Linear Models Review

Vectors in $\mathbb{R}^n$ will be written as ordered n-tuples which are understood to be **column vectors**, or $n \times 1$ matrices. A vector variable will be indicted with bold face, and the prime sign will be used for transpose. A vector $\boldsymbol{v} = (1, 0, 1, 2, 0)'$ is in $\mathbb{R}^5$.

Definition: A linear subspace V of $\mathbb{R}^n$ is a collection of vectors that includes the zero vector, such that if $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are in V, then $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 \in V$, for any real numbers a_1 and a_2.

Examples:

1. Consider the space $\mathbb{R}^3$. The set of vectors $(x_1, x_2, x_3)'$ in $\mathbb{R}^3$ such that $x_3 = 3x_1 - x_2$ is a linear subspace of $\mathbb{R}^3$. This space forms a plane through the origin. We can check the rule for some example vectors. Let $\boldsymbol{v}_1 = (1, 2, 1)'$ and $\boldsymbol{v}_2 = (0, 3, -3)'$. These are both in the subspace, because they follow the $x_3 = 3x_1 - x_2$ rule. Now, any $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2$ is $(a_1, 2a_1 + 3a_2, a_1 - 3a_2)'$, which also follows the rule.

2. Consider the space $\mathbb{R}^3$. The set of vectors $(x_1, x_2, x_3)'$ in $\mathbb{R}^3$ such that $x_3 = 5+3x_1-x_2$ is *not* a linear subspace of $\mathbb{R}^3$. This set forms a plane that does *not* go through the origin. You can see that it is not a subspace by taking the example vectors $\boldsymbol{v}_1 = (1, 2, 6)'$ and $\boldsymbol{v}_2 = (0, 3, 2)'$, which are both in the set, but now $\boldsymbol{v}_1+\boldsymbol{v}_2 = (1, 5, 8)'$ is not in the set.

3. The set of all vectors of the form $(0, 0, 0, a, b)'$ for any real numbers a and b forms a linear subspace of $\mathbb{R}^5$.

4. The set of all vectors of the form $(0, 0, 0, a, b)'$ for *positive* real numbers a and b is not a linear subspace of $\mathbb{R}^5$.

Some Terminology: Any collection of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ in $\mathbb{R}^n$ can form a linear subspace $\mathbb{R}^n$ simply by defining the linear subspace as all linear combinations of these vectors. We write the subspace as $V = \mathcal{L}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. We say "$V$ is the space **spanned by** the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$."

For example, the vectors $\boldsymbol{v}_1 = (0, 0, 0, 1, 0)'$ and $\boldsymbol{v}_2 = (0, 0, 0, 0, 1)'$ form the linear subspace $\mathcal{L}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ that can be described as "All five-tuples where the first three are zero, and the last two can be any numbers." Equivalently, we have "the set of all vectors of the form $(0, 0, 0, a, b)'$ for any real numbers a and b."

Exercise 1 *For each of the following, describe in words the vector space spanned by the given vectors.*

(a) $\boldsymbol{v}_1 = (1,1,1,0,0,0)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$

(b) $\boldsymbol{v}_1 = (1,1,1,1,1,1)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$

(c) $\boldsymbol{v}_1 = (1,1,1,0,0,0)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$, $\boldsymbol{v}_3 = (1,1,1,1,1,1)'$

(d) $\boldsymbol{v}_1 = (1,1,0,0,0,0)'$, $\boldsymbol{v}_2 = (0,0,0,0,1,1)'$

(e) $\boldsymbol{v}_1 = (1,1,1,1,1)'$, $\boldsymbol{v}_2 = (1,2,3,4,5)'$

Exercise 2 *Let* $\boldsymbol{v}_1 = (1,1,1,1,1,1)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$, *and* $\boldsymbol{v}_3 = (1,2,3,1,2,3)'$ *Which of the following are in* $V = \mathcal{L}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$*?*

(a) $(6,5,4,3,2,1)'$

(b) $(1,2,3,11,12,13)'$

(c) $(0,0,0,6,6,6)'$

(d) $(8,6,4,6,4,2)'$

(e) $(1,2,3,5,7,9)'$

Definition: A collection of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ in $\mathbb{R}^n$ is **linearly independent** if no vector in the set is a linear combination of other vectors in the set.

