Linear Algebra Fundamentals

It can be argued that all of linear algebra can be understood using the *four fundamental subspaces* associated with a matrix. Because they form the foundation on which we later work, we want an explicit method for analyzing these subspaces- That method will be the *Singular Value Decomposition* (SVD). It is unfortunate that most first courses in linear algebra do not cover this material, so we do it here. Again, we cannot stress the importance of this decomposition enough- We will apply this technique throughout the rest of this text.

1 Representation, Basis and Dimension

Let us quickly review some notation and basic ideas from linear algebra:

Suppose that the matrix V is composed of the columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, and that these columns form a basis basis for some subspace, H, in $\mathbb{R}^n$.

Side Remark: Notice that this implies $k \leq n$ so that the matrix V must be a "tall" matrix- Why?

By the definition of a basis, every vector in the subspace H can be written as a linear combination of the basis. In particular, if $\boldsymbol{x} \in H$, then we can write:

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k \doteq V\boldsymbol{c}$$

where $\boldsymbol{c}$ is sometimes denoted as $[\boldsymbol{x}]_V$, and is referred to as the *coordinates of* $\boldsymbol{x}$ *with respect to the basis in* V.

Therefore, every vector in our subset of $\mathbb{R}^n$ can be identified with a point in $\mathbb{R}^k$, which gives us a function:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \longleftrightarrow \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \boldsymbol{c}$$

Since $k < n$, we think of $\boldsymbol{c}$ is the **low dimensional representation** of the vector $\boldsymbol{x}$.

Furthermore, we would say that the subspace H (a subspace of $\mathbb{R}^n$) is *isomorphic* to $\mathbb{R}^k$. We'll recall the definition:

Definition 1.1. Any one-to-one (and onto) linear map is called an isomorphism. In particular, any change of coordinates is an isomorphism. Spaces that are isomorphic have essentially the same algebraic structure- adding vectors in one space is corresponds to adding vectors in the second space, and scalar multiplication in one space is the same as scalar multiplication in the second.

Example 1.1. If $\boldsymbol{v}_i, \boldsymbol{v}_j$ are two linearly independent vectors, then the subspace created by their span is *isomorphic* to the xy plane - but is not *equal* to the plane.

Definition 1.2. Let H be a subspace of vector space X. Then H has dimension k if a basis for H requires k vectors.

Generally speaking, to find the coordinates of $\boldsymbol{x}$ with respect to some arbitrary basis (as columns of a matrix V), we have to solve the following system:

$$\boldsymbol{x} = V\boldsymbol{c}$$

Again, in the most general case, this would require row reduction or if the matrix was square, matrix inversion. However, a special case is when the columns of V are orthonormal.

Before continuing, let's look at an example:

Example 1.2. Let the subspace H be formed by the span of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ given below. Given the point $\boldsymbol{x}_1, \boldsymbol{x}_2$ below, find which one belongs to H, and if it does, give its coordinates.

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \quad \boldsymbol{x}_1 = \begin{bmatrix} 7 \\ 4 \\ 0 \end{bmatrix} \quad \boldsymbol{x}_2 = \begin{bmatrix} 4 \\ 3 \\ -1 \end{bmatrix}$$

SOLUTION: Rather than row-reduce twice, we'll do it once on the augmented matrix below.

$$\left[\begin{array}{rr|rr} 1 & 2 & 7 & 4 \\ 2 & -1 & 4 & 3 \\ -1 & 1 & 0 & -1 \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 3 & 2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

How should this be interpreted? The second vector, $\boldsymbol{x}_2$ is in H, as it can be expressed as $2\boldsymbol{v}_1 + \boldsymbol{v}_2$. Its low dimensional representation is thus $[2, 1]^T$.

The first vector, $\boldsymbol{x}_1$, cannot be expressed as a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, so it does not belong to H.

If the basis is orthonormal, we do not need to perform any row reduction. Let us recall a few more definitions:

Definition 1.3. A real $n \times n$ matrix $\mathbb{Q}$ is said to be *orthogonal* if

$$\mathbb{Q}^T\mathbb{Q} = I$$

Be careful with the definition- The way the definition is worded, all *orthogonal* matrices are square. We will often work with non-square matrices with orthonormal columns.

