PART I - Fundamentals

Lecture 1 - *Matrix-Vector Multiplication*

OBJECTIVE:

You should already know how to do matrix multiplication: $\mathbf{b} = \mathbf{A}\mathbf{x}$.

Now we show how to interpret $\mathbf{b}$ *as a linear combination of the columns of* $\mathbf{A}$.

$\diamond$ FAMILIAR DEFINITIONS

Let $\mathbf{x}$ be an n-dimensional column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

Let $\mathbf{A}$ be an $m \times n$ matrix (m rows, n columns)

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

If $\mathbf{b} = \mathbf{A}\mathbf{x}$, then $\mathbf{b} \in \mathbb{R}^m$ where each component of $\mathbf{b}$,

$$b_i = \sum_{j=1}^{n} a_{ij} x_j \qquad i = 1, 2, \dots, m.$$

e.g., If $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 1 \times 8 + 2 \times 9 \\ 3 \times 8 + 4 \times 9 \\ 5 \times 8 + 6 \times 9 \end{pmatrix} = \begin{pmatrix} 26 \\ 60 \\ 94 \end{pmatrix}$$

(verify!)

Note 1. *The text assumes numbers are complex ($\mathbb{C}$). We will only use real numbers ($\mathbb{R}$). But, everything we say about real quantities can be applied to complex quantities.*

We can view $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x}$ as a *linear map*.

i.e., for any (vectors) $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any (scalar) $\alpha \in \mathbb{R}$,

$$\begin{aligned} \mathbf{A}(\mathbf{x} + \mathbf{y}) &= \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y} \\ \mathbf{A}(\alpha\mathbf{x}) &= \alpha\mathbf{A}\mathbf{x} \end{aligned}$$

Exercise: Which side is more expensive to compute?

Conversely, every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be expressed as a multiplication by an $m \times n$ matrix.

$\diamond$ MATRIX-VECTOR MULTIPLICATION

Let $\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{array} \right]$ i.e., $\mathbf{a}_j \in \mathbb{R}^m$ is the j^{th} column of $\mathbf{A}$.

Then, $\mathbf{b} = \mathbf{A}\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j$ i.e., $\mathbf{b}$ *is a linear combination of the columns of* $\mathbf{A}$.

$$
\begin{aligned}
\left[\begin{array}{c} \mathbf{b} \end{array} \right] &= \left[\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{array} \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \left[\begin{array}{c} \mathbf{a}_1 \end{array} \right] + x_2 \left[\begin{array}{c} \mathbf{a}_2 \end{array} \right] + \ldots + x_n \left[\begin{array}{c} \mathbf{a}_n \end{array} \right]
\end{aligned}
$$

Note 2. *This is nothing but a change of viewpoint (and notation).*
Instead of viewing $\mathbf{A}\mathbf{x} = \mathbf{b}$ *as "*$\mathbf{A}$ *acting on* $\mathbf{x}$ *to give* $\mathbf{b}$*", we view as "*$\mathbf{x}$ *acting on* $\mathbf{A}$ *to produce* $\mathbf{b}$*".*

$\diamond$ MATRIX-MATRIX MULTIPLICATION

If $\mathbf{B} = \mathbf{AC}$, then *each column of* $\mathbf{B}$ *is a linear combination of the columns of* $\mathbf{A}$.

Let $\mathbf{A} \in \mathbb{R}^{l \times m}$ and $\mathbf{C} \in \mathbb{R}^{m \times n}$

Then, $\mathbf{B} \in \mathbb{R}^{l \times n}$ with entries $b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj}$ $i = 1, 2, \ldots, l$, $j = 1, 2, ..., m$.

$$\begin{aligned} &\left[\ \mathbf{b}_1 \,\middle|\, \mathbf{b}_2 \,\middle|\, \ldots \,\middle|\, \mathbf{b}_n\ \right] \\ &\qquad = \left[\ \mathbf{a}_1 \,\middle|\, \mathbf{a}_2 \,\middle|\, \ldots \,\middle|\, \mathbf{a}_m\ \right] \left[\ \mathbf{c}_1 \,\middle|\, \mathbf{c}_2 \,\middle|\, \ldots \,\middle|\, \mathbf{c}_n\ \right] \end{aligned}$$

$$\Longrightarrow \qquad \mathbf{b}_j = \mathbf{A}\mathbf{c}_j = \sum_{k=1}^{m} c_{kj} \mathbf{a}_k$$

i.e., $\mathbf{b}_j$ is a linear combination of the columns $\mathbf{a}_j$ with the coefficients c_{kj} (each element of column j in $\mathbf{C}$).

