Projections

Consider the following example. If a matrix $U = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ has orthonormal columns (so if U is $n \times k$, then that requires $k \le n$), then U^TU is $k \times k$, and can be computed as:

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k \end{bmatrix} [\mathbf{u}_1, \ldots, \mathbf{u}_k] = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_k \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_k \\ \vdots & & & \vdots \\ \mathbf{u}_k^T\mathbf{u}_1 & \mathbf{u}_k^T\mathbf{u}_2 & \cdots & \mathbf{u}_k^T\mathbf{u}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_k$$

But UU^T (which is $n \times n$) is NOT the identity if $k \neq n$ (If $k = n$, then the previous computation proves that the inverse is the transpose).

Here is a computation one might make for UU^T (these are OUTER products):

$$UU^T = [\mathbf{u}_1, \ldots, \mathbf{u}_k] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \cdots + \mathbf{u}_k\mathbf{u}_k^T$$

Example 4.1.2. Consider the following computations:

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

If UU^T is not the identity, what is it? Consider the following computation:

$$UU^T\mathbf{x} = \mathbf{u}_1\mathbf{u}_1^T\mathbf{x} + \mathbf{u}_2\mathbf{u}_2^T\mathbf{x} + \cdots + \mathbf{u}_k\mathbf{u}_k^T\mathbf{x}$$

$$= \mathbf{u}_1(\mathbf{u}_1^T\mathbf{x}) + \mathbf{u}_2(\mathbf{u}_2^T\mathbf{x}) + \cdots + \mathbf{u}_k(\mathbf{u}_k^T\mathbf{x})$$

which we recognize as the projection of $\mathbf{x}$ into the space spanned by the orthonormal vectors of U. In the previous section, we called this matrix B, but it is common notation that a matrix with orthonormal vectors is typically denoted by U.

Just to repeat then: If we have a set of orthonormal vectors in a matrix U forming a basis for some subspace H, then if $\mathbf{x} \in H$,

$$\mathbf{x} = U[\mathbf{x}]_U = U(U^T\mathbf{x}) = (UU^T)\mathbf{x}$$

However, if $\mathbf{x}$ is not completely contained in H, then

$$\mathrm{Proj}_H(\mathbf{x}) = UU^T\mathbf{x}$$

And of course, if $n > k$ and U is $n \times k$, then $U^TU = I$.

Notes about Programming

Programming in Matlab

- Find a random matrix with orthonormal columns

  SOLUTION: Use the "QR" decomposition of a matrix A. That is, decompose matrix A as a product of matrix Q (with orthonormal columns) times matrix R.

```
X=randn(9,6);
[Q,R]=qr(X,0);
Q'*Q           % Should be 6 x 6 identity matrix.
```

- Take a matrix X and "normalize" it by making each column have norm 1.

  SOLUTION: Find a row representing the norm of each column, then divide each row of the matrix by that row.

```
X=[1,0;0,1;1,0]
RowNorms=sqrt(sum(X.*X));
C=X./RowNorms
```

- Let $\boldsymbol{x}$ be a random vector in $\mathbb{R}^9$. We're going to project this into the subspace spanned by the first three columns of the matrix Q that we constructed previously.

  - First, what are the coordinates? In linear algebra, the coordinates (a vector in $\mathbb{R}^3$) are: $Q^T\boldsymbol{x}$, where this Q is 9×3.
  - Next, project this vector into the subspace: In linear algebra notation, $QQ^T\boldsymbol{x}$

  Here it is in Matlab, using the Q from the QR decomposition.

```
x=rand(9,1);
Coords=Q(:,1:3)'*x;
Projx=Q(:,1:3)*Coords;
Check=Q*(Q'*Projx);  % Don't compute QQ^T first!
norm(Check-Projx);  %Should be close to zero
```

Programming in Python

- Find a random matrix with orthonormal columns.

  SOLUTION: Use the "QR" decomposition of a matrix A. That is, decompose matrix A as a product of matrix Q (with orthonormal columns) times matrix R.

```
A=np.random.randn(9,6)
q,r=np.linalg.qr(A)
q.shape              #Answer is (9,6)
D=np.matmul(q.T,q)   #Should be 6 x 6 identity
```

- Take a matrix X and "normalize" it by making each column have norm 1.

  SOLUTION: Find a row representing the norm of each column, then divide each row of the matrix by that row.

```
import numpy as np
import numpy.linalg

X=np.array([[1, 0],[0,1],[1,0]])
RowNorms=np.linalg.norm(X,axis=0)
C=X/RowNorms[np.newaxis,:]
Output:
array([[0.70710678, 0.        ],
       [0.        , 1.        ],
       [0.70710678, 0.        ]])
```

- Let $\boldsymbol{x}$ be a random vector in $\mathbb{R}^9$. We're going to project this into the subspace spanned by the first three columns of the matrix Q that we constructed previously.

– First, what are the coordinates? In linear algebra, the coordinates (a vector in $\mathbb{R}^3$) are: $Q^T\boldsymbol{x}$, where this Q is 9×3.

