Matrices

$\mathbb{R}$ denotes the set of all real numbers.

$\mathbb{C}$ denotes the set of all complex numbers.

We write $\mathbb{F}$ for either $\mathbb{R}$ or $\mathbb{C}$.

The numbers in $\mathbb{F}$ are called **scalars**.

A **matrix** is a rectangular array of symbols arranged in rows and columns.

The individual symbols in a matrix are called its entries.

All our matrices will use their entries from $\mathbb{F}$. That is, our matrices will be matrices of scalars.

Matrix-entries

A matrix looks like $\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$.

Call this matrix A. We write $A = [a_{ij}]$, where $a_{ij} \in \mathbb{F}$.

The **size** of this matrix is $m \times n$.

It has m number of rows and n number of columns.

An $m \times m$ matrix is called a square matrix; its **order** is m.

The entry a_{ij} in the ith row and jth column is called (i, j)th entry.

The set of all $m \times n$ matrices with entries from $\mathbb{F}$ will be denoted by $\mathbb{F}^{m \times n}$.

Equality

A **row vector** of size n is a matrix in $\mathbb{F}^{1\times n}$.

A **column vector** of size n is a matrix in $\mathbb{F}^{n\times 1}$.

Both $\mathbb{F}^{1\times n}$ and $\mathbb{F}^{n\times 1}$ are written as $\mathbb{F}^n$.
The vectors in $\mathbb{F}^n$ will be written as $(a_1,\ldots,a_n)$. We will sometimes write a column vector as $[b_1 \quad \cdots \quad b_n]^T$, or as $(b_1,\ldots,b_n)^T$ for saving vertical space.

Two matrices of the same size are considered **equal** when their corresponding entries are equal. That is, if $A=[a_{ij}]\in\mathbb{F}^{m\times n}$ and $B=[b_{ij}]\in\mathbb{F}^{m\times n}$, then

$$A=B \quad \text{iff} \quad a_{ij}=b_{ij} \ \text{ for each } \ i=1,2,\ldots,m,\ j=1,2,\ldots,n.$$

We write the **zero matrix** all whose entries are 0 as 0.
To show the size of the zero matrix, we sometimes write $0_{m\times n}$.

Sum & Scalar multiplication

Sum of two matrices are done entry-wise. That is, if $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{F}^{m\times n}$, then

$$A + B = [a_{ij} + b_{ij}] \text{ for each } i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n.$$

For instance, $\begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 5 & 6 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 8 & 8 \\ 4 & 4 & 4 \end{bmatrix}.$

A matrix is **multiplied by a scalar** entry-wise. That is, $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ and $\alpha \in \mathbb{F}$, then

$$\alpha A = [\alpha a_{ij}] \text{ for each } i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n.$$

For instance, $\quad 4 \begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 12 & 8 \\ 8 & 12 & 4 \end{bmatrix}.$

Some properties

The addition and scalar multiplication for matrices satisfy the following properties:

Let $A, B, C \in \mathbb{F}^{m \times n}$. Let $\alpha, \beta \in \mathbb{F}$. Then

1. $A + B = B + A$.
2. $(A + B) + C = A + (B + C)$.
3. $A + 0 = 0 + A = A$.
4. $A + (-A) = (-A) + A = 0$.
5. $\alpha(\beta A) = (\alpha\beta)A$.
6. $\alpha(A + B) = \alpha A + \alpha B$.
7. $(\alpha + \beta)A = \alpha A + \beta A$.
8. $IA = A$.

Matrix multiplication

Let $A = [a_{ik}] \in \mathbb{F}^{m\times n}$ and $B = [b_{kj}] \in \mathbb{F}^{n\times r}$. Then their **product** AB is a matrix $[c_{ij}] \in \mathbb{F}^{m\times r}$, where the entries are

$$c_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

$$\begin{bmatrix} a_{11} & & a_{1k} & & a_{1n} \\ a_{i1} & \cdots & a_{ik} & \cdots & a_{in} \\ a_{m1} & & a_{mk} & & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{1j} & b_{1r} \\ & \vdots & \\ b_{\ell 1} & b_{\ell j} & b_{\ell r} \\ & \vdots & \\ b_{n1} & b_{nj} & b_{nr} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{1j} & c_{1r} \\ c_{i1} & c_{ij} & c_{ir} \\ c_{m1} & c_{mj} & c_{mr} \end{bmatrix}$$

The ith row of A multiplied with the jth column of B gives the (i,j)th entry in AB.

Notice that AB is defined only when the number of columns in A is equal to the number of rows in B.

Example 1

$$\begin{bmatrix} 3 & 5 & -1 \\ 4 & 0 & 2 \\ -6 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -2 & 3 & 1 \\ 5 & 0 & 7 & 8 \\ 9 & -4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 22 & -2 & 43 & 42 \\ 26 & -16 & 14 & 6 \\ -9 & 4 & -37 & -28 \end{bmatrix}.$$

We may view matrix multiplication in another way.
Look at the above product.

The first column of the product is obtained from the entries in the first column of the second matrix multiplied with the respective columns of the first matrix, and then taken the sum. That is,

$$\begin{bmatrix} 22 \\ 26 \\ -9 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 4 \\ -6 \end{bmatrix} + 5 \begin{bmatrix} 5 \\ 0 \\ -3 \end{bmatrix} + 9 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}.$$

Similarly, verify other columns in the product.

Commutativity?

A peculiarity: Suppose $u \in \mathbb{F}^{1\times n}$ and $v \in \mathbb{F}^{n\times 1}$.
Then $uv \in \mathbb{F}$ but $vu \in \mathbb{F}^{n\times n}$.

$$\begin{bmatrix} 3 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 19 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 3 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 1 \\ 6 & 12 & 2 \\ 12 & 24 & 4 \end{bmatrix}.$$

So, matrix multiplication is not commutative.

Commutativity can break down due to various reasons.

- Even if AB is defined, BA may not be defined.
- AB and BA may not be of the same size.
- Even when they have same size, AB may not be equal to BA.

An example

The third case needs an example. It requires square matrices A and B of the same order where $AB \neq BA$.

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 7 \\ 6 & 11 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 8 & 13 \end{bmatrix}.$$

It does not mean that AB is never equal to BA. In some cases they may be equal.

Unlike numbers, product of two nonzero matrices can be a zero matrix. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Matrix multiplication Cont.

Verify the following properties of matrix multiplication:

1. If $A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times r}$ and $C \in \mathbb{F}^{r\times p}$, then $(AB)C = A(BC)$.
2. If $A, B \in \mathbb{F}^{m\times n}$ and $C \in \mathbb{F}^{n\times r}$, then $(A+B)C = AB + AC$.
3. If $A \in \mathbb{F}^{m\times n}$ and $B, C \in \mathbb{F}^{n\times r}$, then $A(B+C) = AB + AC$.
4. If $\alpha \in \mathbb{F}$, $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times r}$, then $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

We denote by I_n, an **identity matrix** of order n, whose (i,i)th entry is 1 for each i, and all other entries are 0.

Then $AI_n = I_m A = A$ for any matrix $A \in \mathbb{F}^{m\times n}$.

Sometimes, we just write I instead of I_n, if no confusion arises.

Powers of square matrices can be defined inductively by taking

$$A^0 = I \quad \text{and} \quad A^n = AA^{n-1} \ \text{ for } \ n \in \mathbb{N}.$$

Block form

Suppose $A \in \mathbb{F}^{m \times n}$. Write its ith row as $A_{i\star}$ Also, write its kth column as $A_{\star k}$. Then we can write A as a row of columns and also as a column of rows in the following manner:

$$A = [a_{ik}] = \begin{bmatrix} A_{\star 1} & \cdots & A_{\star n} \end{bmatrix} = \begin{bmatrix} A_{1\star} \\ \vdots \\ A_{m\star} \end{bmatrix}.$$

Write $B \in \mathbb{F}^{n \times r}$ similarly as

$$B = [b_{kj}] = \begin{bmatrix} B_{\star 1} & \cdots & B_{\star r} \end{bmatrix} = \begin{bmatrix} B_{1\star} \\ \vdots \\ B_{\star n} \end{bmatrix}.$$

Then their product AB can now be written as

$$AB = \begin{bmatrix} AB_{\star 1} & \cdots & AB_{\star r} \end{bmatrix} = \begin{bmatrix} A_{1\star}B \\ \vdots \\ A_{m\star}B \end{bmatrix}.$$

When writing this way, we ignore the extra brackets [and].

Inverse

A square matrix A of order m is called **invertible** iff there exists a matrix B of order m such that $AB = I = BA$.
Such a matrix B is called an **inverse** of A.

If B and C are inverses of A, then

$$C = CI = C(AB) = (CA)B = IB = B.$$

Therefore, an inverse of a matrix is unique and is denoted by A^{-1}.
We talk of invertibility of square matrices only.
All square matrices are not invertible.
For example, I is invertible but 0 is not. If $AB = 0$ for nonzero square matrices A and B, then neither A nor B is invertible. However,
If $A, B \in \mathbb{F}^{n \times n}$ are invertible matrices, then $(AB)^{-1} = B^{-1}A^{-1}$.

Transpose

Given a matrix $A \in \mathbb{F}^{m\times n}$, its **transpose** is a matrix in $\mathbb{F}^{n\times m}$, denoted by A^T, and is defined by

$$\text{the } (i,j)\text{th entry of } A^T = \text{the } (j,i)\text{th entry of } A.$$

Transpose of a row vector is a column vector, and transpose of a column vector is a row vector.
Transpose of A^T is A.
The transpose has the following properties:

1. $(A^T)^T = A$.
2. $(A+B)^T = A^T + B^T$.
3. $(\alpha A)^T = \alpha A^T$.
4. $(AB)^T = B^T A^T$.
5. If A is invertible, then A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$.

In the above properties, we assume that the matrices are of suitable size so that the operations are allowed.

Adjoint

We write $\overline{\gamma}$ for the **complex conjugate** of a scalar γ. That is, $\overline{\alpha + i\beta} = \alpha - i\beta$ for $\alpha, \beta \in \mathbb{R}$.
The **adjoint** of $A \in \mathbb{F}^{m\times n}$ is denoted as A^*, and is defined by

the (i,j)th entry of A^* = the complex conjugate of (j,i)th entry of A.

When A has only real entries, $A^* = A^T$. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1+i & 2 & 3 \\ 2 & 3 & 1-i \end{bmatrix}^* = \begin{bmatrix} 1-i & 2 \\ 2 & 3 \\ 3 & 1+i \end{bmatrix}.$$

The adjoint has the following properties:

1. $(A^*)^* = A$.
2. $(A+B)^* = A^* + B^*$.
3. $(\alpha A)^* = \overline{\alpha} A^*$.
4. $(AB)^* = B^* A^*$.
5. If A is invertible, then A^* is invertible, and $(A^*)^{-1} = (A^{-1})^*$.

Diagonal matrix

We now define some special types of matrices.

Let $A = [a_{ij}] \in \mathbb{F}^{n \times n}$ be a square matrix of order n. The entries a_{ii} are called as the **diagonal entries** of A. All other entries are called **off-diagonal** entries.

If all off-diagonal entries of A are 0, then A is called a **diagonal matrix**.

For example, I and 0 are diagonal matrices.

If A is a diagonal matrix with $a_{ii} = d_i$, then we write

$$A = \operatorname{diag}(d_1, d_2, \ldots, d_n).$$

Thus $I = \operatorname{diag}(1, 1, \ldots, 1)$.

Special types of matrices

The jth column of I is denoted by e_j.
The column vectors $e_1, \ldots, e_n$ in $\mathbb{F}^{n\times 1}$, are called the **standard basis vectors** of $\mathbb{F}^{n\times 1}$.

A **scalar matrix** is a diagonal matrix with equal diagonal entries.
Thus αI are the only scalar matrices for $\alpha \in \mathbb{F}$.

A square matrix $A = [a_{ij}] \in \mathbb{F}^{n\times n}$ is **upper triangular** when $a_{ij} = 0$ for $i > j$.

A square matrix $A = [a_{ij}] \in \mathbb{F}^{n\times n}$ is **lower triangular** when $a_{ij} = 0$ for $i < j$.

Triangualr matrices

A square matrix $A = [a_{ij}] \in \mathbb{F}^{n \times n}$ is **triangular** if it is upper triangular or lower triangular.

$$L = \begin{bmatrix} 1 & & \\ 2 & 3 & \\ 3 & 4 & 5 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 2 & 3 \\ & 3 & 4 \\ & & 5 \end{bmatrix}.$$

L is lower triangular; U is upper triangular. Both are triangular.

Transpose of a lower triangular matrix is upper triangular; and transpose of an upper triangular matrix is lower triangular.

Special Matrices

Let $A \in \mathbb{F}^{n\times n}$. It is called

- **hermitian** iff $A^* = A$.
- **skew hermitian** iff $A^* = -A$.
- **symmetric** iff $A^T = A$.
- **real symmetric** iff $\mathbb{F} = \mathbb{R}$ and A is hermitian.
- **skew symmetric** iff $A^T = -A$.
- **unitary** iff $A^*A = I = AA^*$.
- **orthogonal** iff $\mathbb{F} = \mathbb{R}$ and $A^TA = I = AA^T$.
- **normal** iff $A^*A = AA^*$.

Special Matrices Cont.

