EECS 16A Designing Information Devices and Systems I
Fall 2023 Lecture Notes Note 4

Mathematical Thinking & Derivation

For many students EECS 16A might be the first time you are asked to "prove" an idea or a concept. This note tries to explain the main ideas behind proofs and give you some tips on how to approach them.

A mathematical proof provides a means for guaranteeing that a statement is true. So what is a proof? A proof is a finite sequence of steps, called logical deductions, which establishes the truth of a desired statement. In particular, the power of a proof lies in the fact that using limited (finite) means, we can guarantee the truth of a statement with infinitely many cases.

More specifically, a proof is typically structured as follows. Recall that there are certain statements, called axioms or postulates, that we accept without proof (we have to start somewhere). Starting from these axioms, a proof consists of a sequence of logical deductions: Simple steps that apply the rules of logic. This results in a sequence of statements where each successive statement is necessarily true if the previous statements were true. This property is enforced by the rules of logic: Each statement follows from the previous statements. These rules of logic are a formal distillation of laws that were thought to underlie human thinking.

In this note, we are going to guide you through the process of developing proofs with a few examples. In particular, we aim to demonstrate the thought process of turning the problem statement into mathematical form and deriving successive mathematically rigorous statements that leads to the desired result.

When we encounter a proof problem, we generally try to understand the problem by asking the following questions:

- "What are the things we can assume based on the problem statement?"
- "What is it that we would like to show?"

The answer to the first question gives us the condition that we are working under and the answer to the second question gives us a clear picture of our goal. Then, we ask the question:

- "How can we utilize what we know to get to what we would like to show under the specified condition?"

To write a proof, the following steps are useful in guiding your thought process and helping you understand the fundamental ideas of the problem.

1. **Carefully read the statement.** Check to see what the direction of the implication is, are you being asked to assume P is true and prove Q is true ($P \implies Q$)? Or are you being asked to assume Q is true and prove P is true ($Q \implies P$)?

2. **Write out what you know from the statement of the theorem.** Say you are asked to prove something like "If P is true, then Q is true." Here, "P is true" is what is given to you/what is known. If the theorem statement has this written out in word, write it out in mathematical notation. Be explicit. Try to simplify any complex notation or jargon.

3. **Write out what you want to prove.** In the example above, what you want to prove is "Q" is true. Again, this might be given to you in words, make sure you translate it into equations. Simplify. Make sure you write this on the paper, even though it seems trivial, it can often help. If there are different ways of writing this – write out both. For example, if you want to prove a set of vectors is linearly dependent, you have two definitions of linear dependance you can work with. Write out both and move forward with whichever definition is more helpful in the context of the proof.

4. **Observe the two statements you just wrote down for what is known and what you want to prove. Find similarities.** Take note of these. It might help to also write them down. How might one of the expressions you wrote be made to look like the other?

5. **Try out a simple example.** Now that you have written out what is given and what you want to prove, try thinking about a simpler version. For example, if you are asked to prove something for an $n \times n$ matrix, first try writing out the same statements for a 2×2 matrix. Check again for similarities between what is known and what you want to prove. See if you can prove the theorem for the simpler case of 2×2 then try generalizing this to an arbitrary $n \times n$ matrix.

   In the process of trying out an example, you may notice a pattern in your working that extends to the general proof. Don't be afraid to try several examples, if just one isn't giving you enough intuition. Trying out examples is also a great way of ensuring that you understand what you want to prove in concrete terms, rather than just as an abstract claim.

6. **Manipulate both sides of the claim. JUSTIFY each step.** After coming up with an algebraic representation of the desired claim, try to manipulate both what you are *given* and what you are *trying* to prove, to see if you can simplify the desired claim. Often, it helps to get rid of complex notation as part of this process - for instance, if you see a summation expressed using $\sum$ notation, it might help to write it out explicitly to get a better understanding of what exactly is being summed.

