Working with Matrices

We introduced matrices as a computational tool for working with linear transformations between finitely generated vector spaces. The general properties of linear transformations and the algebraic operations on them — addition, multiplication by scalars and composition — enabled us to define corresponding algebraic operations on matrices — matrix addition, multiplication by scalars and matrix multiplication.

The axioms governing these algebraic operations on matrices, as well as the restrictions required for their definition, followed immediately from the general properties of linear transformations.

Our analysis also provided an explicit procedure for finding the matrix representing a given linear transformation, as well as for the relationship between different matrices representing the same linear transformation.

We recall the definitions.

**Definition.**

\[ M(m \times n; F) \times M(m \times n; F) \longrightarrow M(m \times n; F), \quad ([a_{ij}], [b_{ij}]) \mapsto [a_{ij} + b_{ij}] \quad (1) \]

\[ F \times M(m \times n; F) \longrightarrow M(m \times n; F), \quad (\lambda, [a_{ij}]) \mapsto [\lambda a_{ij}] \quad (2) \]

\[ M(m \times n; F) \times M(n \times p; F) \longrightarrow M(m \times p; F), \quad ([a_{ij}], [b_{jk}]) \mapsto [c_{ik}] \quad (3) \]

where

\[ c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk} \]

The last definition, expressed in Equation (3), provides the key to many applications.

It is not immediately apparent that we have actually achieved much more than simply providing the means for working computationally with linear transformations. We explore some of the additional aspects and applications.

In the first place, by defintion, \( F(m) = M(m \times 1; F) \) and \( F(n) = M(1 \times n; F) \). Hence, putting \( p = 1 \) in (3), matrix multiplication defined the function

\[ M(m \times n; F) \times F(n) \longrightarrow F(m), \quad (A, x) \mapsto y = Ax \quad (4) \]

If we write out Equation (4) explicitly for \( x = [x_j]_{n \times 1}, A = [a_{ij}]_{m \times n} \) and \( y = [y_i]_{m \times 1} \), we obtain, for all \( i \in \{1, \ldots, m\} \),

\[ y_i = \sum_{j=1}^{n} a_{ij} x_j \quad (5) \]

By Theorem 5.3, \( T: F(n) \longrightarrow F(m) \) is a linear transformation if and only if there constants, \( a_{ij} \in F \) (\( 1 \leq i \leq m, \ 1 \leq j \leq n \)), such that for all \( i \in \{1, \ldots, m\} \),

\[ y_i = \sum_{j=1}^{n} a_{ij} x_j \]

In other words,

**Theorem.** A linear transformation \( T: F(n) \longrightarrow F(m) \), is multiplication (on the left) by a matrix.

**Observation.** The matrix in the last theorem does not rely on the choice of a basis: any linear transformation between vector spaces of the form \( F(n) \) is multiplication on the left by a matrix.

Vector spaces of the form \( F(n) \) are unique in having this property. A linear transformation between other finitely generated vector spaces is not actually multiplication on the left by a matrix, although
we have shown that it can be so represented: the representation arises from choosing bases, one for the domain and one for the co-domain, with the matrix obtained dependent on the bases chosen.

We next observe that the vector space structure on the set of all \( m \times n \) matrices with coefficients in \( F \) defined by (1) and (2), together with the internal structure of matrices, is the basis for defining isomorphisms useful in applications.

Given the \( m \times n \) matrix

\[
\mathbf{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{m \times n}
\]

the vector space structures defined provide an isomorphism

\[
\mathbb{F}^{mn} \longrightarrow \mathbf{M}(m \times n; F), \quad (x_1, \ldots, x_n, x_{n+1}, \ldots, x_{mn}) \longmapsto \begin{bmatrix} x_{(i-1)n+j} \end{bmatrix}_{m \times n}
\] (6)

We may also regard the matrix \( \mathbf{A} \) as comprising \( m \) rows,

\[
\mathbf{r}_i \mathbf{A} := \begin{bmatrix} a_{i1} & \cdots & a_{im} \end{bmatrix}
\] (7)

for \( 1 \leq i \leq m \), or as comprising \( n \) columns

\[
\mathbf{c}_j \mathbf{A} := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}
\] (8)

for \( 1 \leq j \leq n \).