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & 4 \\ 3 & -4 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & -2i & 3 \\ 2i & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 2+i & 3 \\ -2+i & i & 4i \\ -3 & 4i & 0 \end{bmatrix}, \quad F = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}, \quad G = \frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{bmatrix}.$$

B is symmetric, also hermitian.
C is skew-symmetric, also skew-hermitian.
D is hermitian.
E is skew-hermitian.
F is a unitary.
G is an orthogonal matrix. Also, G is unitary.

Special Matrices Cont.

A skew-symmetric matrix must have a zero diagonal.

The diagonal entries of a skew-hermitian matrix must be 0 or purely imaginary.

Reason: $a_{ii} = -\overline{a}_{ii} \Rightarrow 2\text{Re}(a_{ii}) = 0.$

Any square matrix is a sum of a symmetric and a skew-symmetric matrix.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

Also, a sum of a hermitian and a skew hermitian matrices:

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*).$$

Orhogonal matrices

The following are examples of orthogonal 2×2 matrices:

$$O_1 := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad O_2 := \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

$$O_1 \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\cos\theta - b\sin\theta \\ a\cos\theta + b\sin\theta \end{bmatrix}.$$

Thus, O_1 is said to be a *rotation by an angle* θ.

$$O_2 \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\cos\theta + b\sin\theta \\ a\sin\theta - b\cos\theta \end{bmatrix}.$$

Thus, O_2 is called a *reflection by an angle* $\theta/2$ along the x-axis.

Take an angle θ, plot the points $(a,b)^T$, $O_1(a,b)^T$ and $O_2(a,b)^T$ to see the geometry.

Inner product of two vectors

The inner product is a generalization of the familiar dot product in the plane or space.

The **inner product** of two vectors $u = (a_1, \ldots, a_n)$ and $v = (b_1, \ldots, b_n)$ in $\mathbb{F}^n$ is defined as

$$\langle u, v\rangle = a_1\overline{b}_1 + \cdots + a_n\overline{b}_n.$$

In particular, if $\mathbb{F} = \mathbb{R}$, then $u, v \in \mathbb{R}^n$ and $\overline{b}_i = b_i$ so that

$$\langle u, v\rangle = a_1 b_1 + \cdots + a_n b_n.$$

For instance, if $u = (1, 2, 3) \in \mathbb{R}^3$ and $v = (2, 1, 3) \in \mathbb{R}^3$, then their inner product is $\langle u, v\rangle = 1 \times 2 + 2 \times 1 + 3 \times 3 = 13$.

If $x = (1 + i, 2 - i, 1) \in \mathbb{C}^3$ and $y = (1 - i, 1 + i, 1) \in \mathbb{C}^3$, then

$$\begin{aligned}\langle x, y\rangle &= (1 + i)(1 + i) + (2 - i)(1 - i) + 1 \times 1 \\ &= 1 + i^2 + 2i + 2 + i^2 - 3i + 1 = 2 - i.\end{aligned}$$

Using matrix product

When we consider row or column vectors, their inner product can be given via matrix multiplication.

Let $u, v \in \mathbb{F}^{1\times n}$. Then $\langle u, v\rangle = uv^*$.

Reason: Suppose $u = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$ and $v = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}$. Then

$$uv^* = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \overline{b}_1 \\ \vdots \\ \overline{b}_n \end{bmatrix} = a_1\overline{b}_1 + \cdots + a_n\overline{b}_n = \langle u, v\rangle.$$

In particular, if $u, v \in \mathbb{R}^{1\times n}$ then $\langle u, v\rangle = uv^T$.

Using matrix product Contd.

Similarly, if $u, v \in \mathbb{F}^{n\times 1}$ then $\langle u, v\rangle = v^* u$.

Verification: Suppose $u = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $v = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. Then

$$v^* u = \begin{bmatrix} \overline{b}_1 & \cdots & \overline{b}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \overline{b}_1 a_1 + \cdots + \overline{b}_n a_n = \langle u, v\rangle.$$

In particular, when $u, v \in \mathbb{R}^{n\times 1}$, $\langle u, v\rangle = v^T u$.

Notice that the inner product of two vectors in $\mathbb{F}^n$ is a scalar; it is an element of $\mathbb{F}$.

Properties of Inner product

Verify that the inner product satisfies the following properties:

Let $x, y, z \in \mathbb{F}^n$ and let $\alpha, \beta \in \mathbb{F}$.

1. $\langle x, x \rangle \geq 0$.
2. $\langle x, x \rangle = 0$ iff $x = 0$.
3. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
4. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
5. $\langle z, x + y \rangle = \langle z, x \rangle + \langle z, y \rangle$.
6. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
7. $\langle x, \beta y \rangle = \overline{\beta} \langle x, y \rangle$.

Norm of a vector

If $u \in \mathbb{F}^n$, we define its **norm** as $\|u\| = \sqrt{\langle u, u \rangle}$.

Hence, if $u = (a_1, \ldots, a_n)$, then

$$\|u\| = \sqrt{a_1\overline{a}_1 + \cdots + a_n\overline{a}_n} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}.$$

In particular, when $u \in \mathbb{R}^n$, $\|u\| = \sqrt{a_1^2 + \cdots + a_n^2}$.

For instance,

$$\begin{aligned}
u &= \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \Rightarrow \|u\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}. \\
v &= \begin{bmatrix} 2 & 3 & 4 \end{bmatrix}^T \Rightarrow \|v\| = \sqrt{2^2 + 3^2 + 4^2} = \sqrt{29}. \\
x &= \begin{bmatrix} 1+i & 2-i & 1 \end{bmatrix} \Rightarrow \|x\| = \sqrt{|1+i|^2 + |2-i|^2 + |1|^2} = \sqrt{8}. \\
y &= \begin{bmatrix} 1-i & 1+i & 1 \end{bmatrix}^T \Rightarrow \|x\| = \sqrt{|1-i|^2 + |1+i|^2 + |1|^2} = \sqrt{5}.
\end{aligned}$$

Properties of norm

The norm satisfies the following properties:

Let $x, y \in \mathbb{F}^n$ and let $\alpha \in \mathbb{F}$.

1. $\|x\| \geq 0$.
2. $\|x\| = 0$ iff $x = 0$.
3. $\|\alpha x\| = |\alpha| \, \|x\|$.
4. $|\langle x, y \rangle| \leq \|x\| \, \|y\|$. (*Cauchy-Schwartz inequality*)
5. $\|x + y\| \leq \|x\| + \|y\|$. (*Triangle inequality*)

First three properties follow from those of the inner product.

We will prove the last two properties.

Proof of Cauchy-Schwartz

If $y = 0$, then the inequality clearly holds. Else, $\langle y, y \rangle \neq 0$.

Write $\alpha = \frac{\langle x,y \rangle}{\langle y,y \rangle}$. Then $\overline{\alpha} = \frac{\langle y,x \rangle}{\langle y,y \rangle}$ and $\overline{\alpha}\langle x, y \rangle = |\alpha|^2 \|y\|^2$.

That is, $\overline{\alpha}\langle y, y \rangle - \langle y, x \rangle = 0$, and

$$
\begin{aligned}
0 &\leq \langle x - \alpha y, x - \alpha y \rangle \\
&= \langle x, x \rangle - \overline{\alpha}\langle x, y \rangle + \alpha\big(\overline{\alpha}\langle y, y \rangle - \langle y, x \rangle\big) \\
&= \|x\|^2 - \overline{\alpha}\langle x, y \rangle \\
&= \|x\|^2 - |\alpha|^2 \|y\|^2 \\
&= \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^4} \|y\|^2.
\end{aligned}
$$

Hence, $|\langle x, y \rangle| \leq \|x\| \, \|y\|$. $\square$

Proof of Triangle inequality

Notice that

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \langle y, x\rangle.$$

Using Cauchy-Schwartz inequality, we get

$$\|x+y\|^2 \le \|x\|^2 + \|y\|^2 + 2\|x\|\,\|y\| = \big(\|x\| + \|y\|\big)^2.$$

Hence, $\|x+y\| \le \|x\| + \|y\|$. □

Let $x, y \in \mathbb{F}^n$. We say that the vectors x and y are **orthogonal**, and we write this as $x \perp y$, when $\langle x, y\rangle = 0$.
That is, $x \perp y$ iff $\langle x, y\rangle = 0$. Thus $0 \perp x$ for each vector x.

If $x \perp y$, then the above computation shows that
$\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Pythagoras law

Let $x, y \in \mathbb{F}^n$. we have proved the following:

Pythagoras Law: If $x \perp y$, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

The converse of Pythagoras law holds when $\mathbb{F} = \mathbb{R}$.
Reason: Suppose $\mathbb{F} = \mathbb{R}$. Our earlier computation shows that

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2.$$

It follows that $\langle x, y\rangle = 0$.

The converse does not hold, in general, for $\mathbb{F} = \mathbb{C}$.

For instance, in $\mathbb{C}$, take $x = 1$ and $y = i$. Then,

$$\|x+y\|^2 = \|1+i\|^2 = 2 = |1|^2 + |i|^2 = \|x\|^2 + \|y\|^2.$$

But $\langle x, y\rangle = \langle 1, i\rangle = 1(-i) = -i \neq 0$.

Adjoint & inner product

Theorem: Let $A \in \mathbb{F}^{m\times n}$, $x \in \mathbb{F}^{n\times 1}$, and let $y \in \mathbb{F}^{m\times 1}$. Then

$$\langle Ax, y\rangle = \langle x, A^*y\rangle, \quad \langle A^*y, x\rangle = \langle y, Ax\rangle.$$

Proof: Here, $\langle u, v\rangle = v^*u$, $Ax \in \mathbb{F}^{m\times 1}$ and $A^*y \in \mathbb{F}^{n\times 1}$.

We are using the same notation for both the inner products in $\mathbb{F}^{m\times 1}$ and in $\mathbb{F}^{n\times 1}$. Then

$$\langle Ax, y\rangle = y^*Ax = (A^*y)^*x = \langle x, A^*y\rangle.$$

Similarly,

$$\langle A^*y, x\rangle = x^*A^*y = (Ax)^*y = \langle y, Ax\rangle.$$

□

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be a unitary or an orthogonal matrix.

1. For each pair of vectors x, y, $\langle Ax, Ay\rangle = \langle x, y\rangle$. In particular, $\|Ax\| = \|x\|$ for any x.
2. The columns of A are orthogonal and each is of norm 1.
3. The rows of A are orthogonal, and each is of norm 1.

Proof: (1) $\langle Ax, Ay\rangle = \langle x, A^*Ay\rangle = \langle x, y\rangle$.
Take $x = y$ for the second equality.

(2) Since $A^*A = I$, the ith row of A^* multiplied with the jth column of A gives δ_{ij}. However, this product is simply the inner product of the jth column of A with the ith column of A.

(3) It follows from (2).
Also, considering $AA^* = I$, we get this result. □

Linear Combinations

If $v = (a_1, \ldots, a_n)^T \in \mathbb{F}^{n\times 1}$, then $v = a_1 e_1 + \cdots + a_n e_n$.
We generalize and give a name to such an expression.

Let $v_1, \ldots, v_m, v \in \mathbb{F}^n$. We say that v is a **linear combination** of the vectors $v_1, \ldots, v_m$ iff there exist scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that

$$v = \alpha_1 v_1 + \cdots + \alpha_m v_m.$$

In $\mathbb{F}^{2\times 1}$, one linear combination of $v_1 = (1, 1)^T$ and $v_2 = (1, -1)^T$ is $(3, 1)^T$. Why?

$$2\begin{bmatrix}1\\1\end{bmatrix} + 1\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}3\\1\end{bmatrix}.$$

Is $(4, -2)^T$ a linear combination of v_1 and v_2? Yes, since

$$\begin{bmatrix}4\\-1\end{bmatrix} = 1\begin{bmatrix}1\\1\end{bmatrix} + 3\begin{bmatrix}1\\-1\end{bmatrix}.$$

Linear Independence

In fact, every vector in $\mathbb{F}^{2\times 1}$ is a linear combination of v_1 and v_2.
Reason:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{a+b}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{a-b}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

However, every vector in $\mathbb{F}^{2\times 1}$ is not a linear combination of $(1,1)^T$ and $(2,2)^T$. Reason?
Any linear combination of these two vectors is a multiple of $(1,1)^T$.
Then $(1,0)^T$ is not a linear combination of these two vectors.

The vectors $v_1, \ldots, v_m$ in $\mathbb{F}^n$ are called **linearly dependent** iff at least one of them is a linear combination of others.
The vectors are called **linearly independent** iff none of them is a linear combination of others.

A Characterization

For example, $(1,1),\ (1,-1),\ (4,-1)$ are linearly dependent vectors whereas $(1,1),\ (1,-1)$ are linearly independent vectors.

Theorem: The vectors $v_1, \dots, v_m \in \mathbb{F}^n$ are linearly independent iff for $\alpha_1, \dots, \alpha_m \in \mathbb{F}$,

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0 \ \text{ implies } \ \alpha_1 = \cdots = \alpha_m = 0.$$

Proof: If the vectors $v_1, \dots, v_m$ are linearly dependent then one of them is a linear combination of others. That is, we have an $i \in \{1, \dots, m\}$ such that

$$v_i = \alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \cdots + \alpha_m v_m.$$

$$\alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + (-1) v_i + \alpha_{i+1} v_{i+1} + \cdots + \alpha_m v_m = 0.$$

Proof Contd.

