Linear Algebra

Kian Kenyon-Dean

October 2017

Abstract

This reflection is written for the purpose of formalizing my new and improved understanding of linear algebra, and an understanding that will improve even during just the process of writing this reflection. We will explore the essential concepts of linear algebra and understand them in a new light; this will be the geometric interpretation of linear algebra, as opposed to the interpretation based solely on the manipulation of symbols. Through this heightened understanding, many other concepts will be elucidated; in particular, topics in Machine Learning and Deep Learning (particularly with respect to *manifolds*) will be explored in the light of this new understanding.

1 Vectors and Linear Transformations

Abstractly, linearity is defined by two simple axioms, additivity and scalability. If a function or object satisfies these axioms, then it can be understood as linear. If the axioms below in Equation 1 are satified, then L is a linear function.

$$\begin{aligned} &L(\mathbf{a}+\mathbf{b}) = L(\mathbf{a}) + L(\mathbf{b}) \\ &L(c\mathbf{a}) = cL(\mathbf{a}) \end{aligned} \tag{1}$$

In linear algebra, vectors satisfy these properties of linearity. On an abstract level the concept of the derivative is also linear since it follows these axioms. But for our purposes here we will explore vectors and matrices, both of which are linear objects that create linear transformations.

We understand a vector as what some might call a *column vector*, e.g., a matrix $\mathbf{v} \in \mathbb{R}^{n\times 1}$. This vector has a tail at the origin and has a head that points to a point in n-dimensional space (n-space) with coordinates corresponding to its values. A more nuanced understanding, which we will explore later, leads us to understand a vector as a matrix with rank 1 that defines a 1-dimensional manifold in the n-space in which it is defined.

A matrix is commonly represented simply as a rectangle of numbers, some $\mathbf{M} \in \mathbb{R}^{n\times m}$. This is a set of symbols, and there are certain laws, such as multiplication, addition, inverse, determinant that are well known, but moreso as rules for the task of manipulating symbols. This, while perhaps useful for a computational implementation, does not strike at the core essence of matrices in their geometric understanding.

We understand a matrix as a *linear transformation.* It is linear because it satisfies the properties described above in Equation 1; geometrically, this means that a matrix transforms space, but keeps the coordinate lines parallel and equidistant with each other; e.g., in a euclidean coordinate plane you would draw vertical lines to designate $x = 0$, $x = 1$, $x = 2$, etc. – a linear transformation may change their orientation and may also compress or stretch their distance, but they will always remain parallel with each other and each one will have the same distance between each of its neighbors. This will become more clear below.

1.1 The Basics

Another representation of a vector is in terms of the fundamental basis vectors of an n-space. Let $\mathbf{i}_j \in \mathbb{R}^n$ with $j = 1 \dots n$ be defined as the sparse vector where $\mathbf{i}_k = 1$ if $k = j$ else $\mathbf{i}_k = 0$. Any vector $\mathbf{v} \in \mathbb{R}^n$ defined with constants $a_1 \dots a_n$ is thus understood as being a sum of these basic vectors:

$$\mathbf{v} = \sum_{j=1}^{n} a_j \mathbf{i}_j \tag{2}$$

In 3-dimensional space it is common to define the basic vectors as follows: $\mathbf{i} = \langle 1, 0, 0 \rangle$; $\mathbf{j} = \langle 0, 1, 0 \rangle$; $\mathbf{k} = \langle 0, 0, 1 \rangle$. These basic vectors actually define the x, y, z-axes respectively in 3-space, which is why, in this case $\mathbf{v} \in \mathbb{R}^3$ is frequently defined with constants x, y, z:

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{3}$$

We will often remain in the general formulation above as expressed in Equation 2, although, occasionally, for the purpose of establishing a solid grounding, we will explain concepts with respect to the specific cases of dimensions that are understandable and visualizable for humans.

Critical Point. The $\mathbf{i}_j$ basic vectors define the space on which we operate in the most simple way possible. Concatenated together, they define the *identity* matrix $\mathbf{I}_n$, the most important matrix in linear algebra.

$$\mathbf{I}_n = \begin{bmatrix} | & | & \dots & | \\ \mathbf{i}_1 & \mathbf{i}_2 & \dots & \mathbf{i}_n \\ | & | & \dots & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{4}$$

By concatenating these basic vectors above in Equation 4, we see that we have defined the columns of a matrix. This matrix defines a basis in n-space called the canonical basis, a space we call the *canonical space* – it is the space in which everything is initially defined, upon which everything is derived. We will soon come to understand all matrices as linear transformations from the canonical space to another space.

