Linear Dependence

Linear dependence is a very useful concept that is often used to characterize "redundancy" in information in real world applications. We will give two equivalent definitions of linear dependence.

Definition 3.1 (Linear Dependence (I)): A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly dependent if there exist scalars $\alpha_1, \ldots, \alpha_n$ such that $\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n = \vec{0}$ and not all α_i's are equal to zero.

Definition 3.2 (Linear Dependence (II)): A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly dependent if there exist an index i and scalars α_j's such that $\vec{v}_i = \sum_{j\neq i} \alpha_j\vec{v}_j$. In words, a set of vectors is linearly dependent if one of the vectors could be written as a linear combination of the rest of the vectors.

Why did we introduce two equivalent definitions? They could be useful in different settings. For example, it is often easier mathematically to show linear dependence with definition (I). Could you see why? If we would like to prove linear dependence with definition (II), we need to first choose a vector $\vec{v}_i$ and show that it is a linear combination of the other vectors. However, with definition (I), we don't need to try to "single out" a vector to get started with the proof. We can blindly write down the equation $\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n = \vec{0}$ and begin our proof from there. On the other hand, definition (II) gives us a more intuitive way to talk about redundancy. If a vector can be constructed from the rest of the vectors, then this vector does not contribute anything new to the representation power of the rest of vectors.

Now we will show that the two definitions are equivalent. This is the first formal proof in the course! We will walk you through it.[1] First, we ask the question, "What does it mean when we say two definitions are equivalent?" It means that when the condition in definition (I) holds, the condition in definition (II) must hold as well. And when the condition in definition (II) holds, the condition in definition (I) must also hold. So there are two directions that we have to show:
(i) To see how definition (II) implies definition (I), we start from the condition in definition (II) – suppose there exist an index i and scalars α_j's such that $\vec{v}_i = \sum_{j\neq i} \alpha_j\vec{v}_j$. We want to somehow transform this equation into the form that appears in definition (I). How can we achieve that? We can move $\vec{v}_i$ to the right:

$$\vec{0} = -1 \times \vec{v}_i + \sum_{j\neq i} \alpha_j\vec{v}_j. \tag{1}$$

Now setting $\alpha_i = -1$, we have

$$\vec{0} = \alpha_i \times \vec{v}_i + \sum_{j\neq i} \alpha_j\vec{v}_j = \sum_{j} \alpha_j\vec{v}_j. \tag{2}$$

Since $\alpha_i = -1$, at least one of the scalars α's is not zero, the condition in definition (I) is satisfied.
(ii) Now let's show the reverse! Suppose the condition in definition (I) is true. Then there exist scalars

[1] We will also be releasing a note that provides a more in depth treatment on how to approach proofs.

$\alpha_1, \ldots, \alpha_n$ such that $\alpha_1 \vec{v_1} + \cdots + \alpha_n \vec{v_n} = 0$ and not all α_i's are equal to zero. Since at least one of the α_i's is nonzero, let's assume that α_1 is nonzero. Now how do we get the equation into the form identical to that in definition (II)? Observe that if we move $\alpha_1 \vec{v_1}$ to the opposite side of equation and divide both sides by α_1, we have

$$\vec{v_1} = \sum_{j \neq 1} \left(\frac{\alpha_j}{\alpha_1} \right) \vec{v_j}. \tag{3}$$

We see that this is identical to the second definition. (In our proof, we made the assumption that $\alpha_1 \neq 0$. However, notice that we could as well have supposed that $\alpha_2 \neq 0$, $\alpha_3 \neq 0$, or any index i so that $\alpha_i \neq 0$. The convention is to set the first index, in this case 1, to be nonzero. In mathematical texts, we typically write *"Without loss of generality (W.L.O.G.), we let $\alpha_1 \neq 0$."*)

Definition 3.3 (Linear Independence): A set of vectors is linearly independent if it is not linear dependent. What does this mean? From the first definition of linear dependence, we can deduce that a set of vectors $\{\vec{v_1}, \ldots, \vec{v_n}\}$ is linearly independent if $\alpha_1 \vec{v_1} + \cdots + \alpha_n \vec{v_n} = \vec{0}$ implies $\alpha_1 = \cdots = \alpha_n = 0$.

Let's see some simple examples of linear dependence and linear independence.

Example 3.1 (Linear dependence of 2 vectors): Consider vectors $\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. $\vec{a}$ and $\vec{b}$ are linearly dependent because

$$\vec{b} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2 \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2 \times \vec{a}. \tag{4}$$

Example 3.2 (Linear independence of 2 vectors): Consider vectors $\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$. We will show that the two vectors are linearly independent. Consider scalars α, β such that $\alpha \vec{a} + \beta \vec{b} = \vec{0}$. Splitting up the components of the vector, we have

$$\alpha \vec{a} + \beta \vec{b} = \vec{0} \tag{5}$$

$$\Rightarrow \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6}$$

$$\Rightarrow \begin{bmatrix} 2\alpha \\ \alpha \end{bmatrix} + \begin{bmatrix} \beta \\ 5\beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{7}$$

$$\Rightarrow \begin{bmatrix} 2\alpha + \beta \\ \alpha + 5\beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

$$\Rightarrow \begin{cases} 2\alpha + \beta = 0 \\ \alpha + 5\beta = 0 \end{cases} \tag{9}$$

The system of linear equations of two variables has a unique solution $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. By definition, $\vec{a}$ and $\vec{b}$ are linearly independent.

Connecting the concept of linear dependence with systems of linear equations

In the last note, we saw that a system of linear equations can have zero solutions, a unique solution, or infinitely many solutions. Is there a way to tell what kind of solution a system of linear equations $A\vec{x} = \vec{b}$ has without running Gaussian Elimination or explicitly solving for the solution? Yes! We will show that just looking at the columns of the matrix A can help us answer this question.