Conversely, suppose we have scalars $\alpha_1, \ldots, \alpha_m$ not all zero such that

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0.$$

Suppose $\alpha_j \neq 0$. Then

$$v_j = -\frac{\alpha_1}{\alpha_j} v_1 - \cdots - \frac{\alpha_{j-1}}{\alpha_j} v_{j-1} - \frac{\alpha_{j+1}}{\alpha_j} v_{j+1} - \cdots - \frac{\alpha_m}{\alpha_j} v_m.$$

That is, $v_1, \ldots, v_n$ are linearly dependent. □

Caution: To show linear independence, you should start with the assumption that $a_1 v_1 + \cdots + a_n v_n = 0$. Then, conclude that $a_1 = 0, \ldots, a_n = 0$.

Example 2

Are the vectors $(1,1,1),\ (2,1,1),\ (3,1,0)$ linearly independent?

Let

$$a(1,1,1)+b(2,1,1)+c(3,1,0)=(0,0,0).$$

Comparing the components, we have

$$a+2b+3c=0,\ a+b+c=0,\ a+b=0.$$

The last two equations imply that $c=0$. Substituting in the first, we see that $a+2b=0$. This and the equation $a+b=0$ give $b=0$. Then it follows that $a=0$.

We conclude that the given vectors are linearly independent.

LI is helpful

Consider the system of linear equations:

$$\begin{array}{rrrl} x_1 & +2\,x_2 & -3\,x_3 & = 2 \\ 2\,x_1 & -\,x_2 & +2\,x_3 & = 3 \\ 4\,x_1 & +3\,x_2 & -4\,x_3 & = 7 \end{array}$$

Here, we find that the third equation is redundant, since 2 times the first plus the second gives the third. Take the coefficients as row vectors:

$$v_1 = (1,\ 2,\ -3,\ 2), \quad v_2 = (2,\ -1,\ 2,\ 3), \quad v_3 = (4,\ 3,\ -4,\ 7).$$

We see that $v_3 = 2v_1 + v_2$, as it should be.
Here, the vectors v_1, v_2, v_3 are linearly dependent.
But the vectors v_1, v_2 are linearly independent.
Thus, solving the given system of linear equations is the same thing as solving the system with only first two equations.

Orthogonality

Let $v_1, \ldots, v_n \in \mathbb{F}^n$. We say that these vectors are **orthogonal** iff $\langle v_i, v_j \rangle = 0$ for all pairs of indices i, j with $i \neq j$.

Theorem: Any orthogonal list of nonzero vectors in $\mathbb{F}^n$ is linearly independent.

Proof: Let $v_1, \ldots, v_n \in \mathbb{F}^n$ be nonzero vectors.

For scalars $a_1, \ldots, a_n$, let $a_1 v_1 + \cdots + a_n v_n = 0$.

Take inner product of both the sides with v_1.

Since $\langle v_i, v_1 \rangle = 0$ for each $i \neq 1$, we obtain $\langle a_1 v_1, v_1 \rangle = 0$.

But $\langle v_1, v_1 \rangle \neq 0$. Therefore, $a_1 = 0$.

Similarly, by taking inner product of $a_1 v_1 + \cdots + a_n v_n$ with v_i, it follows that each $a_i = 0$. $\square$

Span

We denote the set of all linear combinations of vectors $v_1, \ldots, v_n$ by $\mathrm{span}(v_1, \ldots, v_n)$; and read it as the **span of** the vectors $v_1, \ldots, v_n$.

For example, in $\mathbb{R}^2$, $\mathrm{span}(1, 1) = \{\alpha(1, 1) : \alpha \in \mathbb{R}\}$.

This is a straight line passing through the origin and the point $(1, 1)$.

Similarly, in $\mathbb{R}^3$, $\mathrm{span}\big((1, 1, 0), (0, 1, 2)\big)$ is the plane passing through the origin and the points $(1, 1, 0), (0, 1, 2)$.

In $\mathbb{C}^2$, $\mathrm{span}(1, 1) = \{\alpha(1, 1) : \alpha \in \mathbb{C}\}$.

We cannot describe it as we have done in the case of $\mathbb{R}^2$.

Notice that $\mathbb{F}^n = \mathrm{span}(e_1, \ldots, e_n)$.

Advantage of having orthogonal vectors

Suppose $v_1, \dots v_m$ are orthogonal vectors in $\mathbb{F}^n$. $v \in \text{span}(v_1, \dots, v_m)$ implies $v = a_1 v_1 + \cdots + a_m v_m$. Then $\langle v, v_i \rangle = a_i \langle v_i, v_i \rangle$.

Therefore, assuming that $v_i \neq 0$, we get $a_i = \frac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle}$, i.e., each coefficient can be expressed using v and v_i.

To understand the linear combination and hence the span, it is highly advantageous to have orthogonal vectors.

Suppose we are given with m number of vectors from $\mathbb{F}^n$. How do we construct orthogonal vectors $v_1, \dots, v_k$ such that the span is retained?

How to orthogonalize?

We aim at constructing orthogonal vectors $v_1, \ldots, v_k$ from the given vectors $u_1, \ldots, u_m \in \mathbb{F}^n$ so that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(u_1, \ldots, u_m), \quad k \leq m.$$

Consider just two vectors, say u_1, u_2 on the plane.
Assume that they are linearly independent.

Keep $v_1 = u_1$.
Take out the projection of u_2 on u_1 to get v_2. Then $v_2 \perp v_1$.

What is the projection of u_2 on u_1?
Its length is $\langle u_2, u_1 \rangle$. Its direction is that of u_1, i.e., $u_1/\|u_1\|$. Thus

$$v_1 = u_1, \; v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1.$$

We may continue this process of taking out projections if more than two vectors in $\mathbb{F}^n$ are given.

Gram-Schmidt Orthogonalization

Theorem: Let $u_1, u_2, \ldots, u_m \in \mathbb{F}^n$. Define

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \\ &\vdots \\ v_m &= u_m - \frac{\langle u_m, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \cdots - \frac{\langle u_m, v_{m-1} \rangle}{\langle v_{m-1}, v_{m-1} \rangle} v_{m-1}. \end{aligned}$$

In the above process, if $v_i = 0$, then both u_i and v_i are ignored for the rest of the steps. After ignoring such u_i and v_is suppose we obtain the vectors as $v_{j1}, \ldots, v_{jk}$. Then $v_{j1}, \ldots, v_{jk}$ are orthogonal and $\text{span}(v_{j1}, \ldots, v_{jk}) = \text{span}\{u_1, \ldots, u_m\}$.
Further, if $v_i = 0$ for $i > 1$, then $u_i \in \text{span}\{u_1, \ldots, u_{i-1}\}$.

Sketch of Proof

We verify algebraically our geometric intuition:

$$v_1 = u_1, \quad v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1.$$

Hence

$$\langle v_2, v_1 \rangle = \left\langle u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1, v_1 \right\rangle = \langle u_2, v_1 \rangle - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} \langle v_1, v_1 \rangle = 0.$$

If $v_2 = 0$, then u_2 is a scalar multiple of u_1.

If $v_2 \neq 0$, then u_1, u_2 are linearly independent.

By induction, we can prove that for each $i \geq 1$, v_{i+1} is orthogonal to $v_1, \ldots, v_i$.

We need to prove that both the sets spans the same set.

Sketch of Proof Contd.

If $x_1, \ldots, x_r \in \text{span}(y_1, \ldots, y_s)$, then $\text{span}(x_1, \ldots, x_r) \subseteq \text{span}(y_1, \ldots, y_s)$.

For, if $v = \alpha_1 x_1 + \cdots + \alpha_r x_r$ and $x_i = a_{i1} v_1 + \cdots a_{is} v_s$, then substituting for each x_i in the previous expression and combining terms, we get

$$v = \sum_{i=1}^{s} (\alpha_1 a_{1i} + \cdots + \alpha_r a_{ri}) v_i \in \text{span}(v_1, \ldots, v_s).$$

If u_i is a linear combination of $u_1, \ldots, u_{i-1}$, then $\text{span}(u_1, \ldots, u_{i-1}) = \text{span}(u_1, \ldots, u_i)$.

Now observe inductively that $v_1, \ldots, v_i \in \text{span}(u_1, \ldots, u_i)$.

From the algorithm, it can also be observed, using induction, that $u_1, \ldots, u_i \in \text{span}(v_1, \ldots, v_i)$.

Therefore, $\text{span}(u_1, \ldots, u_i) = \text{span}(v_1, \ldots, v_i)$ for each $i \geq 1$. $\square$

Example 3

Orthogonalize $u_1 = (1,0,0),\ u_2 = (1,1,0),\ u_3 = (1,1,1)$.

$v_1 = (1,0,0)$.

$$v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 = (1,1,0) - \frac{(1,1,0)\cdot(1,0,0)}{(1,0,0)\cdot(1,0,0)}(1,0,0) = (0,1,0).$$

$$v_3 = u_3 - \frac{\langle u_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle u_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2$$

$$= (1,1,1) - (1,1,1)\cdot(1,0,0)(1,0,0) - (1,1,1)\cdot(0,1,0)(0,1,0)$$

$$= (1,1,1) - (1,0,0) - (0,1,0) = (0,0,1).$$

The set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is orthogonal; and span of the new vectors is the same as span of the old ones, which is $\mathbb{R}^3$.

Example 4

Use Gram-Schmidt orthogonalization on the vectors $u_1 = (1, 1, 0, 1)$, $u_2 = (0, 1, 1, -1)$ and $u_3 = (1, 3, 2, -1)$.

$v_1 = (1, 1, 0, 1)$.

$$v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = (0, 1, 1, -1).$$

$$v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2$$
$$= (1, 3, 2, -1) - (1, 1, 0, 1) - 2(0, 1, 1, -1) = (0, 0, 0, 0).$$

Now, u_1, u_2 are already orthogonal, Gram-Schmidt process returned $v_2 = u_2$.
Next, $u_3 = u_1 + 2u_2$.

Example 5

Consider orthogonalizing $u_1 = (1, 2, 2, 1)$, $u_2 = (2, 1, 0, -1)$, $u_3 = (4, 5, 4, 1)$ and $u_4 = (5, 4, 2, -1)$.

$v_1 = (1, 2, 2, 1)$.

$v_2 = (\frac{17}{10}, \frac{2}{5}, -\frac{3}{5}, -\frac{13}{10})$.

$v_3 = (0, 0, 0, 0)$.

So, we ignore v_3; and mark that u_3 is a linear combination of u_1 and u_2. Next, we compute

$$v_4 = u_4 - \frac{\langle u_4, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_4, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 = 0.$$

Therefore, we conclude that u_4 is a linear combination of u_1 and u_2. In fact, $u_3 = 2u_1 + u_2$ and $u_4 = u_1 + 2u_2$. Finally, we obtain the orthogonal vectors v_1, v_2 such that $\text{span}(u_1, u_2, u_3, u_4) = \text{span}(v_1, v_2)$.

Trace

The sum of all diagonal entries of a square matrix is called the **trace** of the matrix.

If $A = [a_{ij}] \in \mathbb{F}^{m\times m}$, then $\text{tr}(A) = a_{11} + \cdots + a_{mm} = \sum_{k=1}^{m} a_{kk}$.

$\text{tr}(I_m) = m$ and $\text{tr}(0) = 0$.
The trace satisfies the following properties:

Let $A, B \in \mathbb{F}^{m\times m}$. Let $\beta \in \mathbb{F}$.

1. $\text{tr}(\beta A) = \beta\,\text{tr}(A)$.
2. $\text{tr}(A^T) = \text{tr}(A)$ and $\text{tr}(A^*) = \overline{\text{tr}(A)}$.
3. $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$ and $\text{tr}(AB) = \text{tr}(BA)$.
4. $\text{tr}(A^*A) = 0$ iff $\text{tr}(AA^*) = 0$ iff $A = 0$.

The last one follows from the observation that

$$\text{tr}(A^*A) = \sum_{i=1}^{m}\sum_{j=1}^{m} |a_{ij}|^2 = \text{tr}(AA^*).$$

Find two square matrices A, B such that $\text{tr}(AB) \neq \text{tr}(A)\,\text{tr}(B)$.

Determinant

Besides trace, one more quantity associated with a square matrix is very helpful. It is called the determinant.

Let $A = [a_{ij}] \in \mathbb{F}^{n\times n}$. Its **determinant**, written as $\det(A)$, is defined inductively as follows:

If $n = 1$, then $\det(A) = a_{11}$.

If $n > 1$, then $\det(A) = \sum_{j=1}^{n} (-1)^{1+j}\, a_{1j} \det(A_{1j})$

where the matrix $A_{1j} \in \mathbb{F}^{(n-1)\times(n-1)}$ is obtained from A by deleting the first row and the jth column of A.

We also use two vertical closing bars to denote the determinant.

Example 6:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (-1)^{1+1} a_{11} \det[a_{22}] + (-1)^{1+2} a_{12} \det[a_{21}] = a_{11}a_{22} - a_{12}a_{21}.$$

Example 7

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{vmatrix}$$

$$= (-1)^{1+1} \times 1 \times \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} + (-1)^{1+2} \times 2 \times \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + (-1)^{1+3} \times 3 \times \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1 \times \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} - 2 \times \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + 3 \times \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= (3 \times 2 - 1 \times 1) - 2 \times (2 \times 2 - 1 \times 3) + 3 \times (2 \times 1 - 3 \times 3)$$

$$= 5 - 2 \times 1 + 3 \times (-7) = -18.$$

Fact: The determinant of a triangular matrix is the product of its diagonal entries.

Adjugate

Let $A \in \mathbb{F}^{n \times n}$.

The sub-matrix of A obtained by deleting the ith row and the jth column is written as A_{ij}.

