if there are rows whose entries are all zero, then they are below the rows having nonzero entries*

Theorem 2 *Every matrix can be put into row echelon form by using elementary row operation 1,2 and 3.*

Definition 9 (Gaussian elimination) *The process of using row operations to transform a linear system into one whose augmented matrix is in row echelon form is called Gaussian elimination*

Terminology 3 *by lead variables we define the unknowns corresponding to the first nonzero elements in each row of the roe echelon form. By free variables we define all other unknowns.*

Definition 10 (reduced row echelon form) *A matrix is said to be in reduced row echelon form if the first nonzero entry in each row is the only nonzero entry in its column.*

Definition 11 (Gauss-Jordan reduction) *The process of transforming a matrix by using elementary row operations into reduced row echelon form is called Gauss-Jordan reduction*

Matrix arithmetic

Definition 12 *Let m, n be natural numbers. An $m \times n$ matrix is an arrangement of real numbers in m rows and n columns.*

The uppercase letter is used to define a matrix. The lowercase letter, instead, is used to define the corresponding value in the matrix=> a_{ij} i =row, j =column.

Terminology 4 *Notation:*

- $\mathbb{R}^{m \times n}$ *denotes the set of all $m \times n$ matrices with real entries;*
- $\mathbb{R}^n$ *denotes the set of all n-column vectors with real entries*
- $\vec{a}_i$ *denotes the ith row of A*
- a_j *denotes the jth row of A*

Definition 13 *A **Column vector** is a matrix with a single column. A **Row vector** $[\vec{a}_i]$ is a matrix with a single row.*

Definition 14 (Scalar multiplication) *If A is an $m \times n$ matrix and α is a scalar, then $\alpha \cdot A$ is the $m \times n$ matrix defined by $\alpha A = (\alpha \cdot a_{ij}$*

Definition 15 *If A and B are both $m \times n$ matrices, then the sum $A + B$ is the $m \times n$ matrix defined by $A + B = (a_{ij} + b_{ij})$*

Matrix multiplication

Consider an $m \times n$ linear system, we can written it in the compact form $Ax = B$ where:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Definition 16 *If $\vec{a}$ is an n-row vector and x is an n-column vector, then the scalar product $\vec{a}x$ is the scalar define by*

$$\vec{a}x = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1x_1 + a_2x_2 + \cdots + a_nx_n \end{bmatrix}$$

Definition 17 *Assume A is an $m \times n$ matrix and B $n \times p$ matrix. Then AB is the $m \times p$ matrix, whose entries are defined by:*

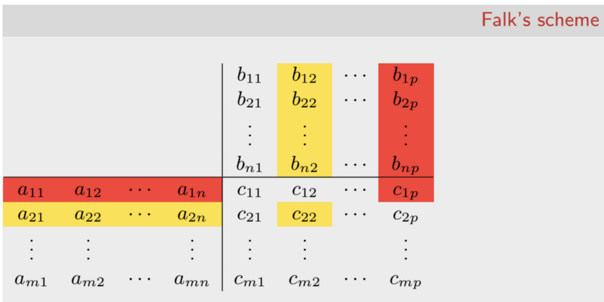
$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj} = \vec{a_i}b_j = \sum_{k=1}^{n} a_{ik}b_{kj}$$

Then:

a. The number of columns of A must be equal the numbers of rows of B;

b. If A is $m \times n$ and B is $n \times p$, then AB will be $m \times p$.

c. The matrix multiplication is associative



Definition 18 *Let A be a square matrix. For each $K \in \mathbb{N}$, we define:*
$A^k = A \cdots A$ for k times.

Matrix algebra

Definition 19 (Transpose) *The transpose of an an $m \times n$ matrix A is the $n \times m$ matrix $B = (b_{ij})$ define by $b_{ij} = a_{ji}$. The transpose of A is denoted by A^T*

Definition 20 (a symmetric square matrix) *A square matrix is called symmetric if $A^T = A$*

Definition 21 (Identity matrix) *The $n \times n$ identity matrix is the matrix $I_n = (\delta_{ij})$ where* $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

Remark! $I_n A = A$ for all $A \in \mathbb{R}^{n \times p}$

Definition 22 (nonsingularity and inverse matrix) *An $n \times n$ matrix A is said to be non-singular or invertible if there exists an $n \times n$ matrix B such that*

$$AB = BA = I_n$$

The matrix B is said to be the multiplicative inverse of A. We say that A is singular if it does not have a multiplicative inverse

Remark! Suppose that B and C are both inverse of A. Then, we have

$$B = BI = B(AC) = (BA)C = IC = C$$

therefore, a matrix can have at most one inverse. We will denote the inverse of a non-singular matrix A by A^{-1}

Theorem 3 *If A and B are non-singular matrices, then AB is also non-singular and $(AB)^{-1} = B^{-1}A^{-1}$*

Elementary Matrices

Definition 23 (Type I) *An elementary matrix of type 1 is matrix obtained by interchanging two rows of I.*

Definition 24 (Type II) *An elementary matrix of type II is matrix obtain by multiplying a row of I by a non-zero number*

Definition 25 (Type III) *An elementary matrix of type III is matrix obtain from I by adding a multiple of one row to another row.*

Theorem 4 (non-singularity of elementary matrices) *If E is an elementary matrix, then E is non-singular and E^{-1} is an elementary matrix of the same type.*

