Lecture 2: Set Notation and Linear Algebra Review

The notes below will briefly refresh (or introduce) some notation, terminology, and concepts reated to set notation and linear algebra that we'll need later on in this course.

Set Notation [Bertsekas, 2008]

A *set* is a collection of objects, called *elements* of the set. If S is a set and x is an element of S we write $x \in S$ where the $\in$ is a symbol for "is an element of". If x is not an element of S we write $x \notin S$ where $\notin$ is a symbol for "is not an element of". A set can have no elements in which case it is called the emptyset, denoted by $\emptyset$. Sets can be specified in a variety of ways. If S contains a finite number of elements, say $x_1, x_2, \ldots, x_n$ we write it as a list of the elements in curly brackets:

$$S = \{x_1, x_2, \ldots, x_n\} \tag{1}$$

Alternatively, we can consider a (possibly infinite) set of all x that have a certain property P and denote it by

$$S = \{x \mid x \text{ satisfies } P\} \tag{2}$$

where the symbol " $\mid$ " (or sometimes " $:$ ") is read as "such that". Occasionally the condition will include the phrase "for all" which is mathematically written using the symbol "$\forall$".

Number Sets

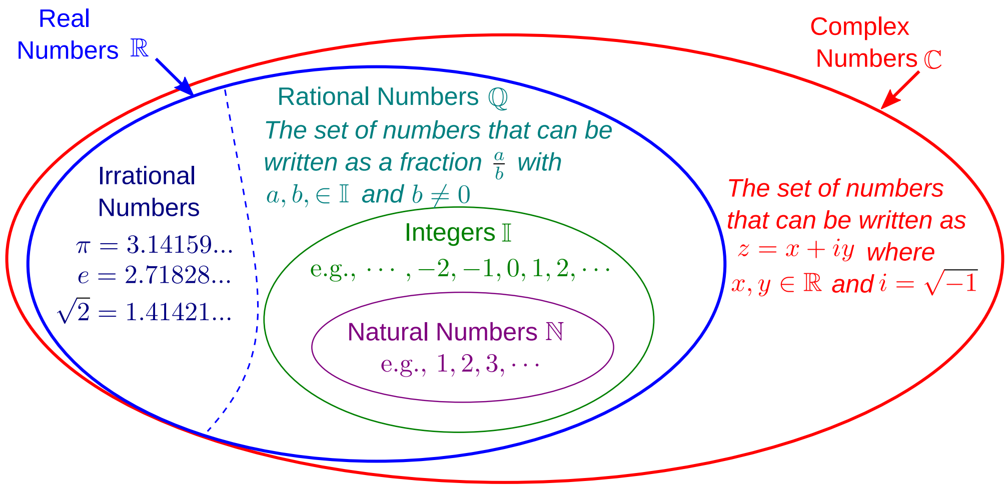
Field (Definition). A field is a set of elements F together with two operations called *addition* (denoted $+$) and *multiplication* (denoted $\cdot$) that map two elements, e.g., a and b in F, to another element c also in F (i.e., the field is *closed* under addition and multiplication since it produces a third element that is also in the field). These binary operations (of addition and multiplication) satisfy the *field axioms*:

- Associativity: $a + (b + c) = (a + b) + c$ and $a \cdot (b + c)$
- Commutativity: $a + b = b + a$ and $a \cdot b = b \cdot a$
- Identity element: for addition the identity element is zero, $a + 0 = a$ and for multiplication the identity element is one, $a \cdot 1 = a$.
- Inverse element: for addition the inverse element of a is $-a$ so that $a + (-a) = 0$ and for multiplication the inverse element of a (assuming $a \neq 0$) is $1/a$ so that $a \cdot (1/a) = 1$.
- Distributivity of multiplication over addition: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

We will not use the axioms much in this course, but we include them here to clarify the concept of a field which is needed to define a vector space (to be discussed shortly).

