LINEAR TRANSFORMATIONS AND MATRICES

1. Vectors

We can identify $n \times 1$ and $1 \times n$ matrices with n-dimensional vectors by taking the entries as the Cartesian coordinates of the head of the (geometric) vector with tail at the origin. When thought of this way we $n \times 1$ matrices are called *(column) vectors* and $1 \times n$ vectors are callled *row vectors*. We denote the space of n-dimensional vectors by $\mathbb{R}^n$ and denote an element with an arrow, e.g., $\vec{v} \in \mathbb{R}^n$.

We can add two vectors by adding their entries

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

this geometrically corresponds to the vectors satisfying a parallelogram law. Similarly, we can scale any vector by a $k \in \mathbb{R}$ by

$$k \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} kx_1 \\ \vdots \\ kx_n \end{bmatrix}.$$

Geometrically, when $k > 0$ this corresponds to stretching the vector by a fact of k. When $k < 0$, this is accompanied by reflecting through the origin.

The *zero vector*, $\vec{0}$, has all entries zero. The *standard vectors*, are the elements $\vec{e}_1, \ldots, \vec{e}_n \in \mathbb{R}^n$ which have entry 1 in the ith row and all other entries 0. Clearly,

$$\vec{v} = v_1\vec{e}_1 + \cdots + v_n\vec{e}_n \text{ for } \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

The *dot product* of two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ is defined to be

$$\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots x_n y_n \text{ where } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

The length of a vector $||\vec{x}||$ satisfies

$$||\vec{x}||^2 = \vec{x} \cdot \vec{x}.$$

Furthermore, for two non-zero vectors, $\vec{x}, \vec{y}$

$$\vec{x} \cdot \vec{y} = ||\vec{x}|| \cdot ||\vec{y}|| \cos\theta$$

where θ is the angle between $\vec{x}$ and $\vec{y}$ (the vectors are non-zero so θ makes sense).

Given a $n \times m$ matrix A it is often convenient to write A in terms of its columns which we may think of as m vectors in $\mathbb{R}^n$. This is expressed as

$$A = \begin{bmatrix} \vec{a}_1 & | & \cdots & | & \vec{a}_m \end{bmatrix}$$

where here $\vec{a}_i \in \mathbb{R}^n$. For instance,

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & | & \vec{a}_2 & | & \vec{a}_3 \end{bmatrix}$$

has columns

$$\vec{a}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \vec{a}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \vec{a}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Given a $n \times m$ matrix and m-dimensional matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & | & \cdots & | & \vec{a}_m \end{bmatrix} \text{ and } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{R}^m,$$

define the product of A and $\vec{x}$ to be

$$A\vec{x} = \begin{bmatrix} x_1 a_{11} + \cdots + x_m a_{1m} \\ \vdots \\ x_1 a_{n1} + \cdots + x_m a_{nm} \end{bmatrix} = x_1 \vec{a}_1 + \cdots + x_m \vec{a}_m.$$

In particular,

$$\vec{a}_i = A\vec{e}_i.$$

That is, the ith column of A is the product of A and the ith standard vector.

Multiplication of a matrix with a vector satisfies:

(1) $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$.
(2) $A(k\vec{x}) = k(A\vec{x})$ for $k \in \mathbb{R}$.

2. Linear Transformations

A *function* (or *transformation*) consists of three things:

(1) A set X called the *domain*;
(2) A set Y called the *target space*;
(3) A rule $f : X \to Y$ that associates to each element $x \in X$ exactly one element $y = f(x)$.

Two sets X and Y and a rule f that associates to each element x of X exactly one element $f(x)$ in Y. An element x in X will be called an *input* and the corresponding value $y = f(x)$ is the *output*.

A transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is said to be a *linear transformation* if the following is true

(1) $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$ for all $\vec{x}, \vec{y} \in \mathbb{R}^m$
(2) $T(k\vec{x}) = kT(\vec{x})$ for all $\vec{x} \in \mathbb{R}^m$ and $k \in \mathbb{R}$.

EXAMPLE: Any $n \times m$ matrix A, gives a linear transformation $T_A : \mathbb{R}^m \to \mathbb{R}^n$

$$T_A(\vec{x}) = A\vec{x}.$$

Indeed, using the algebraic properties from above we have:

$$T_A(\vec{x} + \vec{y}) = A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = T_A(\vec{x}) + T_A(\vec{y})$$

and

$$T_A(k\vec{x}) = A(k\vec{x}) = k(A\vec{x}) = kT_A(\vec{x}).$$

Note that the textbook takes the opposite approach, as they define linear transformations as those given by multiplication by a matrix and then deduce our definition of linear transformation as a property.