The (i,j)th **co-factor** of A is $(-1)^{i+j}\det(A_{ij})$; it is denoted by $C_{ij}(A)$.

The **adjugate** of A is the $n \times n$ matrix obtained by taking transpose of the matrix whose (i,j)th entry is $C_{ij}(A)$; it is denoted by $\operatorname{adj}(A)$.

That is, $\operatorname{adj}(A) \in \mathbb{F}^{n \times n}$ is the matrix whose (i,j)th entry is the (j,i)th co-factor $C_{ji}(A)$.

Denote by $A_j(x)$ the matrix obtained from A by replacing the jth row of A by a (new) row vector $x \in \mathbb{F}^{n \times 1}$.

We list some important facts about the determinant.

Determinant: facts

Let $A \in \mathbb{F}^{n \times n}$. Let $i, j, k \in \{1, \ldots, n\}$. Then the following statements are true:

1. $\det(A) = \sum_i a_{ij} C_{ij}(A) = \sum_i a_{ij} (-1)^{i+j} \det(A_{ij})$ for any fixed j.
2. For any $j \in \{1, \ldots, n\}$,
   $\det(\,A_j(x+y)\,) = \det(\,A_j(x)\,) + \det(\,A_j(y)\,)$.
3. For any $\alpha \in \mathbb{F}$, $\det(\,A_j(\alpha x)\,) = \alpha \det(\,A_j(x)\,)$.
4. For $A \in \mathbb{F}^{n \times n}$, let $B \in \mathbb{F}^{n \times n}$ be the matrix obtained from A by interchanging the jth and the kth columns, where $j \neq k$. Then $\det(B) = -\det(A)$.
5. If a column of A is replaced by that column plus a scalar multiple of another column, then determinant does not change.
6. Columns of A are linearly dependent iff $\det(A) = 0$.
7. All of (2)-(6) are true for rows instead of columns.

Facts Contd.

8. $\det(A) = \sum_j a_{ij}C_{ij}(A) = \sum_j a_{ij}(-1)^{i+j}\det(A_{ij})$ for any fixed i.
9. If A is a triangular matrix, then $\det(A)$ is equal to the product of the diagonal entries of A.
10. $\det(AB) = \det(A)\det(B)$ for any matrix $B \in \mathbb{F}^{n\times n}$.
11. If A is invertible, then $\det(A) \neq 0$ and $\det(A^{-1}) = (\det(A))^{-1}$.
12. If $B = P^{-1}AP$ for some invertible matrix P, then $\det(A) = \det(B)$.
13. A is invertible iff columns of A are linearly independent iff rows of A are linearly independent iff $\det(A) \neq 0$.
14. $\det(A^T) = \det(A)$.
15. $A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)\,I$.

Example 8 (Using row operations)

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{vmatrix} \overset{R1}{=} \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & -1 & -1 & 2 \end{vmatrix}$$

$R1$: Replace 2nd, 3rd, 4th rows with those plus first.

With $R2$: Replace 3rd and 4th with those plus 2nd,

$$\overset{R2}{=} \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -1 & 4 \end{vmatrix} \overset{R3}{=} \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 8 \end{vmatrix} = 8.$$

$R3$: Replace 4th with that plus 3rd.
Then, its values is 8, since we have got a triangular matrix, whose determinant is the product of diagonal entries.

Theorem on Determinant of special matrices

1. If A is hermitian, then $\det(A) \in \mathbb{R}$.
2. If A is unitary, then $|\det(A)| = 1$.
3. If A is orthogonal, then $\det(A) = \pm 1$.

Proof: (1) Let $A \in \mathbb{C}^{n\times n}$ be hermitian. Then $A = A^*$. It implies

$$\det(A) = \det(A^*) = \det(\overline{A}) = \overline{\det(A)}.$$

Hence, $\det(A) \in \mathbb{R}$.

(2) Let $A \in \mathbb{C}^{n\times n}$ be unitary. Then $A^*A = AA^* = I$. Now,

$$1 = \det(I) = \det(A^*A) = \det(\overline{A})\det(A) = \overline{\det(A)}\det(A) = |\det(A)|^2.$$

Hence $|\det(A)| = 1$.

(3) Let $A \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. That is, $A \in \mathbb{R}^{n\times n}$ and A is unitary. Then $\det(A) \in \mathbb{R}$ and by (2), $|\det(A)| = 1$. That is, $\det(A) = \pm 1$. □

Linear Equations

$$\begin{aligned} x_1 + x_2 &= 3 \\ x_1 - x_2 &= 1 \end{aligned}$$

has a unique solution $x_1 = 2,\ x_2 = 1$. What about

$$\begin{aligned} x_1 + x_2 &= 3 \\ x_1 - x_2 &= 1 \\ 2x_1 - x_2 &= 3 \end{aligned}$$

$x_1 = 2,\ x_2 = 1$ satisfies the third. So the extra equation does not put any constraint on the solutions that we obtained earlier. What about

$$\begin{aligned} x_1 + x_2 &= 3 \\ x_1 - x_2 &= 1 \\ 2x_1 + x_2 &= 3 \end{aligned}$$

$x_1 = 2,\ x_2 = 1$ is the solution of first two. But third is not satisfied by it. So, the system has no solution.

Linear equations cont.

What about

$$x_1 + x_2 = 3$$

The old solution $x_1 = 2, \; x_2 = 1$ is still a solution of this system. But $x_1 = 1, \; x_2 = 2$ is also a solution. It has infinitely many solutions. What about

$$\begin{aligned} x_1 + x_2 &= 3 \\ 2x_1 + 2x_2 &= 6 \\ 3x_1 + 3x_2 &= 9 \end{aligned}$$

It again has infinitely many solutions.
We see that the number of equations really does not matter, but the number of *independent* equations does matter.

We will tackle these things in a more systematic way.

Our tools will be matrices.

Three kinds of row operations

While solving a system of linear equations, we add and subtract equations, multiply an equation with a nonzero constant, and exchange two equations.

These heuristics give rise to the row operations on a matrix.

There are three kinds of **Elementary Row Operations** for a matrix $A \in \mathbb{F}^{m\times n}$:

$ER1$. Exchange of two rows.

$ER2$. Multiplication of a row by a nonzero constant.

$ER3$. Adding to a row a nonzero multiple of another row.

Notation

When a matrix B is obtained from A by using an elementary row operation O, we will write $A \xrightarrow{O} B$.

For elementary row operations, let α be a nonzero scalar.

1. $R_i \leftrightarrow R_j$: The ith row and the jth row are exchanged.
2. $R_i \leftarrow \alpha R_i$: The ith row is multiplied by α.
3. $R_i \leftarrow R_i + \alpha R_j$: To the ith row α times the jth row is added.

A finite sequence of elementary row operations is called a **row operation**.

In general, when a matrix B is obtained from A by using a row operation O, we will write $A \xrightarrow{O} B$ as earlier.

In this case, O will be a finite sequence of elementary row operations instead of just one.

Example 1

See the following computation on the first matrix.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We get the second matrix from the first by adding to its third row (-3) times the first row.

Similarly, the third matrix is obtained from the second by adding to its second row (-2) times the first.

Therefore, we may write $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{O} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$,

where $O = R_3 \leftarrow R_3 - 3R_1,\ R_2 \leftarrow R_2 - 2R_1$.

Elementary matrices

We capture elementary row operations as matrix products with the help of three types of **elementary matrices** :

1. $E[i,j]$ is the matrix obtained from I by exchanging its ith and jth rows.
2. $E_\alpha[i]$ is the matrix obtained from I by multiplying α to its ith row; $\alpha \neq 0$.
3. $E_\alpha[i,j]$ is the matrix obtained from I by adding to its ith row α times the jth row; $\alpha \neq 0$.

For instance,

$$E[1,2] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_{-1}[2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_2[3,1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}.$$

Example 2

$$E[1,2]A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix}$$

$$A \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix}$$

$$E_{-3}[2]A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -6 & -6 & -6 \\ 3 & 3 & 3 \end{bmatrix}$$

$$A \xrightarrow{R_2 \leftarrow -3R_2} \begin{bmatrix} 1 & 1 & 1 \\ -6 & -6 & -6 \\ 3 & 3 & 3 \end{bmatrix}$$

Example 2 Contd.

$$E_{-3}[3,1]A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

What do you observe?

Observation

Let $A \in \mathbb{F}^{m\times n}$. Consider $E[i,j],\ E_\alpha[i],\ E_\alpha[i,j] \in \mathbb{F}^{m\times m}$ for $\alpha \neq 0$. Then the following are true:

1. $A \xrightarrow{R_i \leftrightarrow R_j} E[i,j]A.$

   That is, $E[i,j]A$ is the matrix obtained from A by exchanging its ith and jth rows.

2. $A \xrightarrow{R_i \leftarrow \alpha R_i} E_\alpha[i]A.$

   That is, $E_\alpha[i]A$ is the matrix obtained from A by multiplying α to its ith row.

3. $A \xrightarrow{R_i \leftarrow R_i + \alpha R_j} E_\alpha[i,j]A.$

   That is, $E_\alpha[i,j]A$ is the matrix obtained from A by adding to the ith row α times the jth row.

Example 3

$$\begin{bmatrix} 1 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 1 \\ 3 & 5 & 3 & 4 & 3 \\ -1 & 0 & -1 & -3 & -1 \end{bmatrix} \xrightarrow{O1} \begin{bmatrix} 1 & 1 & 1 & 2 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \xrightarrow{O2} \begin{bmatrix} 1 & 1 & 1 & 2 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here, $O1 = R_2 \leftarrow R_2 - R_1,\ R_3 \leftarrow R_3 - 3R_1,\ R_4 \leftarrow R_4 + R_1$ and
$O2 = R_3 \leftarrow R_3 - 2R_2,\ R_4 \leftarrow R_4 - R_2.$

The corresponding elementary matrices are

In $O1$: $E_{-1}[2,1],\ E_{-3}[3,1],\ E_1[4,1]$ and
In $O2$: $E_{-2}[3,2],\ E_{-1}[4,2]$.

Hence the third matrix is equal to
$E_{-1}[4,2]E_{-2}[3,2]E_1[4,1]E_{-3}[3,1]E_{-1}[2,1]$ times the first matrix.

RREF

We use elementary row operations for bringing a matrix to a nice form.

The first nonzero entry (from left) in a nonzero row of a matrix is called a **pivot**. We denote a pivot in a row by putting a box around it. A column where a pivot occurs is called a **pivotal column**.

A matrix $A \in \mathbb{F}^{m \times n}$ is said to be in **row reduced echelon form** iff the following conditions are satisfied:

(1) Each pivot is equal to 1.

(2) The column index of the pivot in the $(i+j)$th row is greater than the column index of the pivot in the ith row.

(3) In a pivotal column, all entries other than the pivot, are zero.

(4) All zero rows are at the bottom.

Example 4

The matrix $\begin{bmatrix} \boxed{1} & 2 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$ is in row reduced echelon form,

whereas the following are not in row reduced echelon form:

$$\begin{bmatrix} 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & \boxed{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \boxed{1} & 3 & 1 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\begin{bmatrix} 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}.$$

Point out which of the defining properties is/are violated.

We will see how to use the elementary row operations and hence elementary matrices to reduce any matrix to RREF.

Reduction to RREF

1. Set the work region R as the whole matrix A.
2. If all entries in R are 0, then stop.
3. If there are nonzero entries in R, then find the leftmost nonzero column. Mark it as the pivotal column.
4. Find the topmost nonzero entry in the pivotal column. Suppose it is α. Box it; it is a pivot.
5. If the pivot is not on the top row of R, then exchange the row of A which contains the top row of R with the row where the pivot is.
6. If $\alpha \neq 1$, then replace top row of R in A by $1/\alpha$ times that row.
7. Bring all entries, except the pivot, in the pivotal column to 0 by replacing each row above and below the top row of R using elementary row operations in A with that row & the top row of R.
8. Find the sub-matrix to the right and below the pivot. If no such sub-matrix exists, then stop. Else, reset the work region R to this sub-matrix, and go to Step 2.

Example 5

$$A = \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 3 & 5 & 7 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \xrightarrow{O1} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix} \qquad O1 = \begin{matrix} R_2 \leftarrow R_2 - 3R_1, \\ R_3 \leftarrow R_3 - R_1, \\ R_4 \leftarrow R_4 - 2R_1. \end{matrix}$$

$$\xrightarrow{R_2 \leftarrow 1/2 R_2} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & \boxed{1} & \frac{1}{2} & \frac{1}{2} \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix} \xrightarrow{O2} \begin{bmatrix} \boxed{1} & 0 & \frac{3}{2} & -\frac{1}{2} \\ 0 & \boxed{1} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \end{bmatrix} \quad O2 = \begin{matrix} R_1 \leftarrow R_1 - R_2, \\ R_3 \leftarrow R_3 - 4R_2, \\ R_4 \leftarrow R_4 - 6R_2. \end{matrix}$$

$$\xrightarrow{R_3 \leftarrow 1/3 R_3} \begin{bmatrix} \boxed{1} & 0 & \frac{3}{2} & -\frac{1}{2} \\ 0 & \boxed{1} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 6 \end{bmatrix} \xrightarrow{O3} \begin{bmatrix} \boxed{1} & 0 & \frac{3}{2} & 0 \\ 0 & \boxed{1} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad O3 = \begin{matrix} R_1 \leftarrow R_1 + 1/2 R_3, \\ R_2 \leftarrow R_2 - 1/2 R_3, \\ R_4 \leftarrow R_4 - 6R_3. \end{matrix}$$

The final matrix is in RREF, and it is equal to

$$E_{-6}[4,3]\, E_{-1/2}[2,3]\, E_{1/2}[1,3]\, E_{1/3}[3]\, E_{-6}[4,2]\, E_{-4}[3,2]$$
$$E_{-1}[2,1]\, E_{1/2}[2]\, E_{-2}[4,1]\, E_{-1}[3,1]\, E_{-3}[2,1]\, A.$$

Example 6

Consider $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Then

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathrm{RREF}(B) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

$$AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathrm{RREF}(AB) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$$\mathrm{RREF}(A)\,\mathrm{RREF}(B) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \neq \mathrm{RREF}(AB).$$

RREF of a product need not be equal to the product of RREFs.

