MTH 264 Introduction to Matrix Algebra - Summer 2023.

LN1B. Matrices and Matrix Operations.

These lecture notes are mostly lifted from the text **Matrix and Power Series, Lee and Scarborough, custom 5th edition**. This document highlights which parts of the text are used in the lecture sessions.

Part 1. Matrices in $\mathbb{R}^{m\times n}$

Definition 1B.1. Matrices

A **matrix** is a doubly-indexed tuple of real numbers. We say that a matrix $\mathbf{A}$ is an $(m \times n)$-matrix (read as an m by n matrix) if the first index goes through $\{1, 2, ..., m\}$ and the second index goes through $\{1, 2, ..., n\}$. Conventionally, we write a matrix $\mathbf{A}$ as the following:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{pmatrix} \text{ with } \mathbf{a_i} = \begin{pmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m,i} \end{pmatrix} \text{ for } i \in \{1, 2, ..., n\}$$

When $\mathbf{A}$ is listed element-wise, the convention is to have the first index represent the **row index** and the second index represent the **column index**. i.e. $\mathbf{A}$ being an $(m \times n)$-matrix states that $\mathbf{A}$ has m rows and n columns.

I do have to remark that the notation for the matrix in terms of its columns may be non-standard. However, we add this notation since it makes it easier (at least for me) to remember how to do matrix operations as you'll see later.

Theorem 1B.2. $\mathbb{R}^{m\times n}$ is a Vector Space

Let $\mathbb{R}^{m\times n}$ denote the set of all $m \times n$ matrices. Then, $\mathbb{R}^{m\times n}$ is a vector space with $\mathbb{R}$ as its set of scalars with **matrix addition** defined as

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{pmatrix}$$

$$= \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \cdots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \cdots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & \cdots & a_{m,n}+b_{m,n} \end{pmatrix} \text{ for all } \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$$

and **scalar multiplication** defined as:

$$k\mathbf{A} = k \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} = \begin{pmatrix} ka_{1,1} & ka_{1,2} & \cdots & ka_{1,n} \\ ka_{2,1} & ka_{2,2} & \cdots & ka_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m,1} & ka_{m,2} & \cdots & ka_{m,n} \end{pmatrix} \text{ for all } k \in \mathbb{R}$$

Observe that our definitions of matrix addition and scalar multiplication matches that of vectors. This is intentional since we want to combine vectors and matrices later on. Additionally, the theorem above lets us use the vector space properties, e.g. we can conclude that matrix addition is commutative and therefore $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. As said previously, it should be relatively obvious that these properties comes from the properties of $\mathbb{R}$ and that knowing the exact list of vector space properties/axioms is not critical.

Theorem 1B.3. Zero Matrix

The **zero matrix**, denoted as $\mathbf{0} = \mathbf{0_{m\times n}} \in \mathbb{R}^{m\times n}$, the matrix with all entries equaling 0. When the dimension of the matrices involved are clear/arbitrary, we usually suppress the subscript. The zero matrix is the matrix such that for all matrices $\mathbf{A}$, $\mathbf{A} + \mathbf{0} = \mathbf{A} = \mathbf{0} + \mathbf{A}$.

Much like the zero vector, this serves as the additive identity for matrix addition.

Definition 1B.4. Column and Row Vectors

A **column vector** with n entries is an $(n \times 1)$-matrix and a **row vector** with m entries is an $(1 \times m)$-matrix.

Convention 1B.5. Notation on Vectors and Matrices

We say that $\mathbf{A} \in \mathbb{R}^{m\times n}$ to refer to $\mathbf{A}$ being an $(m \times n)$-matrix with real number coefficients.

When the matrix has a single uppercase letter as its name, we typically denote its entries by indexing the corresponding lowercase letter. However, when the matrix has a more complicated label, we use the notation $(\mathbf{A})_{i,j}$ to denote the entry of $\mathbf{A}$ at row i and column j.

For this course, when we call vectors in relation to matrices, we assume that vectors $\mathbf{v} \in \mathbb{R}^m$ are presented as **column vectors** unless otherwise specified; and we may write $\mathbf{A} = \begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{pmatrix} \in \mathbb{R}^{m\times n}$ to refer to the columns of A as column vectors $\mathbf{a_i} \in \mathbb{R}^m$ for $i = 1, 2, ..., n$.

The vector space structure is limited to the family of vectors $\mathbb{R}^{m\times n}$ for fixed m and n. That is, if the dimensions of the matrices do not match, matrix addition is not defined.

Looking more generally, we can define multiplication on the family of all matrices. We first start with the matrix product of one matrix against one vector.