You can determine if a set of vectors is linearly independent by forming a matrix for which the vectors are the rows. If standard row-reduction produces at least one row of zeros, then the vectors are not linearly independent.

Exercise 3 *For each of the following, determine if the set of vectors is linearly independent.*

(a) *Let* $\boldsymbol{v}_1 = (1,0,0,0,0)'$, $\boldsymbol{v}_2 = (0,1,0,0,0)'$, *and* $\boldsymbol{v}_3 = (0,0,0,0,1)'$

(b) $\boldsymbol{v}_1 = (1,1,1,1,1)'$, $\boldsymbol{v}_2 = (1,2,3,4,5)'$, $\boldsymbol{v}_3 = (2,3,4,5,6)'$

(c) $\boldsymbol{v}_1 = (1,1,1,1,1)'$, $\boldsymbol{v}_2 = (1,0,0,0,0)'$, *and* $\boldsymbol{v}_3 = (2,3,3,3,3)'$

Definition: A set of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ in $\mathbb{R}^n$ forms a **basis** for a subspace $V \subseteq \mathbb{R}^n$ if the vectors are linearly independent and span V.

A basis for a subspace is not unique. The space spanned by the vectors $\boldsymbol{v}_1 = (1,1,1,0,0,0)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$ is also spanned by the vectors $\boldsymbol{v}_1 = (1,1,1,1,1,1)'$, $\boldsymbol{v}_2 = (0,0,0,1,1,1)'$. However, all bases must have the same number of elements. The **dimension** d of a vector space is the number of vectors in a basis for the vector space.

Exercise 4 *Write a simple basis for the space spanned by the vectors* $\boldsymbol{x}_1 = (1,1,1,1,1,1)'$, $\boldsymbol{x}_2 = (0,0,0,4,4,4)'$, *and* $\boldsymbol{x}_3 = (-2,-2,-2,8,8,8)'$.

Some definitions

- Length of a vector: for $\boldsymbol{x} \in \mathbb{R}^n$,

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

- The distance between two vectors is the length of the difference, i.e., the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\| \boldsymbol{x} - \boldsymbol{y} \| = \| \boldsymbol{y} - \boldsymbol{x} \| .$$

- Inner or dot product: for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}'\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i.$$

  Note that

$$\langle \boldsymbol{x}, a_1\boldsymbol{y}_1 + a_2\boldsymbol{y}_2 \rangle = a_1\langle \boldsymbol{x}, \boldsymbol{y}_1 \rangle + a_2\langle \boldsymbol{x}, \boldsymbol{y}_2 \rangle$$

- Projections: Take $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$. The *projection* of $\boldsymbol{y}$ onto $\boldsymbol{x}$ is written $\Pi(\boldsymbol{y}|\boldsymbol{x})$ and is defined to be

$$\Pi(\boldsymbol{y}|\boldsymbol{x}) = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\| \boldsymbol{x} \|^2}\boldsymbol{x}.$$

  We can alternatively define the projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$ to be the vector in the space spanned by $\boldsymbol{x}$ which is closest (in Euclidean distance) to y.

Exercise 5 *Prove the claim in the last sentence. That is, find the scalar a to minimize* $\| \boldsymbol{y} - a\boldsymbol{x} \|$.