Elementary matrices Invertible?

Theorem

A square matrix is invertible iff it is a product of elementary matrices.

Proof: $E[i,j]$ is its own inverse, $E_{1/\alpha}[i]$ is the inverse of $E_\alpha[i]$, and $E_{-\alpha}[i,j]$ is the inverse of $E_\alpha[i,j]$. So, any product of elementary matrices is invertible.

Conversely, suppose that A is invertible. Let EA^{-1} be the RREF of A^{-1}. If EA^{-1} has a zero row, then $EA^{-1}A$ also has a zero row. That is, E has a zero row. Wrong, since E is invertible. So, EA^{-1} does not have a zero row. Then each row in the square matrix EA^{-1} has a pivot. But the only square matrix in RREF having a pivot at each row is the identity matrix. Therefore, $EA^{-1} = I$. That is, $A = E$, a product of elementary matrices. □

RREF is unique

Suppose A is an $m \times n$ matrix. Instead of using our algorithm for RREF reduction, suppose you use another algorithm for reducing A to RREF. But restrict yourself to using only elementary row operations in any order you like in this reduction. Then, it is guaranteed that you end up with the same matrix in RREF.

This fact is encapsulated by the following result.

Theorem: Let $A \in \mathbb{F}^{m \times n}$. There exists a unique matrix in $\mathbb{F}^{m \times n}$ in row reduced echelon form obtained from A by elementary row operations.

The proof of this fact uses the result that elementary matrices are invertible and their inverses are also elementary matrices. Further, it uses the fact that each invertible matrix is a product of elementary matrices.

See the classnotes for its proof.

Proof of uniqueness of RREF

Observe that elementary matrices are invertible and their inverses are also elementary matrices.
We see that $B = E_1 A$ and $C = E_2 A$ for some invertible matrices $E_1, E_2 \in \mathbb{F}^{m\times m}$.
Now, $B = E_1 A = E_1 (E_2)^{-1} C$.
Write $E = E_1 (E_2)^{-1}$ to have $B = EC$, where E is invertible.

We consider a particular case first, when $n = 1$. Here, B and C are column vectors in RREF.
Thus, they can be zero vectors or e_1.
Since $B = EC$, where E is invertible, it cannot happen that one is the zero vector and the other is e_1.
Hence, either both are zero vectors or both are e_1. In either case, $B = C$.

Proof Cont.

For $n > 1$, assume, on the contrary, that $B \neq C$.
Then there exists a column index, say $k \geq 1$, such that the first $k-1$ columns of B coincide with the first $k-1$ columns of C, respectively; and the kth column of B is not equal to the kth column of C.
Let u be the kth column of B, and let v be the kth column of C.
We have $u = Ev$ and $u \neq v$.

Suppose the pivotal columns that appear within the first $k-1$ columns in C are $e_1, \ldots, e_j$.
Then $e_1, \ldots, e_j$ are also the pivotal columns in B that appear within the first $k-1$ columns.
Since $B = EC$, we have $C = E^{-1}B$; and consequently,

$$e_1 = Ee_1 = E^{-1}e_1, \ldots, e_j = Ee_j = E^{-1}e_j.$$

The column vector u may be a pivotal column in B or a non-pivotal column in B. Similarly, v may be pivotal or non-pivotal in C.

If both u and v are pivotal columns, then both are equal to e_{j+1}. This contradicts $u \neq v$.

Proof Cont.

So, assume that u is non-pivotal in B or v is non-pivotal in C.

If u is non-pivotal in B, then $u = \alpha_1 e_1 + \cdots + \alpha_j e_j$ for some scalars $\alpha_1, \ldots, \alpha_j$. (See it.) Then

$$v = E^{-1}u = \alpha_1 E^{-1} e_1 + \cdots + \alpha_j E^{-1} e_j = \alpha_1 e_1 + \cdots + \alpha_j e_j = u.$$

This contradicts $u \neq v$.

If v is a non-pivotal column in C, then $v = \beta_1 e_1 + \cdots + \beta_j e_j$ for some scalars $\beta_1, \ldots, \beta_j$. Then

$$u = Ev = \beta_1 E e_1 + \cdots + \beta_j E e_j = \beta_1 e_1 + \cdots + \beta_j e_j = v.$$

Here also, $u = v$, which is a contradiction.

Therefore, $B = C$. $\square$

RREF-Observations

Suppose A has been reduced to its RREF.

Let $R_{i1}, \ldots, R_{ir}$ be the rows of A which have become the nonzero rows in the RREF, and other rows have become the zero rows.

Also, suppose $C_{j1}, \ldots, C_{jr}$ for $j1 < \cdots < jr$, be the columns of A which have become the pivotal columns in the RREF, other columns being non-pivotal.

$$\begin{bmatrix} - & R_{i1} & - \\ & \vdots & \\ - & R_{ir} & - \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} * & | & * & | & \cdots & | & * \\ * & C_{j1} & * & C_{j2} & \vdots & C_{jr} & * \\ * & | & * & | & \cdots & | & * \end{bmatrix}.$$

Observations Contd.

Using the last theorem, we see that the following are true:

1. All rows of A other than $R_{i1}, \ldots, R_{ir}$ are linear combinations of $R_{i1}, \ldots, R_{ir}$.
2. The columns $C_{j1}, \ldots, C_{jr}$ have respectively become $e_1, \ldots, e_r$ in the RREF.
3. All columns of A other than $C_{j1}, \ldots, C_{jr}$ are linear combinations of $C_{j1}, \ldots, C_{jr}$.
4. If $e_1, \ldots, e_k$ are all the pivotal columns in the RREF that occur to the left of a non-pivotal column, then the non-pivotal column is in the form $(a_1, \ldots, a_k, 0, \ldots, 0)^T$. Further, if a column C in A has become this non-pivotal column in the RREF, then $C = a_1 C_{j1} + \cdots + a_k C_{jk}$.
5. If A is a square matrix, then A is invertible iff its RREF is I.

Example 7

Consider the matrix and its RREF in Example 5. There, we had

$$A = \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 3 & 5 & 7 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} \boxed{1} & 0 & 3/2 & 0 \\ 0 & \boxed{1} & 1/2 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Our observation implies that

1. R_4 is a linear combination of $R_1,\ R_2,\ R_3$. Indeed, $R_4 = 3R_1 - R_2 + 2R_3$.
2. The pivotal columns in the RREF are $e_1,\ e_2,\ e_3$.
3. $col(3)$ is a liner combination of the pivotal columns, which are $col(1)$, $col(2)$ and $col(4)$.
4. Specifically, $col(3)$ of $A = 3/2\, col(1) + 1/2\, col(2)$.

Rank of a matrix

The number of pivots in the RREF of a matrix A is called the **rank** of A, and it is denoted by $\text{rank}(A)$.

Since RREF of a matrix is unique, rank is well-defined.

Further, a matrix in RREF is invertible iff it is equal to the identity matrix.

Reason: Suppose B is a matrix in RREF.
If B is invertible, then its RREF does not have a zero row.
So, the RREF is equal to I.
But B is already in RREF. So, $B = I$.
Conversely, if $B = I$, then it is in invertible, and also it is in RREF.

Invertibility & Rank

Theorem A square matrix is invertible iff its rank is equal to its order.

Proof: Let A be a square matrix of order n. Let B be the RREF of A. Then $B = EA$, where E is invertible.

Let A be invertible. Then B is invertible.
Since B is in RREF, $B = I$. So, $\text{rank}(A) = n$.

Conversely, suppose $\text{rank}(A) = n$.
Then B has n number of pivots. Thus $B = I$.
In that case, $A = E^{-1}B = E^{-1}$; and A is invertible. $\square$

If not invertible?

Suppose A is an $n \times n$ matrix.
If $\text{rank}(A) = r < n$, then there are r number of linearly independent columns in A and other columns are linear combinations of these r columns.

The linearly independent columns correspond to the pivotal columns in the RREF of A.

Also, there exist r number of linearly independent rows of A such that other rows are linear combinations of these r rows.

The linearly independent rows correspond to the nonzero rows in the RREF of A.

Example 8

$$\text{Let } A = \begin{bmatrix} 1 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 1 \\ 3 & 5 & 3 & 4 & 3 \\ -1 & 0 & -1 & -3 & -1 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 1 & 3 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From the RREF, we conclude that $\text{rank}(A) = 2$.

$row(1),\ row(2)$ are linearly independent, and

$col(1),\ col(2)$ are linearly independent.

$col(3) = col(5) = col(1), \quad col(4) = 3\, col(1) - col(2).$

Notice also that
$row(3) = row(1) + 2\, row(2), \quad row(4) = row(2) - 2\, row(1).$
However, the RREF does not give this information.

Extracting linearly independent vectors

Given m number of vectors from $\mathbb{F}^n$:

$$u_1 = (u_{11}, \ldots, u_{1n}), \ldots, u_m = (u_{m1}, \ldots, u_{mn}).$$

How to extract linearly independent vectors retaining the span?

Form the matrix A with rows as $u_1, \ldots, u_m$.

Reduce A to its RREF, say, B.

Suppose there are r number of nonzero rows in B.

Then the rows corresponding to those rows in A are linearly independent.

The other rows, which have become the zero rows in B, are linear combinations of those r rows.

Example 9

From among the vectors $(1,2,2,1)$, $(2,1,0,-1)$, $(4,5,4,1)$, $(5,4,2,-1)$, find linearly independent vectors; and point out which are the linear combinations of these independent ones.

$$\begin{bmatrix} \boxed{1} & 2 & 2 & 1 \\ 2 & 1 & 0 & -1 \\ 4 & 5 & 4 & 1 \\ 5 & 4 & 2 & -1 \end{bmatrix} \xrightarrow{O1} \begin{bmatrix} \boxed{1} & 2 & 2 & 1 \\ 0 & \boxed{-3} & -4 & -3 \\ 0 & -3 & -4 & -3 \\ 0 & -6 & -8 & -6 \end{bmatrix}$$

$$\xrightarrow{O2} \begin{bmatrix} \boxed{1} & 0 & -2/3 & -1 \\ 0 & \boxed{1} & 4/3 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$O1 = R_2 \leftarrow R_2 - 2R_1,\ R_3 \leftarrow R_3 - 4R_1,\ R_4 \leftarrow R_4 - 5R_1$ and
$O2 = R_2 \leftarrow -3R_2,\ R_1 \leftarrow R_1 - 2R_2,\ R_3 \leftarrow R_3 + 3R_2,\ R_4 \leftarrow R_4 + 6R_2.$

Example 9 Contd.

Notice that no row exchanges have been applied in this reduction.

The nonzero rows are the first and the second rows.

Therefore, the linearly independent vectors are

$$(1,2,2,1),\ (2,1,0,-1).$$

The third and the fourth are linear combinations of these.

Indeed,

$$(4,5,4,1) = 2(1,2,2,1) + 1(2,1,0,-1).$$

$$(5,4,2,-1) = 1(1,2,2,1) + 2(2,1,0,-1).$$

Also, the span of all the four rows is equal to

$$\operatorname{span}\big((1,2,2,1),\ (2,1,0,-1)\big)$$

which is also equal to $\operatorname{span}\big((1,2,2,1),\ (0,1,4/3,1)\big)$.

More is not good

Why there cannot be more than 3 linearly independent vectors in $\mathbb{R}^3$?

Theorem: Let $u_1, \ldots, u_k,\ v_1, \ldots, v_m \in \mathbb{F}^n$. Suppose each of $v_1, \ldots, v_m$ is a linear combination of $u_1, \ldots, u_k$. If $m > k$, then $v_1, \ldots, v_m$ are linearly dependent.

Proof: Consider all vectors as row vectors. Form the matrix A by taking its rows as $u_1, \ldots, u_k, v_1, \ldots, v_m$ in that order. Now, $r = \text{rank}(A) \leq k$.
Similarly, construct the matrix B by taking its rows as $v_1, \ldots, v_m, u_1, \ldots, u_k$, in that order. Since one is obtained from the other by re-ordering the rows, both A and B have the same RREF.
Therefore, $\text{rank}(B) = \text{rank}(A) = r \leq k$.
If $m > k$, then $m > r = \text{rank}(B)$. Thus, out of $v_1, \ldots, v_m$ at most r vectors can be linearly independent. It follows that $v_1, \ldots, v_m$ are linearly dependent. □

Theorem: Given any n vectors in $\mathbb{F}^m$ there exists a unique $r \leq n$ such that some r of these n vectors are linearly independent, and other $n - r$ vectors are linear combinations of these r vectors.

