The Pivotal Role of Zero in Linear Algebra

E. L. Lady

(May 28, 1997)

In linear algebra, as in any other subject, there are lots of buzz words. For instance, one constantly comes across terms like *linearly independent*, *span*, *linear transformation*, and *one-to-one*.

But one of the most crucial words in linear algebra is one you may not even notice, because it is so familiar, namely **zero**. Most often, you find statements that certain things *do not equal zero* or are *non-zero*.

A lot of things in linear algebra depend on various vectors or scalars being non-zero. There are two reasons for this.

• The important thing about a *scalar* being non-zero is that non-zero scalars are precisely those that one can divide by. Thus for scalars, the word "non-zero" really means "one that has a reciprocal."

For practical purposes, this means that any single **non-zero** scalar in an equation might as well be 1.

Consider, for instance, the following.

Theorem. If $v_1, \ldots, v_n$ are vectors which are *not* linearly independent, then one of these can be written as a linear combination of the others.

The proof goes as follows. If $v_1, \ldots, v_n$ are not linearly independent, then there exists an equation

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0},$$

where at least one c_i is not zero. What this really means is, *where it is possible to divide by at least one* c_i.

Now if c_1, for instance, is non-zero, then one can divide the equation through by c_1 to get an equation

$$\mathbf{v}_1 + c_2'\mathbf{v}_2 + \cdots + c_n'\mathbf{v}_n = \mathbf{0},$$

which can be rewritten as

$$\mathbf{v}_1 = -c_2'\mathbf{v}_2 - \cdots - c_n'\mathbf{v}_n,$$

proving that $\mathbf{v}_1$ (or $\mathbf{v}_i$, for whichever $c_i \neq 0$) can be written as a linear combination of the other vectors.

This reasoning would not work if only integers were allowed to be scalars, and in that case the preceding theorem would not be true. In fact, let $\mathbb{Z}$ stand for the set of integers, and consider $\mathbb{Z}^2$. This is like a two-dimensional vector space, where the elements look like $[x, y]$, except that in this case x and y are required to be integers. All scalars used in the theory for $\mathbb{Z}^2$ are also required to be integers. For this quasi-vector space (the technical term is "module") the preceding statement is not true. For instance, the three vectors $[2, 0]$, $[0, 2]$, and $[3, 3]$ are not linearly independent, even in this new sense, because there is an equation

$$3\,[2, 0] + 3\,[0, 2] - 2\,[3, 3] = [0, 0]$$

where the scalars are integers and not all of them are zero. (In fact, none of them are zero.) And yet none of these three vectors can be expressed as a linear combination of the other two if the coefficients are required to be integers.

- The second reason why zero plays such an important role in linear algebra is the following:

Being able to tell whether things are zero or not is the same as being able to tell when things are equal.

This is, of course, because saying that $\mathbf{v} = \mathbf{w}$ is the same as saying that $\mathbf{v} - \mathbf{w} = \mathbf{0}$.

Consider, for instance, the usual, very non-intuitive, definition of linear independence.

Definition. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if no equation

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

is possible except when all the coefficients c_i are 0.

On first seeing this definition, one's first thought is likely to be, "Why would anybody care about a thing like that? So what?" But one can rephrase it as follows:

Alternate Definition. Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if an equation

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$$

can only be true when all the coefficients are the same, i. e. for each i, $a_i = b_i$.

The reasoning here is that if $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$, then $(a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}$, so if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent according to the first definition, then all $a_i - b_i = 0$, showing that $a_i = b_i$ for all i. (An even easier proof shows that if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent according to the second definition then they will also be linearly independent according to the first one.

This second definition makes linear independence seem at least a little more like a concept that might be useful. It's very important to realize, though, that **the first definition is easier to use for proving theorems.**

As a second example of how zero is an important concept in this way, consider the idea of the **kernel** of a linear transformation.

By definition, $\operatorname{Ker} L$ consists of all the vectors $\mathbf{v}$ such that $L(\mathbf{v}) = \mathbf{0}$. We make a big deal about $\operatorname{Ker} L$ because $\mathbf{0}$ is so important in linear algebra. In fact, because L is linear, it follows that $L(\mathbf{v}) = L(\mathbf{w})$ if and only if $L(\mathbf{v} - \mathbf{w}) = 0$. In other words,

Lemma. If L is a linear transformation and $\mathbf{v}$, $\mathbf{w}$ are vectors in the domain of L, then $L(\mathbf{v}) = L(\mathbf{w})$ if and only if $\mathbf{v} - \mathbf{w} \in \operatorname{Ker} L$.

Now see how this relates to the idea of a *one-to-one* linear transformation. By definition, to say that L is one-to-one means that $L(\mathbf{v}) = L(\mathbf{w})$ can only happen when $\mathbf{v} = \mathbf{w}$.

TRANSLATION: L is one-to-one means that $L(\mathbf{v} - \mathbf{w}) = \mathbf{0}$ can only happen when $\mathbf{v} - \mathbf{w} = \mathbf{0}$.

From this (plus the fact that $\mathbf{0} \in \operatorname{Ker} L$ for every possible linear transformation L), the following important theorem becomes obvious:

Theorem. A linear transformation L is one-to-one if and only if $\operatorname{Ker} L = \{\mathbf{0}\}$.