Exercise 6 *Find* $\Pi(\boldsymbol{y}|\boldsymbol{x})$*:*

(a) $\boldsymbol{y} = (1, 4, 2, -1)', \boldsymbol{x} = (2, 2, 1, 1)'$

(b) $\boldsymbol{y} = (6, 4, -2, 1)', \boldsymbol{x} = (1, 1, 0, 0)'$

(c) $\boldsymbol{y} = (6, 4, 2, 6)', \boldsymbol{x} = (1, 1, 1, 1)'$

(d) $\boldsymbol{y} = (6, 4, 2, 6)', \boldsymbol{x} = (1, 2, 3, 4)'$

Definition: Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ are **orthogonal** or perpendicular if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

Exercise 7 *Show that* $\boldsymbol{y} - \Pi(\boldsymbol{y}|\boldsymbol{x})$ *is orthogonal to* $\boldsymbol{x}$*, for any* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* $\mathbb{R}^n$*.*

Exercise 8 *Check to see if the following pairs are orthogonal:*

(a) $\boldsymbol{x}_1 = (0, 0, 0, 0, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (1, 1, 1, 1, 0, 0, 0, 0)'$

(b) $\boldsymbol{y} = (1, 4, 2, -1)', \boldsymbol{x} = (2, 2, 1, 1)'$

(c) $\boldsymbol{x}_1 = (1, 1, 1, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (-3, -2, -1, 0, 1, 2, 3)'$

Definition: A vector $\boldsymbol{x}$ in $\mathbb{R}^n$ is orthogonal to a subspace $V \subseteq \mathbb{R}^n$ if $\langle \boldsymbol{x}, \boldsymbol{v} \rangle = 0$ for all $\boldsymbol{v} \in V$. Note that if $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ forms a basis for V, we have that $\boldsymbol{x}$ is orthogonal to V if $\langle \boldsymbol{x}, \boldsymbol{v}_i \rangle = 0$ for $i = 1, \ldots, k$.

Exercise 9 *Check to see if the vector* $\boldsymbol{y}$ *is orthogonal to the space spanned by* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$*:*

(a) $\boldsymbol{y} = (1, -1, 1, -1, 2, -2, 5, -5)'$, $\boldsymbol{x}_1 = (0, 0, 0, 0, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (1, 1, 1, 1, 0, 0, 0, 0)'$

(b) $\boldsymbol{y} = (4, 3, 2, 1)'$, $\boldsymbol{x}_1 = (1, 2, 3, 4)'$, $\boldsymbol{x}_2 = (1, 1, 1, 1)$.

(c) $\boldsymbol{y} = (-4, 3, 2, -1)'$, $\boldsymbol{x}_1 = (1, 2, 3, 4)'$, $\boldsymbol{x}_2 = (1, 1, 1, 1)$.

Exercise 10 *A physics teacher has his class run an experiment to measure spring coefficients. Each team of students is provided with a spring, which they extend to a given distance and measure the force. The physical relationship between force (y) and distance (x) is given by* $y = kx$*, but there is error in the student's measurements (assume for now that there is no error in the distance measurements). Each student measures force with* $x = 1, 2, 3, 4, 5cm$*. Show that the least-squares estimate of* k *is also the coefficient of the projection of the data* $\boldsymbol{y} = (y_1, y_2, y_3, y_4, y_5)'$ *onto* $\boldsymbol{x} = (1, 2, 3, 4, 5)'$.

Projections onto Subspaces

The projection of $\boldsymbol{y} \in \mathbb{R}^n$ onto a linear subspace V of $\mathbb{R}^n$ is defined as the vector in the subspace that is closest to $\boldsymbol{y}$, in Euclidean distance. Specifically, we denote the projection as $\Pi(\boldsymbol{y}|V)$, and it is the vector $\boldsymbol{v} \in V$ that minimizes $\| \boldsymbol{y} - \boldsymbol{v} \|$.

Suppose we want to project the vector $\boldsymbol{y}$ in $\mathbb{R}^n$ onto a two-dimensional subspace $V \subseteq \mathbb{R}^n$. First, suppose that the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ form a basis for V **and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal**. The projection is a vector $\hat{\boldsymbol{y}} \in V$ such that the distance from $\boldsymbol{y}$ to $\hat{\boldsymbol{y}}$ is minimized. In other words, we find scalars a and b to minimize

$$S = \| \boldsymbol{y} - (a\boldsymbol{x}_1 + b\boldsymbol{x}_2) \|^2 = \sum_{i=1}^{n} [y_i - (ax_{1i} + bx_{2i})]^2.$$

(Note that we're actually minimizing the squared distance, but because the square function is monotone over the positive reals, this amounts to the same thing as minimizing the distance.)