$AB = I \Rightarrow BA = I$

Let A and B be square matrices satisfying $AB = I$.
Let EA be the RREF of A, where E is a suitable product of elementary matrices.
If A is not invertible, then EA has a zero row.
Then EAB also has a zero row.
However, $EAB = E$ does not have a zero row.
Thus A is invertible; $B = A^{-1}$ is invertible, and $BA = I$. □
This is not true for non-square matrices, in general:

$$\begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 1 & -1 \\ 4 & 2 & -3 \end{bmatrix}.$$

A and PAQ

Let $u_1, \ldots, u_r, u \in \mathbb{F}^{m\times 1}$. Let $a_1, \ldots, a_r \in \mathbb{F}$. Observe that

$$u = a_1 u_1 + \cdots + a_r u_r \quad \text{iff} \quad Pu = a_1 Pu_1 + \cdots a_r Pu_r.$$

Taking $u = 0$, we see that the vectors $u_1, \ldots, u_r \in \mathbb{F}^{m\times 1}$ are linearly independent iff $Pu_1, \ldots, Pu_r$ are linearly independent.

Now, if $A \in \mathbb{F}^{m\times n}$, then its columns are vectors in $\mathbb{F}^{m\times 1}$. The above equation implies that if there exist r number of columns in A which are linearly independent and other columns are linear combinations of these r columns, then the same is true for the matrix PA.
Similarly, let $Q \in \mathbb{F}^{n\times n}$ be invertible. If there exist r number of rows of A which are linearly independent and other rows are linear combinations of these r rows, then the same is true for the matrix AQ.

Rank and Linear independence

These facts, our observation on RREF, and the last theorem can be used to prove the following theorem.

Theorem: Let $A \in \mathbb{F}^{m\times n}$. Then

$$\begin{aligned}
\text{rank}(A) &= \text{the maximum number of linearly independent rows in } A \\
&= \text{the maximum number of linearly independent columns in } A \\
&= \text{rank}(A^t) \\
&= \text{rank}(PAQ), \text{ where } P \in \mathbb{F}^{m\times m} \text{ and } Q \in \mathbb{F}^{n\times n} \\
&\quad\ \text{are any invertible matrices.}
\end{aligned}$$

Using elementary row operations

Let $A \in \mathbb{F}^{n \times n}$ be invertible.

Its inverse can be computed using elementary row operations.

Form the augmented matrix $[A \mid I]$.

Apply elementary row operations on the augmented matrix so that the matrix A there reduces to I.

Then the I portion there has been reduced to A^{-1}.

Why does it work?

Since A is invertible, there exists an invertible matrix P such that PA is I. But P is a product of elementary matrices. And

$$P[A \mid I] = [PA \mid PI] = [I \mid P] = [I \mid A^{-1}].$$

Notice that this reduction is same as reducing A to its RREF.

Of course, if A is invertible, then its RREF is I.

And, if A is not invertible, then during this reduction process, the matrix in A portion of the augmented matrix $[A \mid I]$ will have a zero row.

Example 10

Find B^{-1} if it exists, where

$$B = \begin{bmatrix} 1 & -1 & 2 & 0 \\ -1 & 0 & 0 & 2 \\ 2 & 1 & -1 & -2 \\ 0 & -2 & 0 & 2 \end{bmatrix}.$$

The augmented matrix $[B \mid I]$ with the first pivot looks like:

$$\left[\begin{array}{cccc|cccc} \boxed{1} & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & -1 & -2 & 0 & 0 & 1 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right].$$

Example 10 Cntd.

Use elementary row operations. Since $a_{11} = 1$, we leave $row(1)$ untouched. To zero-out the other entries in the first column, we use the sequence of elementary row operations $R_2 \leftarrow R_2 + R_1$, $R_3 \leftarrow R_3 - 2R_1$. It gives

$$\left[\begin{array}{cccc|cccc} \boxed{1} & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 2 & 1 & 1 & 0 & 0 \\ 0 & 3 & -5 & -2 & -2 & 0 & 1 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right].$$

The pivot is -1 in $(2,2)$ position. Use $R_2 \leftarrow -R_2$ to get the pivot as 1.

$$\left[\begin{array}{cccc|cccc} \boxed{1} & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & \boxed{1} & -2 & -2 & -1 & -1 & 0 & 0 \\ 0 & 3 & -5 & -2 & -2 & 0 & 1 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right].$$

Example 10 Cntd.

And then $R_1 \leftarrow R_1 + R_2,\ R_3 \leftarrow R_3 - 3R_2,\ R_4 \leftarrow R_4 + 2R_2$ gives

$$\left[\begin{array}{cccc|cccc} \boxed{1} & 0 & 0 & -2 & 0 & -1 & 0 & 0 \\ 0 & \boxed{1} & -2 & -2 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 4 & 1 & 3 & 1 & 0 \\ 0 & 0 & -4 & -2 & -2 & -2 & 0 & 1 \end{array}\right].$$

Next pivot is 1 in $(3,3)$ position. Now,
$R_2 \leftarrow R_2 + 2R_3,\ R_4 \leftarrow R_4 + 4R_3$ produces

$$\left[\begin{array}{cccc|cccc} \boxed{1} & 0 & 0 & -2 & 0 & -1 & 0 & 0 \\ 0 & \boxed{1} & 0 & 6 & 1 & 5 & 2 & 0 \\ 0 & 0 & \boxed{1} & 4 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 14 & 2 & 10 & 4 & 1 \end{array}\right].$$

Example 10 Cntd.

Next pivot is 14 in $(4,4)$ position. Use $R_4 \leftarrow {}^1/_{14} R_4$ to make it 1:

$$\left[\begin{array}{cccc|cccc} \boxed{1} & 0 & 0 & -2 & 0 & -1 & 0 & 0 \\ 0 & \boxed{1} & 0 & 6 & 1 & 5 & 2 & 0 \\ 0 & 0 & \boxed{1} & 4 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & \boxed{1} & 1/7 & 5/7 & 2/7 & 1/14 \end{array}\right].$$

Use $R_1 \leftarrow R_1 + 2R_4$, $R_2 \leftarrow R_2 - 6R_4$, $R_3 \leftarrow R_3 - 4R_4$ to zero-out the entries in the pivotal column:

$$\left[\begin{array}{cccc|cccc} \boxed{1} & 0 & 0 & 0 & 2/7 & 3/7 & 4/7 & 1/7 \\ 0 & \boxed{1} & 0 & 0 & 1/7 & 5/7 & 2/7 & -3/7 \\ 0 & 0 & \boxed{1} & 0 & 3/7 & 1/7 & -1/7 & -2/7 \\ 0 & 0 & 0 & \boxed{1} & 1/7 & 5/7 & 2/7 & 1/14 \end{array}\right].$$

Example 10 Cntd.

Therefore,

$$B^{-1} = \frac{1}{7}\begin{bmatrix} 2 & 3 & 4 & 1 \\ 1 & 5 & 2 & -3 \\ 3 & 1 & -1 & -2 \\ 1 & 5 & 2 & \frac{1}{2} \end{bmatrix}.$$

Verify: $B^{-1}B = BB^{-1} = I$.

Example 11

$$\begin{bmatrix} 1 & -1 & 2 & 0 \\ -1 & 0 & 0 & 2 \\ 2 & 1 & -1 & -2 \\ 1 & -2 & 4 & 2 \end{bmatrix}$$

We want to find the inverse of this matrix if at all it is invertible.

Augment it with an identity matrix to get

$$\left[\begin{array}{cccc|cccc} \boxed{1} & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & -1 & -2 & 0 & 0 & 1 & 0 \\ 1 & -2 & 4 & 2 & 0 & 0 & 0 & 1 \end{array}\right].$$

Use elementary row operations. Since $a_{11} = 1$, we leave $row(1)$ untouched. To zero-out the other entries in the first column, we use an appropriate sequence of elementary row operations.

Example 11 Cntd.

Use $R_2 \leftarrow R_2 + R_1$, $R_3 \leftarrow R_3 - 2R_1$, $R_4 \leftarrow R_4 - R_1$ to obtain

$$\left[\begin{array}{cccc|cccc} \boxed{1} & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 2 & 1 & 1 & 0 & 0 \\ 0 & 3 & -5 & -2 & -2 & 0 & 1 & 0 \\ 0 & -1 & 2 & 2 & -1 & 0 & 0 & 1 \end{array}\right].$$

The pivot is -1 in $(2,2)$ position. Use $R_2 \leftarrow -R_2$ to make it 1.

$$\left[\begin{array}{cccc|cccc} 1 & -1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & \boxed{1} & -2 & -2 & -1 & -1 & 0 & 0 \\ 0 & 3 & -5 & -2 & -2 & 0 & 1 & 0 \\ 0 & -1 & 2 & 2 & -1 & 0 & 0 & 1 \end{array}\right].$$

Example 11 Cntd.

Use $R_1 \leftarrow R_1 + R_2$, $R_3 \leftarrow R_3 - 3R_2$, $R_4 \leftarrow R_4 + R_2$ to zero-out all non-pivot entries in the pivotal column to 0:

$$\left[\begin{array}{cccc|cccc} \boxed{1} & 0 & 0 & -2 & 0 & -1 & 0 & 0 \\ 0 & \boxed{1} & -2 & -2 & -1 & -1 & 0 & 0 \\ 0 & 0 & \boxed{1} & 4 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1 & 0 & 1 \end{array}\right].$$

Since a zero row has appeared in the portion where the original matrix was, we conclude that the given matrix is not invertible.

Linear System

A system of linear equations, also called a **linear system** with m equations in n unknowns looks like:

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

As you know, using the abbreviation

$$x = (x_1, \ldots, x_n)^T, \quad b = (b_1, \ldots, b_m)^T, \quad A = [a_{ij}],$$

the system can be written in the compact form:

$$Ax = b.$$

Solvability

If $A \in \mathbb{F}^{m\times n}$, $x \in \mathbb{F}^{n\times 1}$ and $b \in \mathbb{F}^{m\times 1}$, then the system $Ax = b$ has *m number of equations* and *n number of unknowns*.

The system $Ax = b$ is **solvable**, also said *to have a solution*, iff there exists a vector $u \in \mathbb{F}^{n\times 1}$ such that $Au = b$.

Thus, the system $Ax = b$ is solvable iff b is a linear combination of columns of A.
Also, $Ax = b$ has a unique solution iff b is a linear combination of columns of A and the columns of A are linearly independent.

Homogeneous System

The **homogeneous system** corresponding to the system $Ax = b$ is the system

$$Ax = 0.$$

The homogeneous system always has a solution, namely, $x = 0$.

It has infinitely many solutions iff it has a nonzero solution. Reason?

If u is a solution, so is αu for any scalar α.

To study linear systems, we use the augmented matrix $[A \mid b] \in \mathbb{F}^{m\times(n+1)}$ which has its first n columns as those of A in the same order, and the $(n+1)$th column is b.

We mention some results on Linear systems and postpone their proofs to Lecture-12.

Theorem on Linear Systems

Let $A \in \mathbb{F}^{m\times n}$ and let $b \in \mathbb{F}^{m\times 1}$. The following are true:

1. If $[A' \mid b']$ is obtained from $[A \mid b]$ by applying a finite sequence of elementary row operations, then each solution of $Ax = b$ is a solution of $A'x = b'$, and vice versa.

2. **(Consistency)** $Ax = b$ has a solution iff $\text{rank}([A \mid b]) = \text{rank}(A)$.

   In view of this, we say that a linear system $Ax = b$ is **consistent** iff $\text{rank}([A \mid b] = \text{rank}(A)$. Only consistent systems have solutions.

3. If u is a (particular) solution of $Ax = b$, then each solution of $Ax = b$ is given by $u + y$, where y is a solution of the homogeneous system $Ax = 0$.

Theorem Contd.

Let $A \in \mathbb{F}^{m \times n}$ and let $b \in \mathbb{F}^{m \times 1}$. The following are true:

4. If $r = \text{rank}([A \mid b]) = \text{rank}(A) < n$, then there are $n - r$ unknowns which can take arbitrary values; and other r unknowns can be determined from the values of these $n - r$ unknowns.
5. If $m < n$, then the homogeneous system has infinitely many solutions.
6. $Ax = b$ has a unique solution iff $\text{rank}([A \mid b]) = \text{rank}(A) = n$.
7. If $m = n$, then $Ax = b$ has a unique solution iff $\det(A) \neq 0$.
8. **(Cramer's Rule)** If $m = n$ and $\det(A) \neq 0$, then the solution of $Ax = b$ is given by $x_j = \det(\, A_j(b)\,)/\det(A)$ for $1 \leq j \leq n$.

Example 12

Is the following system of linear equations consistent?

$$\begin{aligned} 5x_1 + 2x_2 - 3x_3 + x_4 &= 7 \\ x_1 - 3x_2 + 2x_3 - 2x_4 &= 11 \\ 3x_1 + 8x_2 - 7x_3 + 5x_4 &= 8 \end{aligned}$$

Reduce the augmented matrix to its RREF:

$$\left[\begin{array}{cccc|c} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{array}\right] \xrightarrow{O1}$$

With $O_1 = R_1 \leftarrow {}^1\!/_5 R_1,\ R_2 \leftarrow R_2 - R_1,\ R_3 \leftarrow R_3 - 3R_1$, we get

Example 12 Cntd.

$$\xrightarrow{O1} \left[\begin{array}{cccc|c} \boxed{1} & 2/5 & -3/5 & 1/5 & 7/5 \\ 0 & -17/5 & 13/5 & -11/5 & 48/5 \\ 0 & 34/5 & -26/5 & 22/5 & -19/5 \end{array}\right]$$

$$\xrightarrow{O2} \left[\begin{array}{cccc|c} \boxed{1} & 0 & -5/17 & -1/17 & 43/17 \\ 0 & \boxed{1} & -13/17 & 11/17 & -48/17 \\ 0 & 0 & 0 & 0 & \boxed{77/5} \end{array}\right]$$

Here, $O2 = R_2 \leftarrow {}^{-5}/_{17}R_2,\ R_1 \leftarrow R_1 - {}^{2}/_{5}R_2,\ R_3 \leftarrow R_3 - {}^{34}/_{5}R_2.$

So, $\text{rank}\big([A \mid b]\big) > \text{rank}(A)$.