Example 1.3. Let U be a non-square matrix with orthonormal columns (so it is tall - why?). Then we show using a small example that U^TU is the identity matrix, but UU^T is not: Let $U = [\boldsymbol{u}_1] = [1, 0]^T$. Then $U^TU = 1$ (the one by one identity), but the other way around, we do not get the identity:

$$UU^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Having an orthonormal basis nice makes it easy to compute the coordinates. For example, if the column vectors $\boldsymbol{u}_k$ are orthonormal and the matrix U is formed from them, we can write a vector $\boldsymbol{x}$ in the columnspace of U as:

$$\boldsymbol{x} = U\boldsymbol{c} \quad \Rightarrow \quad U^T\boldsymbol{x} = U^TU\boldsymbol{c} = \boldsymbol{c}$$

We see that the coordinates of $\boldsymbol{x}$ with respect to the basis in the columns of U are simply: $U^T\boldsymbol{x}$. NOTE: No matrix inversion, no row reduction necessary!

Coordinates as Projections

The matrix equation we looked at earlier can be written as:

$$\boldsymbol{x} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_k\boldsymbol{u}_k$$

And taking the dot product of each side with $\boldsymbol{u}_j$, we get:

$$\boldsymbol{u}_j^T\boldsymbol{x} = 0 + 0 + \cdots + c_j\boldsymbol{u}_j^T\boldsymbol{u}_j + 0 + \cdots + 0$$

Therefore, each coordinate can be expressed as:

$$c_j = \frac{\boldsymbol{u}_j^T\boldsymbol{x}}{\boldsymbol{u}_j^T\boldsymbol{u}_j}$$

Which we interpret in terms of projections below.

Recall that the orthogonal projection of $\boldsymbol{x}$ onto a vector $\boldsymbol{u}$ is the following:

$$\text{Proj}_{\boldsymbol{u}}(\boldsymbol{x}) = \frac{\boldsymbol{u}_j \cdot \boldsymbol{x}}{\boldsymbol{u} \cdot \boldsymbol{u}}\boldsymbol{u}$$

In fact, we might also recall the *scalar projection* of $\boldsymbol{x}$ onto a vector $\boldsymbol{u}$ is the following:

$$\frac{\boldsymbol{u}_j \cdot \boldsymbol{x}}{\|\boldsymbol{u}\|}$$

In particular, if $\|\boldsymbol{u}\| = 1$, then the scalar projection of $\boldsymbol{x}$ onto $\boldsymbol{u}$ is the coordinate, and things simplify.

Additionally, the **projection of $\boldsymbol{x}$** onto the subspace spanned by the (orthonormal) columns of a matrix U is found by projecting $\boldsymbol{x}$ onto each column of U and summing:

$$\text{Proj}_U(\boldsymbol{x}) = \text{Proj}_{u_1}(\boldsymbol{x}) + \cdots + \text{Proj}_{u_k}(\boldsymbol{x})$$

or, in matrix form:

$$\text{Proj}_U(\boldsymbol{x}) = U\boldsymbol{c} = UU^T\boldsymbol{x} \tag{1}$$

Example 1.4. Let $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ be given. Find the coordinates of $\mathbf{x}_1$ with respect to this basis.

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \qquad \boldsymbol{x}_1 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

SOLUTION:

$$\boldsymbol{c} = U^T\boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{c} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} & 0 \\ 2/\sqrt{6} & -1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} -1 \\ 8 \\ -2 \end{bmatrix} = \begin{bmatrix} 3\sqrt{5} \\ -2\sqrt{6} \end{bmatrix}$$

The reader should verify that this is accurate.