$\diamond$ RANGE AND NULLSPACE

Definition 1. $\text{range}(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{x} = \mathbf{Ay}$ *for some* $\mathbf{y}\}$
i.e., the set of all vectors that can be expressed as $\mathbf{Ay}$ *for some vector* $\mathbf{y}$.

Theorem 1. $\text{range}(\mathbf{A})$ *is the space spanned by the columns of* $\mathbf{A}$.

Note 3. *The range of* $\mathbf{A}$ *is also called the* column space *of* $\mathbf{A}$.

Definition 2. $\text{null}(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{Ax} = \mathbf{0}\}$
i.e., the set of vectors $\mathbf{x}$ *that map to the zero vector via* $\mathbf{A}$.

$\rightarrow$ each vector $\mathbf{x} \in \text{null}(\mathbf{A})$ gives the expansion coefficients of the zero vector as a linear combination of columns of $\mathbf{A}$:

$$\mathbf{0} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n$$

$\diamond$ RANK

column rank of a matrix = dimension of space spanned by its columns

row rank of a matrix = dimension of space spanned by its rows

ROW RANK ALWAYS EQUALS COLUMN RANK!

So, we often just refer to as "rank".

$\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to have *full rank* if it has *maximal rank*.
i.e., rank($\mathbf{A}$) $= \min(m, n)$

> e.g., if $m \geq n$, a matrix with full rank must have n linearly independent columns.

Theorem 2. *A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ has full rank if and only if it maps no two distinct vectors to the same vector.*

$\rightarrow$ $\mathbf{A}$ is a one-to-one mapping.

$\diamond$ MATRIX INVERSE

A *nonsingular* or invertible matrix is a square matrix with full rank.

i.e., the m columns of a nonsingular $m \times m$ matrix $\mathbf{A}$ span (form a basis) for the *whole space* $\mathbb{R}^m$
$\leftrightarrow$ any vector in $\mathbb{R}^m$ can be expressed as a linear combination of the columns of $\mathbf{A}$.

In particular, expand the canonical basis vector

$$\mathbf{e}_j = \begin{array}{cc} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{c} \\ \\ \\ \\ \leftarrow j \\ \\ \\ \\ \end{array} \end{array}$$

$$\mathbf{e}_j = \sum_{i=1}^{m} z_{ij}\mathbf{a}_i = \mathbf{A}\mathbf{z}_j$$

If we now place all the $\mathbf{e}_j$ for $j = 1, 2, \ldots, m$ in a matrix, we obtain

$$\left[\begin{array}{c|c|c|c} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_m \end{array}\right] = \mathbf{I} = \mathbf{A}\mathbf{Z}$$

$\mathbf{I}$ is called the $m \times m$ *identity matrix*; $\mathbf{Z}$ is called the *inverse of* $\mathbf{A}$ ($\mathbf{A}^{-1}$).

Every nonsingular (square) matrix $\mathbf{A}$ has a unique inverse $\mathbf{A}^{-1}$ that satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Here are some useful equivalences concerning nonsingular matrices:

a) $\mathbf{A}$ has a unique inverse $\mathbf{A^{-1}}$

b) rank$(\mathbf{A}) = m$

c) range$(\mathbf{A}) = \mathbb{R}^m$

d) null$(\mathbf{A}) = \mathbf{0}$

e) 0 is not an eigenvalue of $\mathbf{A}$

f) 0 is not a singular value of $\mathbf{A}$

g) det$(\mathbf{A}) \neq 0$

Note 4. *Although it is a convenient theoretical notion, the determinant is rarely used in practice.*

Do not take the formula $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ literally!

WE BASICALLY NEVER FIND $\mathbf{A}^{-1}$ IN PRACTICE
$\rightarrow$ CERTAINLY NOT FOR SOLVING $\mathbf{Ax} = \mathbf{b}$!

So, instead of thinking of hitting $\mathbf{b}$ with $\mathbf{A}^{-1}$ to get $\mathbf{x}$, think of $\mathbf{x}$ as the vector of coefficients required to uniquely expand $\mathbf{b}$ in the basis of columns of $\mathbf{A}$.

Multiplying by $\mathbf{A}^{-1}$ is a *change of basis* operation:

$\mathbf{b}$: coefficients of expansion using canonical basis $\{e_1, e_2, ..., e_m\}$	$\xrightarrow{\mathbf{A}^{-1}}$ $\xleftarrow[\mathbf{A}]{}$	$\mathbf{A}^{-1}\mathbf{b}$: coefficients of expansion using basis of columns of $\mathbf{A}$