Introduction

OK, so I want to explore something fundamental about matrices:

Why does the rank of the columnspace equal the rank of the rowspace?

I vaguely remember proofs involving row reduction, but those do not satisfy me as they involve this "machinery".

"Machinery" remark: I remember reading Axler's Linear Algebra Done right proof of Rank Nullity theorem feeling so natural. And I saw there was a footnote stating that this proof is more elegant than pivot column proofs using row reduction.

So I want a similar, direct proof. And I believe I've come up with some insight about something I'll dub "preservation". Which ultimately is a very simple property of linearity which unsurprisingly, I say linear algebra comes down to.

TODO: note on intuition/lack of rigor/handwaviness...

Rank 1 Matrices: Scaling Preserves Ratios

Consider the column vector $a_1 = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$. This vector can denote the side lengths of your classic 3-4-5 right triangle. Now scale it by some scalar, say 3. Then we have $a_2 = \begin{pmatrix} 9 \\ 12 \\ 15 \end{pmatrix}$. a_2 also has a geometric interpretation of a 3-4-5 right triangle except scaled by 3. We also have the relation: $a_1 = 3 * a_2$. We can construct the matrix:

$$A = (a_1 \ a_2) = \begin{pmatrix} 3 & 9 \\ 4 & 12 \\ 5 & 15 \end{pmatrix}$$

Note that the rank of the columnspace of A is 1 as the column space is simply span$\{a_1\}$. What's the rank of the row space? Well, also 1 as all the rows are scalar multiples of $\begin{pmatrix} 3 \\ 9 \end{pmatrix}^T$ That is, the second row is 4/3 of it and the third row is 5/3 of it. The key idea is that scaling a_1 by 3 preserves the ratios of its entries and this is at the heart of my article. And geometrically, it makes sense, scaling the triangle does not change the ratios of the side lengths.

TODO: add bonus conditional probability definition as scaling and the ratios of the constituents of P(AandB) and P(notAandB)

Establish (in R^3) properties of linearly independent vectors

This is a detour to establish a crucial result for the remaining section where I attempt to generalize the previous section. I find along the way to prove this, I come across a key, fundamental concept of linear algebra which is wonderful to express simply on it's own right: the relationship of linear dependence and redundancy mandating a non-trivial null space. For example back to the elegant Rank Nullity Theorem proof in Linear Algebra Done Right, this redundancy and null space connection is at the heart of that proof.

3 linearly independent vectors in R^3 span entirely of R3. Intuition, first one span(a_1) line. Second one span(a_1, a_2) plane. Third one span(a_1, a_2, a_3) everything, all of R^3. First one is a line. Second one and together (a_1, a_2) are lin indep so not on the same line so define a plane. And the third one not in the span(a_1, a_2) plane so we get R^3. I know this is handwavy, so TODO express this better. Conclusion: these 3 vectors serve as a basis for R^3.

Unique Coordinate Representation of All Vectors Given Basis

Now to proceed to the key result. Consider the equation $A * \vec{x} = b \forall b \in \mathbb{R}^3$ x denotes a linear combination of the vectors a_1, a_2, a_3 such that they result in (b). The key result is that not only does x always exist but that it is unique. Let me reiterate, the key result is that, given any b in R^3, we can express it as a unique linear combination of a_1, a_2, a_3 where (a_1, a_2, a_3) are linearly independent vectors in R^3.

Generalizing: Linear Combinations Preserving Linear Combinations

For the rest of this article, I will work with the matrix A = [a_1, a_2, a_3]. m by n matrix n = 3 columns and m >= 3 rows. Let's assume this A abides by rank colspace = rank rowspace = 3 And for aesthetic reasons, let the first 3 rows of A be linearly independent and the rest of the rows below be linear combinations of the top 3 rows. So we have:

$$A = (\vec{a_1} \ \vec{a_2} \ \vec{a_3}) = \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,3} & a_{2,3} & a_{3,3} \\ \vdots & \vdots & \vdots \\ a_{1,i} & a_{2,i} & a_{3,i} \\ \vdots & \vdots & \vdots \end{pmatrix}$$

And $\{\vec{a_1}, \vec{a_2}, \vec{a_3}\}$ is linearly independent with rank 3 and it's span is the column space of A, a subspace of RR^M with rank 3. And the top 3 rows: $\begin{pmatrix} a_{1,1} \\ a_{2,1} \\ a_{3,1} \end{pmatrix}^T , \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ a_{3,2} \end{pmatrix}^T , \begin{pmatrix} a_{1,3} \\ a_{2,3} \\ a_{3,3} \end{pmatrix}^T$ are likewise linearly independent with rank 3 and form a basis for the row space of A, which is RR^3. So we have ranks of the column and row spaces both being 3. A consequence of the top 3 rows being a basis for the row space is that all rows below the top 3 are redundant in that they can be expressed as a unique linear combination, $L_{r,i}$, of the top 3 rows as proved in the prior groundworks section. That is, the ith row $\begin{pmatrix} a_{1,i} \\ a_{2,i} \\ a_{3,i} \end{pmatrix}^T$ is uniquely expressed as a linear combination, $L_{r,i}$, of the top 3 rows $\forall m \geq i > 3$. Let's spell out $L_{r,i} = \vec{x}^T = (x_1 \ x_2 \ x_3)$, which is purely notational and allows me to express the ith row as a linear combination of the top 3 rows formally as this left side matrix multiplication:

$$(x_1 \ x_2 \ x_3) * \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,3} & a_{2,3} & a_{3,3} \end{pmatrix} = (a_{1,i} \ a_{2,i} \ a_{3,i})$$

Or to be super clear and to literally spell this out, this means there are 3 equalties:

$$a_{1,i} = x_1 * a_{1,1} + x_2 * a_{1,2} + x_3 * a_{1,3}$$

$$a_{2,i} = x_1 * a_{2,1} + x_2 * a_{2,2} + x_3 * a_{2,3}$$

$$a_{3,i} = x_1 * a_{3,1} + x_2 * a_{3,2} + x_3 * a_{3,3}$$

Adding Redundant Column Vector

As I saw in the first, rank 1 section, appending a scalar multiple column to a_1 kept the rank of the matrix $(a_1 \ a_2) = (a_1 \ 3*a_1) = \begin{pmatrix} 3 & 9 \\ 4 & 12 \\ 5 & 15 \end{pmatrix}$ both column and row wise at 1. Now this subsection explores a more general case of appending a redundant a_4 vector to our $A = (a_1 \ a_2 \ a_3)$, redundant meaning

that I can express a_4 as some linear combination, $L_c = \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$ of a_1, a_2, a_3. That is, $\vec{a_4} = y_1 * \vec{a_1} + y_2 * \vec{a_2} + y_3 * \vec{a_3}$. So adding a_4 to A gives:

$$\begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} & a_{4,1} \\ a_{1,2} & a_{2,2} & a_{3,2} & a_{4,2} \\ a_{1,3} & a_{2,3} & a_{3,3} & a_{4,3} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1,i} & a_{2,i} & a_{3,i} & a_{4,i} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

What affect does appending a_4 have on the rank of the column space and row space? Well, since a_4 is redundant and is in span$\{a_1, a_2, a_3\}$ as specified by $L_c = \vec{y}$, the column space remains unchanged so it's rank remains 3. But what about the rank of the row space? I would aim to prove that it, too, stays rank 3. First off, just purely visually, appending an extra column has the affect of extending each rows by 1. Each of the m rows now has 4 elements insteand of 3. Now the top 3 rows which, as I just saw, form a linear independent set before adding a_4, will continue to form a linear independent set with a rank of 3, except now in $\mathbb{R}^4$. (Why? Simple proof by contradiction on the initial 3 rows not being independent to begin with as if, by adding 1 element to the end of those 3 rows, one of the rows can be written as a linear combination of the other 2 rows, that is, these 3 extended rows are linearly dependent, then we have reached a contradiction as that same linear combination relationship would apply to the initial 3 rows showing that they were not linearly independent to begin with.) The question, then, is whether the remaining rows below the top 3, now with an extra element tacked on at the end, will continue being redundant. That is, can the ith row, $\forall m \geq i > 3$, continue to be expressed as the linear combination, $L_{r,i} = \vec{x}^T$, of the top 3 rows?

We already know these hold from before, reproduced: adding another column does not change this:

$$a_{1,i} = x_1 * a_{1,1} + x_2 * a_{1,2} + x_3 * a_{1,3}$$

$$a_{2,i} = x_1 * a_{2,1} + x_2 * a_{2,2} + x_3 * a_{2,3}$$

$$a_{3,i} = x_1 * a_{3,1} + x_2 * a_{3,2} + x_3 * a_{3,3}$$

For the ith row to continue to respect $L_{r,i}$ all that is left to prove is 1 additional relation:

$$a_{4,i} = x_1 * a_{4,1} + x_2 * a_{4,2} + x_3 * a_{4,3}$$

This can be readily shown purely algebraically. With the relation $\vec{a_4} = y_1 * \vec{a_1} + y_2 * \vec{a_2} + y_3 * \vec{a_3}$ in mind, I will take the three established relations and multiply both sides by y_i (and intentionally "pushing in" the y_i after the x_i) as follows:

$$y_1 * a_{1,i} = x_1 * y_1 * a_{1,1} + x_2 * y_1 * a_{1,2} + x_3 * y_1 * a_{1,3}$$

$$y_2 * a_{2,i} = x_1 * y_2 * a_{2,1} + x_2 * y_2 * a_{2,2} + x_3 * y_2 * a_{2,3}$$

$$y_3 * a_{3,i} = x_1 * y_3 * a_{3,1} + x_2 * y_3 * a_{3,2} + x_3 * y_3 * a_{3,3}$$

And if we add these 3 equations, the total LHS becomes $y_1 * a_{1,i} + y_2 * a_{2,i} + y_3 * a_{3,i}$ which from a_4 being a linear combo $L_c = y$ of a_1, a_2, a_3, this LHS simplifies to $a_{4,i}$. And the total RHS similarly simplifies via factoring out the x_i to $x_1 * a_{4,1} + x_2 * a_{4,2} + x_3 * a_{4,3}$

I just symbolically proved that the a4 column abides by the $L_{r,i}$ row relation across this column. But ok, so that was simply algebra, but I want to present a bit more intuition to hammer this preservation point in. As I feel like something simple, yet perhaps even deep, just happened.

To this end, let me take a step back. If a row r_4 is a linear combination of some other rows x_1*r_1 + x_2*r_2 + x_3*r_3, then if we stack these rows and build a matrix then we can view it both row-wise and column-wise. The rows define columns (and vice versa). The relation, x, can be viewed across every column. A row is a vector and the usual linear combination, linearity, operations of scaling and adding rows, these operations when applied to row vectors or more generally applying these operations to vectors in general do it itemwise in "lockstep" to all the constituent elements. Say we have:

$$
\begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \\ \hline r_4^T \end{pmatrix} = \left(\begin{array}{ccc} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \\ \hline r_{4,1} & r_{4,2} & r_{4,3} \end{array} \right)
$$

Note that the indexing has changed to be row-first, then column, which at least, is how I'm used to denoting 2D arrays in programming. This matrix gives me 3 columns:

$$
\begin{pmatrix} r_{1,1} \\ r_{2,1} \\ r_{3,1} \\ r_{4,1} \end{pmatrix}, \begin{pmatrix} r_{1,2} \\ r_{2,2} \\ r_{3,2} \\ r_{4,2} \end{pmatrix}, \begin{pmatrix} r_{1,3} \\ r_{2,3} \\ r_{3,3} \\ r_{4,3} \end{pmatrix}
$$

The ith column can be written as $\begin{pmatrix} r_{1,i} \\ r_{2,i} \\ r_{3,i} \\ r_{4,i} \end{pmatrix}$ Let me say that the fourth row, r_4, has a "x-flavor" of $\vec{x}$ and this "x-flavor" manifests itself across all the columns. The ith column's bottom entry can be expressed as this flavor $r_{4,i} = x_1 * r_{1,i} + x_2 * r_{2,i} + x_3 * r_{3,i}$. So we can explicitly write out our "x-flavored" ith column as $\begin{pmatrix} r_{1,i} \\ r_{2,i} \\ r_{3,i} \\ x_1 * r_{1,i} + x_2 * r_{2,i} + x_3 * r_{3,i} \end{pmatrix}$

Now for the fun part: linearity. There are 2 principle linear operators over a vectorspace: scaling and addition. I believe, and I may be wrong, that this "x-flavor" defines an "x-flavored" vector space closed under these 2 linear operations. And thus if I append a new column to the matrix that is a linear combination of the column vectors, the resulting column will also be "x-flavored". And by extension, the redundant 4rth r_4 row will continue to be redundant and continue to be the same "x-flavored" linear combination of r_1, r_2, r_3.

This sounds fancy, but the idea is really simple. The first operation is scaling. If we scale our ith column vector by k, we get:

$$
k * \begin{pmatrix} r_{1,i} \\ r_{2,i} \\ r_{3,i} \\ x_1 * r_{1,i} + x_2 * r_{2,i} + x_3 * r_{3,i} \end{pmatrix} = \begin{pmatrix} k * r_{1,i} \\ k * r_{2,i} \\ k * r_{3,i} \\ x_1 * k * r_{1,i} + x_2 * k * r_{2,i} + x_3 * k * r_{3,i} \end{pmatrix}
$$

And this resultant vector is also "x-flavored" or closure under scaling. That is, the bottom entry is a linear combination (dot product) of the upper entries, weighted by $\vec{x}$.

The second operation is closure under addition. That is, the sum of two "x-flavored" vectors is also "x-flavored". I'll borrow a jth column vector and so I have:

$$\begin{pmatrix} r_{1,i} \\ r_{2,i} \\ r_{3,i} \\ x_1 * r_{1,i} + x_2 * r_{2,i} + x_3 * r_{3,i} \end{pmatrix} + \begin{pmatrix} r_{1,j} \\ r_{2,j} \\ r_{3,j} \\ x_1 * r_{1,j} + x_2 * r_{2,j} + x_3 * r_{3,j} \end{pmatrix} = \begin{pmatrix} r_{1,i} + r_{1,j} \\ r_{2,i} + r_{2,j} \\ r_{3,i} + r_{3,j} \\ x_1 * \left(r_{1,i} + r_{1,j} \right) + x_2 * \left(r_{2,i} + r_{2,j} \right) + x_3 * \left(r_{3,i} + r_{3,j} \right) \end{pmatrix}$$

And to belabor the point, the summed vector is again "x-flavored" in that the bottom entry is the sum of its upper entries weighted by $\vec{x}$.

So the preservation point is that the column vector linear operation $L_c = \vec{y}$ used to generate a redundant a_4 weighting the columns a_1, a_2, a_3 by $\vec{y}$ preserves the row vector linear operation $L_{r,i} = x$ for the ith redundant row so it remains redundant in the exact same manner it was redundant weighting the top three rows except now the last column ith entry weights the top three entries of the last column by $\vec{x}$ to preserve the action of $L_{r,i}$. Done!

The other direction

Transpose argument really. A and A^T. I believe my work is essentially done in the previous section, but I'll opine further to further flesh it out.

So we've shown adding a vector a_4 that is a linear combination of a_1, a_2, a_3 mandates that there is no change in the row space's rank, the non-top 3 rows stay redundant linear combinations of the top 3 rows and thus both column space and row space remain rank 3 except the row space is a subspace of RR^4 and not RR^3. The question now is if this if-then relationship is an if-and-only-if one, that is, a_4 being a linear combination of a_1, a_2, a_3 is the only cause for no change in row space. That is, if by adding some a_4 to A, the rank of the row space remains unchanged, then a_4 must be a linear combination of a_1, a_2, a_3. If so, they I'm done. (Why? If a_4 is not a combination of a_1, a_2, a_3 then it can't be that the row space's rank is unchanged and remains rank 3 because there's contradiction on rank of rowspace remaining unchanged mandating a_4 being a combination of a_1, a_2, a_3. So by contradiction, I know adding such an a_4 that increases the rank of the column space by 1 to rank 4 (because a_4 is not in span(a_1, a_2, a_3)) must change the row space's rank. So the question now is how will the row space change? I want to show that it's rank must increase from 3 to 4 to maintain rank(colspace) = rank(rowspace) equivalency. Again, adding a_4 extends all entries of the rows by 1, and this certainly won't cause any of the top 3 rows being linearly independent. (Why? Simple proof by contradiction on the initial 3 rows not being independent to begin with as if, by adding 1 element to the end of those 3 rows, one of the rows can be written as a linear combination of the other 2 rows, then we have reached a contradiction as that same linear combination relationship would apply to the initial 3 rows and they are not linearly independent to begin with) So adding a_4 certainly won't decrease the rank of the rowspace, the top 3 rows will continue to form a linearly independent set. And we know from a prior (only-if) contradiction that adding this a_4 that's not a linear combination of a_1, a_2, a_3 must change the rank of the rowspace. So the only forced option is that the rank of the rowspace has to increase from rank 3 to some higher rank. But the row vectors now extended by 1 entry to the right are in RR^4 so 4 is the highest rank the row space can span, all of RR^4, so from rank 3, we go to rank 4 rowspace. And done.)

Again let me quickly restate the remaining objective. Appending the a_4 to A extends all the existing rows by 1. The top 3 rows will obviously remain independent as I've shown by contradiction. But my aim now is to show that if the remaining lower rows remain the same linear combinations of the top 3, then this mandates a_4 to in turn, be a combination of a_1, a_2, a_3.

Recall my naming of the top 3 entries of a_4: $a_{4,1}, a_{4,2}$ $a_{4,3}$. When I append a_4, note that the column vector $\begin{pmatrix} a_{4,1} \\ a_{4,2} \\ a_{4,3} \end{pmatrix}$ is uniquely expressed as a linear combination, L_c, of the 3 columns of the truncated matrix [a_1, a_2, a_3] where we only take the top 3 rows. That is:

$$\begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,3} & a_{2,3} & a_{3,3} \end{pmatrix} * \vec{x} = \begin{pmatrix} a_{4,1} \\ a_{4,2} \\ a_{4,3} \end{pmatrix}$$

where $\vec{x}$ specifies L_c.

Now back to the full version of A, appended with a_4, reproduced from the prior subsection:

$$\begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} & a_{4,1} \\ a_{1,2} & a_{2,2} & a_{3,2} & a_{4,2} \\ a_{1,3} & a_{2,3} & a_{3,3} & a_{4,3} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1,i} & a_{2,i} & a_{3,i} & a_{4,i} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

So the ith row is redundant, that is, it is some linear combination, $L_{r,i}$ of the first 3 rows. And this redundancy applies for all rows below the top 3 rows, but let me just consider this ith row. The same concept from the prior section applies, and I can identically apply that to show that $a_{4,i}$ is the same linear combination of $a_{1,i}, a_{2,i}, a_{3,i}$ specified by L_c. Again, that is, $\begin{pmatrix} a_{1,i} \\ a_{2,i} \\ a_{3,i} \end{pmatrix}^T * \vec{x} = a_{4,i}$. And this can be done for all the redundant, non-top 3 rows. Thus, these redundant rows force L_c to be respected for all $a_{4,i}$ and that forces a_4 to be a linear combination of (a_1, a_2, a_3) as specified by L_c.

And done, this essentially the same argument as the previous section except transposed.