Since each \( \mathbf{r}_i \mathbf{A} \in \mathbb{F}^n \) and \( \mathbf{c}_j \mathbf{A} \in \mathbb{F}^m \), the vector space structures provide isomorphisms

\[
\mathbf{M}(m \times n; F) \longrightarrow \bigoplus_{i=1}^m \mathbb{F}^{(n)} \quad \mathbf{A} \longmapsto (\mathbf{r}_1 \mathbf{A}, \ldots, \mathbf{r}_m \mathbf{A})
\] (9)

\[
\mathbf{M}(m \times n; F) \longrightarrow \bigoplus_{j=1}^n \mathbb{F}^{(m)} \quad \mathbf{A} \longmapsto (\mathbf{c}_1 \mathbf{A}, \ldots, \mathbf{c}_n \mathbf{A})
\] (10)

The three isomorphisms, (6), (9) and (10) provide not only flexibility in computing, but broaden the range of both practical and theoretical applications.

**Observation.** The isomorphism

\[
\mathbf{M}(1; F) \longrightarrow \mathbb{F}, \quad [\mathbf{A}] \longmapsto a
\] (11)

allows us to identify \( \mathbf{M}(1; F) \) with \( F \) and to restate definition of the matrix product \( \mathbf{A} \mathbf{B} \) with \( \mathbf{A} \in \mathbf{M}(m \times n; F) \) and \( \mathbf{B} \in \mathbf{M}(n \times p; F) \) as the matrix \( [c_{ik}] \in \mathbf{M}(m \times p; F) \) given by

\[
c_{ik} := \mathbf{r}_i \mathbf{A} \mathbf{c}_k
\] (12)

We next apply Equation (3) to establish the relation between the rows and columns of the product of two matrices and the rows and columns of the two matrices multiplied.

Take \( \mathbf{A} = [a_{ij}] \in \mathbf{M}(m \times n; F) \) and \( \mathbf{B} = [b_{jk}] \in \mathbf{M}(n \times p; F) \).

The \( i \)th row of \( \mathbf{A} \mathbf{B} \) is obtained by fixing \( i \) and varying \( k \) in Equation (3):

\[
\mathbf{r}_i \mathbf{A} \mathbf{B} = \begin{bmatrix} \sum_{j=1}^n a_{ij} b_{j1} & \cdots & \sum_{j=1}^n a_{ij} b_{jn} \end{bmatrix}
\]

\[= \sum_{j=1}^n a_{ij} \begin{bmatrix} b_{j1} & \cdots & b_{jn} \end{bmatrix}
\]

\[= \sum_{j=1}^n a_{ij} \mathbf{r}_j \mathbf{B}
\]
On the other hand,

\[
\begin{bmatrix}
\sum_{j=1}^{n} a_{ij} b_{j1} & \cdots & \sum_{j=1}^{n} a_{ij} b_{jn}
\end{bmatrix}
= \begin{bmatrix}
[a_{i1}]_{n} & \cdots & [a_{im}]_{n}
\end{bmatrix}
\begin{bmatrix}
b_{11} & \cdots & b_{1p}
\vdots & \ddots & \vdots 
\hline
b_{n1} & \cdots & b_{np}
\end{bmatrix}
= r_{A} B
\]

It is left to the reader to verify in detail that similarly, by fixing \(k\) and varying \(i\),

\[
c_{A} B = \sum_{j=1}^{n} c_{A} b_{jk} = A c_{k}
\]

We summarise these calculations.

**Theorem (Matrix Multiplication Reconsidered).** The \(i\)th row of the product of two matrices is the linear combination of the rows of the right-hand matrix given by the entries in the \(i\)th row of the left-hand matrix and can be computed by multiplying on the right the \(i\)th row of the left-hand matrix by the right-hand matrix.

The \(k\)th column of the product of two matrices is the linear combination of the columns of the left-hand matrix given by the entries in the \(k\)th column of the right-hand matrix and can be computed by multiplying on the left the \(k\)th column of the right-hand factor by the left-hand matrix.