Common Number Sets. There are several commonly used sets of numbers that are organized according to what values are contained therein, as illustrate by the Venn diagram below. For example, the natural numbers $\mathbb{N}$ are the well known counting numbers (1,2,3, etc.) and are considered to be a subset of the integers $\mathbb{Z}$ which includes zero and negative values (e.g., -3, -2,

-1 0, 1, 2, etc.). The set of real numbers $\mathbb{R}$ consists of all rational numbers and irrational number. Moreover, real numbers are contained with the set of complex numbers $\mathbb{C}$ that also include the imaginary number. The imaginary number is usually denoted with the symbol i or j and is equal to $i = \sqrt{-1}$. From this definition it follows that $i^2 = \sqrt{-1}\sqrt{-1} = -1$. When we defome a variable it is useful to denote what number set it belongs to. For example if $z = 3 + 1i$ is a complex number, we write $z \in \mathbb{C}$ meaning "z is in the set of complex numbers $\mathbb{C}$". Note that the complex numbers, real numbers, and rational numbers are each a field (according to our definition above). However, the integers, natural numbers, and irrational numbers are not (e.g., because they lack a multiplicative inverse element).



Linear Algebra

Vector Space. A *vector space* (or *linear space*) is defined over a scalar field F (usually either the field of reals $\mathbb{R}$ or complex numbers $\mathbb{C}$), and is a set of vectors which are closed under addition and closed under scalar multiplication by elements of the scalar field. The elements of the vector space must satisfy a set of axioms (similar to the field axioms described above). Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ be vectors in the vector space V and $\alpha, \beta \in F$ be scalars.

- Associativity of vector addition: $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
- Commutativity: $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
- Identity element: there exists an element $\boldsymbol{0}$ called the *zero vector* such that $\boldsymbol{v} + \boldsymbol{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$.
- Inverse element: for every vector $\boldsymbol{v}$ there exists an inverse element of $-\boldsymbol{v}$ such that $\boldsymbol{v} + (-\boldsymbol{v}) = \boldsymbol{0}$
- Scalar multiplication and field multiplication: $\alpha(\beta \boldsymbol{v}) = (\alpha\beta)\boldsymbol{v}$
- Identity element of scalar multiplication $1\boldsymbol{v} = \boldsymbol{v}$
- Distributivity of scalar multiplication: $\alpha(\boldsymbol{v} + \boldsymbol{w}) = \alpha\boldsymbol{v} + \alpha\boldsymbol{w}$ and $(\alpha + \beta)\boldsymbol{v} = \alpha\boldsymbol{v} + \beta\boldsymbol{v}$.

A vector space with the field of scalars being real numbers is called a *real vector space*. Again, the axioms for a vector space and field that were introduced above will not be directly used in this course. But they are helpful to remind ourselves of the properties of real-valued vectors we will encounter frequently.

Vectors. A vector is an element of a n-dimensional real vector space, denoted $\mathbb{R}^n$. This is simply a collection of n real numbers x_1, x_2 and so on to x_n where all $x_i \in \mathbb{R}$ for $i = 1, \ldots, n$ along with the operations of vector addition and scalar multiplication that satisfy the axioms of a vector space described above. A vector space is usually denoted with bold lowercase (e.g., $\boldsymbol{x}$). A common convention (which we adopt in this course) is that vectors are assumed to be a column of numbers (rather than a row). For example,

$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \in \mathbb{R}^3 . \tag{3}$$

Taking the transpose (denoted by a superscript $^{\mathrm{T}}$) of a column vector converts it into a row vector. For example, using $\boldsymbol{x}$ from above:

$$\boldsymbol{x}^{\mathrm{T}} = \begin{bmatrix} 2 & 3 & 1 \end{bmatrix} \tag{4}$$

which we sometimes write (equivalently) with commas as $\boldsymbol{x}^{\mathrm{T}} = [2, 3, 1]$. Some authors might leave the row/column nature ambiguous and simply write $\boldsymbol{x} = (2, 3, 1)$ with parantheses. Note that taking the transpose twice returns the original column vector, $\boldsymbol{x} = (\boldsymbol{x}^{\mathrm{T}})^{\mathrm{T}}$.