Definition 26 (Row equivalence) *A matrix B is row equivalent A if there exist elementary matrices $E_1, E_2, ..., E_k$ such that $B = E_k E_{k-1} \cdots E_2 E_1 A$*

Remark! If B is row equivalent to A is row equivalent to B: symmetric. Moreover, its transetivety

Computation: Transform the matrix $[A \quad I]$ into reduced row echelon form. If A is nonsingular then you will obtain $[I \quad A^{-1}]$:

$$E_k \cdots E_1 [A \quad I] = [I \quad A^{-1}]$$

Theorem 5 *Let A be a square matrix. Then the following statements are equivalent:*

a. *A is non-singular*

b. *$Ax = 0$ has only the trivial solution*

c. *A is row equivalent to I*

Theorem 6 (Uniqueness of solution to linear square system) *The linear system $Ax = b$ of n equations in n unknowns has a unique solution if and only if A is non-singular*

Definition 27 (Triangular matrices) *A square matrix A is said to be*

a. upper triangular *if $a_{ij} = 0$ for $i > j$*

b. lower triangular *if $a_{ij=0}$ for $j > i$*

c. triangular *if either upper ot lower triangular*

d. diagonal *if it is upper triangular and lower triangular*

e. strict upper(lower) triangular *if is upper(lower) triangular and every diagonal entry is nonzero*

Definition 28 (LU factorization) *If a square matrix A can be reduced to strict upper triangular form by using only row operation III, then it can be written as a product of a Lower and Upper triangular matrix. Such a factorization is called **LU factorization***

Partitioned matrices

Let A be a $n \times n$ matrix and B be $n \times p$ matrix.

- if $B = [B_1 \quad B_2]$ where $B_1 \in \mathbb{R}^{n\times q}$ and $B_2 \in \mathbb{R}^{n\times(p-q)}$:

$$AB = A[B_1 \quad B_2] = [AB_1 \quad AB_2]$$

- if $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ where $A_1 \in \mathbb{R}^{k\times n}$ and $A_2 \in \mathbb{R}^{(m-k)\times n}$

$$AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} A_1B \\ A_2B \end{bmatrix}$$

- if $A = [A_1 \quad A_2]$ where $A_1 \in \mathbb{R}^{m\times r}$ and $A_2 \in \mathbb{R}^{m\times(n-r}$ and $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ where $B_1 \in \mathbb{R}^{r\times p}$ and $B_2 \in \mathbb{R}^{(n-r)\times p}$:

$$AB = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = A_1B_1 + A_2 + B_2$$

Therefore, you can perform the multiplication as the matrix were scalar.

Determinants

Let $A = (a_{ij})$ be an $n \times n$ matrix and let M_{ij} denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting the row and column containing a_{ij}

Definition 29 *The determinant of an $n \times n$ matrix A, denoted by $det(A)$, is scalar defined recursively by*

$$det(A) = \begin{cases} a_{11} & \text{if } n = 1 \\ a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} & \text{if } n > 1 \end{cases}$$

where $A_{ij} = (-1)^{i+j} det(M_{ij})$

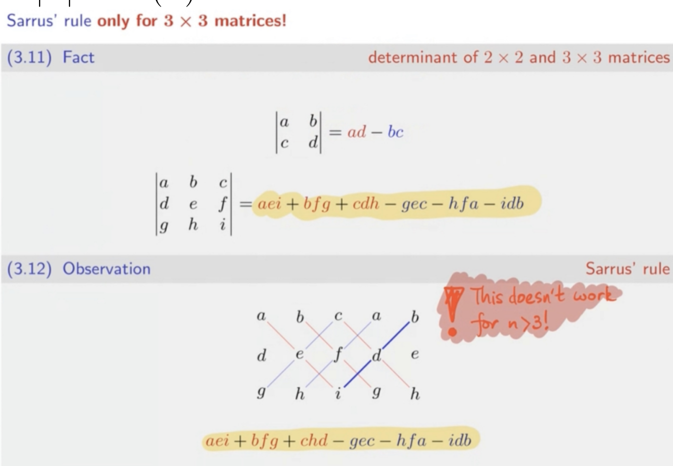
Terminology 5

- *$det(M_{ij})$ is called the minor of a_{ij}*
- *A_{ij} is called the cofactor of a_{ij}*
- *$det(A) = a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n}$ is called the cofactor expansion*

Theorem 7 *if $A \in \mathbb{R}^{n \times n}$ with $n \geq 2$, then $det(A)$ can be expressed as a cofactor expansion along any row or any column of A, that is*

$$\begin{aligned} det(A) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} \end{aligned}$$

Note:$|A| = det(A)$



Theorem 8 *Let A be a square matrix. Then, the following statements hold:*

a. *If A has a zero row or zero column, then $det(A) = 0$*

b. *if A has two identical rows or two identical columns, then $det(A) = 0$*

c. *$det(A^T) = det(A)$*

d. *if A is a triangular matrix, then $det(A)$ equals the product of its diagonal elements*

Theorem 9 *Let A be an nn matrix. if $i \neq j$, then $a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{in} = 0$*

Properties of Determinants

In summation, if E is an elementary matrix, then

$$det(EA) = det(E)det(A)$$

where

$$det(E) = \begin{cases} -1 & \text{if } E \text{ is of type I} \\ \alpha \neq 0 & \text{if } E \text{ is of type II} \\ 1 & \text{if } E \text{ is of type III} \end{cases}$$