Definition 1B.6. Matrix-Vector Multiplication

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and let $\mathbf{v} = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$. Describe $\mathbf{A}$ using **column vectors** $\mathbf{a_1}, ..., \mathbf{a_n} \in \mathbb{R}^n$. We define the matrix-vector product $\mathbf{Av} \in \mathbb{R}^m$ by

$$\mathbf{Av} = \begin{pmatrix} | & | & & | \\ \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \\ | & | & & | \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1\mathbf{a_1} + v_2\mathbf{a_2} + ...v_n\mathbf{a_n}$$

This definition also gives us the definition of row vector by a column vector multiplication.

Observe that if the number of columns in the right matrix does not match the number of rows in the left column vector, the matrix-vector product is **not** defined.

As a note, it is possible to express the matrix-vector product using row vectors. However, we introduce matrix multiplication this way since having column vectors allow us to express linear transformations in terms of left multiplication by a matrix (as we'll see later). If we were to use row vectors, we need to do right multiplication by a matrix.

However, most of the theory (at least in the text) of linear transformations are expressed typically by left multiplication matrices. Having row vectors requires us to translate the results and theorems in terms of right multiplication matrices – which, while possible, involves a lot of work.

We extend the previous definition to get the matrix product.

Definition 1B.7. Matrix Multiplication

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and let $\mathbf{B} \in \mathbb{R}^{n \times r}$ with **column vectors** $\mathbf{b_1}, ..., \mathbf{b_r}$. The product $\mathbf{AB} \in \mathbb{R}^{m \times r}$ is defined as

$$\mathbf{AB} = \mathbf{A} \begin{pmatrix} | & | & & | \\ \mathbf{b_1} & \mathbf{b_2} & \cdots & \mathbf{b_r} \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \mathbf{Ab_1} & \mathbf{Ab_2} & \cdots & \mathbf{Ab_r} \\ | & | & & | \end{pmatrix}$$

Observe that the product $\mathbf{AB}$ is only defined when the number of rows for $\mathbf{A}$ match the number of columns for $\mathbf{B}$. One way to think about this is to look at their dimensions side by side:

$$\overbrace{(m \times n)}^{\text{(dimension of } \mathbf{A}\text{)}} \underbrace{(n \times r)}_{\text{(dimension of } \mathbf{B}\text{)}}$$

The matrix product $\mathbf{AB}$ is defined if and only if the inner dimensions, as written above, are the same.

We also add another way to view matrix multiplication that might be easier to use when calculating matrix products by hand.

Theorem 1B.8. Equivalent Definition of Matrix-Vector Multiplication

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and let $\mathbf{v} \in \mathbb{R}^n$ be a column vector. Describe $\mathbf{A}$ using **row vectors** $\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_m} \in \mathbb{R}^n$. Then,

$$\mathbf{Av} = \begin{pmatrix} — & \mathbf{a_1} & — \\ — & \mathbf{a_2} & — \\ & \vdots & \\ — & \mathbf{a_m} & — \end{pmatrix} \mathbf{v} = \begin{pmatrix} \mathbf{a_1 v} \\ \mathbf{a_2 v} \\ \vdots \\ \mathbf{a_m v} \end{pmatrix}$$

Observe that $\mathbf{a_i v}$ is the multiplication of a row vector by a column vector.

Theorem 1B.9. Equivalent Definition of Matrix Multiplication

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with **row vectors** $\mathbf{a_1}, ..., \mathbf{a_m}$ and let $\mathbf{B} \in \mathbb{R}^{n \times r}$ with **column vectors** $\mathbf{b_1}, ..., \mathbf{b_r}$. Observe that $\mathbf{a_i}$ and $\mathbf{b_j}$ all have n entries each. Define the matrix product $\mathbf{AB} \in \mathbb{R}^{m \times r}$ by the following:

$$\mathbf{AB} = \begin{pmatrix} — & \mathbf{a_1} & — \\ — & \mathbf{a_2} & — \\ & \vdots & \\ — & \mathbf{a_m} & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ \mathbf{b_1} & \mathbf{b_2} & \cdots & \mathbf{b_r} \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \mathbf{a_1 b_1} & \mathbf{a_1 b_2} & \cdots & \mathbf{a_1 b_r} \\ \mathbf{a_2 b_1} & \mathbf{a_2 b_2} & \cdots & \mathbf{a_2 b_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_m b_1} & \mathbf{a_m b_2} & \cdots & \mathbf{a_m b_r} \end{pmatrix}$$

Observe that in general, **matrix multiplication is not commutative**. Therefore, when we multiply matrices, we need to specify whether we mean **left multiplication** or **right multiplication**. For example, multiplying $\mathbf{A}$ by $\mathbf{B}$ on the left refers to the matrix product $\mathbf{BA}$.