Vector Span. The span of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ is denoted

$$\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\} \tag{5}$$

and is the subspace (i.e., a subset of vectors in the vector space) consisting of all linear combinations of the vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$. In other words, the subspace is equal to the set of all vectors $\boldsymbol{u} \in V$ that are contained in

$$V_1 = \{\boldsymbol{u} \in V \mid \boldsymbol{u} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_p \boldsymbol{v}_p\} \tag{6}$$

for some set of scalars $\alpha_1, \alpha_2, \ldots, \alpha_p$. The above equation illustrates a common way that sets are introduced in mathematical texts. In English, the above statement translates to "V_1 is the set of vectors $\boldsymbol{u}$ that are elements of V such that $\boldsymbol{u}$ is equal to $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_p \boldsymbol{v}_p$ for some set of scalars $\alpha_1, \alpha_2, \ldots, \alpha_p$.

Example. Let

$$V_1 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\}$$

be a subspace of the vector space $\mathbb{R}^3$. Are the vectors

$$\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 6 \end{bmatrix} \quad \text{and} \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

in V_1? By inspection, we note that

$$\boldsymbol{x}_1 = 2\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Thus $\boldsymbol{x}_1 \in V_1$. For $\boldsymbol{x}_2$, there exists no such scalars and $\boldsymbol{x}_2 \notin V_1$.

Linear Independence. The set of vectors $\boldsymbol{v}_1, \ldots \boldsymbol{v}_p$ is linear independent if the only scalars $\alpha_1, \ldots, \alpha_p$ for which

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_p \boldsymbol{v}_p = 0 \tag{7}$$

are $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. Conversely, the set of vectors $\boldsymbol{v}_1, \ldots \boldsymbol{v}_p$ are linearly dependent if they are not linearly independent.

Example. Determine if the following set of vectors are linearly dependent or independent:

$$V_1 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 8 \end{bmatrix} \right\} \qquad \text{and} \qquad V_2 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\} \tag{8}$$

For V_1, by inspection, we can note that

$$-4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1\begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{9}$$

which implies linear dependence. For V_2 it is clear that there are no scalars α_1 and α_2 such that

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

hence the vectors in V_2 are linearly independent.

Vector Multiplication (Dot/Inner Product). The dot product of two vectors (of the same size) $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ and $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$ is defined as

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_N = \sum_{i=1}^{n} x_i y_i \tag{11}$$

The dot product is also sometimes denoted $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ or, using matrix multiplication, as $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$.

Vector Norm. The square of the norm of a vector $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ is written $||\boldsymbol{x}||^2$ and is equal to the sum of the squares of the individual elements $||\boldsymbol{x}||^2 = x_1^2 + x_2^2 + \cdots + x_n^2$. This norm can also be defined by multiplying the transpose of the vector by the vector itself:

$$\begin{aligned}
||\boldsymbol{x}||^2 &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \\
&= [x_1 \ x_2 \ \cdots x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1x_1 + x_2x_2 + \cdots + x_nx_n \\
&= x_1^2 + x_2^2 + \cdots + x_n^2
\end{aligned}$$

Linear System of Algebraic Equations (Standard Form). A linear system of m equations with n variables $\{x_1, x_2, \cdots, x_n\}$ is written in *standard form* as

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

where a_{ij} and b_i are coefficients in which the index i indicates the equation number (from 1 to m) and the index j indicates the corresponding variable x_j (from 1 to n). The right hand side (RHS) contains all of the constants and the left hand side (LHS) includes all the variables and their coefficients summed in increasing order from x_1 to x_n.

Linear System of Algebraic Equations (Matrix Form). A linear system of equations written in standard form can be converted into *matrix form* as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\boldsymbol{b}} \tag{12}$$

where $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ is the matrix of coefficients with rows and columns and $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} = \mathbb{R}^m$ are vectors. We say that a matrix is of size $(m \times n)$ (read as "m by n") to indicate that there are m rows and n columns. $\mathbb{R}^{n\times m}$ denotes the set of all matrices of this size that have real-valued elements. An element of the matrix $\boldsymbol{A}$ is referred to using index notation as a_{ij} to indicate that it is the element corresponding to the i-th row and the j-th column. The vectors we defined previously can be viewed as matrices with just one row or column.