– Next, project this vector into the subspace: In linear algebra notation, $QQ^T\boldsymbol{x}$

```
import numpy as np
import numpy.linalg

x=np.random.rand(9,1)  #Random vector in R^9
A=np.random.rand(9,3)  #We'll go with a 3-d subspace in R^9
Q,R=np.linalg.qr(A)

Q.shape   #This checks the dimensions- This is 9 x 3

Coords=np.matmul(Q.T,x)  #This is a 3 x 1 vector, Q^Tx
xProjected=np.matmul(Q,Coords)  #xProjected is 9 x 1

#Optional: Check by projecting the projection (shouldn't change)
Check=np.matmul(Q,np.matmul(Q.T,xProjected))
np.linalg.norm(xProjected-Check)  #Returns something like 2 x 10^(-16)
```

Programming in R

- Find a random matrix with orthonormal columns

  SOLUTION: Use the "QR" decomposition of a matrix A. That is, decompose matrix A as a product of matrix Q (with orthonormal columns) times matrix R.

```
A<-matrix(rnorm(54),nrow=9)   # rnorm are random, normal dist.
X=qr(A)                       # X is a QR-object
Q<-qr.Q(X)                    # Extract the matrix Q
dim(Q)                        # Check dimensions on Q (returns (9,6))
H<- t(Q) %*% Q                # Compute Q^TQ to get 6 x 6 identity
```

- Take a matrix X and "normalize" it by making each column have norm 1.

  SOLUTION: Find a row representing the norm of each column, then divide each row of the matrix by that row.

```
X<-matrix(c(1,1,0,0,1,0),3,2)
N<-sqrt(colSums(X^2)
C<-sweep(X,2,N,FUN='/')
```

- Let $\boldsymbol{x}$ be a random vector in $\mathbb{R}^9$. We're going to project this into the subspace spanned by the first three columns of the matrix Q that we constructed previously.

  – First, what are the coordinates? In linear algebra, the coordinates (a vector in $\mathbb{R}^3$) are: $Q^T\boldsymbol{x}$, where this Q is 9×3.

  – Next, project this vector into the subspace: In linear algebra notation, $QQ^T\boldsymbol{x}$

```
> x<-matrix(runif(9),ncol=1)
> A<-matrix(runif(27),ncol=3)
> X<-qr(A)
> Q<-qr.Q(X)            #This is a bit awkward
```

```
> Coords<-t(Q)%*%x
> xProjected<-Q %*% Coords
> Check=Q %*% (t(Q) %*% xProjected)
> y=xProjected-Check
> sqrt(sum(y^2))
[1] 5.360492e-16
```

Exercises

1. 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. (NOTE: The basis vectors are NOT orthonormal)

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

2. 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 $\mathbb{R}^2$.

3. 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 matrix (UU^T in our notes) 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 $\boldsymbol{x}$.

4. 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)

5. 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 $\boldsymbol{x}$?

6. 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 $\boldsymbol{a}$.

7. Refer to one of the programming languages (Matlab/Python/R), and reproduce finding an arbitrary 10×4 matrix with orthonormal columns. Use a random $\boldsymbol{x} \in \mathbb{R}^{10}$, and first find the coordinates of $\boldsymbol{x}$ with respect to the four columns in Q, then compute the orthogonal projection of $\boldsymbol{x}$ into the subspace spanned by the first four columns of Q.

8. To prove that we have an *orthogonal* projection, the vector $\text{Proj}_u(\boldsymbol{x}) - \boldsymbol{x}$ should be orthogonal to $\boldsymbol{u}$. Use this definition to show that our earlier formula was correct- that is,

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

is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$.

9. Continuing with the last exercise, show that $UU^T\mathbf{x}$ is the *orthogonal* projection of $\mathbf{x}$ into the space spanned by the columns of U by showing that $(UU^T\boldsymbol{x} - \boldsymbol{x})$ is orthogonal to $\mathbf{u}_i$ for any $i = 1, 2, \cdots, k$.

4.2 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 4.2.1. 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,

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

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

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

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

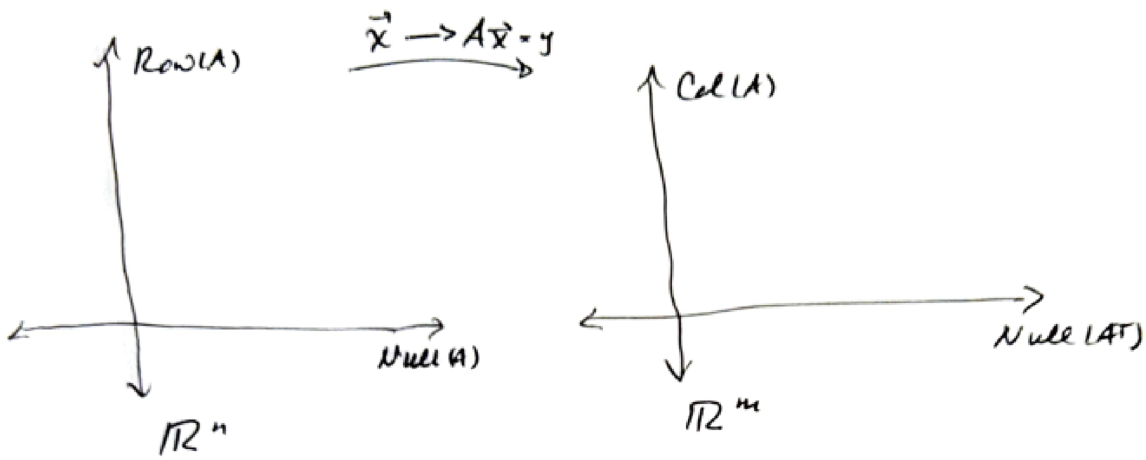
$$\text{Col}(A) = \{\boldsymbol{y} \in \mathbb{R}^m \,|\, y = A\boldsymbol{x} \quad \in \mathbb{R}^n\}$$