The system is inconsistent. It does not have a solution.

Gauss-Jordan Elimination

Gauss-Jordan elimination converts the augmented matrix to its RREF for solving linear systems.

Example 13: We change the last equation in Example 12 as follows:

$$\begin{aligned} 5x_1 + 2x_2 - 3x_3 + x_4 &= 7 \\ x_1 - 3x_2 + 2x_3 - 2x_4 &= 11 \\ 3x_1 + 8x_2 - 7x_3 + 5x_4 &= -15 \end{aligned}$$

We start with the augmented matrix $[A \mid b]$ and reduce it to its RREF.

Example 13 Cntd.

Reduction to RREF gives

$$\left[\begin{array}{cccc|c} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & -15 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{cccc|c} \boxed{1} & 0 & -5/17 & -1/17 & 43/17 \\ 0 & \boxed{1} & -13/17 & 11/17 & -48/17 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This expresses the fact that the third equation is redundant.

Now, solving the new system in RREF is easier.

Example 13 Cntd.

Writing as equations, we have

$$\begin{array}{lllll} \boxed{1}\,x_1 & & -\frac{5}{17}x_3 & -\frac{1}{17}x_4 & = \frac{43}{17} \\ & \boxed{1}\,x_2 & -\frac{13}{17}x_3 & +\frac{11}{17}x_4 & = -\frac{48}{17} \end{array}$$

The unknowns corresponding to the pivots are called the **basic variables** and the other unknowns are called the **free variable**.

We assign free variables to arbitrary numbers, say x_i to α_i; and express the basic variables in terms of the free variables to get a solution of the equations.

Here, the basic variables are x_1 and x_2; and the unknowns x_3, x_4 are free variables. We assign x_3 to α_3 and x_4 to α_4. The solution is written as follows:

$$x_1 = \frac{43}{17} + \frac{5}{17}\alpha_3 + \frac{1}{17}\alpha_4, \;\; x_2 = -\frac{48}{17} + \frac{13}{17}\alpha_3 - \frac{11}{17}\alpha_4, \;\; x_3 = \alpha_3, \;\; x_4 = \alpha_4.$$

Result 1

In all the following results, we consider a linear system $Ax = b$, where A as an $m \times n$ matrix and b is a column vector of size m.

1. If $[A' \mid b']$ is obtained from $[A \mid b]$ by applying a finite sequence of elementary row operations, then each solution of $Ax = b$ is a solution of $A'x = b'$, and vice versa.

Proof: If $[A' \mid b']$ has been obtained from $[A \mid b]$ by a finite sequence of elementary row operations, then $A' = EA$ and $b' = Eb$, where E is the product of corresponding elementary matrices. The matrix E is invertible. Now, $A'x = b'$ iff $EAx = Eb$ iff $Ax = E^{-1}Eb = b$. $\square$

Result 2

2. $Ax = b$ has a solution iff $\text{rank}([A \mid b]) = \text{rank}(A)$.

Proof: Due to (1), we assume that $[A \mid b]$ is in RREF. Suppose $Ax = b$ has a solution. If there is a zero row in A, then the corresponding entry in b is also 0. Therefore, there is no pivot in b. Hence $\text{rank}([A \mid b]) = \text{rank}(A)$.

Conversely, suppose that $\text{rank}([A \mid b]) = \text{rank}(A) = r$. Then there is no pivot in b. That is, b is a non-pivotal column in $[A \mid b]$. Thus, b is a linear combination of pivotal columns, which are some columns of A. Therefore, $Ax = b$ has a solution. $\square$

Results 3 & 4

3. If u is a (particular) solution of $Ax = b$, then each solution of $Ax = b$ is given by $u + y$, where y is a solution of the homogeneous system $Ax = 0$.

Proof: Let u be a solution of $Ax = b$. Then $Au = b$. Now, z is a solution of $Ax = b$ iff $Az = b$ iff $Az = Au$ iff $A(z - u) = 0$ iff $z - u$ is a solution of $Ax = 0$. That is, each solution z of $Ax = b$ is expressed in the form $z = u + y$ for a solution y of the homogeneous system $Ax = 0$. □

4. If $r = \text{rank}([A \mid b]) = \text{rank}(A) < n$, then there are $n - r$ unknowns which can take arbitrary values; and other r unknowns can be determined from the values of these $n - r$ unknowns.

Proof of Result 4

Proof: Let $\text{rank}([A \mid b]) = \text{rank}(A) = r < n$. By (2), there exists a solution. Due to (3), we consider solving the corresponding homogeneous system. Due to (1), assume that A is in RREF.
There are r number of pivots in A and $m - r$ number of zero rows.
Omit all the zero rows; it does not affect the solutions.
Write the system as linear equations. Rewrite the equations by keeping the unknowns corresponding to pivots on the left hand side, and taking every other term to the right hand side.
The unknowns corresponding to pivots are now expressed in terms of the other $n - r$ unknowns.
For obtaining a solution, we may arbitrarily assign any values to these $n - r$ unknowns, and the unknowns corresponding to the pivots get evaluated by the equations. □

Result 5

5. If $m < n$, then the homogeneous system has infinitely many solutions.

Proof: Let $m < n$. Then $r = \text{rank}(A) \leq m < n$. Consider the homogeneous system $Ax = 0$. By (4), there are $n - r \geq 1$ number of unknowns which can take arbitrary values, and other r unknowns are determined accordingly. Each such assignment of values to the $n - r$ unknowns gives rise to a distinct solution resulting in infinite number of solutions of $Ax = 0$. □

Result 6 & 7

6. $Ax = b$ has a unique solution iff $\text{rank}([A \mid b]) = \text{rank}(A) = n$.

Proof: Recall that (4) says: If $r = \text{rank}([A \mid b]) = \text{rank}(A) < n$, then there are $n - r$ unknowns which can take arbitrary values; and other r unknowns can be determined from the values of these $n - r$ unknowns. This would give rise to non-unique solutions. On the other hand, if $\text{rank}([A \mid b]) = \text{rank}(A) = n$, then A is invertible, and we have the unique solution as $x = A^{-1}b$. $\square$

7. If $m = n$, then $Ax = b$ has a unique solution iff $\det(A) \neq 0$. In this case, the unique solution is given by $x = A^{-1}b$.

Proof: If $A \in \mathbb{F}^{n \times n}$, then it is invertible iff $\text{rank}(A) = n$ iff $\det(A) \neq 0$. Then use (6). $\square$

Result 8

8. If $m = n$ and $\det(A) \neq 0$, then the solution of $Ax = b$ is given by $x_j = \det(\,A_j(b)\,)/\det(A)$ for each $j \in \{1, \ldots, n\}$.

Proof: Recall that $A_j(b)$ is the matrix obtained from A by replacing the jth column of A with the vector b. Since $\det(A) \neq 0$, by (6), $Ax = b$ has a unique solution, say $y \in \mathbb{F}^{n\times 1}$.

Write the identity $Ay = b$ in the form:

$$y_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} + \cdots + y_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix} + \cdots + y_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

This gives

$$y_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} + \cdots + \begin{bmatrix} (y_j a_{1j} - b_1) \\ \vdots \\ (y_j a_{nj} - b_n) \end{bmatrix} + \cdots + y_n \begin{bmatrix} a_{1n} \\ \cdots \\ a_{nn} \end{bmatrix} = 0.$$

In this sum, the jth vector is a linear combination of other vectors, where $-y_j$s are the coefficients.

Proof of (8) Cont.

Therefore,

$$\begin{vmatrix} a_{11} & \cdots & (y_j a_{1j} - b_1) & \cdots & a_{1n} \\ & & \vdots & & \\ a_{n1} & \cdots & (y_j a_{nj} - b_n) & \cdots & a_{nn} \end{vmatrix} = 0.$$

From Property (6) of the determinant, it follows that

$$y_j \begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ & & \vdots & & \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix} - \begin{vmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n} \\ & & \vdots & & \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{vmatrix} = 0.$$

Therefore, $y_j = \det(\,A_j(b)\,)/\det(A)$. □

A fixed line

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Here, $A : \mathbb{R}^{2\times 1} \to \mathbb{R}^{2\times 1}$.
It transforms straight lines to straight lines or points.

Get me a straight line which is transformed to itself by A.

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

Thus, the line $\{(x, x) : x \in \mathbb{R}\}$ never moves.

Also the line $\{(x, -x) : x \in \mathbb{R}\}$. does not move.

Observe:

$A \begin{bmatrix} x \\ x \end{bmatrix} = 1 \begin{bmatrix} x \\ x \end{bmatrix}$ and $A \begin{bmatrix} x \\ -x \end{bmatrix} = (-1) \begin{bmatrix} x \\ -x \end{bmatrix}$.

Eigenvalues & Eigenvectors

Let $A \in \mathbb{C}^{n\times n}$. A scalar $\lambda \in \mathbb{C}$ is called an **eigenvalue** of A iff there exists a non-zero vector $v \in \mathbb{C}^{n\times 1}$ such that $Av = \lambda v$.

Such a vector v is called an **eigenvector of A for** (or, associated with, or, corresponding to) the eigenvalue λ.

Example 1: Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

It has an eigenvector $(1, 0, 0)^T$ associated with the eigenvalue 1.

Is $(2, 0, 0)^T$ an eigenvector associated with the same eigenvalue 1?

Corresponding to an eigenvalue, there are infinitely many eigenvectors.

Characteristic Polynomial

Theorem: Let $A \in \mathbb{C}^{n\times n}$. A complex number λ is an eigenvalue of A iff $\det(A - \lambda I) = 0$.

Proof: Let $v \in \mathbb{C}^{n\times 1},\ v \neq 0$. Then,
v is an eigenvector of A for the eigenvalue $\lambda \in \mathbb{C}$
iff v is a nontrivial solution of the homogeneous system $(A - \lambda I)x = 0$
iff $\text{rank}(A - \lambda I) < n$ iff $\det(A - \lambda I) = 0$. □

The polynomial $\det(A - tI)$ is called the **characteristic polynomial** of the matrix A.

Fundamental Theorem of Algebra says that each polynomial of degree n with complex coefficients has exactly n complex zeros.

The zeros of the characteristic polynomial are the eigenvalues of A.

Note: If $\alpha + i\beta$ is an eigenvalue of a matrix with real entries, where $\beta \neq 0$, then $\alpha - i\beta$ is also an eigenvalue of this matrix.

Example 2

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$.

The characteristic polynomial of A is

$$\det(A - tI) = \begin{vmatrix} 1-t & 0 & 0 \\ 1 & 1-t & 0 \\ 1 & 1 & 1-t \end{vmatrix} = (1-t)^3.$$

The zeros are $1, 1, 1$. These are the eigenvalues of A.

To get an eigenvector, we solve $A(a, b, c)^T = 1\,(a, b, c)^T$ or that

$$a = a,\; a + b = b,\; a + b + c = c.$$

It gives $a = b = 0$ and $c \in \mathbb{F}$ can be arbitrary.
All eigenvectors are given by $(0, 0, c)^T$, for $c \neq 0$.

Example 3

Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Its characteristic polynomial is $t^2 + 1 = 0$.

Then i and $-i$ are its eigenvalues.

The corresponding eigenvectors are obtained by solving

$$A(a,b)^T = i(a,b)^T \text{ and } A(a,b) = -i(a,b)^T.$$

For $\lambda = i$, we have $b = ia$, $-a = ib$. Thus, $(a, ia)^T$ is an eigenvector for $a \neq 0$.

For the eigenvalue $-i$, the eigenvectors are $(a, -ia)$ for $a \neq 0$.

The maximum k such that $(t - \lambda)^k$ divides the characteristic polynomial is called the **algebraic multiplicity** of the eigenvalue λ.

In Example 2, the algebraic multiplicity of the eigenvalue 1 is 3.

Some results

1. A and A^T have the same eigenvalues.

Reason: $\det(A^T - tI) = \det((A - tI)^T) = \det(A - tI)$.

Matrices $A, B \in \mathbb{C}^{n \times n}$ are called **similar** iff there exists an invertible matrix $P \in \mathbb{C}^{n \times n}$ such that $P^{-1}AP = B$.

2. Similar matrices have the same eigenvalues.

Reason: $\det(P^{-1}AP - tI) = \det(P^{-1}(A - tI)P)$
$= \det(P^{-1})\det(A - tI)\det(P) = \det(A - tI)$.

3. If A is a diagonal or an upper triangular or a lower triangular matrix, then its diagonal elements are precisely its eigenvalues.

Reason: In all these cases, $\det(A - tI) = (a_{11} - t) \cdots (a_{nn} - t)$.

Trace & Determinant

4. Let $A \in \mathbb{C}^{n\times n}$. Then $\det(A)$ equals the product and $\mathrm{tr}(A)$ equals the sum of all eigenvalues of A.

Proof: Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, not necessarily distinct. Now,

$$\det(A - tI) = (\lambda_1 - t)\cdots(\lambda_n - t).$$

Put $t = 0$. It gives $\det(A) = \lambda_1 \cdots \lambda_n$.