We summarize our discussion with the following theorem:

Change of Basis Theorem. Suppose H is a subspace of $\mathbb{R}^n$ with orthonormal basis vectors given by the k columns of a matrix U (so U is $n \times k$). Then, given $\boldsymbol{x} \in H$,

- The **low dimensional representation** of $\boldsymbol{x}$ with respect to U is the vector of coordinates, $\boldsymbol{c} \in \mathbb{R}^k$:

$$\boldsymbol{c} = U^T\boldsymbol{x}$$

- The **reconstruction** of $\boldsymbol{x}$ as a vector in $\mathbb{R}^n$ is:

$$\hat{\boldsymbol{x}} = UU^T\boldsymbol{x}$$

  where, if the subspace formed by U contains $\boldsymbol{x}$, then $\boldsymbol{x} = \hat{\boldsymbol{x}}$- Notice in this case, the projection of $\boldsymbol{x}$ into the columnspace of U is the same as $\boldsymbol{x}$.

This last point may seem trivial since we started by saying that $\boldsymbol{x} \in U$, however, soon we'll be loosening that requirement.

Example 1.5. Let $\boldsymbol{x} = [3, 2, 3]^T$ and let the basis vectors be $\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}[1, 0, 1]^T$ and let $\boldsymbol{u}_2 = [0, 1, 0]^T$. Compute the low dimensional representation of $\boldsymbol{x}$, and its reconstruction (to verify that $\boldsymbol{x}$ is in the right subspace).

SOLUTION: The low dimensional representation is given by:

$$\boldsymbol{c} = U^T\boldsymbol{x} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} \\ 2 \end{bmatrix}$$

And the reconstruction (verify the arithmetic) is:

$$\hat{\boldsymbol{x}} = UU^T\mathbf{x} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$$

For future reference, you might notice that UU^T is **not** the identity, but U^TU is the 2×2 identity:

$$U^TU = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Projections are important part of our work in modeling data- so much so that we'll spend a bit of time formalizing the ideas in the next section.

Figure 1: Projections P_1 and P_2 in the first and second graphs (respectively). Asterisks denote the original data point, and circles represent their destination, the projection of the asterisk onto the vector $[1,1]^T$. The line segment follows the direction $P\boldsymbol{x} - \boldsymbol{x}$. Note that P_1 does not project in an orthogonal fashion, while the second matrix P_2 does.

2 Special Mappings: The Projectors

In the previous section, we looked at projecting one vector onto a subspace by using Equation 1. In this section, we think about the projection as a function whose domain and range will be subspaces of $\mathbb{R}^n$.

The defining equation for such a function comes from the idea that if one projects a vector, then projecting it again will leave it unchanged.

Definition 2.1. A *Projector* is a square matrix $\mathbb{P}$ so that:

$$\mathbb{P}^2 = \mathbb{P}$$

In particular, $\mathbb{P}\boldsymbol{x}$ is the projection of $\boldsymbol{x}$.

Example 2.1. The following are two projectors. Their matrix representations are given by:

$$P_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad P_2 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Some samples of the projections are given in Figure 1, where we see that both project to the subspace spanned by $[1,1]^T$.

Let's consider the action of these matrices on an arbitrary point:

$$P_1\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix}, P_1(P_1\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ x \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix}$$

$$P_2\boldsymbol{x} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{bmatrix} = \frac{x+y}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

You should verify that $P_2^2\boldsymbol{x} = P_2(P_2(\boldsymbol{x})) = \boldsymbol{x}$.

You can deduce along which direction a point is projected by drawing a straight line from the point $\boldsymbol{x}$ to the point $\mathbb{P}\boldsymbol{x}$. In general, this direction will depend on the point. We denote this direction by the vector $\mathbb{P}\boldsymbol{x} - \boldsymbol{x}$.

From the previous examples, we see that $\mathbb{P}\boldsymbol{x} - \boldsymbol{x}$ is given by:

$$P_1\boldsymbol{x} - \boldsymbol{x} = \begin{bmatrix} 0 \\ x - y \end{bmatrix}, \text{ and } P_2\boldsymbol{x} - \boldsymbol{x} = \begin{bmatrix} \frac{-x+y}{2} \\ \frac{x-y}{2} \end{bmatrix} = \frac{x-y}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

You'll notice that in the case of P_2, $P_2\boldsymbol{x} - \boldsymbol{x} = (P_2 - I)\boldsymbol{x}$ is orthogonal to $P_2\boldsymbol{x}$.