  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. Below we give a "cartoon" of the relationship between the four subspaces. On the left is the domain of the matrix mapping, and it represents $\mathbb{R}^n$. On the right is the codomain, and it represents $\mathbb{R}^m$. The spaces can apparently be split into two subspaces each. In the domain, these are the Row and Null spaces. In the codomain, these are the Column space and the null space of A^T (we don't use this one much).



In the diagram, notice that the axes are representing subspaces. The picture suggests that the two spaces that split our domain and range are actually orthogonal subspaces- and that is true.

Theorem 4.2.1. *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: We'll prove the first statement, the second statement is almost identical to the first. To prove the first statement, we have to show that if we take any vector $\mathbf{x}$ from nullspace of A and any vector $\mathbf{y}$ from the row space of A, then $\mathbf{x} \cdot \mathbf{y} = 0$.

Alternatively, if we can show that $\mathbf{x}$ is orthogonal to each and every row of A, then we're done as well (since $\mathbf{y}$ is a linear combination of the rows of A).

In fact, now we see a strategy: Write out what it means for $\mathbf{x}$ to be in the nullspace using the rows of A. For ease of notation, let $\mathbf{a}_j$ denote the j^{th} **row** of A, which will have size $1 \times n$. Then:

$$A\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1\mathbf{x} \\ \mathbf{a}_2\mathbf{x} \\ \vdots \\ \mathbf{a}_m\mathbf{x} \end{bmatrix} = \mathbf{0}$$

Therefore, the dot product between any row of A and $\mathbf{x}$ is zero, so that $\mathbf{x}$ is orthogonal to every row of A. Therefore, $\mathbf{x}$ must be orthogonal to any linear combination of the rows of A, so that $\mathbf{x}$ is orthogonal to the row space of A. $\square$

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 4.2.1. Construct a basis for the column space, row space and nullspace of the matrix A below that is row equivalent to the matrix beside it, RREF(A):

$$A = \begin{bmatrix} 2 & 0 & -2 & 2 \\ -2 & 5 & 7 & 3 \\ 3 & -5 & -8 & -2 \end{bmatrix} \qquad \text{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

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

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:

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

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 4.2.2. 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 4.2.2. 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.

4.3 Exercises

In the exercises below, recall that the usual norm for a vector is the Euclidean norm, or the $2-norm$, which is defined as:

$$\|\mathbf{x}\| = \sqrt{x_1^1 + x_2^2 + \cdots x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

1. Show that $\text{Null}(A^T) \perp \text{Col}(A)$. Hint: You may use what we already proved.

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 $\{\boldsymbol{x}_i\}$

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

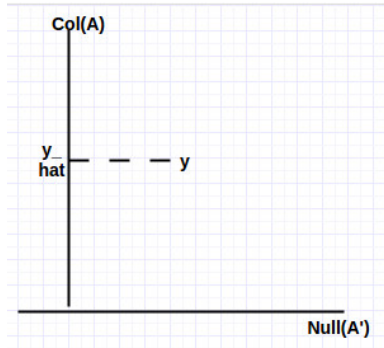
   The case where $n = 1$ is trivial, so you might look at $n = 2$ first. Try starting with

   $$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) = \cdots$$

   and then simplify to get $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. Now try the induction step on your own.

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

   The following image shows the relevant subspaces:



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

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

$$A^T(\hat{\boldsymbol{y}} - \boldsymbol{y}) = \boldsymbol{0} \tag{4.3}$$

(c) Show that there exists $\boldsymbol{x} \in \mathbb{R}^n$ so that Equation (4.3) can be written as:

$$A^T(A\boldsymbol{x} - \boldsymbol{y}) = 0 \tag{4.4}$$

(d) Argue that A^TA (which is $n \times n$) is invertible, so that Equation (13.2) implies that

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

(e) Finally, show that this implies that

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

Note: If A has rank $k \neq n$, 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 $\boldsymbol{w} = (\boldsymbol{x} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \ldots + (\boldsymbol{x} \cdot \boldsymbol{u}_p)\boldsymbol{u}_p$, 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.