In symbols, using the notation above

\[
\begin{align*}
l_{A} B &= \sum_{j=1}^{n} a_{ij} l_{j} B \\
&= r_{A} B \\
l_{k} A B &= \sum_{j=1}^{n} c_{A} b_{jk} \\
&= A c_{k}
\end{align*}
\]

**Definition.** The row space of the \(m \times n\) matrix \(A\) is the vector subspace of \(F^{(n)}\) generated by \(\{l_{1}, \ldots, l_{n}\}\) and its column space is the vector subspace of \(F^{(m)}\) generated by \(\{c_{1}, \ldots, c_{n}\}\).

**Application.** If a matrix, \(A\), represents the linear transformation, \(T\), then the column space of \(A\) is isomorphic with the image of \(T\), with the columns of \(A\) providing an explicit set of generators for the image of \(T\).

**Proof.** Let \(B = \{e_{1}, \ldots, e_{n}\}\) and be a basis for \(V\), \(C = \{f_{1}, \ldots, f_{m}\}\) a basis for \(W\) and \(T: V \to W\) a linear transformation.

Since \(B\) is a basis for \(V\), \(\{T(e_{1}), \ldots, T(e_{n})\}\) generates \(\text{im}(T)\), so that \(w \in \text{im}(T)\) if and only if

\[
w = \sum_{j=1}^{n} x_{j} T(e_{j})
\]

for suitable \(x_{1}, \ldots, x_{n} \in F\).

As \(A = [a_{ij}] \in M(m \times n; F)\) is the matrix of \(T\) with respect to \(B\) and \(C\) if and only if for each \(j \in \{1, \ldots, n\}\)

\[
T(e_{j}) = \sum_{i=1}^{n} a_{ij} f_{i}
\]
the set of $n$ vectors
$$\left\{ \sum_{i=1}^{n} a_{1i} f_i, \ldots, \sum_{i=1}^{n} a_{ni} f_i \right\}$$
generates $\text{im}(T)$.

Moreover, $w \in \text{im}(T)$ if and only if
$$w = \sum_{j=1}^{n} x_j T(e_j)$$
$$= \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} x_j \right) f_i,$$
if and only if the co-ordinate vector of $w$ with respect to $C$ is
$$y = \begin{bmatrix}
\sum_{j=1}^{n} a_{1j} x_j \\
\vdots \\
\sum_{j=1}^{n} a_{mj} x_j
\end{bmatrix}
= \begin{bmatrix}
a_{11} \\
\vdots \\
a_{m1}
\end{bmatrix} x_1 + \cdots + \begin{bmatrix}
a_{1n} \\
\vdots \\
a_{mn}
\end{bmatrix} x_n
= x_1 e^A_1 + \cdots + x_n e^A_n$$

This is the case if and only if $y$ is in the column space of $A$.

By the definition of the vector space operations on $\mathbb{F}_m$, this bijection is an isomorphism. \qed

**Application.** The system of $m$ simultaneous linear equations in $n$ unknowns
$$a_{11} x_1 + \cdots + a_{1n} x_n = b_1$$
$$\vdots$$
$$a_{m1} x_1 + \cdots + a_{mn} x_n = b_m$$
is represented by the matrix equation
$$Ax = b$$
(18)

where
\[
A = \begin{bmatrix} a_{ij} \end{bmatrix}_{m \times n} \in \mathbb{M}(m \times n; \mathbb{F})
\]
$$x = \begin{bmatrix} x_1 \\
\vdots \\
x_n \end{bmatrix} \in \mathbb{F}_n$$
$$b = \begin{bmatrix} b_1 \\
\vdots \\
b_m \end{bmatrix} \in \mathbb{F}_m$$

and, therefore, has a solution if and only if $b$ is in the column space of $A$. 

The Gauß-Jordan Algorithm. Gauß-Jordan Elimination provides an algorithm for finding all solutions to the system of equations (17).

While it is essentially a numerical procedure, it has important consequences for the theory of linear algebra of finitely generated vector spaces and for calculation with matrices.

This is suggested by the fact that the system of equations (17) is equivalent to the matrix equation Equation (18).

We use matrix multiplication to re-formulate the Gauß-Jordan algorithm for solving the system of equations (17) adapted to solving Equation (18) and illustrate applications.

Gauß-Jordan Elimination rests on three observations.

**GJ1** The set of all solutions of a system of $m$ linear equations in $n$ unknowns is not affected when we multiply any single equation by a non-zero scalar.