Definition 2.2. $\mathbb{P}$ is said to be an *orthogonal* projector if it is a projector, and the range of $\mathbb{P}$ is orthogonal to the range of $(I - \mathbb{P})$. We can show orthogonality by taking an arbitrary point in the range, $\mathbb{P}\boldsymbol{x}$ and an arbitrary point in $(I - \mathbb{P})$, $(I - \mathbb{P})\boldsymbol{y}$, and show the dot product is 0.

There is a property of real projectors that make them nice to work with: They are also symmetric matrices:

Theorem 2.1. *The (real) projector $\mathbb{P}$ is an orthogonal projector iff $\mathbb{P} = \mathbb{P}^T$. For a proof, see for example, [**?**].*

Caution: An orthogonal projector need not be an orthogonal matrix. Notice that the projector P_2 from Figure 1 was not an orthogonal matrix (that is, $P_2 P_2^T \neq I$).

We have two primary sources for projectors:

Projecting to a vector: Let $\boldsymbol{a}$ be an arbitrary, real, non-zero vector. We show that

$$\mathbb{P}_{\boldsymbol{a}} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\|\boldsymbol{a}\|^2}$$

is a rank one orthogonal projector onto the span of $\boldsymbol{a}$:

- The matrix $\boldsymbol{a}\boldsymbol{a}^T$ has rank one, since every column is a multiple of $\boldsymbol{a}$.
- The given matrix is a projector:

$$\mathbb{P}^2 = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\|\boldsymbol{a}\|^2} \cdot \frac{\boldsymbol{a}\boldsymbol{a}^T}{\|\boldsymbol{a}\|^2} = \frac{1}{\|\boldsymbol{a}\|^4}\boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{a})\boldsymbol{a}^T = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\|\boldsymbol{a}\|^2} = \mathbb{P}$$

- The matrix is an orthogonal projector, since $\mathbb{P}^T = \mathbb{P}$.

Projecting to a Subspace: Let $Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k]$ be a matrix with orthonormal columns. Then

$$\mathbb{P} = QQ^T$$

is an orthogonal projector to the column space of Q. This generalizes the result of the previous exercise. Note that if Q was additionally a square matrix, $QQ^T = I$.

Note that this is exactly the property that we discussed in the last example of the previous section.

Exercises

1. Show that the plane H defined by:

$$H = \left\{ \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ such that } \alpha_1, \alpha_2 \in \mathbb{R} \right\}$$

is isormorphic to $\mathbb{R}^2$.

2. Let the subspace G be the plane defined below, and consider the vector $\boldsymbol{x}$, where:

$$G = \left\{ \alpha_1 \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \text{ such that } \alpha_1, \alpha_2 \in \mathbb{R} \right\} \qquad \boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

   (a) Find the projector P that takes an arbitrary vector and projects it (orthogonally) to the plane G.

   (b) Find the orthogonal projection of the given $\boldsymbol{x}$ onto the plane G.

   (c) Find the distance from the plane G to the vector $\boldsymbol{x}$.

3. If the low dimensional representation of a vector $\boldsymbol{x}$ is $[9, -1]^T$ and the basis vectors are $[1, 0, 1]^T$ and $[3, 1, 1]^T$, then what was the original vector $\boldsymbol{x}$? (HINT: it is easy to compute it directly)

4. If the vector $\boldsymbol{x} = [10, 4, 2]^T$ and the basis vectors are $[1, 0, 1]^T$ and $[3, 1, 1]^T$, then what is the low dimensional representation for $\boldsymbol{x}$?

5. Let $\boldsymbol{a} = [-1, 3]^T$. Find a square matrix P so that $P\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto the span of $\boldsymbol{a}$.

3 The Four Fundamental Subspaces

Given any $m \times n$ matrix A, we consider the mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ by:

$$\boldsymbol{x} \to A\boldsymbol{x} = \boldsymbol{y}$$

The four subspaces allow us to completely understand the domain and range of the mapping. We will first define them, then look at some examples.