Expand $\det(A - tI)$ and equate the coefficients of t^{n-1} to get

Coeff of t^{n-1} in $\det(A - tI)$ = Coeff of t^{n-1} in $(a_{11} - t) \cdot A_{11}$

(Here, A_{11} is the determinant of the matrix obtained from $A - tI$ deleting its first row and first column.)

$= \cdots =$

$=$ Coeff of t^{n-1} in $(a_{11} - t) \cdot (a_{22} - t) \cdots (a_{nn} - t) = (-1)^{n-1}\mathrm{tr}(A)$.

But Coeff of t^{n-1} in $\det(A - tI) = (-1)^{n-1} \cdot \sum \lambda_i$. □

Cayley-Hamilton Theorem

Theorem: Any square matrix satisfies its characteristic polynomial.

Proof: Let $A \in \mathbb{C}^{n\times n}$. Let $p(t) = c_0 + c_1 t + \cdots + c_n t^n$ be the characteristic polynomial of A. We show that $p(A) = 0$, the zero matrix. Now, $p(t)\,I = \det(A - tI)\,I = \left[\text{adj}\,(A - tI)\right](A - tI)$. The entries in $\text{adj}\,(A - tI)$ are polynomials in t of degree at most $n - 1$. Write $\text{adj}\,(A - tI) := B_0 + tB_1 + \cdots + t^{n-1}B_{n-1}$, where $B_0, \ldots, B_{n-1} \in \mathbb{C}^{n\times n}$. Then

$$c_0 I + c_1 I t + \cdots + c_n I t^n = p(t)I = (B_0 + tB_1 + \cdots t^{n-1}B_{n-1})(A - t\,I).$$

Comparing the coefficients of t^k, we obtain

$$c_0 I = B_0 A,\ c_1 I = B_1 A - B_0, \ldots, c_{n-1} I = B_{n-1}A - B_{n-2},\ c_n I = -B_{n-1}.$$

Then, substituting these values in $p(A)$, we have

$$\begin{aligned} p(A) &= c_0 I + c_1 A + \cdots + c_n A^n = c_0 I + c_1 IA + \cdots + c_n IA^n \\ &= B_0 A + (B_1 A - B_0)A + \cdots + (B_{n-1}A - B_{n-2})A^{n-1} - B_{n-1}A^n \\ &= 0. \end{aligned}$$

□

Two applications

Suppose that a matrix $A \in \mathbb{C}^{n\times n}$ has the characteristic polynomial

$$a_0 + a_1 t + \cdots + a_{n-1}t^{n-1} + (-1)^n t^n.$$

By Cayley-Hamilton theorem, $a_0 I + a_1 A + \cdots + (-1)^n A^n = 0$. Then

$$A^n = (-1)^{n-1}(a_0 I + a_1 A + \cdots + a_{n-1}A^{n-1}).$$

1. So, $A^n, A^{n+1}, \ldots$ can be reduced to computing $A,\ A^2,\ \ldots,\ A^{n-1}$.

2. If A is invertible, then $\det(A) \neq 0$. So, 0 is not an eigenvalue of A. So, $a_0 \neq 0$. Then

$$a_0 I + A(a_1 I + \cdots + a_{n-1}A^{n-2} + (-1)^n A^{n-1}) = 0.$$

Multiplying A^{-1} and simplifying, we obtain

$$A^{-1} = -\frac{1}{a_0}\left(a_1 I + a_2 A + \cdots + a_{n-1}A^{n-2} + (-1)^n A^{n-1}\right).$$

Hermitian matrices

Let $A \in \mathbb{C}^{n \times n}$.

Let $\lambda \in \mathbb{C}$ be any eigenvalue of A with an eigenvector $v \in \mathbb{C}^{n \times 1}$.

Now, $Av = \lambda v$.

Pre-multiplying with v^*, we have $v^* A v = \lambda v^* v \in \mathbb{C}$.

Taking adjoint, $v^* A^* v = \overline{\lambda} v^* v$.

1. All eigenvalues of a Hermitian matrix are real.

Reason: If A is Hermitian, then $A^* = A$.

So, $\lambda v^* v = \overline{\lambda} v^* v$, where $v \neq 0$.

Thus $\lambda = \overline{\lambda}$. That is, $\lambda \in \mathbb{R}$.

Real symmetric

A real symmetric matrix is a hermitian matrix.

So, all its eigenvalues are real.

In addition, we have the following result.

2. If A is a real symmetric $n \times n$ matrix, then a real vector corresponding to a (real) eigenvalue can always be chosen.

Reason: Suppose $A(x+iy) = \lambda(x+iy)$, with $\lambda \in \mathbb{R},\ x, y \in \mathbb{R}^{n\times 1}$.

Then $Ax = \lambda x$ and $Ay = \lambda y$.

Since $x + iy \neq 0$, at least one of x, y is nonzero.

If $x \neq 0$, then it is an eigenvector for the eigenvalue λ; else, choose y.

Skew-hermitian

3. All eigenvalues of a skew-hermitian or a real skew-symmetric matrix are zero or purely imaginary.

Reason: When A is skew-hermitian, $A^* = -A$.

Earlier we had $v^*Av = \lambda v^*v$ and $v^*A^*v = \overline{\lambda}v^*v$.

So, $\overline{\lambda}v^*v = -\lambda v^*v$.

Again since $v \neq 0$, we have $\overline{\lambda} = -\lambda$.

That is, $2\text{Re}(\lambda) = 0$.

Unitary & Orthogonal

4. Each eigenvalue of a unitary or an orthogonal matrix has absolute value 1.

Reason: Let A be unitary. That is, $A^*A = I = AA^*$.

Now, $Av = \lambda v$, $v \neq 0$ implies $v^*A^* = (\lambda v)^* = \overline{\lambda} v^*$. Then

$$v^*v = v^*Iv = v^*A^*Av = \overline{\lambda}\lambda v^*v = |\lambda|^2 v^*v.$$

Since $v^*v \neq 0$, $|\lambda| = 1$.

The determinant of A is the product of its eigenvalues.

So, the determinant of a unitary matrix has absolute value 1.

Since an orthogonal matrix is a real unitary matrix, its determinant is a real number, and its absolute value is 1.

Hence, the determinant of an orthogonal matrix is either 1 or -1.

Distinct eigenvalues

Theorem: Eigenvectors associated with distinct eigenvalues of an $n \times n$ matrix are linearly independent.

Proof: Let $\lambda_1, \dots, \lambda_m$ be all the distinct eigenvalues of $A \in \mathbb{C}^{n \times n}$. Let $v_1, \dots, v_m$ be corresponding eigenvectors. We use induction on $i \in \{1, \dots, m\}$.
For $i = 1$, since $v_1 \neq 0$, $\{v_1\}$ is linearly independent,
Induction Hypothesis: for $i = k$ suppose $\{v_1, \dots, v_k\}$ is linearly independent. We want to show that $v_1, \dots, v_k, v_{k+1}$ are linearly independent. Towards this, assume that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k + \alpha_{k+1} v_{k+1} = 0. \tag{1}$$

Proof Cont.

Then, $A(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k + \alpha_{k+1} v_{k+1}) = 0$. Since $Av_j = \lambda_j v_j$, we have

$$\alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_k \lambda_k v_k + \alpha_{k+1} \lambda_{k+1} v_{k+1} = 0. \qquad (2)$$

Multiply (1) with λ_{k+1}. Subtract from (2) to get:

$$\alpha_1(\lambda_1 - \lambda_{k+1})v_1 + \cdots + \alpha_k(\lambda_k - \lambda_{k+1})v_k = 0.$$

By the Induction Hypothesis, $\alpha_j(\lambda_j - \lambda_{k+1}) = 0$ for each $j = 1, \ldots, k$. Since $\lambda_1, \ldots, \lambda_{k+1}$ are distinct, we conclude that $\alpha_1 = \cdots = \alpha_k = 0$. Then, from (1), it follows that $\alpha_{k+1} v_{k+1} = 0$. As $v_{k+1} \neq 0$, we have $\alpha_{k+1} = 0$. $\square$

Using linearly independent eigenvectors

Suppose an $n \times n$ matrix A has n linearly independent eigenvectors $v_1, \dots, v_n$. Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. We find that

$$Av_1 = \lambda_1 v_1, \quad \dots, Av_n = \lambda_n v_n.$$

Construct the matrix $P \in \mathbb{C}^{n \times n}$ by taking its columns as the eigenvectors $v_1, \dots, v_n$. That is, let

$$P = \begin{bmatrix} v_1 & v_2 & \cdots & v_{n-1} & v_n \end{bmatrix}.$$

Also, construct the diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. That is,

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Using LI eigenvectors Contd.

With $P = \begin{bmatrix} v_1 \; v_2 & \cdots & v_{n-1} \; v_n \end{bmatrix}$ and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, we can write the products $Av_1 = \lambda_1 v_1, \; \ldots, Av_n = \lambda_n v_n$ as the single equation

$$AP = PD.$$

Now, $\text{rank}(P) = n$. So, P is an invertible matrix. Then

$$P^{-1}AP = D.$$

Let $A \in \mathbb{C}^{n \times n}$. We call A to be **diagonalizable** iff there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.
(That is, A is similar to a diagonal matrix).
We also say that A is **diagonalizable by the matrix** P iff $P^{-1}AP = D$.

It follows that if an $n \times n$ matrix has n linearly independent eigenvectors, then it is diagonalizable.

Diagonalizability

We ask whether diagonalizability of an $n \times n$ matrix implies that it has n number of linearly independent eigenvectors.

Let $A \in \mathbb{F}^{n \times n}$ be diagonalizable.
So, let $P = \begin{bmatrix} v_1, & \cdots & v_n \end{bmatrix}$ be an invertible matrix so that
$P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.
Then $AP = A\begin{bmatrix} v_1, & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1, & \cdots & v_n \end{bmatrix}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.
Consequently, $Av_1 = \lambda_1 v_1, \ \ldots, \ Av_n = \lambda_n v_n$.
That is, $v_1, \ldots, v_n$ are eigenvectors of A.
Moreover, P is invertible implies that $v_1, \ldots, v_n$ are linearly independent.

We have proved the following result.

Theorem: An $n \times n$ matrix is diagonalizable iff it has n linearly independent eigenvectors.

Diagonalizability

We know that if an $n \times n$ matrix has n distinct eigenvalues, then the corresponding eigenvectors are linearly independent.

Also, if an $n \times n$ matrix has n linearly independent eigenvectors, then it is diagonalizable.

It follows that if an $n \times n$ matrix has n distinct eigenvalues, then it is diagonalizable.

Another sufficient condition for diagonalizability is given by the following theorem.

Spectral Theorem:

1. A is a normal matrix iff A is diagonalizable by a unitary matrix.
2. If A is real symmetric, then A is diagonalizable by an orthogonal matrix.

Comments

Recall that A is a normal matrix iff $A^*A = AA^*$. Thus, hermitian matrices and real symmetric matrices are normal matrices.

Spectral theorem says that if A is a normal matrix, then there exists a unitary matrix P such that $P^{-1}AP$ is a diagonalizable matrix.

Since P is unitary, we have $P^* = P^{-1}$. Hence, P^*AP is a diagonal matrix.

In this case, $P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, $P^*AP = \text{diag}(\lambda_1, \ldots, \lambda_n)$, and v_i is an eigenvector associated with the eigenvalue λ_i of A.

The Spectral theorem says that the vectors v_i can be chosen in such a way that they form an orthonormal set.

Similarly, when A is real-symmetric, these orthonormal vectors can be chosen to be in $\mathbb{R}^{n\times 1}$.

Diagonalization

In each of these cases, our procedure of diagonalization is the same.

We find eigenvalues.

We choose the corresponding linearly independent eigenvectors.

Next, we form the matrix P by taking these eigenvectors as columns.

Then $P^{-1}AP$ is a diagonal matrix with diagonal entries as the eigenvalues of A.

If A is normal or real-symmetric, then we may orthogonalize the eigenvectors; and divide each with its norm; and then form P.

Then $P^{*}AP$ will be the diagonal matrix with the eigenvalues of A on its diagonal.

Example 4

Let $A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$.

It is real symmetric. It has eigenvalues $-1,\ 2,\ 2$.

We must find out the associated eigenvectors, by solving $Ax = \lambda x$.

For the eigenvalue -1, the system $Ax = -x$ gives

$$x_1 - x_2 - x_3 = -x_1,\ -x_1 + x_2 - x_3 = -x_2,\ -x_1 - x_2 + x_3 = -x_3.$$

Then $x_1 = x_2 = x_3$. One eigenvector is $(1, 1, 1)^T$.

For the eigenvalue 2, we have the equations as

$$x_1 - x_2 - x_3 = 2x_1,\ -x_1 + x_2 - x_3 = 2x_2,\ -x_1 - x_2 + x_3 = 2x_3.$$

It gives $x_1 + x_2 + x_3 = 0$.

Example 4 Contd.

Since $x_1 + x_2 + x_3 = 0$, we can have two linearly independent eigenvectors such as

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}.$$

Along with the earlier eigenvector $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$, we see that the three eigenvectors are orthogonal; we divide their norms to get the orthonormal eigenvectors as

$$\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}.$$

Example 4 Contd.

Next, we take

$$P = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}.$$

We see that $P^{-1} = P^T$, that is, P is an orthogonal matrix. And,

$$P^{-1}AP = P^TAP = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

the diagonal matrix similar to A.