1.2 Bases and Linear Transformations

The question we are led to ask is the following: can we use vectors other than $\mathbf{i}_j$ to define our n-space? Indeed, we may want to, for example, build a representation of 2-space such that, given any vector $\mathbf{v}$ defined by basis vectors $\mathbf{i}_1, \mathbf{i}_2$, we could determine, via a general form, where $\mathbf{v}$ would be oriented given a 90-degree rotation of the canonical space.

1.2.1 Basis and Span

To understand a basis, we must first understand the concept of **span**. The span of a vector $\mathbf{v} \in \mathbb{R}^n$ is simply the set of all vectors that can be represented by a scaling of $\mathbf{v}$; e.g., $\{a\mathbf{v} : \ \forall a \in \mathbb{R}\}$. This span defines a line in n-space; this line is, by definition, 1-dimensional, but it exists in n-space because $\mathbf{v}$ is defined in n-space. We can also generalize to understand the span of a set of vectors $S = \{\mathbf{v}_1 \dots \mathbf{v}_k\}$, which is the set of all possible linear combinations of the vectors in S; e.g., $\text{span}(S) = \{\sum_{i=1}^{k} a_i \mathbf{v}_i | k \in \mathbb{N}, a_i \in \mathbb{R}, \mathbf{v}_i \in S\}$. A vector $\mathbf{v}_i \in S$ is redundant if it can be expressed as a linear combination of the other vectors $\mathbf{v}_j \in S$ $(j \neq i)$ – this means that $\mathbf{v}_i$ is linearly dependent on the other vectors in S and therefore does not add any representative power to S; e.g. $\text{span}(S) = \text{span}(S - \{\mathbf{v}_i\})$. This is the essence of the concept of linear independence. Essentially, if a vector $\mathbf{v}$ adds representative power to S then it is linearly independent from the rest of the vectors in S; e.g., it cannot be expressed via a linear combination of the other vectors in S, meaning that $\text{span}(S \cup \{\mathbf{v}\}) \neq \text{span}(S)$.

The definition of a **basis** for $\mathbb{R}^n$ is a set S of vectors such that S is a span of $\mathbb{R}^n$; e.g., every possible vector in $\mathbb{R}^n$ can be represented as a linear combination of the vectors in S. It is easy to prove that S can at most be a basis for $\mathbb{R}^{|S|}$ – therefore, we thus recognize that a single vector $\mathbf{v} \neq \mathbf{0} \in \mathbb{R}^n$ only spans $\mathbb{R}^1$, but it defines a line that lies within $\mathbb{R}^n$.

Every matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ has columns (and rows, but we will use columns only for consistency) that define a basis in some k-space with $k \leq n$. In the next subsection we will consider what happens when $k = n$, and then we will explore when $k < n$. Note that the maximum value of k that $\mathbf{M}$ defines a basis for is called the **rank** of $\mathbf{M}$. The following statements are equivalent:

- $\mathbf{M}$ has rank k;
- $\mathbf{M}$ as k linearly independent columns;
- $\mathbf{M}$ spans $\mathbb{R}^k$.
- The largest n-space for which $\mathbf{M}$ defines a basis is $\mathbb{R}^k$;
- The smallest n-space for which $\mathbf{M}$ cannot define a basis is $\mathbb{R}^{k+1}$;
- *Kian's definition:* $\mathbf{M}$'s maximal representative power occurs when it is used to express $\mathbb{R}^k$.

1.2.2 Matrices as Linear Transformations

We will now consider when $\mathbf{M} \in \mathbb{R}^{n \times n}$ has full rank; e.g., $\text{rank}(\mathbf{M}) = n$. $\mathbf{M}$ is fundamentally an expression of a linear transformation from one space to another; in particular, on its own it is the transformation from the canonical basis of $\mathbb{R}^n$ to the $\mathbf{M}$-basis of $\mathbb{R}^n$. In a word, it expresses a new space in terms of linear combinations of the columns of $\mathbf{I}_n$.