We can do some calculus:

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - (ax_{1i} + bx_{2i}))x_{1i}$$

which is zero when

$$\sum_{i=1}^{n} x_{1i} y_i = a \sum_{i=1}^{n} x_{1i}^2 + b \sum_{i=1}^{n} x_{1i} x_{2i}.$$

Note that because $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal, the last term is zero, and we have

$$a = \frac{\sum_{i=1}^{n} x_{1i} y_i}{\sum_{i=1}^{n} x_{1i}^2} = \frac{\langle \boldsymbol{x}_1, \boldsymbol{y} \rangle}{\| \boldsymbol{x}_1 \|^2}.$$

Similarly,

$$b = \frac{\langle \boldsymbol{x}_2, \boldsymbol{y} \rangle}{\| \boldsymbol{x}_2 \|^2},$$

so that *if* $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal, the projection $\hat{\boldsymbol{y}}$ of $\boldsymbol{y}$ onto the space spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is the sum of the projections onto the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ individually:

$$\hat{\boldsymbol{y}} = \Pi(\boldsymbol{y}|\boldsymbol{x}_1) + \Pi(\boldsymbol{y}|\boldsymbol{x}_2).$$

This result can be generalized to a larger set of orthogonal spanning vectors: If $\boldsymbol{x}_1, \ldots \boldsymbol{x}_k$ form an orthogonal set (i.e., $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 0$ for $i \neq j$), then the projection of $\boldsymbol{y}$ onto the space spanned by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ is the sum of the projections of $\boldsymbol{y}$ onto the $\boldsymbol{x}_i$ individually.

Theorem 1 *If the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ in $I\!R^n$ form an orthogonal set, then the projection $\hat{\boldsymbol{y}}$ of any vector $\boldsymbol{y}$ onto $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ is*

$$\hat{\boldsymbol{y}} = \sum_{j=1}^{k} \Pi(\boldsymbol{y}|\boldsymbol{x}_j).$$

Proof: Any vector in $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ can be written as $\sum_{j=1}^{k} a_j \boldsymbol{x}_j$, so we can write the quanitity to minimize as

$$S = \| \boldsymbol{y} - \sum_{j=1}^{k} a_j \boldsymbol{x}_j \|^2 = \sum_{i=1}^{n} [y_i - \sum_{j=1}^{k} a_j \boldsymbol{x}_j]^2.$$

As in the two-dimensional example, we take the derivative of S with respect to each a_j, set these to zero, and solve the resulting k equations and k unknowns. For each j,

$$\frac{\partial S}{\partial a_l} = -2 \sum_{i=1}^{n} (y_i - \sum_{j=1}^{k} a_j \boldsymbol{x}_j) x_{li},$$

and setting to zero gives the equation

$$\sum_{i=1}^{n} y_i x_{ji} = \sum_{j=1}^{k} a_j \left[\sum_{i=1}^{n} x_{ji} x_{li} \right] = \sum_{j=1}^{k} a_j \langle \boldsymbol{x}_l, \boldsymbol{x}_j \rangle.$$

Because the $\boldsymbol{x}$ vectors form an orthogonal set, $\langle \boldsymbol{x}_l, \boldsymbol{x}_j \rangle = 0$ for $l \neq j$, and the equations become

$$\langle \boldsymbol{y}, \boldsymbol{x}_l \rangle = a_l \langle \boldsymbol{x}_l, \boldsymbol{x}_l \rangle = \| \boldsymbol{x}_l \|^2 .$$

This gives $a_l = \langle \boldsymbol{y}, \boldsymbol{x}_l \rangle / \| \boldsymbol{x}_l \|^2$ for each $l = 1, \ldots, k$. ♣

In summary, if we want to project the vector $\boldsymbol{y}$ in $\mathrm{I\!R}^n$ onto a k-dimensional subspace $V \subseteq \mathrm{I\!R}^n$, with the orthogonal vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ forming a basis for V, we have

$$\hat{\boldsymbol{y}} = \Pi(\boldsymbol{y}|\boldsymbol{x}_1) + \cdots + \Pi(\boldsymbol{y}|\boldsymbol{x}_k).$$

Remember that this doesn't work for a basis which is not an orthogonal set.