**GJ2** The set of all solutions of a system of $m$ linear equations in $n$ unknowns is not affected when we add any multiple of any single equation to any other of the equations.

**GJ3** The set of all solutions of a system of $m$ linear equations in $n$ unknowns is not affected by the order in which the equations are listed.

We typically perform Gauß-Jordan elimination by representing the system of equations by the (augmented) matrix, in which the $i$th row comprises the coefficients in the $i$th equation, and then applying **elementary row operations** corresponding to the observations just made.

- **ERO1:** Multiply the $i$th row by $\lambda \neq 0$.
- **ERO2:** Add $\mu$ times the $j$th row to the $i$th row.
- **ERO3:** Swap the $i$th row and the $j$th row.

Using our theorem which expresses the rows of a product of matrices in terms of the rows of the right-hand matrix, we see that the elementary row operations can be performed by multiplying the matrix in question on the left by a suitable matrix. We list these matrices first, and then illustrate them with concrete examples.

**GJ1/ERO1:**

$$M(i; \lambda) = [x_{k\ell}]_{m \times m} \quad \text{where} \quad x_{k\ell} := \begin{cases} \lambda & \text{for } k = \ell = i \\ 1 & \text{for } k = \ell \neq i \\ 0 & \text{otherwise} \end{cases}$$ (19)

**GJ2/ERO2:**

$$A(i; \mu, j) = [x_{k\ell}]_{m \times m} \quad \text{where} \quad x_{k\ell} := \begin{cases} \mu & \text{for } k = i, \ell = j \\ 1 & \text{for } k = \ell \\ 0 & \text{otherwise} \end{cases}$$ (20)

**GJ3/ERO3:**

$$S(i; j) = [x_{k\ell}]_{m \times m} \quad \text{where} \quad x_{k\ell} := \begin{cases} 1 & \text{for } k = i, \ell = j \\ 1 & \text{for } k = \ell, i = j \\ \lambda & \text{for } i \neq k, \ell \neq j \\ 0 & \text{otherwise} \end{cases}$$ (21)

We illustrate these using the $3 \times 2$ matrix, leaving it to the reader to check the details.

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}.$$
1) We multiply the second row by $\lambda$.
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
a & b \\
c & d \\
e & f
\end{bmatrix} =
\begin{bmatrix}
a & b \\
\lambda c & \lambda d \\
e & f
\end{bmatrix}
\]

2) We add $\mu$ times the third row to the second row.
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & \mu \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
a & b \\
c & d \\
e & f
\end{bmatrix} =
\begin{bmatrix}
a & b \\
c + \mu e & d + \mu f \\
e & f
\end{bmatrix}
\]

3) We swap the first row and the third row.
\[
\begin{bmatrix}
0 & 0 & 1 \\
0 & 1 & \mu \\
1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
a & b \\
c & d \\
e & f
\end{bmatrix} =
\begin{bmatrix}
e & f \\
c & d \\
a & b
\end{bmatrix}
\]

The elementary row operations are performed until the (unaugmented) matrix is brought into reduced row echelon form, which is when

(i) Each row has a 1 as first non-zero coefficient (from the left).

(ii) The first 1 is any row is further to the right than in any row above the row considered.

(iii) Every other entry in a column, that contains the first 1 of a given row, is 0.

(iv) If a row has every coefficient 0, then the same is also true of every row below it.

Example. Of the two matrices
\[
\begin{bmatrix}
0 & 1 & 2 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
1 & 0 & 3 & 2 \\
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]
the former is in reduced row echelon form, but the latter is not.

The transformation of matrices to reduced row echelon form is central to applications.

The theory we have developed provides the basis for the Gauß-Jordan algorithm’s applications. In particular, the various isomorphisms involving the vector space of all $m \times n$ matrices, and the algebraic properties of matrices, allow us to change perspective and interpret the various calculations accordingly.

We compute an example and then explain the theoretical basis for it.

Example. We determine the image and kernel of the real linear transformation $T: V \rightarrow W$, whose matrix with respect to the bases $B = \{e_1, e_2, e_n\}$ for $V$ and $C = \{f_1, f_2, f_3\}$ for $W$ is
\[
\begin{bmatrix}
1 & 2 & 3 \\
4 & 8 & 13 \\
5 & 10 & 14
\end{bmatrix}
\]
(22)

The vector $w = rf_1 + sf_2 + tf_3 \in W$ is in the image of $T$ if and only if we can find a vector $v = xe_1 + ye_2 + ze_3 \in V$ with $T(v) = w$.