Observation: since $det(B^T) = det(B)$, we have

$$det(AE) = det((AE)^T) = det(E^T A^T) = det(E^T)det(A^T) = det(E)det(A)$$

Theorem 10 *A square matrix A is nonsigular if only if* $det(A) \neq 0$

Theorem 11 *If A and B are* $n \times n$ *matrices, then*

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

Definition 30 *Let A be an* $n \times n$ *matrix. Its adjoint is defined by (where* $A_{ij} = (-1)^{i+j} det(M_{ij})$*)*

$$adj A = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

Fact:

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn} = \begin{cases} det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note: $A(adj(A)) = det(A)I$. If $det(A) \neq 0$ then:

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

Theorem 12 (Cramer's rule) *Let* $A \in \mathbb{R}^{n\times n}$ *and let* $b \in \mathbb{R}^n$*. Let* A_i *be the matrix obtained from A by replacing the ith column by b. If x is the unique solution of* $Ax = b$*, then* $x_1 = \frac{det(A_i)}{det(A)}$ *for* $i = 1, 2, ..., n$

Vectors Spaces

Definition 31 (Vector spaces) *Let V be a set and F be the set of scalars (*$\mathbb{R}$ *or* $\mathbb{C}$*). Also, let*

$$\oplus : V \times V \to V \quad \text{and } \odot : F \times V \to V$$

Be, respectively addition and scalar multiplication operations, that is

$$x \in V \quad \text{and } y \in V \Rightarrow x \oplus y \in V$$
$$\alpha \in F \quad \text{and } x \in V \Rightarrow \alpha \odot x \in V$$

We say that $(V, F, \oplus, \odot)$ *form a vector space if the following* ***axioms*** *are satisfied:*

A1. $x \oplus y = y \oplus x$ *for all* $x, y \in V$

A2. $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ *for all* $x, y, z \in V$

A3. There exists an element $0 \in V$ *such that* $x \oplus 0 = x$ *for all* $x \in V$

A4. For each $x \in V$*, there exists an element* $-x \in V$ *such that* $x \oplus (-x) = 0$

A5. $\alpha \odot (x \oplus y) = (\alpha \odot x) \oplus (\alpha \odot y)$ *for each scalar* α *and any* $x, y \in V$

A6. $(\alpha + \beta) \odot x = (\alpha \odot x) \oplus (\beta \odot x)$ *for any scalar* α, β *and any* $x \in V$

A7. $(\alpha\beta) \odot x = \alpha \odot (\beta \odot x)$ *for any scalars* α, β *and any* $x \in V$

A8. $1 \odot x = x$ *for all* $x \in V$

Definition 32 (The vector space C[a,b]) *Let* $C[a, b]$ *denote the set of all real-valued functions that are defined and continuous on the closed interval* $[a, b]$*. Let* $f, g \in C[a, b]$ *and* $\alpha \in \mathbb{R}$

$$(f \oplus g)(x) := f(x) + g(x) \quad \text{for all } x \in [a, b]$$
$$(\alpha \odot f)(x) := \alpha f(x) \quad \text{for all } x \in [a, b]$$

Definition 33 (The Vector Space P_n) *Let* P_n *denote the set of all polynomials of degree less than* n*.*

Theorem 13 *If* V *is a vector space and* $x \in V$*, then*

i. $0 \odot x = 0$

ii. $x \oplus y = 0 \Rightarrow y = -x$

iii. $(-1) \odot x = -x$

Notation: We write $x + y$ and x meaning $x \oplus y$ and $\alpha \odot x$

Definition 34 (subspace) *Let* S *be a subset of a vector space* V*. We say that* S *is a subspace of* V *if*

- S *is nonempty*
- $x \in S$ *and* $\alpha \in F \Rightarrow \alpha x \in S$
- $x \in S$ *and* $y \in S \Rightarrow x + y \in S$

Observation:

- $\{0\}$ and V are subspaces of V
- All other subspaces of V are refereed to as **proper subspaces**
- We refer to $\{0\}$ as the **zero subspace**
- If S is a subspace, then $0 \in S$

Null Space

Let A be an $n \times n$ matrix. Let $N(A)$ denote the set of all solution to the homogeneous system $Ax = 0$, thus

$$N(A) = \{x \in \mathbb{R}^n | Ax = 0\}$$

Therefore, $N(A)$ is a subspace of R^n. That is, the set all solutions of the homogeneous system $Ax = 0$ forms a subspace of $\mathbb{R}^n$. The subspace $N(A)$ is called the Null space of A

Span

Definition 35 (Span) *Let $v_1, v_2, ..., v_n$ be vectors in a vector space V*

- *For given scalars $\alpha_1, \alpha_2, ..., \alpha_n$, a sum of the form*

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

*is called **a linear combination** of $v_1, v_2, ..., v_n$*

- *The set of all linear combinations of $v_1, v_2, ..., v_n$ is called the **span** of $v_1, v_2, ..., v_n$ and is denoted by **span(**$v_1, v_2, ..., v_n$**)**. Alternatively,*

$$span(v_1, v_2, ..., v_n) = \{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n | \alpha_1, \alpha_2, ..., \alpha_n \text{ scalars}\}$$