Exercise 11 *For each of the following, determine if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal, and if so, find the projection of $\boldsymbol{y}$ onto the space spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.*

(a) $\boldsymbol{y} = (4, 6, 5, 11, 6, 7)'$, $\boldsymbol{x}_1 = (0, 0, 0, 1, 1, 1)'$, $\boldsymbol{x}_2 = (1, 1, 1, 0, 0, 0)'$

(b) $\boldsymbol{y} = (4, 6, 5, 11, 6, 7)'$, $\boldsymbol{x}_1 = (1, 1, 0, 0, 0, 0)'$, $\boldsymbol{x}_2 = (0, 0, 0, 0, 1, 1)'$

(c) $\boldsymbol{y} = (4, 6, 5, 6, 8)'$, $\boldsymbol{x}_1 = (1, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (1, 2, 3, 4, 5)'$

(d) $\boldsymbol{y} = (4, 6, 5, 6, 8)'$, $\boldsymbol{x}_1 = (1, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (-2, -1, 0, 1, 2)'$

Exercise 12 *Consider the ANOVA problem with random samples of size four from each of three populations. Let the parameters β_1, β_2, and β_3 be the population means, and assume that the populations all have the same variance. Write the model to estimate the parameters in matrix form, and find the least-squares estimates of the parameters.*

If we want to project $\boldsymbol{y}$ in R^n onto the subspace spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, but the latter are not orthogonal, we can use calculus to get the answer. Let $\hat{\boldsymbol{y}} = a\boldsymbol{x}_1 + b\boldsymbol{x}_2$ be a vector in the subspace; now find a and b to minimize the distance from $\boldsymbol{y}$ to $\hat{\boldsymbol{y}}$. The squared distance from $\boldsymbol{y}$ to $\hat{\boldsymbol{y}}$ is

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a\boldsymbol{x}_{1i} + b\boldsymbol{x}_{2i})^2;$$

we minimize this with respect to a and b by taking partial derivatives:

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a\boldsymbol{x}_{1i} + b\boldsymbol{x}_{2i})\boldsymbol{x}_{1i}; \quad \text{and}$$

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (y_i - a\boldsymbol{x}_{1i} + b\boldsymbol{x}_{2i})\boldsymbol{x}_{2i}.$$

This produces two linear equations and two unknowns, which may have a unique solution.

Exercise 13 *(a) What is the condition for the two above equations to have a unique solution, in terms of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$?*

(b) Find a formula for the projection of $\boldsymbol{y}$ onto the subspace spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

This method may be extended to projections of the data vector $\boldsymbol{y}$ onto a subspace spanned by k vectors in R^n. If the set $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ is linearly independent, then the procedure gives a set of k linear equations and k unknowns that have a unique solution.

An alternative method for projecting $\boldsymbol{y}$ onto a subspace spanned by an arbitrary set of k vectors in R^n uses the Gram-Schmidt orthogonalization to produce a set of orthogonal vectors that span the same subspace as the original set, then the projection of $\boldsymbol{y}$ onto the subspace is the sum of the projections of $\boldsymbol{y}$ onto the individual vectors in the orthogonal set.

Given any basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ for a subspace V, we can create an **orthonormal basis** $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ for the same subspace. An orthonormal basis is such that all basis vectors are orthogonal to each other and have length equal to unity.

