

Introduction

In this note(s), I will work with a matrix A with m rows and n columns. A can be viewed as a function that maps a vector, \vec{v} , in \mathbb{R}^n to $A\vec{v}$ in \mathbb{R}^m . I have two, closely related, goals for this note concerning this matrix A . The first is to prove the Rank Nullity Theorem, $N = \text{Dim}(\text{Range}(A)) + \text{Dim}(\text{Nullspace}(A))$. And the second is to discuss decomposing the domain of A (viewed as a function), \mathbb{R}^n , as the direct sum of the nullspace of A and any complementary subspace to it in \mathbb{R}^n . Note that for the second goal, one such complementary subspace to the nullspace is the rowspace, its orthogonal complement in \mathbb{R}^n .

Poor Man's Method

When I ponder about the Rank Nullity Theorem, I often end up arriving at this constructive approach I've dubbed the Poor Man's Method. In the next paragraph, I'll provide a high-level overview into this construction, and then I will explore an example to illustrate this idea more concretely.

Say the matrix A I am working with has a rank of k , that is, $k = \text{Dim}(\text{Range}(A))$. Then k column vectors of A form a basis for the range of A , and the remaining $n - k$ column vectors are redundant. Let me say that the first k columns of A are linearly independent and span the entire range of A . My Poor Man's idea, then, is to manually create a basis for \mathbb{R}^n with two classes of basis vectors that correspond to the two classes of column vectors of A , the first k columns and the rest. The first class are k standard-basis-like vectors that correspond to the first k columns. Particularly, where the i th such vector corresponds to the i th column vector in that the i th basis vector of \mathbb{R}^n contains a 1 at the i th entry and zeroes everywhere else. Somewhat similarly, the second class are $n - k$ vectors that correspond to the last $n - k$ redundant columns. Like the first class, a vector of this class has a 1 corresponding to a given redundant column. But rather than all other entries being zero, the top k entries will have some non-zero entry(ies) that capture the redundancy of the corresponding column vector: so the column vector is redundant in that it can be expressed as a linear combination of the first k columns with some weights and thus the top k entries would be those same weights (but negative). These $n - k$ second class basis vectors are in linearly independent and are all in the nullspace of A by design, and the k first class basis vectors are not in the nullspace of A (that would violate the presumption of the first k columns of A forming a linearly independent set- read the Appendix if this is unclear). Together these two classes of basis vectors are n in number and form a linearly independent set, confirming their designation as a basis of \mathbb{R}^n , so it must be that the $n - k$ second class basis vectors form a basis for the nullspace of A and thus the Rank Nullity Dimensions add up: $n = k + (n - k)$.

Let me start with two vectors, \vec{a}_1 and \vec{a}_2 in \mathbb{R}^3 that are non-zero vectors and are not scalar multiples of each other. Then I can concatenate them to form the 3 by 2 matrix $A = (a_1 \ a_2)$. The range of A has a dimension of 2 and visually, it is the plane that is $\text{span}(\vec{a}_1, \vec{a}_2)$. And the nullspace of A is empty except for the trivial zero vector, so the dimension of the nullspace is 0 (See the Appendix if you are not comfortable with this). So Rank Nullity holds as $n = 2 = \text{Dim}(\text{Range}(A))$.

The Proper Method

Visual Remarks

You are an explorer in the world \mathbb{R}^n Let me take the familiar case where I am working with a matrix A with first two columns independent and the third column a linear combination of the first two. The domain is thus \mathbb{R}^3 , which you are exploring. You are at some point in this domain. If you traverse along the nullspace, so along n , that is an infinite line in a particular direction, you get no

reward. So you pick some direction \vec{v}_1 where \vec{v}_1 is a unit vector and you take one step in that direction. You obtain the reward vector $A\vec{v}_1$. You take k steps and obtain the reward vector $kA\vec{v}_1$. So this line you walk along maps to a line in the reward space, or the range of A .

You note that for some $kA\vec{v}_1$ reward, you can actually reach it in an infinite number of ways: take k steps along \vec{v}_1 and take any number of steps along \vec{n} . (Side remark: a key idea in proof, to me, is that taking steps through along the nullspace is not just a way, but in fact it is the only way to generate duplicate rewards). So in fact now you have a plane, $\text{span}(\vec{v}_1, \vec{n})$, which you can traverse to get all rewards of the form, $kA\vec{v}_1$, or a line in the reward space.

This plane is 2-dimensional and you are in a 3-dimensional domain, and the reward space, or the range, is 2-dimensional but, confined to the exploring the input plane of $\text{span}(\vec{v}_1, \vec{n})$, you can only reach one dimension in the reward space, $A\vec{v}_1$, but you know there is one more dimension to the reward space you have not explored. And you know there exists a direction that is outside the plane that you have not explored

So as an explorer who wants to explore, you repeat the procedure and take some unit vector direction, \vec{v}_2 , that is outside this plane. Before actually stepping along this direction you ask yourself if this will let you get some, non-zero, reward that is outside $\text{span}(A\vec{v}_1)$? Note that with this choice of \vec{v}_2 , the set of vectors $\{\vec{n}, \vec{v}_1, \vec{v}_2\}$ form a basis for your input domain, \mathbb{R}^3 . Again, the reward space, or the range, is 2-dimensional and traversing the plane only takes you along one line in the reward space. So there must exist some “interesting” direction you can take to get a reward outside of this line, or else the dimension of the range would be 1 and not 2. Whatever this heretofore unexplored “interesting” direction must be, as $\{\vec{n}, \vec{v}_1, \vec{v}_2\}$ form a basis for all the directions you can explore, this direction must be able to be represented as a linear combination of that basis or some (uninteresting) steps along the plane and some (interesting) steps along \vec{v}_2 . So \vec{v}_2 itself is an “interesting” direction as well, as traversing it must then let you reach a reward outside $\text{span}(A\vec{v}_1)$!

Appendix: Linear Independence, Dependence, Redundancy, Nullspace

Let me start by reviewing the definition(s) of linear independence and dependence.

A set of vectors $\{a_1, a_2, \dots, a_n\}$ is linearly independent if the only way to form the zero vector, $\vec{0}$, by taking a linear combination of the them is when all the weights are 0. That is, $x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n = \vec{0}$ only when all the x_i are themselves 0. All the x_i being 0 is known as the trivial solution to this equation. Note, the above equation can be written in matrix form as $A\vec{x} = \vec{0}$ where $A =$

$(a_1 \ a_2 \ \dots \ a_n)$ and $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Then, if the only solution to this equation is $x = \vec{0}$, $\{a_1, a_2, \dots, a_n\}$ is

linearly independent. And matrix A has a trivial nullspace only containing the zero vector.

And if a set of vectors is not linearly independent, then it is linearly dependent. This means that there exists a nontrivial solution to $x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n = \vec{0}$.

An upshot of this is that at least one of the a_i can be expressed as a linear combination of the remaining vectors. To see this for a set of linearly independent vectors, consider the nontrivial

solution, $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. At least one of the x_i is not 0 so from $x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n = \vec{0}$,

$$x_i\vec{a}_i = \sum_{j \neq i} -x_j\vec{a}_j$$

$$\vec{a}_i = \sum_{j \neq i} -\left(\frac{x_j}{x_i}\right) \vec{a}_j$$

Observe that this upshot means that every vector in a linear independent set cannot be expressed as a linear combination of the other vectors. Visually, it “juts out” of the span of the other vectors.

Another upshot is that the matrix A , again formed by concatenating the a_i , has a nontrivial nullspace. Since $A\vec{x} = \vec{0}$ for a non-trivial \vec{x} , the nullspace of A contains that non-zero \vec{x} at the very least. Moreover, the nullspace contains all scalar multiples of \vec{x} as well, $\text{span}(\vec{x})$ or $k\vec{x}$. I can show this by taking the equation, $x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n = \vec{0}$, and multiplying both sides by k to get $kx_1\vec{a}_1 + kx_2\vec{a}_2 + \dots + kx_n\vec{a}_n = \vec{0}$. $k\vec{0} = \vec{0}$ and thus $k\vec{x}$ also satisfies the equation $A(k\vec{x}) = \vec{0}$.

Linear Independence And Uniqueness

I will show that $A\vec{x} = \vec{b}$ has exactly one solution when the columns of A form a linearly independent set. If $\vec{b} = \vec{0}$, this is evident by the definition of linear independence, as the solution is $\vec{x} = \vec{0}$.

Suppose for contradiction that $A\vec{x} = \vec{b}$ has two solutions, \vec{u} and \vec{v} where $\vec{u} \neq \vec{v}$, then:

$$A\vec{u} = A\vec{v} \Rightarrow A\vec{u} - A\vec{v} = \vec{0} \Rightarrow A(\vec{u} - \vec{v}) = \vec{0}$$

But this means I have found a non-trivial vector in the nullspace of A , $\vec{u} - \vec{v}$. Which means the columns of A did not form a linearly independent set to begin with.

Now this is a very terse proof that I would like to delve into a bit more and provide some visual intuition for. This might get a bit intimidating, but I assure the reader that I will add concrete and simple examples to provide solid intuition. So don't worry if this and the next paragraph are hard to immediately understand. The key step to try to understand is, by linearity, $A\vec{u} - A\vec{v} = A(\vec{u} - \vec{v})$. \vec{u} and \vec{v} are weights to the columns of A and for a particular column, a_i , its weight will be the i th entry of \vec{u} minus the i th entry of \vec{v} . Visually, $A\vec{u} - A\vec{v}$ is an offset vector that, when added to $A\vec{v}$ using the “head to tail” vector addition method, restores the vector $A\vec{u}$. Each column of A visually, is an axis, and $A\vec{v}$ means start the i th entry units of \vec{v} along the a_i axis. Then to get to $A\vec{u}$, along every axis a_i , I need to add the i th entry of \vec{u} minus the i th entry of \vec{v} units along a_i to arrive at where $A\vec{u}$ rests along that axis. And as this must be done over all the axes that are the columns of A , the offset vector is $A(\vec{u} - \vec{v})$.

And the approach is to set the offset equal to $\vec{0}$. As $\vec{u} \neq \vec{v}$, at least one of the entries of the offset vector, $A(\vec{u} - \vec{v})$, is non-zero. Let's say the i th entry of the offset vector is non-zero. So the offset to $A\vec{v}$, which must be $\vec{0}$, is some non-zero along a_i . But this means for the offset to be $\vec{0}$, there must be contributions along the remaining axes that cancel out this non-zero contribution along a_i . But this would imply that the axis a_i is redundant and falls along the span of the other axes.

Now I can only visualize things in 2 or at most 3 dimensions. So, for a simple and concrete toy example, say I have two vectors in \mathbb{R}^2 , \vec{a}_1 and \vec{a}_2 , that are linearly independent. Let $\vec{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and let $\vec{a}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. I intentionally set \vec{a}_1 to be on the x-axis, to show that \vec{a}_2 has some y-component “jutting out” that is orthogonal to the x-axis. As \vec{a}_1 and \vec{a}_2 are two linearly independent vectors in \mathbb{R}^2 , together they span all of \mathbb{R}^2 . So consider a general vector, $\begin{pmatrix} x \\ y \end{pmatrix}$. I want to show how it is uniquely specified by some vector $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ as weights for a linear combination of the a_i , so where $\begin{pmatrix} x \\ y \end{pmatrix} = x_1\vec{a}_1 + x_2\vec{a}_2$.

Due to the choice of \vec{a}_1 being on the x-axis, this is quite easy. I will need x_2 to be y because only a_2 has a y component. Note that x_2 must be exactly y because any more and this overshoots vertically,

and there is no way by traversing any distance along a_1 , the x-axis, horizontally to undo the overshooting. Likewise, if x_2 is less than y , there is no way when strictly confined to movement along the x-axis to make up the deficit along the y-axis. Plugging $x_2 = y$ into $\begin{pmatrix} x \\ y \end{pmatrix} = x_1 a_1 + x_2 a_2$,

$$\begin{pmatrix} x \\ y \end{pmatrix} = x_1 a_1 + y(11)$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = x_1 a_1 + (yy)$$

$$\begin{pmatrix} x - y \\ 0 \end{pmatrix} = x_1 a_1$$

$$\begin{pmatrix} x - y \\ 0 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x - y \\ 0 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$

$$x_1 = x - y$$

That is, $x_2 = y$ accounts for the y component, but x_2 of a_2 also contributes y along the x-axis, thus only $x - y$ contribution along the x-axis is needed so $x_1 = x - y$.

And all of this is forced. Let me extend this to 3 dimensions. Say I have a \vec{a}_1 , \vec{a}_2 , and \vec{a}_3 that form a linearly independent set where \vec{a}_1 and \vec{a}_2 lay on, and span, the x-y plane, and so \vec{a}_3 has some z-component that juts out. This similarly forces x_3 to account for the z-component, and now I have reduced the problem to a subproblem but with 2 vectors, \vec{a}_1 and \vec{a}_2 and 2 weights, x_1 and x_2 .

Back to the \mathbb{R}^2 case, where I have some \vec{a}_1 and \vec{a}_2 that span all of \mathbb{R}^2 , but I want to be general. Sure I could rotate the system and treat \vec{a}_1 as the x-axis. But let me not and instead try to invoke classic “head to tail” vector addition. Say I have the vector $x_1 \vec{a}_1 + x_2 \vec{a}_2$ that is a linear combination of the basis with weights $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. I want to show that there cannot be a different vector, $\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$ such that $x_1 \vec{a}_1 + x_2 \vec{a}_2 = x'_1 \vec{a}_1 + x'_2 \vec{a}_2$. As $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$, at least one of the pairwise entries must be different. Without loss of generality, say $x_1 \neq x'_1$. And again, without loss of generality let $x_1 > x'_1$ (if $x_1 < x'_1$, swap the roles of $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$). Now $x_1 \vec{a}_1 + x_2 \vec{a}_2$ can be viewed as x_1 units along the axis a_1 and then x_2 units along the a_2 axis. That is using $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ as weights. But now consider $\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$ as weights, where $x'_1 < x_1$ and x'_2 is unknown. First, move x'_1 along the a_1 axis. Observe that there is a deficit of $x_1 - x'_1$ units along the a_1 axis. The question is, what should x'_2 be? If I set to be x_2 , the deficit remains unaccounted for. No matter what, movement along the a_2 axis cannot account for the $x_1 - x'_1$ unit deficit along the a_1 axis, because of the linear independence: a_1 is not on the a_2 axis. Purely algebraically, rearrange $x_1 \vec{a}_1 + x_2 \vec{a}_2 = x'_1 \vec{a}_1 + x'_2 \vec{a}_2$ to $(x'_1 - x_1) \vec{a}_1 + (x'_2 - x_2) \vec{a}_2 = (x_1 - x'_1) \vec{a}_1$. And by linear independence, I’m done, since no matter what x'_2 is, movement along the a_2 axis cannot restore the deficit along the a_1 axis.