Theorem 14 (span of vectors is a subspace) *Let V be a vector space and $v_1, v_2, ..., v_n \in V$, then $span(v_1, v_2, ..., v_n)$ is a subspace*

Definition 36 (spanning set) *The set $\{v_1, v_2, ..., v_n\}$ is a spanning set for V if every vector of V can be written as a linear combination of the vectors $v_1, v_2, ..., v_n$. In other words, if $span(v_1, v_2, ..., v_n) = V$*

Terminology:

- We say that $span(v_1, v_2, ..., v_n)$ is **spanned by** $v_1, v_2, ..., v_n$
- In case $\{v_1, v_2, ..., v_n\}$ is a spanning set for V, we say that the vectors $v_1, v_2, ..., v_n$ **span** V

Theorem 15 *Let $v_1, v_2, ..., v_n$ belong to a vector space V*

a. if $v_1, v_2, ..., v_n$ span V and one of them can be written as a linear combination of the other $n-1$ vectors, then those $n-1$ vectors span V

*b. One of the vectors $v_1, v_2, ..., v_n$ is a linear combination of the other $n-1$ vectors **if and only if** there exist scalars, $c_1, c_2, ..., c_n$, not all zero, such that*

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

Definition 37 (linear independence) *The vectors $v_1, v_2, ..., v_n$ in a vector space V are said to be*

- ***linearly independent*** *if*

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$$

  implies $c_1 = c_2 = \cdots = c_n = 0$

- ***linearly dependent*** *if there exists scalars* $c_1, c_2, ..., c_n$ *not all zero, such that*

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$$

Theorem 16 (linear dependence/independence in $\mathbb{R}^n$) *let* $x_1, x_2, ..., x_n$ *be vectors in* $\mathbb{R}^n$ *and let* $X = [x_1 \quad x_2 \quad \cdots \quad x_n]$. *The vectors* $x_1, x_2, ..., x_n$ *are linearly dependent if and only if* X *is singular.*

Theorem 17 *Let* $v_1, v_2, ..., v_n$ *be vectors in a vector space* V. *Every vector in span*$(v_1, v_2, ..., v_n)$ *can be written* ***uniquely*** *as a linear combination of* $v_1, v_2, ..., v_n$ *if and only if* $v_1, v_2, ..., v_n$ *are linearly independent*

Basis and dimension

Definition 38 (basis) *The vectors* $v_1, v_2, ..., v_n$ *form a basic for a vector space* V *if*

a. $v_1, v_2, ..., v_n$ *are* ***linearly independent***

b. $v_1, v_2, ..., v_n$ *span* V

Theorem 18 *let* $\{v_1, v_2, ..., v_n\}$ *be a spanning set for a vector space* V *and* m *be a positive integer with* $m > n$. *Then, any collection of* m *vectors in* V *is linearly dependent*

Corollary 1 *if both* $\{v_1, v_2, ..., v_n\}$ *and* $\{v_1, v_2, ..., v_m\}$ *are bases for a vectors space* V, *then* $n = m$

Definition 39 (Dimension) *let* V *be a vector space*

- *If* V *has a basis consisting of* n *vectors, we say that* V *has* ***dimension*** n
- *The subspace* $\{0\}$ *of* V *has dimension 0*
- V *is said to be* ***finite dimensional*** *if there is a finite set of vectors that spans* V
- *Otherwise, we say that* V *is* ***infinite dimensional.***

Theorem 19 *Let* V *be a vector space of dimension* $n > 0$. *Then,*

a. *any set of* n *linearly independent vectors spans* V

b. *any* n *vectors that span* V *are linearly independent*

c. *no set of fewer than* n *vectors can span* V

d. *any subset of fewer than* n *lin. ind. vectors can be completed to form a basis for* V

e. *any spanning set with more than* n *vectors can be trimmed down to form a basis for* V

Definition 40 (coordinates and coordinate vectors) *Let V be a vector space and $E = (v_1, v_2, ..., v_n)$ be an ordered basis for V. If $x \in V$, then we have*

$$x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

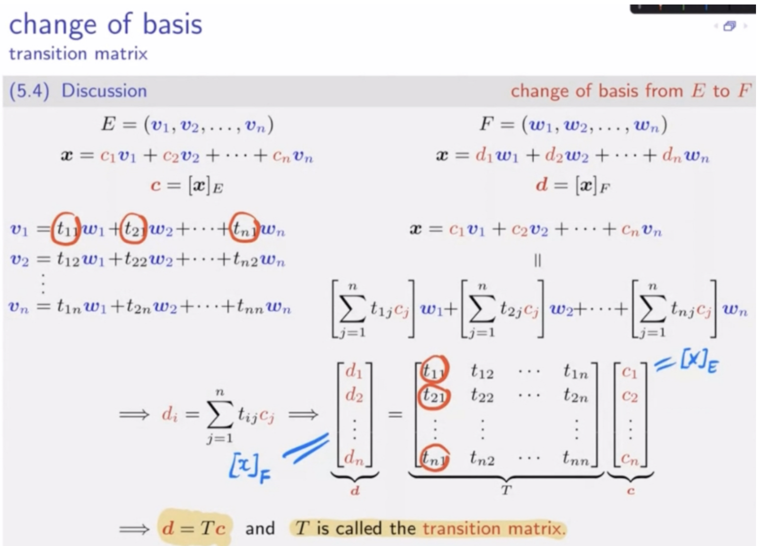
where $c_1, c_2, ..., c_n$ are scalars. As such can associate with each vector x a unique vector

$$c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

This vector c is called ***coordinate vector*** *of x* ***with respect to*** *E and is denoted by:*

$$[x]_E$$

The scalars $c_1, c_2, ..., c_n$ are called the ***coordinates*** *of x* ***relative to*** *E*