   Still, at this stage it is important to ensure that your manipulations are valid - it's no use making an amazing simplification if it turns out to be wrong! It can never hurt to break down complex steps into multiple smaller ones, and always ask yourself what conditions are needed for each step to be true. For instance, if you are dividing two quantities, then you should ensure that you are not dividing by zero!

7. **Different approaches.** There are different ways to prove a statement. You might encounter some of the following types of proofs.

   - *Direct proofs.* This style of proof directly shows what is to be proven from what is known by doing a series of mathematical and logical steps.
   - *Constructive proofs.* Essentially, when you are asked to prove that a certain object "exists", you can prove the statement by explicitly constructing the object.
   - *Proof by contradiction.* Another common technique when asked to prove a claim is to try and show that it is impossible for the claim to *not* be true. This technique tends to be useful when the negation of the desired claim is easier to express algebraically than the claim itself.

   No one approach is better or more powerful than any other approach — there is no hierarchy. When you are thinking about proofs try out different approaches to see which one might work.

Solving proof problems are very similar to solving design problems, which is one of the focuses of this course. In a proof, there is something we want to show, and in a design problem, there are specifications we want our design to meet. Both cases are open ended, and we frequently have to integrate many ideas and explore several different possibilities to reach a solution. Think of the different proof techniques we have introduced as a set of tools in a toolbox that can be used in different ways to solve different problems.

The key to getting better at doing proofs is by doing a lot of proofs! For the rest of this Note, we'll look at some examples based on the concepts introduced in Note 3.

Example 4.1 (Constructive Proof): Prove that $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\} = \mathbb{R}^2$

First, let's figure out exactly what the question is asking us to prove. The span of two vectors is a set of all vectors that can be formed as a linear combination of those two vectors. $\mathbb{R}^2$ is the set of all (Cartesian) vectors with two real components.

We want to show that these two sets are equal, meaning that no element can be in one set but not the other. Thus, we need to show

1. That every vector in the given span is inside $\mathbb{R}^2$, and that
2. Every vector in $\mathbb{R}^2$ is in the given span.

Let's prove each of these statements in order. First, how do we show that every vector in $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$ is contained within $\mathbb{R}^2$? Let's express what we are given algebraically. Consider an arbitrary vector $\vec{u} \in \text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$. By the definition of span and linear combinations, we know that we can write

$$\vec{u} = \alpha \begin{bmatrix}1\\1\end{bmatrix} + \beta \begin{bmatrix}1\\-1\end{bmatrix}$$

for some real scalar coefficients α and β. By the rules of vector arithmetic, we can multiply in our scalar coefficients and simplify to obtain

$$\vec{u} = \begin{bmatrix}\alpha+\beta\\ \alpha-\beta\end{bmatrix}.$$

Why is this vector in $\mathbb{R}^2$? Well, since α and β are both real numbers, their sums and differences are both also real, as real numbers are closed under addition. And $\vec{u}$ clearly has exactly two components. Thus, $\vec{u} \in \mathbb{R}^2$, for arbitrary real scalars α and β. Since $\vec{u}$ could have been *any* vector in $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$, we have shown that *any* vector in the given span is inside $\mathbb{R}^2$, which is the first thing we wanted to show.

But we're not done yet! All we've done is shown that $\text{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\} \subseteq \mathbb{R}^2$ (i.e. the given span is a subset of the set of points on the real 2D plane) - there might still exist a vector outside of the given span that is still within $\mathbb{R}^2$! Let's now try to show that *every* vector in $\mathbb{R}^2$ lies in the given span, by writing any vector in $\mathbb{R}^2$ as a linear combination of the two given vectors.

We'll start with an arbitrary vector $\vec{u} \in \mathbb{R}^2$. By definition, $\vec{u}$ has two real components, so it can be written as

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where u_1 and u_2 are both real scalars. We want to know if $\vec{u}$ can be written as a linear combination of the two given vectors, and so must ask whether we can find real scalars α and β (that will depend on u_1 and u_2) such that

$$\alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}?$$

Manipulating the above equation, we obtain the linear system

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

We want to determine whether there will always exist real solutions α and β to the above linear system, no matter what u_1 and u_2 are. And we know how to do this - Gaussian elimination!

