Definition 1. An \( m \times n \) matrix \( A \) is an array \((a_{ij})\) of real (or complex) numbers, indexed by natural numbers \( i \) and \( j \), with \( 1 \leq i \leq m \) and \( 1 \leq j \leq n \), written like this:

\[
\begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\]

A column vector is simply an \( m \times 1 \) matrix for some \( m \), whereas a row vector is a \( 1 \times n \) matrix for some \( n \). \( m \) is called the height of the column vector and \( n \) the width of the row vector. We’ll write \( \mathbb{R}^{m \times n} \) for the set of all \( m \times n \) matrices; we’ll often identify \( \mathbb{R}^{m \times 1} \) with \( \mathbb{R}^m \) and \( \mathbb{R}^{1 \times n} \) with \( \mathbb{R}^n \).

If \( A \) is an \( m \times n \) matrix and \( B \) is an \( n \times p \) matrix, then their product \( AB \) is the \( m \times p \) matrix \( C \) with entries

\[
C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \quad 1 \leq i \leq m \quad 1 \leq j \leq p
\]

If \( A \) is \( m \times n \) and \( B \) is \( p \times q \) where \( n \neq p \), we leave \( AB \) undefined. If \( \lambda \in \mathbb{R} \) and \( A \) is a matrix with entries \( a_{ij} \), then \( \lambda A \) is the matrix with entries \( \lambda a_{ij} \).

The \( n \times n \) identity matrix \( I_n \) is the matrix

\[
\begin{pmatrix}
  1 & 0 & \cdots & 0 \\
  0 & 1 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & 1
\end{pmatrix}
\]

We will often drop the \( n \) subscript when there is no possibility of confusion (and sometimes even when there is).

We will often be working with linear combinations of column (or row) vectors, ie, expressions of the form

\[
\lambda_1 a_1 + \cdots + \lambda_n a_n
\]
where $\lambda_i \in \mathbb{R}$ for each $i$, and $a_1, \ldots, a_n$ are all column vectors of the same height. It’s very useful to note that if $A$ is an $m \times n$ matrix and $x$ is a column vector of height $n$, then $Ax$ is a linear combination of the columns $a_1, \ldots, a_n$ of $A$:

$$x_1 a_1 + \cdots + x_n a_n$$

**Fact 1.** The following hold for all matrices $A$, $B$, and $C$ (so long as the sizes make sense), and all $\lambda \in \mathbb{R}$.

1. $(AB)C = A(BC)$. (Associativity.)
2. $A(B + C) = AB + AC$. (Distributivity.)
3. $\lambda(AB) = (\lambda A)B = A(\lambda B)$. (Commutativity of scalar multiplication.)
4. $\lambda(A + B) = \lambda A + \lambda B$. (Distributivity of scalar multiplication.)
5. If $A$ is $m \times n$, then $AI_n = A$ and $I_mA = A$. (Identity.)

**Proof of (1).** First, notice that (for either product to make sense) the sizes of $A$, $B$, and $C$ must be $m \times n$, $n \times p$, and $p \times q$ respectively, for some $m, n, p, q$. The products $(AB)C$ and $A(BC)$ both have size $m \times q$. Now for any $i \leq m$ and $j \leq q$, we have

$$((AB)C)_{ij} = \sum_{k=1}^{p} (AB)_{ik} C_{kj}$$

$$= \sum_{k=1}^{p} \left( \sum_{\ell=1}^{n} A_{i\ell} B_{\ell k} \right) C_{kj}$$

$$= \sum_{k=1}^{p} \sum_{\ell=1}^{n} A_{i\ell} B_{\ell k} C_{kj}$$

$$= \sum_{\ell=1}^{n} \sum_{k=1}^{p} A_{i\ell} B_{\ell k} C_{kj}$$

$$= \sum_{\ell=1}^{n} A_{i\ell} \left( \sum_{k=1}^{p} B_{\ell k} C_{kj} \right)$$

$$= \sum_{\ell=1}^{n} A_{i\ell} (BC)_{\ell j}$$

$$= (A(BC))_{ij}$$

\[\square\]

**Proof of (5).** I’ll just prove that $AI_n = A$. The other equation is similar (though you should work it out on your own anyway). First notice that $AI_n$ has size $m \times n$. Now if $i \leq m$ and $j \leq n$,

$$(AI)_{ij} = \sum_{k=1}^{n} A_{ik} I_{kj}$$
(Here I’ve dropped the subscript on $I_n$.) By definition, $I_{kj} = 1$ if $k = j$, and $I_{kj} = 0$ if $k \neq j$. Hence the only term in the above sum which is nonzero is the term where $k = j$, and that term is $A_{ij}$. Hence $(AI)_{ij} = A_{ij}$ for all $i$ and $j$.

**Fact 2.** For all $n$, there are $n \times n$ matrices $A$ and $B$ such that $AB \neq BA$.

A system of linear equations can be written as a single equation involving matrices and vectors; namely, the system

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

can be written as $Ax = b$, where $A$ is the $m \times n$ matrix with entries $a_{ij}$, $x$ is the column vector with entries $x_i$, and $b$ is the column vector with entries $b_i$. We also sometimes represent the entire system as a matrix;

\[
\begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\
  a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn} & b_m
\end{pmatrix}
\]

We write $(A \mid b)$ for this matrix, and call it *augmented* (by $b$).

You can perform row operations on a matrix, and it’s just as you would expect from the above. The interesting thing is that each row operation can be implemented by multiplication by a certain matrix. For instance, if $E$ is the following matrix, where the zeros along the diagonal occur in rows $i$ and $j$ (and all blank entries are zero), then $EA$ is the result of swapping rows $i$ and $j$ in $A$.

\[
\begin{pmatrix}
  1 & & & & \\
  & \ddots & & & \\
  & & 0 & 1 & \\
  & & 1 & \ddots & \\
  & & & & 1
\end{pmatrix}
\]
The other row operations are realized by the following matrices.

\[
\begin{pmatrix}
1 & & \\
& \ddots & \\
& & \lambda \\
& & \\
& & 1
\end{pmatrix}
\quad
\begin{pmatrix}
1 & & \\
& \ddots & \\
& & 1 \\
& & \lambda \\
& & 1
\end{pmatrix}
\]

**Question.** Which matrix implements which row operation?

We call these *elementary matrices*. They are all *square*; that is, they have size \( n \times n \) for some \( n \). In each case, to perform a row operation on an \( m \times n \) matrix \( A \), we multiply \( A \) on the left by its corresponding \( m \times m \) elementary matrix \( E \), to get \( EA \). Multiplication by \( E \) on the right (assuming \( m = n \); otherwise this doesn’t even make sense) would perform a *column operation* on \( A \). If Fact 2 did not convince you to be careful of which way you multiply matrices, then this should.

**Fact 3.** If \( E \) is an \( m \times m \) elementary matrix, then there is an \( m \times m \) elementary matrix \( F \) such that \( FE = I \).

**Proof.** We saw in the proof of Theorem 1, from day one, that every row operation is reversible. So let \( F \) be the elementary matrix which implements the reverse of the row operation that \( E \) implements. Then we have \( F(EA) = A \) for all \( m \times n \) matrices \( A \), for all \( n \). By associativity, this means \((FE)A = A \) for all \( m \times n \) matrices \( A \), for all \( n \). It’s not too hard to show from this that \( FE = I \). (It suffices, in fact, to consider just one matrix \( A \), of size \( m \times m \). Which one is it?)