Matrix (Mapping View). In the above example, the vector $\boldsymbol{b}$ was a constant vector from the linear system of equations. We can also view the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ as a *mapping* from an n-dimensional vector space to a m-dimensional one, which is denoted:

$$\boldsymbol{A} \ : \ \mathbb{R}^n \mapsto \mathbb{R}^m \tag{13}$$

Note: when introduce some type of operator as we have done above it is common to write the variable denoting this operator (e.g., $\boldsymbol{A}$) followed by a colon ":", input space (e.g., $\mathbb{R}^m$, then an arrow (e.g., $\to$ or $\mapsto$), followed by the output space (e.g., $\mathbb{R}^m$). Usually an operator that generates a scalar is called a *function* and uses the symbol $\to$, whereas an operator that does not generate a scalar (e.g., it generates a vector) is called a *mapping* and uses the symbol $\mapsto$.

When the matrix $\boldsymbol{A}$ operates on a vector $\boldsymbol{x} \in \mathbb{R}^n$ it produces a new vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$. In this course we will often pre-multiply a column vector $\boldsymbol{x}$ with n elements, by a matrix $\boldsymbol{A}$ of size $(m \times n)$ to produce this new vector $\boldsymbol{y}$ which has m elements (i.e., a $(m \times 1)$ matrix). This equation is written

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrix-by-Matrix Multiplication. To multiply a matrix $\boldsymbol{A}$ by a matrix $\boldsymbol{B}$ their respective sizes must be compatible. For example, for the operation $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ to be well defined, the number of columns of $\boldsymbol{A}$ should be equal to the number of rows of $\boldsymbol{B}$. This condition is satisfied if $\boldsymbol{A}$ is size $(m \times n)$ and $\boldsymbol{B}$ is size $(n \times p)$ (notice that they have the n in common). The result of the multiplication will be a new matrix $\boldsymbol{C}$ of size $(m \times p)$.

A convenient way to ensure a matrix operation is valid, and to determine the size of the resulting matrix, is to write out the respective sizes of each matrix, side-by-side, and strike out the "canceling terms". For example, in the above case we can write

$$(m \cancel{\times n)(n} \times p) = (m \times p)$$

and strike out the two n terms since they "cancel". This is particularly useful for a complex expression involving several matrices. Suppose we have $\boldsymbol{Q} \in \mathbb{R}^{n \times q}$, $\boldsymbol{R} \in \mathbb{R}^{q \times r}$, $\boldsymbol{M} \in \mathbb{R}^{r \times n}$. The equation $\boldsymbol{D} = \boldsymbol{Q}\boldsymbol{R}\boldsymbol{M}$ is well defined because

$$(n \cancel{\times q)(q \times r)(r} \times n) = (n \times n)$$

which implies the result $\boldsymbol{D}$ has n rows and n columns. (Matrices with the same number of rows and columns are called *square*.) However, the product in reverse order $\boldsymbol{M}\boldsymbol{R}\boldsymbol{Q}$ is not well defined. In general, matrix multiplication is not *commutative* (i.e., the order of matrix multiplication matters). For clarity, we will sometimes say $\boldsymbol{A}$ *pre-multiplies* or *post-multiplies* $\boldsymbol{B}$ to denote that we are referring to the product $\boldsymbol{A}\boldsymbol{B}$ or $\boldsymbol{B}\boldsymbol{A}$, respectively.

Example: Determining the size of a matrix resulting from a matrix multiplication

Let

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 \\ 7 & 2 & 0 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 9 & 1 & 3 \\ 1 & 2 & 1 \\ -5 & 0 & 10 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{C} = \begin{bmatrix} 3 & 2 \\ -1 & 8 \\ 0 & 5 \end{bmatrix},$$

The size of $\boldsymbol{A}$ is (2×3), the size of $\boldsymbol{B}$ is (3×3), the size of $\boldsymbol{C}$ is (3×2). Thus, the matrix $\boldsymbol{D} = \boldsymbol{ABC}$ is of size

$$(2 \cancel{\times 3)(3 \times 3)(3} \times 2) \rightarrow (2 \times 2)$$

Note that the following operations are valid $\boldsymbol{AB}$, $\boldsymbol{AC}$, $\boldsymbol{CA}$, $\boldsymbol{BC}$ whereas the following are not $\boldsymbol{BA}$, $\boldsymbol{CB}$.