The process of creating an orthonormal basis from any given basis is called **Gram-Schmidt orthogonalization**. We can set $\boldsymbol{u}_1$ to be $\boldsymbol{v}_1 / \parallel \boldsymbol{v}_1 \parallel$. Now, create $\boldsymbol{u}_2$ from $\boldsymbol{v}_2$ and $\boldsymbol{u}_1$ by "subtracting out" the component of $\boldsymbol{v}_2$ that is in the space spanned by $\boldsymbol{u}_1$. First compute

$$\boldsymbol{v}_2 - \Pi(\boldsymbol{v}_2|\boldsymbol{u}_1)$$

which is orthogonal to $\boldsymbol{u}_1$, and then "normalize," so that

$$\boldsymbol{u}_2 = \frac{\boldsymbol{v}_2 - \Pi(\boldsymbol{v}_2|\boldsymbol{u}_1)}{\parallel \boldsymbol{v}_2 - \Pi(\boldsymbol{v}_2|\boldsymbol{u}_1) \parallel}$$

Now we continue. Note that the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ span the same space as $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. To get $\boldsymbol{u}_3$, we take $\boldsymbol{v}_3$ and subtract off the projection of $\boldsymbol{v}_3$ onto $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Because the set of $\boldsymbol{v}_i$s was linearly independent, there is a residual from the projection. We get $\boldsymbol{u}_3$ by normalizing this projection. In general,

$$\boldsymbol{u}_i = \frac{\boldsymbol{v}_i - [\Pi(\boldsymbol{v}_i|\boldsymbol{u}_1) + \cdots + \Pi(\boldsymbol{v}_i|\boldsymbol{u}_{i-1})]}{\parallel \boldsymbol{v}_i - [\Pi(\boldsymbol{v}_i|\boldsymbol{u}_1) + \cdots + \Pi(\boldsymbol{v}_i|\boldsymbol{u}_{i-1})] \parallel}. \tag{1}$$

Note that because $\boldsymbol{u}_i$ is a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i$, for all $i = 1, n$, the linear space spanned by the $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_i$ is that spanned by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i$. If at any step, the denominator of (**??**) is zero, then at that step, the vector $\boldsymbol{v}_i$ is a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}$, and we conclude that the original set $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_i$ is not linearly independent. We can remove $\boldsymbol{u}_i$ from the set and continue when this happens, and we still get an orthonormal basis for the space spanned by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, but it will be of fewer than k dimensions.

This procedure is very useful, but we're not going to have to actually *do* it very often. In deriving things, we just *say*, let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be an orthonormal basis, and we know from the Gram-Schmidt orthogonalization theory that we can do this.

Exercise 14 *For each of the following sets of vectors, find an orthonormal basis for the space spanned by the vectors. (*Hint*: it might be easier to find an orthogonal basis and then normalize, instead of following (**??**) exactly.)*

(a) $\boldsymbol{x}_1 = (0, 0, 0, 0, 1, 1, 1, 1)'$, $\boldsymbol{x}_2 = (1, 1, 1, 1, 1, 1, 1, 1)'$.

(b) $\boldsymbol{x}_1 = (1,1,1,1,1,1,1)'$, $\boldsymbol{x}_2 = (1,2,3,4,5,6,7)'$.

We can project a vector $\boldsymbol{y}$ onto the space spanned by a set of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ by summing the projections of $\boldsymbol{y}$ onto the individual $\boldsymbol{x}_i$, *if* the vectors $\boldsymbol{x}_i$ form an orthogonal set. The orthogonalization gives us a way to project a vector $\boldsymbol{y}$ onto a subspace when the given basis is not orthogonal. We can first find an orthogonal basis, then project $\boldsymbol{y}$ onto each of these basis vectors, and add the projections.

Exercise 15 *For each of the following sets of vectors, use the results of Exercise* **??** *to project* $\boldsymbol{y} = (2,3,1,4,6,4,5,9)'$ *onto the space spanned by the vectors.*

(a) $\boldsymbol{x}_1 = (0,0,0,0,1,1,1,1)'$, $\boldsymbol{x}_2 = (1,1,1,1,1,1,1,1)'$.

(b) $\boldsymbol{x}_1 = (1,1,1,1,1,1,1)'$, $\boldsymbol{x}_2 = (1,2,3,4,5,6,7)'$.