(1) If the system of linear equations $A\vec{x} = \vec{b}$ has infinite number of solutions, then the columns of A are linearly dependent.

Let's see why this is the case: If the system has infinite number of solutions, it must have at least two distinct solutions. Let's call them $\vec{y}_1$ and $\vec{y}_2$. Then $\vec{y}_1$ and $\vec{y}_2$ must satisfy

$$A\vec{y}_1 = \vec{b} \tag{10}$$

$$A\vec{y}_2 = \vec{b}. \tag{11}$$

Subtracting the first equation from the second equation, we have $A(\vec{y}_2 - \vec{y}_1) = \vec{0}$. Let $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \vec{y}_2 - \vec{y}_1$.

Because $\vec{y}_1$ and $\vec{y}_2$ are distinct, not all α_i's are equal to zero. Let the columns of A be $\vec{a}_1, \ldots, \vec{a}_n$. Then, $A(\vec{y}_2 - \vec{y}_1) = \sum_{i=1}^{n} \alpha_i \vec{a}_i = \vec{0}$. By definition, the columns of A are linearly dependent.

(2) If the columns of A in the system of linear equations $A\vec{x} = b$ are linearly dependent, then the system does not have a unique solution.

To show this, let the columns of A be $\{\vec{a}_1, \ldots, \vec{a}_n\}$. If A has linearly dependent columns, then there exist scalars $\alpha_1, \ldots, \alpha_n$ not all zeros such that $\sum_i^n \alpha_i \vec{a}_i = 0$. Define $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$. Then $A\vec{\alpha} = \vec{0}$. Now we can break the problem down into two cases: the system has either no solutions or at least one solution. If the system has no solutions, then the system must not have a unique solution. To show the other case, suppose the system has at least one solution $\vec{y}$. Then

$$A\vec{y} = \vec{b} = \vec{b} + \vec{0} = A\vec{y} + A\vec{\alpha} = A(\vec{y} + \vec{\alpha}). \tag{12}$$

$\vec{y} + \vec{\alpha}$ is also a solution to the system. Since both $\vec{y}$ and $\vec{y} + \vec{\alpha}$ are solutions, the system has infinite number of solutions.

For system of linear equations with at least one solution, the above essentially tells us that a system $A\vec{x} = \vec{b}$ has infinite number of solutions if and only if A has linearly dependent columns. Let's think about why this makes sense intuitively. In an experiment, each column in matrix A represents the influence of a variable on the collection of measurements. The fact that we cannot pin down the values of the variables (having infinite number of solutions) is the same as saying that the influence of each variable to the set of measurements cannot be disambiguated.

Example 3.3 (Intuition): Suppose we have a black and white image with two pixels. We cannot directly see the color of each pixel, but we are able to measure how much light the two pixels absorb in total. Could we figure out the color of each pixel? Let's model this as a system of linear equations. Suppose pixel 1 absorbs x_1 units of light and pixel 2 absorbs x_2 units of light. Our measurement indicates that total amount of light absorbed by the image are 10 units of light. Then we could write down the equation,

$$x_1 + x_2 = 10. \tag{13}$$

Written in matrix form, we have

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10 \end{bmatrix}. \tag{14}$$

We see that the columns are $\begin{bmatrix} 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \end{bmatrix}$. The total amount of light absorbed is influenced by 1 unit of x_1 and 1 unit of x_2. However, we cannot pin down the exact influence by x_1 and x_2 because if pixel 1 absorbs c units less, we can just have pixel 2 absorbs c units more. This is connected with the fact that the two columns of linearly dependent — if one pixel absorbs less, it is possible to find a way such that the other pixel absorbs more to make up for the loss (the column of that pixel can be written as a linear combination of the columns of the other pixels).

The result has important implications to the design of engineering experiments. Often times, we can't directly measure the values of the variables we're interested in. However, we can measure the total weighted contribution of each variable. The hope is that we can fully recover each variable by taking several of such measurements. Now we ask the questions: "What is the minimum number of measurements we need to fully recover the solution?" and "How do we design our experiment so that we can fully recover our solution with the minimum number of measurements?" Take tomography for example. We are confident that we can figure out the configuration of the stack when the columns of the lighting pattern matrix A in $A\vec{x} = \vec{b}$ are linearly independent. On the other hand, if the columns of the lighting pattern matrix are linearly dependent, we know that we don't yet have enough information to figure out the configuration. Checking whether the columns are linearly independent essentially gives us a way to validate whether we've effectively designed our experiment.

Now let's look at how the rows of A in the system of linear equations $A\vec{x} = \vec{b}$ relate to its solution.
(3)If the system of linear equations $A\vec{x} = \vec{b}$ has infinite number of solutions and the number of rows of A is greater or equal to the number of columns (A is a square or a tall matrix), then the rows of A are linearly dependent.
Let's see why this is true. Since $A\vec{x} = \vec{b}$ has infinite number of solutions and A is a square or tall matrix, we must end up with at least one row of zeros after running Gaussian Elimination. What does it mean if we end up with a row of zeros when we run Gaussian Elimination? At least one of the rows can be written as a linear combination of the rest of the rows (because row operations correspond exactly to computing linear combinations of the rows). Thus, the rows of A must be linearly dependent.
Now let's think about this intuitively. In an experiment, each row of A represents a measurement. If the number of measurements taken is at least the number of variables and we still cannot completely determine the variables, then at least one of our measurements must be redundant (it doesn't give us any new information).

In summary, looking from the perspective of columns, linear dependency gives us a way to characterize ambiguity in the influence of each variable to our collection of measurements; looking from the perspective of rows, linear dependency gives us a way to characterize redundancy in the measurements.