After determining the size of the matrix resulting from a matrix multiplication we can sketch the result using index notation. For example, after multiplying $\boldsymbol{A}$ of size $(m \times n)$ by $\boldsymbol{B}$ of size $(n \times p)$ we can sketch the result $\boldsymbol{C} = \boldsymbol{AB}$ of size $(m \times p)$ as follows:

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

Each element c_{ij} is a result of multiplying the i-th row of $\boldsymbol{A}$ by the j-th column of $\boldsymbol{B}$

$$\begin{aligned} c_{ij} &= \begin{bmatrix} i\text{-th row of } \boldsymbol{A} \end{bmatrix} \begin{bmatrix} j\text{-th} \\ \text{column} \\ \text{of} \\ \boldsymbol{B} \end{bmatrix} \\ &= \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} \\ &= (a_{i1} b_{1j}) + (a_{i2} b_{2j}) + \cdots + (a_{in} b_{nj}) \end{aligned}$$

Then by evaluating for all of the values c_{ij} we arrive at the resulting matrix $\boldsymbol{C}$.

Example: Evaluating the matrix resulting from a matrix multiplication
Following the previous example, let

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 \\ 7 & 2 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 9 & 1 & 3 \\ 1 & 2 & 1 \\ -5 & 0 & 10 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{C} = \begin{bmatrix} 3 & 2 \\ -1 & 8 \\ 0 & 5 \end{bmatrix}.$$

The product $\boldsymbol{D} = \boldsymbol{AB}$ is of size (2×3) and written in index notation as

$$\boldsymbol{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \end{bmatrix}$$

we can evaluate each element as follows:

$$d_{11} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 1 \\ -5 \end{bmatrix} = 0(9) + 1(1) + 1(-5) = -4$$

$$d_{12} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = 0(1) + 1(2) + 1(0) = 2$$

Repeating this procedure for the remaining indices (d_{13}, d_{21}, d_{22}, and d_{23}) we arrive at the final result

$$\boldsymbol{D} = \begin{bmatrix} -4 & 2 & 11 \\ 65 & 11 & 23 \end{bmatrix}$$

If a matrix multiplication involves several matrices, we can iterate this process for every pair of matrices until we compute the entire product. For example, to evaluate $\boldsymbol{E} = \boldsymbol{ABC}$ we can use our result above in which $\boldsymbol{AB} = \boldsymbol{D}$ and then evaluate $\boldsymbol{E} = \boldsymbol{DC}$.

Matrix addition and subtraction. Matrix addition and subtraction requires that both matrices are of the same size and the operation occurs element-wise. For example, $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ implies that $\boldsymbol{A}$ and $\boldsymbol{B}$ are the same size and each element of $\boldsymbol{C}$ is obtained by adding the corresponding element of $\boldsymbol{A}$ and $\boldsymbol{B}$, that is $c_{ij} = a_{ij} + b_{ij}$. In general, if

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

then

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$$

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix} = \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) & \cdots & (a_{1n} + b_{1n}) \\ (a_{21} + b_{21}) & (a_{22} + b_{22}) & \cdots & (a_{2n} + b_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (a_{m1} + b_{m1}) & (a_{m2} + b_{m2}) & \cdots & (a_{mn} + b_{mn}) \end{bmatrix}$$

And similarly for subtraction:

$$\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$$

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix} = \begin{bmatrix} (a_{11} - b_{11}) & (a_{12} - b_{12}) & \cdots & (a_{1n} - b_{1n}) \\ (a_{21} - b_{21}) & (a_{22} - b_{22}) & \cdots & (a_{2n} - b_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (a_{m1} - b_{m1}) & (a_{m2} - b_{m2}) & \cdots & (a_{mn} - b_{mn}) \end{bmatrix}$$