Rewriting the above as an augmented matrix and performing row operations without comment, we obtain

$$\begin{aligned}
&\left[\begin{array}{cc|c} 1 & 1 & u_1 \\ 1 & -1 & u_2 \end{array}\right] \\
\iff &\left[\begin{array}{cc|c} 1 & 1 & u_1 \\ 0 & -2 & u_2 - u_1 \end{array}\right] \\
\iff &\left[\begin{array}{cc|c} 1 & 1 & u_1 \\ 0 & 1 & \frac{u_1 - u_2}{2} \end{array}\right] \\
\iff &\left[\begin{array}{cc|c} 1 & 0 & \frac{u_1 + u_2}{2} \\ 0 & 1 & \frac{u_1 - u_2}{2} \end{array}\right].
\end{aligned}$$

Rewriting our end result in matrix-vector form once again, we obtain

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{u_1 + u_2}{2} \\ \frac{u_1 - u_2}{2} \end{bmatrix},$$

which can be expanded and rearranged as

$$\begin{aligned}
\alpha &= \frac{u_1 + u_2}{2} \\
\beta &= \frac{u_1 - u_2}{2}.
\end{aligned}$$

What have we shown? Since we proved in an earlier note that Gaussian elimination cannot introduce spurious solutions, we've shown that for all u_1 and u_2, we can produce coefficients α and β such that

$$\alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \vec{u}.$$

So all $\vec{u} \in \mathbb{R}^2$ can be written as a linear combination of our two vectors! Thus,

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \supseteq \mathbb{R}^2,$$

meaning that the set of points on the real 2D plane is also a subset of the given span, which when combined with our earlier result proves that the two sets are equal, as desired! This completes our proof. □[1]

Notice that the second half of this proof was a "constructive proof" since, in order to show that the coefficients α and β existed for all $\vec{u}$, we were actually able to construct them explicitly!

Example 4.2 (Definition equivalence): Recall that in Note 3 we gave two definitions of linear dependence (repeated below) that we claimed were equivalent. Here, we will formally prove that they are.

(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.

(II) A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is **linearly dependent** if there exist scalars $\alpha_1, \ldots, \alpha_n$ and an index i such that $\vec{v}_i = \sum_{j \neq i} \alpha_j \vec{v}_j$. In words, a set of vectors is linearly dependent if any one of the vectors could be written as a linear combination of the rest of the vectors.

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 scalars $\alpha_1, \ldots, \alpha_n$ and an index i 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:

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

We see that if we set $\alpha_i = -1$, we can make the right hand side look like the form given in definition (I):

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

Since $\alpha_i = -1$, at least one of the α_j terms is not zero, and the condition in definition (I) is satisfied.

(ii) Now let's show the reverse — that definition (I) implies definition (II). First, suppose the condition in definition (I) is true. Then, there exist scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n = \vec{0}, \text{ and not all } \alpha_i\text{'s are equal to zero.} \tag{3}$$

Let's assume that α_1 is one of the nonzero ones (since we can always reorder terms in the summation because addition is commutative) [2]. 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

$$\vec{v}_1 = \sum_{j \neq 1} \left(\frac{\alpha_j}{\alpha_1} \right) \vec{v}_j. \tag{4}$$

[1] Note that this small empty box is used to (satisfyingly) denote the end of a proof.

[2] Notice that we could have also chosen $\alpha_2 \neq 0$, $\alpha_3 \neq 0$, 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, you may see *"Without loss of generality (W.L.O.G.), we let* $\alpha_1 \neq 0$*."*

We see that we have constructed a linear combination exactly in the form of the second definition, completing the second half of the proof. With these two directions of proof, we have now proven that the two definitions are equivalent. □

Next, we will prove Theorems 3.1 and 3.2 from Note 3 about the connection between linear dependence and the number of solutions to a system of linear equations.