Row and Column spaces

Definition 41 (row/column spaces) *Let $A \in \mathbb{R}^{m \times n}$. The subspace of $\mathbb{R}^{1 \times n}$ spanned by the rows of A is called the* ***row space*** *of A*

The subspace of $\mathbb{R}^{m \times 1}$ spanned by the columns of A is called the ***column space*** *of A*

Theorem 20 *Two row equivalent matrices have the same row space*

Definition 42 (rank of a matrix) *The rank of a matrix A, denoted by* ***rank(A)****, is the dimension of the row space of A*

Observation:
To determine the rank of a matrix, we can first transform it to row echelon form. Since the nonzero rows of the echelon form is a basis for the row space, the number of nonzero rows is the rank.

Theorem 21 (consistency theorem for linear systems) *A linear system $Ax = b$ is consistent if and only if b is a linear combination of the columns of A*

Theorem 22 *A linear system $Ax = b$ is consistent if and only if b is in the column space of A*

Theorem 23 (linear systems) *Let $A \in \mathbb{R}^{m\times n}$:*

a. The linear system $Ax = b$ is ***consistent*** *for every $b \in \mathbb{R}^m$ if and only if the column vectors of A span $\mathbb{R}^m$*

b. The linear system $Ax = b$ ***has at most one solution*** *for every $b \in \mathbb{R}^m$ if and only if the column vectors of A are linearly independent*

Definition 43 (nullity of a matrix) *The dimension of the null space of A is called nullity of A and is denoted by* ***null(A)***

Theorem 24 (rank-nullity theorem) *If $A \in \mathbb{R}^{m\times n}$, the rank of A plus nullity of A equals n*

Note: $dim\, N(A) = n - r$ where $n = \#$ rows, and $r = \#$ lead variables. That is, $N(A) =$ the number of free variables

Theorem 25 *For every matrix, the dimension of the row space and that of the column space are equal*

Definition 44 *The Range of $A \in \mathbb{R}^m \times n$ is denoted by*

$$R(A) := \{b \in \mathbb{R}^m | b = Ax \text{ for some } x \in \mathbb{R}^n\}$$

Note: The range of A is the column space of A

Linear Transformation

Definition 45 *A mapping L from a vector space V into a vector space W is said to be a linear trasformation if*

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)$$

for all vectors $x, y \in V$ and scalars α, β

Terminology:

- $L : V \to W$ a mapping L from a vector space V into a vector space W
- $L : V \to V \Rightarrow L$ is a operator

Note:

$$C^k[a, b] := \{f : [a, b] \to \mathbb{R} | f^k \in C[a, b]\}$$

Example:
Let $A \in \mathbb{R}^{m\times n}$ and $L_A : \mathbb{R}^n \to \mathbb{R}^m$ given by $L_A(v) = Av$ for $v \in \mathbb{R}^n$ is linear:

$$L_A(\alpha x + \beta y) = A(\alpha x + \beta y) = \alpha Ax + \beta Ay = \alpha L_A(x) + \beta L_A(y)$$

Observation:
Let $L : V \to W$ be a linear transformation. Then,

a. $L(0_V) = 0_W$

b. $L(-v) = -L(v)$ for all $v \in V$

c. $v_1, v_2, ..., v_n$ are vectors of V and $\alpha_1, \alpha_2, ..., \alpha_n$ are scalars, then

$$L(\alpha_1, v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) = \alpha_1 L(v_1) + \alpha_2 L(v_2) + \cdots + \alpha_n L(v_n)$$

Definition 46 *Let $L : V \to W$ be a linear transformation. The karnel of L is defined by*

$$ker(L) := \{v \in V | L(v) = 0_W\}$$

Definition 47 *Let $L : V \to W$ be a linear transformation and let S be a subspace of V. The image of S under L, denoted by $L(S)$ is defined by*

$$L(S) := \{w \in W | w = L(v) \textit{ for some } v \in S\}$$

The image of $V, L(V)$, is called the range of L

Theorem 26 *if $L : V \to W$ be a linear transformation and let S be a subspace of V, then both $kerL$ and $L(S)$ are subspaces*

Theorem 27 (Matrix representation) *Let $E = (v_1, v_2, ..., v_n)$ and $F = (w_1, w_2, ..., w_n)$ be ordered bases for vector spaces V and W, respectively, and let $L : V \to W$ be a linear trasformation. Define the $m \times n$ matrix A by*

$$a_j = [L(v_j)]_F \qquad \textit{for } j = 1, 2, ..., n$$

Then,

$$[L(v)]_F = A[v]_E \qquad \textit{for all} v \in V$$

The matrix A is called the ***matrix representation*** *of L relative to the bases E and F*

Theorem 28 *Let $V \to V$ be a linear operator, $E = (v_1, v_2, ..., v_n)$ and $F = (w_1, w_2, ..., w_m)$ be ordered bases for the vector space V. Also, let S be the transition matrix representing the basis change from F to E. If A and B are the matrices representing L w.r.t E and F, respectively, then $B = S^{-1}AS$*