Matrix-by-scalar Multiplication. Scalar multiplication also occurs element-wise. If the matrix $\boldsymbol{A}$ is multiplied by a scalar α then the result is obtained by multiplying each element a_{ij} by α. Thus,

$$\alpha\boldsymbol{A} = \alpha \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{bmatrix}$$

Matrix Transpose. The transpose of a matrix is denoted with the superscript symbol $^{\mathrm{T}}$ and is obtained by interchanging the rows and columns of the matrix. Thus, the transpose of the matrix

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

is

$$\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Example: Transposing matrices and vectors

The "skinny" matrix $\boldsymbol{A}$ becomes a "wide" matrix $\boldsymbol{A}^{\mathrm{T}}$ after transposing:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 7 & -2 \\ -1 & 3 \\ 9 & 5 \end{bmatrix} \qquad \boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 0 & 7 & -1 & 9 \\ 1 & -2 & 3 & 5 \end{bmatrix}$$

Similarly, the column vector becomes a row vector after transposing

$$\boldsymbol{b} = \begin{bmatrix} 4 \\ 3 \\ -1 \\ 10 \end{bmatrix} \qquad \boldsymbol{b}^{\mathrm{T}} = \begin{bmatrix} 4 & 3 & -1 & 10 \end{bmatrix}$$

For a square matrix (with an equal number of rows and columns) the diagonal elements (i.e., those on the line connecting the upper left to bottom right corner of each matrix) remain unchanged when transposing. In the following example, the elements on the diagonal $(0, 2, 3, 1)$ remain unchanged:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 5 \\ 7 & 2 & 0 & 3 \\ -1 & 3 & 3 & -3 \\ 9 & 5 & -1 & 1 \end{bmatrix} \qquad \boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 0 & 7 & -1 & 9 \\ 1 & 2 & 3 & 5 \\ 1 & 0 & 3 & -1 \\ 5 & 3 & -3 & 1 \end{bmatrix}$$

Symmetric Matrix. The matrix $\boldsymbol{A}$ is symmetric if it remains unchanged by taking the transpose (i.e., $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$).

Transpose of a Matrix Product. Suppose we have a matrix product $\boldsymbol{ABC}$, then the transpose of this product is $(\boldsymbol{ABC})^{\mathrm{T}}$. However, if distributing this transpose to the terms inside the product we must reverse the order of matrix multiplication:

$$(\boldsymbol{ABC})^{\mathrm{T}} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \tag{14}$$

and this is, in general, *not* equivalent to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}$. One exception to this rule is when the matrix product results in a scalar. For example, suppose $\boldsymbol{y}$ and $\boldsymbol{p}$ are vectors while $\boldsymbol{A}$ is a matrix and

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p} = c$$

where c is a scalar. The transpose of a scalar c (which can be viewed as a 1×1 matrix) is also c. Then, it follows that:

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = c$$

and we can replace $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p} = \boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

Identity matrix. The *identity* matrix is a square $n \times n$ matrix, denoted $\boldsymbol{I}_n$ (or sometimes just $\boldsymbol{I}$ when the size is clear from context). The identity matrix has all zero entries, except for ones on the diagonal:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

An identity matrix has the property that when it pre-multiplies a vector $\boldsymbol{x}$ (with n rows), or another matrix $\boldsymbol{A}$ (also with n rows and n columns) they remain unchanged. In other words,

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{x} \qquad \text{and} \qquad \boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$$

Example: Using the identity matrix

For $n = 2$ the identity matrix is

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then if $\boldsymbol{x} = [2, \quad 3]^{\mathrm{T}}$, we see that

$$\boldsymbol{I}\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1(2)+0(3) \\ 0(2)+1(3) \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \boldsymbol{x}$$

Similarly, if

$$\boldsymbol{A} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$$

Then

$$\boldsymbol{I}\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 1(3)+0(5) & 1(4)+0(6) \\ 0(3)+1(5) & 0(4)+1(6) \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} = \boldsymbol{A}$$