Example 1B.9.1. Let $\mathbf{A} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and let $\mathbf{B} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then, $\mathbf{AB} = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} = \mathbf{BA}$.

Example 1B.9.2. Let $\mathbf{R} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and let $\mathbf{Q} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

Then, $\mathbf{RQ} = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$ but $\mathbf{QR} = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$, i.e. $\mathbf{RQ} \neq \mathbf{QR}$.

Theorem 1B.10. Properties of Matrix Multiplication

Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be matrices. The following identities apply when the relevant products are defined.

- **(a) Associativity.** $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. With this, we can state $\mathbf{ABC}$ without ambiguity.
- **(b) Left Distributivity over Addition.** $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.
- **(c) Right Distributivity over Addition.** $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$.
- **(d) Compatibility with Scalar Multiplication.** For all $k \in \mathbb{R}$, $k(\mathbf{AB}) = (k\mathbf{A})\mathbf{B} = \mathbf{A}(k\mathbf{B})$. With this, we can write $k\mathbf{AB}$ without ambiguity.

Lastly, we then introduce some more notation to make our discussions easier.

Definition 1B.11. Square Matrices

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then, $\mathbf{A}$ is a **square matrix** if and only if $m = n$. Equivalently, $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a square matrix.

Part 2. Matrix Inverses.

Recall that for real numbers: multiplication can be undone by a division operation. For example, multiplication by 2 can be undone by multiplying by $\frac{1}{2}$ (equivalently, dividing by two). Here, we say that $\frac{1}{2}$ is the **multiplicative inverse** of 2 since $(2)(\frac{1}{2}) = 1$ and 1 is the **multiplicative identity** of the real numbers.

We then extend this concept to the family of matrices since we do have a notion of multiplication. However, first, we must introduce the multiplicative identity for the family of matrices.

Definition 1B.12. Identity Matrix

The **identity matrix** $\mathbf{I_n} \in \mathbb{R}^{n\times n}$ is the square matrix with 1's on the main diagonal and 0's everywhere else. That is,

$$\mathbf{I_n} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Theorem 1B.13. Multiplicative Identity of the Family of Matrices

For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{I_m A} = \mathbf{A}$ and $\mathbf{A I_n} = \mathbf{A}$. Therefore, $\mathbf{I_n}$ is the **multiplicative identity** of matrix multiplication. In other words, any matrix multiplied by the appropriate identity matrix (either on the left or on the right) yields the same matrix. When the relevant dimension is clear from context, we may suppress the subscript and write $\mathbf{I}$.

Finally, we can introduce the notion of multiplicative inverses.

Definition 1B.14. Matrix Inverses

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. If there exists a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I_n}$, then $\mathbf{A}$ is called **invertible** and $\mathbf{B}$ is called an **inverse** of $\mathbf{A}$.

There are two important notes about this definition:

1. There is **no** analog for matrix inverses for **non-square** matrices. If $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \neq n$, the identity for left multiplication is $\mathbf{I_m}$ but the identity for right multiplication is $\mathbf{I_n}$. We would need another definition (if we can make a reasonable definition, that is) for non-square matrices. That is not covered in this course.
2. **Not all matrices are invertible** – which is why we've given matrices that have an inverse a name. We'll see more theorems later that would tell us if an matrix is invertible or not.

Observe, from the definition, that there are two conditions for invertibility: for a square matrix $\mathbf{A}$, $\mathbf{AA}^{-1} = \mathbf{I}$ **and** $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. So, to call a matrix $\mathbf{B}$ the inverse of $\mathbf{A}$, we need to check $\mathbf{AB} = \mathbf{I}$ **and** $\mathbf{BA} = \mathbf{I}$. To make our calculations easier, we can use the following theorem.

Theorem 1B.15. Uniqueness of Inverse Matrices

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an invertible matrix.

- **(a)** The matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I_n}$ is uniquely determined by $\mathbf{A}$. Therefore, we can call $\mathbf{B}$ **the** inverse of $\mathbf{A}$ and denote the inverse as $\mathbf{A}^{-1} := \mathbf{B}$.
- **(b)** If we've found a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{AB} = \mathbf{I_n}$, then $\mathbf{A}^{-1} = \mathbf{B}$.
- **(c)** If we've found a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{BA} = \mathbf{I_n}$, then $\mathbf{A}^{-1} = \mathbf{B}$.