Definition 3.1. The Four Fundamental Subspaces

- The **row space** of A is a subspace of $\mathbb{R}^n$ formed by taking all possible linear combinations of the rows of A. Formally,

$$\text{Row}(A) = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid x = A^T\boldsymbol{y} \;\; y \in \mathbb{R}^m \right\}$$

- The **null space** of A is a subspace of $\mathbb{R}^n$ formed by

$$\text{Null}(A) = \{ \boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{0} \}$$

- The **column space** of A is a subspace of $\mathbb{R}^m$ formed by taking all possible linear combinations of the columns of A.

$$\text{Col}(A) = \{ \boldsymbol{y} \in \mathbb{R}^m \mid y = A\boldsymbol{x} \quad \in \mathbb{R}^n \}$$

  The column space is also the image of the mapping. Notice that $A\boldsymbol{x}$ is simply a linear combination of the columns of A:

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

- Finally, we define the **null space** of A^T can be defined in the obvious way (see the Exercises).

The fundamental subspaces subdivide the domain and range of the mapping in a particularly nice way:

Theorem 3.1. *Let A be an $m \times n$ matrix. Then*

- *The nullspace of A is orthogonal to the row space of A*
- *The nullspace of A^T is orthogonal to the columnspace of A*

Proof: See the Exercises.

Before going further, let us recall how to construct a basis for the column space, row space and nullspace of a matrix A. We'll do it with a particular matrix:

Example 3.1. Construct a basis for the column space, row space and nullspace of the matrix A below:

$$A = \begin{bmatrix} 2 & 0 & -2 & 2 \\ -2 & 5 & 7 & 3 \\ 3 & -5 & -8 & -2 \end{bmatrix}$$

SOLUTION: The row reduced form of A is:

$$\begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The first two columns of the original matrix form a basis for the columnspace (which is a subspace of $\mathbb{R}^3$):

$$\mathrm{Col}(A) = \mathrm{span}\left\{ \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix} \right\}$$

A basis for the row space is found by using the row reduced rows corresponding to the pivots (and is a subspace of $\mathbb{R}^4$). You should also verify that you can find a basis for the null space of A, given below (also a subspace of $\mathbb{R}^4$). If you're having any difficulties here, be sure to look it up in a linear algebra text:

$$\mathrm{Row}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} \qquad \mathrm{Null}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We will often refer to the dimensions of the four subspaces. We recall that there is a term for the dimension of the column space- That is, the rank.

Definition 3.2. The *rank* of a matrix A is the number of independent columns of A.

In our previous example, the rank of A is 2. Also from our example, we see that the rank is the dimension of the column space, and that this is the same as the dimension of the row space (all three numbers correspond to the number of pivots in the row reduced form of A). Finally, a handy theorem for counting is the following.

The Rank Theorem. Let the $m \times n$ matrix A have rank r. Then

$$r + \dim(\text{Null}(A)) = n$$

This theorem says that the number of pivot columns plus the other columns (which correspond to free variables) is equal to the total number of columns.

Example 3.2. The Dimensions of the Subspaces.

Given a matrix A that is $m \times n$ with rank k, then the dimensions of the four subspaces are shown below.

- $\dim(\text{Row}(A)) = k$
- $\dim(\text{Col}(A)) = k$
- $\dim(\text{Null}(A)) = n - k$
- $\dim\left(\text{Null}(A^T)\right) = m - k$

There are some interesting implications of these theorems to matrices of data- For example, suppose A is $m \times n$. With no other information, we do not know whether we should consider this matrix as n points in $\mathbb{R}^m$, or m points in $\mathbb{R}^n$. In one sense, it doesn't matter! The theorems we've discussed shows that the dimension of the columnspace is equal to the dimension of the rowspace. Later on, we'll find out that if we can find a basis for the columnspace, it is easy to find a basis for the rowspace. We'll need some more machinery first.