Definition 48 *Let A and B be $n \times n$ matrices. We say that B is similar to A if there exists a nonsingular matrix S such that $B = S^{-1}AS$*

Note: Similarity is symmetric

Orthogonality

Definition 49 ***Scalar product*** *Let x and y be two column vectors in $\mathbb{R}^n$. The product $x^T y$ is called the scalar product of x and y*

Definition 50 ***Length*** *The Euclidean length of a vector $x \in \mathbb{R}^n$ is defined by*

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

Observation:

$$\|x\| = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{if } x \in \mathbb{R}^2 \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{if } x \in \mathbb{R}^3 \end{cases}$$

Definition 51 ***Distance*** *The distance between two vectors $x, y \in \mathbb{R}^n$ is define as $\|x - y\|$*

Definition 52 ***Orthogonality*** *Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal if*

$$x^T y = 0$$

we write $x \perp y$

Theorem 29 ***Pythagorean law*** *if $x \perp y$, then*

$$\|x\|^2 + \|y\|^2 = \|x + y\|^2$$

Theorem 30 ***angle between two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$*** *if x and y are two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and θ is the angle between them, then*

$$x^T y = \|x\| \, \|y\| \cos\theta$$

Theorem 31 ***Cauchy-Schwartz inequality*** *Let x and y be two vectors in $\mathbb{R}^n$. Then,*

$$|x^T y| \leq \|x\|\|y\|$$

Observation: The angle between vectors

$$-1 \leq \frac{x^T y}{\|x\|\|y\|} \leq 1$$

Definition 53 ***Orthogonal subspaces*** *Two subspaces X and Y of $\mathbb{R}^n$ are said to be orthogonal if $x^T y = 0$ for every $x \in X$ and $y \in Y$*

Definition 54 ***Orthogonal complement*** *Let X be a subspace of $\mathbb{R}^n$. Define*

$$X^\perp := \{y \in \mathbb{R}^n | x^T y = 0 \text{ for all } x \in X\}$$

The set $X^\perp$ is called orthogonal complement of X

Observation:

- if $X \perp Y$, then $X \cap Y = \{0\}$
- $X^\perp$ is a subspace

Theorem 32 ***Fundamental subspace theorem*** *Let $A \in \mathbb{R}^{m \times n}$. Then,*

$$N(A) = R(A^T)^\perp \quad \text{and} \quad N(A^T) = R(A)^\perp$$

Observation:

- The column space of $R(A^T)$ is essentially the same as the row space of A

- The $N(A)$ contains all the vectors that are $\perp$ to the row space. That is, every x in the nullspace of A is orthogonal to the row space of A
- Hence, $N(A) \perp R(A^T)$. That is, $N(A) = R(A^T)^\perp$
- In the same way $N(A^T) \perp R(A)$. That is, every y in the nullspace of A^T is orthogonal to the column space of A, thus $N(A^T) = R(A)^\perp$
- By knowing that the dimension of $N(A)$ is $n - r$, we deduce that the dimension of $R(A^T)$ is r.
- In the same way by know that the dimension of $N(A^T)$ is $m - r$, then the dimension of $R(A)$ is r

Theorem 33 *If S is a subspace of $\mathbb{R}^n$, then $dim(S) + dim(S^\perp) = n$. Moreover, if $\{x_1, ..., x_r\}$ is a basis for S and $\{x_{r+1}, ..., x_n\}$ is a basis for $S^\perp$, then $\{x_1, ..., x_n\}$ is a basis for $\mathbb{R}^n$*

Definition 55 *Let U and V be subspaces of the vector space W. If each $w \in W$ can be written uniquely as a sum of $u + v$ where $u \in U$ and $v \in V$, then we say that W is a direct sum of U and V and write $W = U \oplus V$*

Theorem 34 ***Direct sum*** *If S is a subspace of $\mathbb{R}^n$, then $\mathbb{R}^n = S \oplus S^\perp$*

Theorem 35 *If S is a subspace of $\mathbb{R}^n$, then $(S^\perp)^\perp = S$*

Least squares problems

A standard technique in mathematical and statistical modeling is to find a *least squares* fit to a set of data points in the plane. The least squares curve is usually the graph of a standard type of function, such as a linear function, a polynomial, or a trigonometric polynomial. Since the data may include errors in measurement or experiment-related inaccuracies, we do not require the curve to pass through all the data points. Instead, *we require the curve to provide an optimal approximation in the sense that the sum of squares of errors between the y values of the data points and the corresponding y values of the approximating curve are minimized.*
Given $A \in \mathbb{R}^m \times n$ with $m > n$ and $b \in \mathbb{R}^m$. For each $x \in \mathbb{R}^n$, define the **residual**

$$r(x) = Ax - b$$

The distance between b and Ax is given by

$$\|b - Ax\| = \|r(x)\|$$

We wish to find a vector $\hat{x} \in \mathbb{R}^n$ for which $\|r(\hat{x})\|$ will be a minimum. Thus,

$$\|r(\hat{x})\| \leq \|r(x)\| \tag{1}$$

for all $x \in \mathbb{R}^n$. In other words we want to to minimize $\|r(x)\|$, equivalently $\|r(x)\|^2$.
A vector $\hat{x}$ satisfying (1) is said to be a **least squares solution** of the system $Ax = b$.