Every linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is of the form $T = T_A$ for some $n \times m$ matrix A. Indeed, for any linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ define the *matrix of* T which we indicate by $[T]$ to be the $n \times m$ matrix given by

$$[T] = \begin{bmatrix} T(\vec{e}_1) & | & \cdots & | & T(\vec{e}_m) \end{bmatrix}$$

so the ith column of $[T]$ is the vector $T(\vec{e}_i)$ (i.e., the output of T given input $\vec{e}_i$). Clearly, $[T]\vec{e}_i = T(\vec{e}_i)$. By linearity and properties of multiplication of a matrix and a vector it follows that $[T]\vec{x} = T(\vec{x})$ for each $\vec{x} \in \mathbb{R}^m$. Indeed, write $\vec{x} = x_1\vec{e}_1 + \cdots + x_m\vec{e}_m$ and observe

$$\begin{aligned} T(\vec{x}) &= T(x_1\vec{e}_1 + \cdots + x_m\vec{e}_m) \\ &= x_1T(\vec{e}_1) + \cdots + x_mT(\vec{e}_m) && \text{by linearity of } T \\ &= x_1[T]\vec{e}_1 + \cdots x_m[T]\vec{e}_m && \text{definition of } [T] \\ &= [T](x_1\vec{e}_1 + \cdots + x_m\vec{e}_m) && \text{algebraic properties} \\ &= [T]\vec{x} && . \end{aligned}$$

In other words, if $A = [T]$, then $T = T_A$. You should think of a matrix as a way to (numerically) represent a linear transformation just as a column vector is a way to numerically represent a geometric vector.

EXAMPLE: Let $I_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n$ with $I_{\mathbb{R}^n}(\vec{x}) = \vec{x}$ be the identity transform. It is easy to see this is linear and that

$$[I_{\mathbb{R}^n}] = I_n$$

where here I_n is the $n \times n$ identity matrix (i.e. the matrix with 1 on the diagonal and all other entries 0).

EXAMPLE: Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates a vector counter-clockwise by θ-radians. Geometrically, clear this is a linear transformation.

$$R_\theta\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \text{ and } R_\theta\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}.$$

Hence,

$$[R_\theta] = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$$

and so

$$R_\theta\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}x_1\cos\theta - x_2\sin\theta\\ x_1\sin\theta + x_2\cos\theta\end{bmatrix}.$$

3. Matrix Multiplication and Composition of Linear Transforms

If B is a $n \times p$ matrix and A is a $p \times m$ matrix, then the matrix product, BA, is

$$BA = \begin{bmatrix} B\vec{a}_1 & | & \cdots & | & B\vec{a}_m \end{bmatrix}$$

where

$$A = \begin{bmatrix} \vec{a}_1 & | & \cdots & | & \vec{a}_m \end{bmatrix}$$

has columns $\vec{a}_j \in \mathbb{R}^p$. This is equivalent to the following: if

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pm} \end{bmatrix}, B = \begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{bmatrix} \text{ and } C = \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{bmatrix},$$