4 Exercises

The Best Approximation Theorem If W is a subspace of $\mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$, then the point closest to $\boldsymbol{x}$ in W is the orthogonal projection of $\boldsymbol{x}$ into W. We prove this in the exercises below.

1. Show that $\mathcal{N}(A) \perp \mathcal{R}(A^T)$. You must show that, for arbitrary $\boldsymbol{x}_1 \in \mathcal{N}(A)$ and $\boldsymbol{x}_2 \in \mathcal{R}(A^T)$, we have $(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$. Hint: Write A in terms of its rows.
2. If A is $m \times n$, how big can the rank of A possibly be?
3. Show that multiplication by an orthogonal matrix preserves lengths: $\|\mathbb{Q}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ (Hint: Use properties of inner products). Conclude that multiplication by $\mathbb{Q}$ represents a rigid rotation.

4. Prove the Pythagorean Theorem by induction: Given a set of n orthogonal vectors $\{\boldsymbol{x}_i\}$

$$\|\sum_{i=1}^{n} \boldsymbol{x}_i\|_2^2 = \sum_{i=1}^{n} \|\boldsymbol{x}_i\|_2^2$$

5. Let A be an $m \times n$ matrix where $m > n$, and let A have rank n. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, such that $\boldsymbol{y}$ is the orthogonal projection of $\boldsymbol{x}$ onto the columnspace of A. We want a formula for the projector $\mathbb{P} : \mathbb{R}^m \to \mathbb{R}^m$ so that $\mathbb{P}\boldsymbol{x} = \boldsymbol{y}$.

   (a) Why is the projector not $\mathbb{P} = AA^T$?

   (b) Since $\boldsymbol{y} - \boldsymbol{x}$ is orthogonal to the range of A, show that

   $$A^T(\boldsymbol{y} - \boldsymbol{x}) = \mathbf{0} \tag{2}$$

   (c) Show that there exists $\boldsymbol{v}$ so that Equation (2) can be written as:

   $$A^T(A\boldsymbol{v} - \boldsymbol{x}) = 0 \tag{3}$$

   (d) Argue that A^TA is invertible, so that Equation (3) implies that

   $$\boldsymbol{v} = \left(A^TA\right)^{-1} A^T\boldsymbol{x}$$

   (e) Finally, show that this implies that

   $$\mathbb{P} = A\left(A^TA\right)^{-1} A^T$$

   Note: If A has rank $k < m$, then we will need something different, since A^TA will not be full rank. The missing piece is the singular value decomposition, to be discussed later.

6. The Orthogonal Decomposition Theorem: if $\boldsymbol{x} \in \mathbb{R}^n$ and W is a (non-zero) subspace of $\mathbb{R}^n$, then $\boldsymbol{x}$ can be written *uniquely* as

   $$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{z}$$

   where $\boldsymbol{w} \in W$ and $\boldsymbol{z} \in W^{\perp}$.

   To prove this, let $\{\boldsymbol{u}_i\}_{i=1}^{p}$ be an orthonormal basis for W, define $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{u}_1)\boldsymbol{u}_1 + \ldots + (\boldsymbol{x}, \boldsymbol{u}_p)\boldsymbol{u}_p$, and define $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{w}$. Then:

   (a) Show that $\boldsymbol{z} \in W^{\perp}$ by showing that it is orthogonal to every $\boldsymbol{u}_i$.

   (b) To show that the decomposition is unique, suppose it is not. That is, there are two decompositions:

   $$\boldsymbol{x} = \boldsymbol{w}_1 + \boldsymbol{z}_1, \quad \boldsymbol{x} = \boldsymbol{w}_2 + \boldsymbol{z}_2$$

   Show this implies that $\boldsymbol{w}_1 - \boldsymbol{w}_2 = \boldsymbol{z}_2 - \boldsymbol{z}_1$, and that this vector is in both W and $W^{\perp}$. What can we conclude from this?

7. Use the previous exercises to prove the **The Best Approximation Theorem** If W is a subspace of $\mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$, then the point closest to $\boldsymbol{x}$ in W is the orthogonal projection of $\boldsymbol{x}$ into W.