Using results **(b)** and **(c)** reduces the number of equations we need to check and is incredibly useful. For example, we will later apply Gaussian elimination as a method of finding inverses of matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$. That results in a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{BA} = \mathbf{I_n}$. Using the theorem above, we can conclude that $\mathbf{A}^{-1} = \mathbf{B}$.

Lastly, we introduce properties of the matrix inverse as an operator on the space of invertible matrices.

Theorem 1B.16. Properties of Matrix Inverses

- **(a)** **(Inverse of Identity Matrices).** For all identity matrices $\mathbf{I_n}$, $\mathbf{I_n}$ is invertible and $\mathbf{I_n}^{-1} = \mathbf{I_n}$.
- **(b)** **(Involution).** For all invertible matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$, $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
- **(c)** **(Inverse Compatibility with Scalar Multiplication).** For all invertible matrices $\mathbf{A} \in \mathbb{R}n \times n$ and nonzero scalars $k \in \mathbb{R}$, $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1}$.
- **(d)** **(Co-Distributivity over Matrix Multiplication).** For all invertible matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, the product $\mathbf{AB}$ is invertible and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Do note that some of these labels (i.e. inverse compatibility and co-distributivity) are somewhat non-standard terms. We add them here for the following reasons: (1) The names make sense in some contexts. Compatibility with scalar multiplication generally means you can "factor" out the scalar outside the operator, similarly with the derivative and anti-derivative operator on function spaces. We add inverse as a prefix since pulling out the scalar involves taking its multiplicative inverse. With distributivity, the co- prefix implies that we have to reverse the order of the matrices upon distribution; (2) It makes it easier to refer to a property when explaining things.

Finally, finding matrix inverses and even checking whether a matrix is invertible or not is a non-trivial problem in high dimensions. We will cover more results about invertibility later as we introduce more concepts and the naive approach of finding inverses can be computationally expensive. Fortunately, we do have an easy formula and check in the case of $\mathbb{R}^{2\times 2}$.

Theorem 1B.17. Matrix Inverses in $\mathbb{R}^{2\times 2}$

Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$. Then, $\mathbf{A}$ is invertible if and only if $ad - bc \neq 0$. In the case where $\mathbf{A}$ is invertible, its inverse $\mathbf{A}^{-1}$ is given by

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

Later, we'll discuss how to use Gaussian Elimination to find inverses.

Part 3. Matrix Transpose.

Lastly, we'll introduce the **transpose** of a matrix.

Definition 1B.18. Transpose of a Matrix

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The **transpose** of $\mathbf{A}$, denoted as $\mathbf{A}^T$, is the matrix $\mathbf{A}^T \in \mathbb{R}^{n\times m}$ where $\left(\mathbf{A}^T\right)_{j,i} = \mathbf{A}_{i,j}$.

Another way to think about the transpose is in terms of the main diagonal, defined below.

Definition 1B.19. Main Diagonal of a Matrix

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The **main diagonal** of $\mathbf{A}$ are the entries $(\mathbf{A})_{i,i}$ with index by $i = 1, 2, ..., \min\{m, n\}$. Elements not on the main diagonal are called **off-diagonal elements**.

Do note that when we use the main diagonal as a term, we usually imply that we're working with square matrices (i.e. $\mathbf{A} \in \mathbb{R}^{n\times n}$). However, this definition extends nicely to non-square matrices to help us understand transposes. Here, you can think of the transpose as **flipping** the matrix about its main diagonal where columns become rows and rows become columns.

The transpose, as an operator on the space of matrices, also has some useful properties.

Theorem 1B.20. Properties of the Matrix Transpose

- **(a) (Involution).** For all matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$, $(\mathbf{A}^T)^T = \mathbf{A}$.
- **(b) (Distributivity over Matrix Addition).** For all matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
- **(c) (Co-Distributivity over Matrix Multiplication).** For all matrices $\mathbf{A}, \mathbf{B}$ such that the product $\mathbf{AB}$ is defined, $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

(d) **(Compatibility with Scalar Multiplication)** For all matrices $\mathbf{A}$ and scalars k, $(k\mathbf{A})^T = k\mathbf{A}^T$.

(e) **(Commutativity over Matrix Inverses).** For all invertible matrices $\mathbf{A}$, $\mathbf{A}^T$ is invertible with inverse $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

(f) **(Products of Transposes)** For all matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$, the matrix products $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m \times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n \times n}$ are both defined.

These properties may be useful for us later when dealing with calculations that involve transposes.