then $C = BA$ means
$$c_{ij} = \sum_{k=1}^{p} b_{ik} a_{kj}.$$

EXAMPLE:
$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 5 & -1 \\ 1 & 2 & 0 \end{bmatrix}$$
is equivalent to
$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Matrix multiplication's definition makes it compatible with composition of linear transformations. Specifically, suppose $T : \mathbb{R}^m \to \mathbb{R}^p$ and $S : \mathbb{R}^p \to \mathbb{R}^n$ are both linear transformations. Their composition $S \circ T : \mathbb{R}^m \to \mathbb{R}^n$ is defined by $(S \circ T)(\vec{x}) = S(T(\vec{x}))$. It is easy to check that $S \circ T$ is linear. For example,
$$\begin{aligned}(S \circ T)(\vec{x} + \vec{y}) &= S(T(\vec{x} + \vec{y})) = S(T(\vec{x}) + T(\vec{y})) \\ &= S(T(\vec{x})) + S(T(\vec{x})) = (S \circ T)(\vec{x}) + (S \circ T)(\vec{y}).\end{aligned}$$
As such, it makes sense to consider $[S \circ T]$, the matrix associated to $S \circ T$. The definition of matrix multiplication ensures that:
$$[S \circ T] = [S][T].$$
To see this observe that,
$$\begin{aligned}[S \circ T] &= \begin{bmatrix} S(T(\vec{e}_1)) & | & \cdots & | & S(T(\vec{e}_m)) \end{bmatrix} && \text{definition of } [S \circ T] \\ &= \begin{bmatrix} [S]T(\vec{e}_1)) & | & \cdots & | & [S]T(\vec{e}_m)) \end{bmatrix} && [S] \text{ is the matrix of } S \\ &= [S] \begin{bmatrix} T(\vec{e}_1)) & | & \cdots & | & T(\vec{e}_m)) \end{bmatrix} && \text{definition of matrix multiplication} \\ &= [S][T]\end{aligned}$$

4. Invertible Matrices

A linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is *invertible* (with *inverse* T^{-1}) if for each $\vec{y} \in \mathbb{R}^n$ the equation
$$T(\vec{x}) = \vec{y} \tag{1}$$
has exactly one solution. This solution is $\vec{x} = T^{-1}(\vec{y})$ which allows us to think of
$$T^{-1} : \mathbb{R}^n \to \mathbb{R}^m$$
as a transformation sending $\vec{y} \in \mathbb{R}^n$ and to $T^{-1}(\vec{y})$, the unique solution to (1). One readily checks that T^{-1} is linear. For instance, as $\vec{x} = T^{-1}(\vec{y})$ solves (1), the linearity of T means $\vec{x} = kT^{-1}(\vec{y})$ solves $T(\vec{x}) = k\vec{y}$. Hence, $kT^{-1}(\vec{y}) = T^{-1}(k\vec{y})$.

The easiest way to check if a candidate transformation, S, is the inverse of T is to use the following fact: If $S : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform that satisfies $S \circ T = I_{\mathbb{R}^m}$ (such S is said to be a *left* inverse of T) and $T \circ S = I_{\mathbb{R}^n}$ (such S is said to be a *right* inverse of T), then T is invertible and $S = T^{-1}$ (e.g., T^{-1} is both a left and right inverse and so is sometimes called a *two-sided* inverse).

To understand why this is so, first observe that if $T \circ S = I_{\mathbb{R}^n}$, then (1) has at least one solution given by $\vec{x} = S(\vec{y})$, but could have more solutions. Conversely, if $S \circ T = I_{\mathbb{R}^m}$, then (1) can have at most one solution, but may have no solutions. In other words, a right inverse ensures existence of some solution while a left inverse ensures uniqueness of any given solution.

A $n \times m$ matrix A is invertible if T_A is invertible and the *inverse matrix* is $A^{-1} = [T_A^{-1}]$. In similar fashion to the above, if B is $m \times n$ matrix and $AB = I_n$ and $BA = I_m$, then A is invertible and $A^{-1} = B$.

EXAMPLE: Consider, R_θ rotation counterclockwise by θ. Geometrically, $I_{\mathbb{R}^2} = R_{-\theta} \circ R_\theta = R_\theta \circ R_{-\theta}$ so $R_\theta^{-1} = R_{-\theta}$. Moreover,

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and can check

$$[R_\theta][R_{-\theta}] = I_2 = [R_{-\theta}][R_\theta].$$

EXAMPLE: Let $T : \mathbb{R}^2 \to \mathbb{R}$ be linear transform $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_1 + x_2$. The matrix of T is $[T] = \begin{bmatrix} 1 & 1 \end{bmatrix}$. If $R(x_1) = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$ can check $T(R(x_1)) = x_1$. That is, R is a right inverse. However, there is no left inverse. Indeed, let $L : \mathbb{R} \to \mathbb{R}^2$ be an arbitrary linear map, so $[L] = \begin{bmatrix} a \\ b \end{bmatrix}$.

$$[L \circ T] = [L][T] = \begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} a & a \\ b & b \end{bmatrix} \neq I_2$$

for any a, b.

5. Calculating the inverse of a matrix

We wish to determine how we can compute A^{-1} for agiven matrix $n \times m$ matrix A. As a first step, recall that A is invertible means $A\vec{x} = \vec{y}$ has a unique solution for each $\vec{y}$. By properties of Gauss-Jordan elimination, this means rref(A)

(1) Has a pivot in each column (ensuring uniqueness of the solution)
(2) Has a pivot in each row (ensuring existence).