Determinant. The determinant of a matrix $\boldsymbol{A}$ is denoted either as $\det(\boldsymbol{A})$ or with vertical bars as $|\boldsymbol{A}|$. For square 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{15}$$

the determinant is computed as

$$\det(\boldsymbol{A}) = ad - bc \tag{16}$$

and for a 3×3 matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{17}$$

the determinant is computed as

$$\det(\boldsymbol{A}) = a \det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b \det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c \det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right) \tag{18}$$

$$= (aei - afh) - (bdi - bfg) + (cdh - ceg) \tag{19}$$

The process can be generalized to larger matrices using *Laplace expansion*. For any $i = 1, 2, \ldots, n$ (i.e., for any row of the matrix):

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} (-1)^{i+j} (-1)^{i+j} a_{ij} \det(\boldsymbol{D}_{ij}) \tag{20}$$

where $\boldsymbol{D}_{ij}$ is a $(n-1) \times (n-1)$ submatrix generated from $\boldsymbol{A}$ by removing the ith row and jth column. The quantity $\boldsymbol{M}_{ij} = \det(\boldsymbol{D}_{ij})$ is called the *minor* of matrix $\boldsymbol{A}$ and the quantity $(-1)^{i+j}\boldsymbol{M}_{ij}$ is called a *cofactor* of a_{ij}. Thus the determinant can be viewed as a weighted sum of cofactors (weighted by the a_{ij} values) and the Laplace expansion (20) can be applied recursively until we have sufficiently small matrices (e.g., 2×2 or 3×3) and the determinant can be evaluated.

Adjugate. The *adjugate* of $\boldsymbol{A}$ is denoted $\text{adj}(\boldsymbol{A})$ and is equal to the transpose of the *cofactor matrix* $\boldsymbol{C}$ of $\boldsymbol{A}$ where the entries of $\boldsymbol{C}$ are computed as

$$c_{ij} = (-1)^{i+j} \boldsymbol{M}_{ij} \tag{21}$$

where $\boldsymbol{M}_{ij}$ is the minor matrix we introduced a moment ago. That is, $\text{adj}(\boldsymbol{A}) = \boldsymbol{C}^T$.

Example. Let

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

then, for example,

$$c_{12} = (-1)^{1+2} |\boldsymbol{M}_{12}| = -\left|\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right| = a_{21}a_{33} - a_{23}a_{31} \tag{23}$$

and this process can be repeated for all c_{ij} to give $\boldsymbol{C}$ so that

$$\text{adj}(\boldsymbol{A}) = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \tag{24}$$

Inverse. Now returning to the equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, we often find ourselves in a situation where $\boldsymbol{A}$ and $\boldsymbol{y}$ are known and we wish to solve for $\boldsymbol{x}$. In scalar algebra, if we were presented with the analogous equation $y = ax$, we could easily solve for x by dividing both sides by $1/a = a^{-1}$ to obtain $x = y/a$. Similarly, in matrix algebra we define the *matrix inverse* $\boldsymbol{A}^{-1}$ with the $^{-1}$ superscript. $\boldsymbol{A}^{-1}$ has the property that $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$. In other words, when a matrix is multiplied by its inverse, the product becomes the identity matrix. Thus when we multiply both sides of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ by the inverse $\boldsymbol{A}^{-1}$, we obtain the solution $\boldsymbol{x}$:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{A}x \\ \boldsymbol{A}^{-1}\boldsymbol{y} &= \boldsymbol{A}^{-1}\boldsymbol{A}x \\ \boldsymbol{A}^{-1}\boldsymbol{y} &= \boldsymbol{I}_n\boldsymbol{x} \\ \boldsymbol{A}^{-1}\boldsymbol{y} &= \boldsymbol{x} \end{aligned}$$

In MATLAB, the inverse function is implemented as `inv(A)`. However, the matrix inverse $\boldsymbol{A}^{-1}$ only exists if the $\det(\boldsymbol{A}) = 0$. If it does exist, then

$$\boldsymbol{A}^{-1} = \frac{\text{adj}(\boldsymbol{A})}{\det(\boldsymbol{A})} \tag{25}$$