Since the matrix of $T$ is $\mathbf{A}$, this is equivalent to solving the system of simultaneous linear equations
\[
\begin{align*}
x + 2y + 3z &= r \\
4x + 8y + 13z &= s \\
5x + 10y + 14z &= t
\end{align*}
\]
(23)

We transform the augmented matrix of this system of equations to reduced row echelon form.
\[
\begin{bmatrix}
1 & 2 & 3 \\
4 & 8 & 13 \\
5 & 10 & 14 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 & 3 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 & 3 \\
0 & 0 & -1 \\
0 & 0 & 0 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

\[R_2 - 4R_1 \]
\[R_3 - 5R_1 \]

From this it follows that

1) The vector \( w = rf_1 + sf_2 + tf_3 \in W \) is in the image of \( T \) if and only if \( 9r = s + t \).

2) The vectors \( f_1 + 4f_2 + 5f_3 \) and \( 3f_1 + 13f_2 + 14f_3 \) form a basis for the image of \( T \).

3) The vector \( xe_1 + ye_2 + ze_3 \) is in the kernel of \( T \) if and only if \( x + 2y = z = 0 \), so that \( 2e_1 - e_2 \) is a basis for the kernel of \( T \).

We explain how this information follows from the reduced row echelon form of the augmented matrix,

\[
E = \begin{bmatrix}
1 & 2 & 0 & 13r - 3s \\
0 & 0 & 1 & -4r + s \\
0 & 0 & 0 & -9r + s + t \\
\end{bmatrix}
\]  \[\text{(24)}\]

1) It is immediate from the third row of \( E \), that the system of equations (23) has a solution if and only if \( 9r = s + t \).

2) It is immediate from (24) that the first and third column of \( E \) form a basis for the column space of \( E \).

While this is not the same as the column space of \( A \), the two are isomorphic, because, being reversible, each step in the Gauß-Jordan Algorithm defines an isomorphism between the column space of the matrix being acted on, and the column space of the resulting matrix.

Since isomorphisms map bases to bases, the corresponding columns of \( A \) — the first and third — form a basis for the column space of \( A \).

These are, respectively, the co-ordinate vectors of \( f_1 + 4f_2 + 5f_3 \) and \( 3f_1 + 13f_2 + 14f_3 \).

Since the image of \( T \) is isomorphic with the column space of \( A \) under the isomorphism defined by choosing the basis, these form a basis for the image of \( T \).

3) The choice of the basis for \( V \) defines an isomorphism between the kernel of \( T \) and the null space of \( A \).

Since each step in the Gauß-Jordan elimination is reversible, it preserves null spaces — the null space of the matrix being acted on is the same as the null space of the resulting matrix. Thus, the null spaces of \( A \) and \( E \) coincide.

To find the null space, we solve the system of equations (23) for \( r = s = t \).

We see from the rows of \( E \) that the solutions are given by \( x + 2y = z = 0 \).

To make the geometry of this even more explicit, we consider the case \( V = W = \mathbb{R}^3 \) and take each of \( B \) and \( C \) to be the standard basis, which is the case typically studied at school.

Our analysis shows that, in this case,

\[T: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad (x, y, z) \mapsto (r, s, t)\]
maps $\mathbb{R}^3$ onto the plane given by the equation
\[ 9r - s - t = 0 \]
and its kernel is the line given by the equations
\[ x + 2y = z = 0 \]
The parametric representation of the image is
\[ \{(\alpha + 3\beta, 4\alpha + 13\beta, 5\alpha + 14\beta) \mid \alpha, \beta \in \mathbb{R}\} \]
or
\[ r = \alpha + 3\beta \quad s = 4\alpha + 13\beta \quad t = 5\alpha + 14\beta \quad (\alpha, \beta \in \mathbb{R}) \]
and that of the kernel is
\[ \{(2\gamma, -\gamma, 0) \mid \gamma \in \mathbb{R}\} \]
or
\[ x = 2\gamma \quad y = -\gamma \quad z = 0 \quad (\gamma \in \mathbb{R}) \]