In other words, $m = n$ and rref$(A) = I_n$. This is equivalent to A being $n \times n$ and $rank(A) = n$. Observe, this immediately means that if $T : \mathbb{R}^m \to \mathbb{R}^n$ is an invertible linear map, then $m = n$.

EXAMPLE: Is $A = \begin{bmatrix} 2 & 4 \\ 1 & -1 \end{bmatrix}$ invertible?

$$\text{rref} \begin{bmatrix} 2 & 4 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so answer is yes.

Suppose now A is an invertible $n \times n$ matrix. The columns of A^{-1} are

$$\vec{v}_1 = A^{-1}\vec{e}_1, \dots, \vec{v}_n = A^{-1}\vec{e}_n$$

One determines the $\vec{v}_i$ by solving

$$A\vec{x} = \vec{e}_i$$

for each $i = 1, \dots, n$. This requires solving n different systems of n equations in n unknowns. As the coefficient matrix the same for each system, you only need to apply Gauss-Jordan elimination once. This is because you can augment n additional columns (instead of just one) corresponding to each standard vector. In this case the augemented matrix is

$$\begin{bmatrix} A & | & \vec{e}_1 & | & \cdots & | & \vec{e}_n \end{bmatrix} = \begin{bmatrix} A & | & I_n \end{bmatrix}$$

and one has (for invertible A)

$$\operatorname{rref}\begin{bmatrix} A & | & I_n \end{bmatrix} = \begin{bmatrix} I_n & | & \vec{v}_1 & | & \cdots & | & \vec{v}_n \end{bmatrix} = \begin{bmatrix} I_n & | & A^{-1} \end{bmatrix}.$$

EXAMPLE: Compute inverse of $\begin{bmatrix} -1 & 2 \\ 2 & -5 \end{bmatrix}$

$$\operatorname{rref}\begin{bmatrix} -1 & 2 & | & 1 & 0 \\ 2 & -5 & | & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & | & -5 & -2 \\ 0 & 1 & | & -2 & -1 \end{bmatrix}$$

so

$$\begin{bmatrix} -1 & 2 \\ 2 & -5 \end{bmatrix}^{-1} = \begin{bmatrix} -5 & -2 \\ -2 & -1 \end{bmatrix}.$$

NON-EXAMPLE:

$$\operatorname{rref}\begin{bmatrix} 2 & 1 & | & 1 & 0 \\ -4 & -2 & | & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & | & 0 & -\frac{1}{4} \\ 0 & 0 & | & 1 & \frac{1}{2} \end{bmatrix}$$

first 2×2 matrix not I_2 so $\begin{bmatrix} 2 & 1 \\ -4 & -2 \end{bmatrix}$ is not invertible.

6. Properties of Matrix Multiplication and the Matrix Inverse

Here are some properties of matrix multiplication and the matrix inverse:

(1) Matrix multiplication is *non-commutative*, e.g., in general $AB \neq BA$. Reflects fact that, in general, $S \circ T \neq T \circ S$.

(2) I_n is the *multiplicative identity*. That is, if A is $n \times m$ matrix, then

$$I_n A = A = A I_m.$$

(3) Matrix multiplication is *associative*

$$(AB)C = A(BC) \Rightarrow ABC \text{ makes sense.}$$

(4) Matrix multiplication *distributes* over matrix addition

$$A(C + D) = AC + AD \text{ and } (A + B)C = AC + BC$$

(5) If A is invertible, then so is A^{-1} and $(A^{-1})^{-1} = A$.

(6) If A and B are $n \times n$ matrices and $AB = I_n$ (or $BA = I_n$), then $BA = I_n$ (or $AB = I_n$) and so $B = A^{-1}$. In other words for square matrices, it is enough to check that B is either a right or a left inverse.

(7) Suppose A and B are invertible $n \times n$ matrices, then so is AB and $(AB)^{-1} = B^{-1}A^{-1}$. Matrix multiplication is not commutative so the order matters.

EXAMPLE: Item (7) follows from (2),(3) and (6). Indeed, using (2) and (3)

$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}I_n B = B^{-1}B = I_n$$

and so, by (6),

$$(AB)^{-1} = B^{-1}A^{-1}.$$