Example 4.3 (Theorem 3.1): If the system of linear equations $A\vec{x} = \vec{b}$ has an infinite number of solutions, then the columns of A are linearly dependent.

If the system has infinite number of solutions, it must have at least two distinct solutions. Let's call them $\vec{x_1}$ and $\vec{x_2}$. (Note that $\vec{x_1}, \vec{x_2}$ are full vectors, not vector elements.) Then $\vec{x_1}$ and $\vec{x_2}$ must satisfy

$$A\vec{x_1} = \vec{b} \tag{5}$$

$$A\vec{x_2} = \vec{b}. \tag{6}$$

Subtracting the first equation from the second equation, we have $A(\vec{x_2} - \vec{x_1}) = \vec{0}$. Let $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \vec{x_2} - \vec{x_1}$. Because $\vec{x_1}$ and $\vec{x_2}$ 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, $A\vec{\alpha} = \sum_{i=1}^{n} \alpha_i \vec{a_i} = \vec{0}$ (see below property of matrix multiplication). By definition (I) of linear dependence, the columns of A are linearly dependent. □

Note that in this proof, we used the property of matrix multiplication that $A\vec{\alpha} = \sum_{i=1}^{n} \alpha_i \vec{a_i}$. We scale each column and add them together. In other words, matrix-vector multiplication is a linear combination of columns. This property is often a useful way to think about matrix multiplication. The following example might help:

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \alpha_1 a_1 + \alpha_2 b_1 + \alpha_3 c_1 \\ \alpha_1 a_2 + \alpha_2 b_2 + \alpha_3 c_2 \\ \alpha_1 a_3 + \alpha_2 b_3 + \alpha_3 c_3 \end{bmatrix} = \alpha_1 \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \alpha_2 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \alpha_3 \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Example 4.4 (Theorem 3.2 (Proof by contradiction)): 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.

If a theorem statement asks you to prove that something does not exist (e.g. the inverse of a matrix does not exist, or a unique solution does not exist etc.) it is often useful to consider a proof by contradiction. Assume that the thing does exists, and then show that making that assumption leads to some contradiction, therefore proving that it cannot exist in the first place.

Let's walk through this proof step by step: we'll start by assuming we have a matrix A with linearly dependent columns, and then we will show that this means that the system does not have a unique solution.

Since we are interested in the columns of A, let's start by explicitly defining the columns of A:

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \ldots & \vec{a}_n \\ | & | & & | \end{bmatrix},$$

Now, first, let us write out what is known to us.

By the definition of linear dependence, there exist scalars $\alpha_1, \ldots, \alpha_n$ such that $\alpha_1\vec{a}_1 + \ldots + \alpha_n\vec{a}_n = \vec{0}$ where not all of the α_i's are zero. We can put these α_i's in a vector

$$\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

and by the definition of matrix-vector multiplication, we can compactly write the expression above:

$$A\vec{\alpha} = \vec{0}$$

where $\vec{\alpha} \neq \vec{0}$.

What do we want to prove?

We are trying to show that the system of equations $A\vec{x} = \vec{b}$ does not have a unique solution. Let us assume the opposite, that we do have a unique solution, and then arrive at a contradiction.

Let's call our unique solution $\vec{x_0}$:

$$\begin{aligned} A\vec{x_0} &= \vec{b} \\ A\vec{x_0} + \vec{0} &= \vec{b} \\ A\vec{x_0} + A\vec{\alpha} &= \vec{b} \\ A(\vec{x_0} + \vec{\alpha}) &= b \end{aligned}$$

Therefore, $\vec{x_0} + \vec{\alpha}$ is also a solution to the system of equations! Since $\vec{\alpha} \neq \vec{0}$, we know that $\vec{x_0} \neq \vec{x_0} + \vec{\alpha}$. Therefore, $\vec{x_0}$ cannot be a unique solution and we have reached a contradiction! □

Note that we can add any multiple of $\vec{\alpha}$ to $\vec{x}$ and it will still be a solution – therefore, if there is at least one solution to the system and the columns of A are linearly dependent, then there are infinite solutions.