Some useful properties for square matrices:

$$\begin{aligned} (\boldsymbol{A}\boldsymbol{B})^{-1} &= \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \\ (\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} &= \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \\ (\boldsymbol{A}^{-1})^{\mathrm{T}} &= (\boldsymbol{A}^{\mathrm{T}})^{-1} \end{aligned}$$

Example. For a 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{26}$$

The matrix inverse is

$$\boldsymbol{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{27}$$

Rank. The rank of a matrix $\boldsymbol{A}$, denoted $\text{rank}(\boldsymbol{A})$, is the number of linearly independent columns in $\boldsymbol{A}$. Interestingly, the rank of a matrix remains unchanged after taking the transpose, $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}^{\mathrm{T}})$, which means rank can be defined equivalently in terms of the independence of rows or columns. It follows that the rank cannot be greater than the minimum of the number of columns or rows. For example, a 2×3 matrix can have rank no greater than 2. A square $n \times n$ matrixhas *full rank* if its rank is n.

Nullspace (Kernel). The nullspace of $\boldsymbol{A}$ is the subspace defined by the vectors $\boldsymbol{x}$ that satisfy $\boldsymbol{A}\boldsymbol{x} = 0$. That is, if $\boldsymbol{A} : \mathbb{R}^n \mapsto \mathbb{R}^m$ then

$$\text{null}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = 0\} \tag{28}$$

When a matrix is full rank the only vector in the nullspace is the zero vector.

Trace. The trace of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements and is denoted by $\text{trace}(\boldsymbol{A})$ or $\text{Tr}(\boldsymbol{A})$:

$$\text{Tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii} \tag{29}$$

Example.

$$\text{Tr}\left(\begin{bmatrix} 0 & 1 & 3 \\ 7 & -2 & 1 \\ -1 & 3 & 5 \end{bmatrix}\right) = 0 + (-2) + 5 = 3 \tag{30}$$

It is also true that $\boldsymbol{x}^T\boldsymbol{x} = \text{Tr}(\boldsymbol{x}\boldsymbol{x}^T)$.

Example. Suppose

$$\boldsymbol{x}^T\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + x_2^2 + x_3^2 \tag{31}$$

then

$$\boldsymbol{x}\boldsymbol{x}^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix} \tag{32}$$

and it follows that $\boldsymbol{x}^T\boldsymbol{x} = \text{Tr}(\boldsymbol{x}\boldsymbol{x}^T)$.

Eigenvalues. Suppose $\boldsymbol{A}$ is a square $n \times n$ matrix with n linearly independent eigenvectors $\boldsymbol{q}_i$ for $i = 1, \ldots, n$ and corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. The eigenvalue equation states that

$$\boldsymbol{A}\boldsymbol{q}_i = \lambda_i \boldsymbol{q}_i \tag{33}$$

for each eigenvalue/eigenvector pair.

Useful Matrix Properties Below is a summary of some of the matrix properties we've covered along with a few new ones. Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be matrices and $\boldsymbol{x}, \boldsymbol{y}$ be vectors of appropriate size. Also, let α, β be scalars.

- $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$
- $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$
- $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{BC}$
- $(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{A} = \boldsymbol{BA}+\boldsymbol{CA}$
- $\alpha(\boldsymbol{AB}) = (\alpha\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(\alpha\boldsymbol{B})$
- $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$
- $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $\det(\boldsymbol{A}) = \lambda_1\lambda_2\cdots\lambda_n$

Further Resources. An excellent resource with many handy matrix identities is the *Matrix Cookbook* by Kaare Petersen and Michae Pedersen available here. Other references include *Linear Algebra and Its Applications* by Gilbert Strang and *Matrix Analysis* by Roger Horn and Charles Johnson.

References

[Bertsekas, 2008] Bertsekas, D. (2008). *Introduction to Probability*. Athena Scientific.

[Stilwell, 2012] Stilwell, D. (2012). Lecture notes in ECE 5744: Linear Systems Theory.

[Strang, 1993] Strang, G. (1993). *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA.