

Dimensionality

Consider a vector \vec{x} in \mathbb{R}^2 what makes it different from say a vector in \mathbb{R}^5 ? For starters, vectors in \mathbb{R}^5 are longer (contain more parameters) than vectors in \mathbb{R}^2 . Let's build on this. For any vector in \mathbb{R}^2 we would need at least two parameters to uniquely describe any vector in that space, and for any vector in \mathbb{R}^5 we would need exactly five parameters to uniquely define any vector. In this sense we can look at the **dimension** of a space as the fewest amount of parameters needed to describe an element or member of that space. The dimensionality can also be thought of as the degrees of freedom of your space, that is the number of parameters that can be varied when describing a member of that space.

Range & Span

Let's assume A is a matrix in $\mathbb{R}^{n \times m}$. Looking at this matrix as a linear operator that acts on vectors, then we are taking in vectors that live in \mathbb{R}^m (an m -dimensional space) and outputting vectors that live in \mathbb{R}^n (an n -dimensional space). We say that the **range** of an operator is the space of all outputs that the operator can map to. What is the range of our matrix operator? To answer this we write our matrix in terms of its columns,

$$A = \left[\begin{array}{c|c|c|c} | & | & \dots & | \\ \vec{a}_1 & \vec{a}_2 & & \vec{a}_m \\ | & | & & | \end{array} \right], \quad (1)$$

where the vectors \vec{a} 's live in \mathbb{R}^n . The matrix A operates on any vector \vec{x} that live in \mathbb{R}^m , where the operation on \vec{x} is just $A\vec{x}$. As a member of \mathbb{R}^m , \vec{x} can be written as

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}. \quad (2)$$

From this we have

$$A\vec{x} = \left[\begin{array}{c|c|c|c} | & | & \dots & | \\ \vec{a}_1 & \vec{a}_2 & & \vec{a}_m \\ | & | & & | \end{array} \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \sum_{k=1}^m x_k \vec{a}_k, \quad (3)$$

so we can conclude that the range of the operator A is the space of all possible linear combinations of its columns, another name for this is the **span** of (the columns of) A , which we can write as

$$\text{span}(A) = \{ \vec{v} \mid \vec{v} = \sum_{i=1}^m x_i \vec{a}_i, \text{ where } x_i \text{'s are scalars} \} \quad (4)$$

Dimension of the Span

What is the dimensionality of $\text{span}(A)$? A reasonable assumption would be to say n since after all the vectors in our span live in \mathbb{R}^n , and thus $\text{span}(A)$ is certainly a subset of \mathbb{R}^n (meaning it is contained in \mathbb{R}^n). In general the matrix operator A will not be able to output every vector in \mathbb{R}^n , so our span will not be equal to \mathbb{R}^n . The dimension cannot be greater than n , since $\text{span}(A)$ is a subset of \mathbb{R}^n , but it can certainly be less. Say for example that A is made of all zeros, then its output would be zero-dimensional (it can only output $\vec{0}$). In addition, we need to remember that the dimension of a space is determined by the minimum number of parameters that we need to describe a vector in that space. So continuing with the same example if we are dealing with a space that only contains one vector (in this case $\vec{0}$), no parameters are needed to be specified to distinguish that vector from any other vectors in that space since there are none. Hence, the dimension of $\text{span}(A)$ where A is a zero matrix is just zero.

We want to find the dimension of the $\text{span}(A)$, and if we look at the definition in equation (4) we see that we only get to choose m parameters, x_1, x_2, \dots, x_m , so the dimension cannot be greater than m . This is true even if m is less than n . Now you might be asking how can that be the case when the vectors in the $\text{span}(A)$ each have n components? The answer is in defining our span we have constrained the kinds of vectors that can live in our space, and because of this we may not need as many parameters as components to identify the vectors in this space. As an example say I was born in Columbus, Ohio and I tell you that I was born in the city of Columbus, if you are thinking in the space of the United States of America then you could reply "Ohio or Georgia?", because for that space I have not given you enough information to identify my birth place. I still need to give you the state parameter. However, if you are thinking just in the space of Ohio, then saying I was born in the city of Columbus would be sufficient information, and this is because the space has already constrained the state parameter, and thus it no longer needs to be specified. It does not change the fact that I was born in Columbus, Ohio (i.e. the vector is the same) but depending on the space we are thinking in I do not need as much information to convey this (i.e. how the vector is represented is different).

Given our discussion thus far, we might be tempted to say that the span is $\min(m, n)$, the minimum between m and n , but this is not completely true. In some cases the columns of A are linearly dependent, which means that some of the vectors are actually redundant. Any vector in the $\text{span}(A)$ can always be represented as a linear combination of the linearly independent columns of A . For example take

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 3 & 2 & 5 \\ 5 & 1 & 6 \\ 2 & 2 & 4 \end{bmatrix} \quad (5)$$

Now clearly the last column is not linearly independent as it can be obtained by adding the first two columns. Let us take the following linear combination

$$2 \begin{bmatrix} 2 \\ 3 \\ 5 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 5 \\ 6 \\ 4 \end{bmatrix} \quad (6)$$

You can verify that

$$2 \begin{bmatrix} 2 \\ 3 \\ 5 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 5 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 12 \\ 32 \\ 37 \\ 26 \end{bmatrix} = 6 \begin{bmatrix} 2 \\ 3 \\ 5 \\ 2 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix}. \quad (7)$$

More generally you can verify

$$x_1 \begin{bmatrix} 2 \\ 3 \\ 5 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 5 \\ 6 \\ 4 \end{bmatrix} = \tilde{x}_1 \begin{bmatrix} 2 \\ 3 \\ 5 \\ 2 \end{bmatrix} + \tilde{x}_2 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \quad \tilde{x}_1 = x_1 + x_3, \text{ and } \tilde{x}_2 = x_2 + x_3, \quad (8)$$

Since the dimension is given by the fewest number of parameters that you need to identify any element in the space, it turns out that the dimension of $\text{span}(A)$ is equal to the number of linearly independent columns of A , which will be less than or equal to the $\min(m, n)$.

$$\dim(\text{span}(A)) \leq \min(m, n). \quad (9)$$

Now let's introduce the term **rank**. The rank of a matrix is the dimension of the span of its columns, i.e., $\text{rank}(A) = \dim(\text{span}(A))$. For example, the rank of the matrix

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 3 & 2 & 5 \\ 5 & 1 & 6 \\ 2 & 2 & 4 \end{bmatrix} \quad (10)$$

defined previously is equal to 2 since it has two linearly independent columns.

Matrix Inversion

Now that we have been introduced to the concepts of linear independence, span, and dimension, we have all the tools we need to tackle matrix inversion. First, let us define what it means for a matrix to be invertible and what a matrix inverse is.

Definition 6.1 (Inverse): A square matrix A is said to be invertible if there exists an matrix B such that

$$AB = BA = I. \quad (11)$$

where I is the identity matrix. In this case, we call the matrix B the inverse of the matrix A , which we denote as A^{-1} .

Example 6.1 (Matrix inverse): Consider the 2×2 matrix $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$. Then $A^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$. We can verify that the following holds

$$AA^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (12)$$

$$A^{-1}A = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (13)$$

Let's show an important property of matrix inverses: **If A is an invertible matrix, then its inverse must be unique.**

Proof. Suppose B_1 and B_2 are both inverses of the matrix A . Then we have

$$AB_1 = B_1A = I \quad (14)$$

$$AB_2 = B_2A = I \quad (15)$$

Now take the equation

$$AB_1 = I. \quad (16)$$

Multiplying both sides of the equation by B_2 from the left, we have

$$B_2(AB_1) = B_2I. \quad (17)$$

Notice that by associativity of matrix multiplication, the left hand side of the equation above becomes

$$B_2(AB_1) = (B_2A)B_1 = IB_1 = B_1. \quad (18)$$

Hence we have

$$B_1 = B_2. \quad (19)$$

We see that B_1 and B_2 must be equal. Thus the inverse of any invertible matrix is unique. \square

In discussion, you will see a few more useful properties of matrix inverses!

Now the natural questions to ask are:

- How do we know if a matrix is invertible or not?
- If a matrix is invertible, how do we go about finding its inverse?

It turns out Gaussian Elimination could help us answer these questions!

Finding inverses with Gaussian Elimination

A square matrix M and its inverse M^{-1} will always satisfy the following conditions $MM^{-1} = I$ and $M^{-1}M = I$, where I is the identity matrix.

$$\text{Let } M = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \text{ and } M^{-1} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

We want to find the values of b_{ij} such that the equation $MM^{-1} = I$ would be satisfied.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since we mathematicians are lazy, we can write the above as an **augmented matrix**, which joins the left and right numerical matrices together and hides the variable matrix, as shown below.

$$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array} \right]$$

Now, to find the inverse matrix M^{-1} using Gaussian Elimination, we have to transform the left numerical matrix (left half of the augmented matrix) to the identity matrix, then the right numerical matrix (right half of the augmented matrix) becomes our solution. In equation form $MM^{-1} = I$, we are transforming M and I simultaneously using row operations so that the equation becomes $IM^{-1} = A$, where A is the resulting numerical matrix from the Gaussian Elimination. Since M^{-1} is multiplied by the identity matrix I , the resulting numerical matrix A must equal to M^{-1} , and we have the values for the elements in our inverse matrix. We will now do the actual computation below:

$$\begin{aligned} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array} \right] &\Rightarrow R_2 - 2R_1 \Rightarrow \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{array} \right] \Rightarrow -1(R_2) \Rightarrow \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & -1 \end{array} \right] \\ &\Rightarrow R_1 - R_2 \Rightarrow \left[\begin{array}{cc|cc} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 \end{array} \right] \end{aligned}$$

M^{-1} is the right half of the augmented matrix. Therefore $M^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$. More generally, for any $n \times n$ matrix M , we can perform Gaussian Elimination on the augmented matrix

$$\left[\begin{array}{c|c} M & I_n \end{array} \right]$$

If at termination of Gaussian Elimination, we end up with an identity matrix on the left, then the matrix on the right is the inverse of the matrix M .

$$\left[\begin{array}{c|c} I_n & M^{-1} \end{array} \right]$$

If we don't end up with an identity matrix on the left after running Gaussian Elimination, we know that the matrix is not invertible. What does this say about the rows of A ? The rows of M are linearly dependent. Conversely, we also know that if the rows of M are linearly dependent, running Gaussian Elimination on the matrix will give us at least one row of zeros. Hence we can conclude that **a matrix M is invertible if and only if its rows are linearly independent.**

Connecting invertibility of a matrix with its columns

In the previous section, we connected invertibility of a matrix with its rows. Alternatively, we can look at whether a square matrix $A \in \mathbb{R}^{n \times n}$ has an inverse from the column perspective.

A matrix A is invertible if and only if its columns are linearly independent.

Let's think about this intuitively. Consider A as an operator on any vector $\vec{x} \in \mathbb{R}^n$. What does it mean for A to have an inverse? It suggests that we can find a matrix that "undoes" the effect of matrix A operating on any vector $\vec{x} \in \mathbb{R}^n$. What property should A have in order for this to be possible? A should map any two distinct vectors to distinct vectors in \mathbb{R}^n , i.e., $A\vec{x}_1 \neq A\vec{x}_2$ for vectors \vec{x}_1, \vec{x}_2 such that $\vec{x}_1 \neq \vec{x}_2$. Since A is a square matrix, there should be a bijective mapping between vectors in \mathbb{R}^n and vectors in $\text{range}(A)$. Hence we must have $\text{span}(A) = \mathbb{R}^n$. Recall that the dimension of the span of a matrix is the number of linearly independent columns. Because $\text{span}(A) = \mathbb{R}^n$, the number of linearly independent columns of A is equal to n . However, since A has exactly n columns, we know then that the columns of A must be linearly independent.

Example 6.2 (Invertibility intuition):

Is the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ invertible? Intuitively, it is not because A can map two distinct vectors into the same vector.

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 1 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \quad (20)$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}. \quad (21)$$

(22)

We cannot recover the vector uniquely after it is operated by A . This is connected with the fact that the columns are linearly dependent – different weighted combinations of columns could generate the same vector.

With the above in mind, we can infer that a square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if there exists a unique solution to the system of linear equations $A\vec{x} = \vec{b}$ for any vector $\vec{b} \in \mathbb{R}^n$.