Consider the particular example of this phenomenon in 2-space and the differences in expression of a vector $\mathbf{v} \in \mathbb{R}^2$, when considering a new basis defined by a matrix $\mathbf{M}$, which we express as the transformation from the canonical basis $\mathbf{I}_2 = \langle \mathbf{i}, \mathbf{j} \rangle$:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} m_{11}\mathbf{i} + m_{21}\mathbf{j} & m_{12}\mathbf{i} + m_{22}\mathbf{j} \end{bmatrix}$$

Equation 5 expresses the canonical basis representation of $\mathbf{v}$, while Equation 6 expresses the $\mathbf{M}$-basis representation of $\mathbf{v}$. This expressivity is understood as the left multiplication of the vector by the matrix. In other words, $\mathbf{M}$ represents $\mathbf{v}$ in a new space in terms of its basis vectors; this is what multiplying a matrix times a vector is, and it can be understood as the projection of $\mathbf{v}$ from the canonical space onto the space defined by $\mathbf{M}$.

$$\mathbf{I}_2\mathbf{v} = a\mathbf{i} + b\mathbf{j} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{5}$$

$$\mathbf{M}\mathbf{v} = a(m_{11}\mathbf{i} + m_{21}\mathbf{j}) + b(m_{12}\mathbf{i} + m_{22}\mathbf{j}) = \begin{bmatrix} am_{11} + bm_{12} \\ am_{21} + bm_{22} \end{bmatrix} \tag{6}$$

The concept of invertibility now arises. When we do a transformation $\mathbf{M}$ we would like to, essentially, "undo" the transformation – this is expressed by $\mathbf{M}^{-1}$. We thus have the property

$\mathbf{M}\mathbf{M}^{-1} = \mathbf{I}_n$. This makes sense because $\mathbf{M}$ transforms us to a new space, but $\mathbf{M}^{-1}$ brings us back to the original one; again, in linear algebra, the original space is the canonical space defined by $\mathbf{I}_n$. This property of invertibility is what differentiates matrices with full rank from those with $\text{rank}(\mathbf{M}) < n$, and it gives rise to a geometric understanding of the determinant, which we describe later on.

When the rank of $\mathbf{M} = k < n$ then we have a matrix with linearly dependent columns. Indeed, it is incapable of spanning $\mathbb{R}^n$, only being able to span $\mathbb{R}^k$. We might say that $\mathbf{M}$ expresses a k-dimensional "slice" of $\mathbb{R}^n$, since it is still defined over that space. For example, a vector in $\mathbb{R}^2$ defines a line in the 2-space; that is, a 1-dimensional slice of $\mathbb{R}^2$ that is within $\mathbb{R}^2$. While only having one-dimensional expressivity power, it is still expressed in the context of the higher dimension. Similarly, a rank 2 matrix in $\mathbb{R}^3$ can only span $\mathbb{R}^2$, meaning that it thus expresses a plane in $\mathbb{R}^3$, meaning that it again slices the space, but does not span it. For example, consider the matrix below defined in 3-space:

$$\mathbf{M} = \begin{bmatrix} | & | & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 3 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{7}$$

This matrix has rank 2, we see clearly that $\mathbf{c}_1$ and $\mathbf{c}_3$ are linearly dependent ($\mathbf{c}_3 = 2\mathbf{c}_1$). This means that it can only express an $\mathbb{R}^2$ subspace of $\mathbb{R}^3$, what (I believe) we can call a linear manifold within $\mathbb{R}^3$. In fact, we find that the columns $\mathbf{c}_1, \mathbf{c}_2$ of $\mathbf{M}$ span an $\mathbb{R}^2$-subspace defined on a plane present in $\mathbb{R}^3$ that represents all linear combinations of $\mathbf{c}_1, \mathbf{c}_2$; e.g., $\{a\mathbf{c}_1 + b\mathbf{c}_2 | \forall a, b \in \mathbb{R}\}$.

Therefore, when you multiply a linear transformation $\mathbf{W} \in \mathbb{R}^{n\times n}$ (with $\text{rank}(\mathbf{W}) = n$) by $\mathbf{M}$, you are projecting W into a space of weaker representative power, it is collapsed into a lower dimension despite the fact that $\mathbf{M} \in \mathbb{R}^{n\times n}$. *This is because a transformation of* $\boldsymbol{W}$ *to* $\boldsymbol{M}$*-space re-expresses* $\boldsymbol{W}$ *with fewer linearly independent vectors than it needs to have full expressivity.* This is why $\text{rank}(XY) = \min(\text{rank}(X), \text{rank}(Y))$. By collapsing $\mathbf{W}$ into k-space by multiplying it by M (but expressed in $\mathbb{R}^{n\times n}$, in n-space) you have forever doomed it to a realm of lower dimensional representation; no linear transformation will ever be able to send it back up to n-space representation. *This is why a matrix in* $\mathbb{R}^{n\times n}$ *with rank* $k < n$ *is* ***not invertible****; you cannot use a linear transformation to go back up into a new dimension after collapsing it down.*