Is there always a least squares solution? Yes, its has it is state in the following theorem

Theorem 36 *Let S be subspace of $\mathbb{R}^m$. For every $b \in \mathbb{R}^m$,* ***there is a unique*** *vector p of S that is closet to b in the sense that*

$$\|b - p\| < \|b - y\|$$

for all $y \neq p$ in S. Moreover, a vector p is S is the closet to a given vector b if and only if $b - p \in S^{\perp}$

Terminology:
The Vector p is said to the **projection of b onto S**
Vector projection of x onto y:

$$p = \frac{x^T y}{y^T y} y$$

Scalar projection of x onto y

$$\alpha = \frac{x^T y}{\|y\|}$$

Now lets prove Theorem 36:

Theorem 34 (Direct sum) implies that each $b \in \mathbb{R}^m$ can be written uniquely as a sum $b = p + z$ where $p \in S$ and $z \in S^{\perp}$. If $y \neq p$ in S,

$$\|b - y\|^2 = \|(b - p) + (p - y)\|^2$$

Since $b - p = z \in S^{\perp}$ and $p - y \in S$,

$$\|b - y\|^2 = \|b - p\|^2 + \|p - y\|^2$$

due to Pythagorean law. Since $y \neq p$, $\|p - y\| > 0$. As such the theorem holds.

Observation:
If $\hat{x}$ is a least squares solution of the system $Ax = b$ and $p = A\hat{x}$, then p is the vector in $R(A)$ that is the closet to b

Moreover, in a solution of a least squares problem we can observe the following:

- In view of the previous observation, take $S = R(A)$ in the theorem 36
- a vector $\hat{x}$ is a least squares solution of the system $Ax = b$ if and only if
- $p = A\hat{x}$ is the vector in $R(A)$ that is closest to b if and only if
- $b - p = b - A\hat{x} \in R(A)^{\perp} = N(A^T)$ if and only if
- $A^T(b - A\hat{x}) = 0$ if and only if
- $A^T A\hat{x} = A^T b$ that is called the **normal equation**

Theorem 37 ***Uniqueness of least squares solution*** *If $A \in \mathbb{R}^{m \times n}$ is of $rank(A) = n$, then the normal equations*

$$A^T A\hat{x} = A^T b$$

have a unique solution

$$\hat{x} = (A^T A)^{-1} A^T b$$

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

Proof: It is enough to prove that A^TA is nonsingular:

- Let z be a vector s.t $A^TAz = 0$
- Then, we have $0 = z^TA^TAz = \|Az\|^2$
- This means that $Az = 0$
- Since $rank(A) = n$, the coulumn of A are linearly independent
- Therefore, we get $z = 0$
- Consequently, A^TA is nonsingular

How to solve a least squares problem:

Given a table of data we wish to find a linear function

$$y = c_0 + c_1x$$

that best fits the data in the least squares sense. If we require that

$$y_i = c_0 + c_1x_i \quad \text{for} \quad i = 1, ..., m$$

we get a system of m equations in two unknowns

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

The linear function whose coefficients are the least squares solution is said to be the best least squares fit to the date by a linear function.

If the data do not resemble a linear function, we could usa a higher degree polynomial. To find the coefficients $c_0, c_1, ..., c_n$ of the best least squares fit to the data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

by a polynomial of degree n, we must find the least squares solution the system

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Eigenvalues

Definition 56 ***Eigenvalues/eigenvectors*** *Let A be a square matrix. A scalar λ is said to be an* ***eigenvalue*** *of A if there exists a nonzero vector x such that*

$$Ax = \lambda x$$

The vector x is said to be an ***eigenvector*** *corresponding to λ*

Note: Powers of a matrix have the same eigenvectors
Proof:

$$A^2x = A(\lambda x) = \lambda(Ax) = \lambda^2 x \qquad A^n x = \lambda^n x$$

Observations: The following statements are equivalent:

- λ is an eigenvalue of A
- $(\lambda I - A)x = 0$ has a nontrivial solution
- $N(\lambda I - A) \neq \{0\}$
- $(\lambda I - A)$ is singular
- $det(\lambda I - A) = 0$

Terminology:

- If λ is an eigenvalue of A, then $N(\lambda I - A)$ is called **eigenspace** corresponding to λ
- $p_A(\lambda) = det(\lambda I - A)$ is called the **characteristic polynomial**
- if $A \in \mathbb{R}^{n\times n}$, then $p_A(\lambda)$ is a polynomial of degree n

Observations:

- If A is a square matrix with **real entitries**, then its characteristic polynomial has real coefficients
- As such, all its nonreal eigenvalues occur in **conjugate pairs**
- Also, the eigenvectors occur in **conjugate pairs**

$$Az = \lambda z \Rightarrow A\bar{z} = \bar{A}\bar{z} = \bar{Az} = \bar{\lambda z}$$

Note: Assume that we have a symmetric matrix $S = S^T$, then the eigenvectors of S are orthogonal.
Proof:

- $Sx = \lambda x$ $Sy = \alpha y$ $\lambda \neq \alpha$ $S^T = S$.
- Transponse to $x^T S^T = \lambda x^T$ and use $S^T = S$ thus $x^T S y = \lambda x^T y$
- We can also multiply $Sy = \alpha y$ by x^T, thus $x^T S y = \alpha x^T y$
- Nowe $\lambda x^T y = \alpha x^T y$. Since $\alpha \neq \lambda$, $x^T y$ must be zero

The product and Sum of the Eigenvalues

$$p_A(\lambda) = det(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

$$= \lambda^n + p_{n-1}\lambda^{n-1} + \cdots + p_1\lambda + p_0$$

$$= (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

$$= \lambda^n - (\lambda_1 + \lambda_2 + \cdots + \lambda_n)\lambda^{n-1} + \cdots + (-1)^n \lambda_1\lambda_2\cdots\lambda_n$$

If we expand it

$$p_A(\lambda) = (\lambda - a_{11})(\lambda - a_{22})\cdots(\lambda - a_{nn}) + q(\lambda) \qquad \text{where } deg(q) \leq n-2$$

$$= \lambda^n - (a_{11} + a_{22} + \cdots + a_{nn})\lambda^{n-1} + \bar{q}(\lambda) \qquad \text{where } deg(\bar{q}) \leq n-2$$

$$tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

$$p_A(0) = p_0 = det(-A) = (-1)^n det(A) = (-1)^n \lambda_1\lambda_2\cdots\lambda_n$$

$$det(A) = \lambda_1\lambda_2\cdots\lambda_n$$

Theorem 38 *Let A and B be $n \times n$ matrices. If A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues*

Diagonalization

Definition 57 *A squares matrix A is said to be diagonalizible if there exits a nonsingular matrix X and a diagonal matrix D such that*

$$X^{-1}AX = D \tag{2}$$

In case A is diagonalizable, we say that X diagonalizes A if (2) holds

Observation: $X^{-1}AX = D \Leftrightarrow AX = XD \Leftrightarrow A = XDX^{-1}$.
From this follows that $A^k = XD^kX^{-1}$

Theorem 39 *An $n \times n$ matrix is diagonalizable if and only if it has n linearly independent eigenvectors*

Observation: If A is diagonalizable, that is $X^{-1}AX = D$ for a nonsingular matrix X and a diagonal matrix D, then

- the columns vectors of X are **eigenvectors** of A
- the diagonal elements of D are **eigenvalues** of A
- X and D are not unique

Theorem 40 *If $\lambda_1, \lambda_2, ..., \lambda_k$ are distinct eigenvalues ($\lambda_i \neq$ for $i \neq j$) of A with the corresponding eigenvectors $x_1, x_2, ..., x_k$, then $x_1, x_2, ..., x_k$ are linearly independent*

Theorem 41 *Any squares matrix with distinct eigenvalues is diagonalizable*

Linear differential equations

A system of linear differential equations is of the form

$$x'(t) = Ax(t)$$

where ' denotes the derivative w.r.t time variable t, $x : \mathbb{R} \to \mathbb{R}^n$ is a vector-valued function, and A is an $n \times n$ matrix.

Definition 58 ***Initial valued problem*** *An initial valued problem amounts to finding a solution to*

$$x'(t) = Ax(t) \quad x(0) = x_0$$

for a given $n \times n$ matrix A and given n-vector x_0
The solution of the initial value problem is given by $x(t) = e^{tA}x_0$

Definition 59 ***Matrix exponential***

$$e^A := I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

Therefore, we deduce that if A is diagonizable then

$$\begin{aligned} A^k &= XD^kX^{-1} \qquad \text{for } k = 1, 2, ... \\ e^A &= X\left(I + D + \frac{1}{2!}D^2 + \frac{1}{3!}D^3 + \cdots\right)X^{-1} \\ &= Xe^DX^{-1} \end{aligned}$$

That is, the initial value problem is:

$$\begin{aligned} y_1'(t) &= a_{11}y_1 + \cdots + a_{1n}y_n \\ &\vdots \\ y_n(t) &= a_{n1}y_1 + \cdots + a_{nn}y_n \end{aligned}$$

Where we define

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

and

$$Y'(t) = AY \qquad Y(0) = Y_0$$

Which is it

$$Y' = e^{tA}Y_0 \tag{3}$$

that is

$$y(t) = c_1e^{\lambda_1 t}x_1 + c_2e^{\lambda_2 t}x_2 + \cdots + c_ne^{\lambda_n t}x_n \tag{4}$$

where $x_1, ..., x_n$ are the eigenvectors corresponding to the eigenvalues $\lambda_1, ..., \lambda_n$ and $c_1, ..., c_n$ are constant. Then, $y(t)$ is the solution of the system.

If A is diagonizable we can write (3) in the form

$$\begin{aligned} Y &= Xe^{tD}X^{-1}Y_0 \\ &= c_1e^{\lambda_1 t}x_1 + c_2e^{\lambda_2 t}x_2 + \cdots + c_ne^{\lambda_n t}x_n \qquad (c = X^{-1}Y_0) \end{aligned}$$

Theorem 42 *Let A and B be $n \times n$ matrices. If A and B are similar, that is $B = S^{-1}AS$ for some $n \times n$ nonsingular matrix S, then*

$$e^B = S^{-1}e^AS$$

If a given matrix A is similar to a diagonal matrix D, then $A = S^{-1}DS$ for some nonsingular matrix S and $e^A = S^{-1}e^DS$