Example 4.5 (Constructive proof): 1. Let $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ be a set of linearly dependent vectors in $\mathbb{R}^n$. Take any matrix $A \in \mathbb{R}^{m \times n}$. Prove that the set of vectors $\{A\vec{v_1}, A\vec{v_2}, \ldots, A\vec{v_n}\}$ is linearly dependent.

Proof: i). What do we know? Based on the problem statement, we know that $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ is a set of linearly dependent vectors. How do we translate this into mathematical form? Recall one of the two definitions of linear dependence we introduced in the previous note – the set of vectors $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ is linearly dependent if there exists an index i and scalars α_j's such that

$$\vec{v_i} = \sum_{j \neq i} \alpha_j \vec{v_j}. \tag{7}$$

ii). What would we like to show? We would like to show that the set of vectors $\{A\vec{v_1}, A\vec{v_2}, \ldots, A\vec{v_n}\}$ is linearly dependent. Again using the definition of linear dependence, we can translate it into a mathematical statement – we would like to show that there exist index k and scalars β_l's such that

$$A\vec{v}_k = \sum_{l \neq k} \beta_l \left(A\vec{v}_l\right). \tag{8}$$

iii). Now, how do we use what we know mathematically from (1) to prove the mathematical statement in (2)? We somehow would like to get vectors of the form $A\vec{v}$. How could we do that? Let's multiply both sides of equation (7) by the matrix A:

$$A\vec{v}_i = A\left(\sum_{j \neq i} \alpha_j \vec{v}_j\right). \tag{9}$$

By distributivity of matrix-vector multiplication, we know that

$$A\left(\sum_{j \neq i} \alpha_j \vec{v}_j\right) = \sum_{j \neq i} A\left(\alpha_j \vec{v}_j\right) = \sum_{j \neq i} \alpha_j \left(A\vec{v}_j\right). \tag{10}$$

Now, we have that

$$A\vec{v}_i = \sum_{j \neq i} \alpha_j \left(A\vec{v}_j\right), \tag{11}$$

which is in exactly the mathematical form we would like to show in (8). So what are the values of the β's we should choose in (8)? We have $\beta_l = \alpha_l$ for all l.

Hence, we have completed our proof by explicitly finding a linear combination of the columns of matrix A that gives another column of the matrix.

Example 4.6 (Direct proof): If $\vec{v}_1$**,** $\vec{v}_2$**, and** $\vec{v}_1 + \vec{v}_2$ **are all solutions to the system of linear equation** $A\vec{x} = \vec{b}$**, prove that** $\vec{b}$ **must be the zero vector.**
Proof: What does it mean for $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_1 + \vec{v}_2$ to be the solutions to $A\vec{x} = \vec{b}$? It means these vectors must satisfy the following equations:

$$A\vec{v}_1 = \vec{b} \tag{12}$$
$$A\vec{v}_2 = \vec{b} \tag{13}$$
$$A\left(\vec{v}_1 + \vec{v}_2\right) = \vec{b} \tag{14}$$

Notice that using distributivity of matrix-vector multiplication, equation (8) can be rewritten as

$$A\vec{v}_1 + A\vec{v}_2 = \vec{b}. \tag{15}$$

Now from equation (6) and (7), we can substitute $A\vec{v}_1$ and $A\vec{v}_2$ with the vector $\vec{b}$, which leads us to

$$\vec{b} + \vec{b} = \vec{b}. \tag{16}$$

Subtracting $\vec{b}$ from both sides of the equation above, we have

$$\vec{b} = \vec{0}. \tag{17}$$

Hence $\vec{b}$ is the zero vector, as desired. □

This note has covered the major proof techniques that will be useful and relevant for this course, along with many examples. It also serves as an introduction to the more complex proof techniques that you will encounter in CS 70.