Example: Consider the simple linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon},$$

or

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Then the vector $\hat{\boldsymbol{y}}$ is the projection of $\boldsymbol{y}$ onto the space V spanned by $\boldsymbol{v}_1 = (1,\ldots,1)'$ and $\boldsymbol{v}_2 = (x_1,\ldots,x_n)'$. Let's find an orthogonal basis $\boldsymbol{u}_1, \boldsymbol{u}_2$ for V. Let $\boldsymbol{u}_1 = \boldsymbol{v}_1$, and let $\boldsymbol{u}_2 = \boldsymbol{v}_2 - \Pi(\boldsymbol{v}_2|\boldsymbol{u}_1)$, that is, $\boldsymbol{u}_2 = (x_1 - \bar{x}, \ldots, x_n - \bar{x})'$. (For ease of presentation we won't normalize the $\boldsymbol{u}$ vectors.) The matrix equation is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \\ 1 & x_n - \bar{x} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

The projection of $\boldsymbol{y}$ onto the space spanned by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ can be written $\hat{\boldsymbol{y}} = \hat{\alpha}_0 \boldsymbol{u}_1 + \hat{\alpha}_1 \boldsymbol{u}_2$ and is the sum of the individual projections.

Exercise 16 (a) *Compute* $\hat{\alpha}_0$ *and* $\hat{\alpha}_1$.

(b) What are β_0 and β_1 in terms of α_0 and α_1?

This is a lot of work (especially for models with more than two parameters), so let's derive a matrix equation for projections onto subspaces. There is a particular result from linear algebra that we find useful here.

Review of Linear Algebra: Solving Quadratic Equations

Let $\boldsymbol{x}$ be a vector variable taking values in $\mathbb{R}^k$. A quadratic function in $\boldsymbol{x}$ can be written

$$f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} - 2\boldsymbol{b}'\boldsymbol{x} + c,$$

where $\boldsymbol{Q}$ is a constant $k \times k$ matrix, $\boldsymbol{b}$ is a constant k-dimensional (column) vector, and c is a scalar. If $\boldsymbol{Q}$ is *positive definite* there is a vector $\boldsymbol{x}$ that minimizes f; if $\boldsymbol{Q}$ is *negative definite* there is a vector $\boldsymbol{x}$ that maximizes f.

Proposition 1 *Suppose $\boldsymbol{Q}$ is positive definite. Then the value of $\boldsymbol{x}$ that minimizes f is $\boldsymbol{Q}^{-1}\boldsymbol{b}$.*

Proof: Let's *complete the square*. Because $\boldsymbol{Q}$ is nonnegative definite, we can do a *Cholesky decomposition*, that is, find a lower triangular matrix $\boldsymbol{B}$ such that $\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{B}'$ (see Appendix). If $\boldsymbol{Q}$ is positive definite, then $\boldsymbol{B}$ is nonsingular. Now, we plug this in to the quadratic equation. If we make the substitution $\boldsymbol{\theta} = \boldsymbol{B}'\boldsymbol{x}$, we have to minimize

$$\begin{aligned} f(\boldsymbol{\theta}) &= \boldsymbol{\theta}'\boldsymbol{\theta} - 2\boldsymbol{b}'(\boldsymbol{B}')^{-1}\boldsymbol{\theta} + c \\ &= \| \boldsymbol{\theta} - (\boldsymbol{B})^{-1}\boldsymbol{b} \|^2 + (\text{constant}). \end{aligned}$$

This expression is minimized when $\boldsymbol{\theta} = (\boldsymbol{B})^{-1}\boldsymbol{b}$, or making the substituion back to $\boldsymbol{x}$, when $\boldsymbol{B}'\boldsymbol{x} = (\boldsymbol{B})^{-1}\boldsymbol{b}$, or when $\boldsymbol{x} = (\boldsymbol{B}')^{-1}(\boldsymbol{B})^{-1}\boldsymbol{b} = \boldsymbol{Q}^{-1}\boldsymbol{b}$. ♣

Exercise 17 *Verify that $\boldsymbol{Q}$ is positive definite and minimize the function $f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + 2\boldsymbol{b}'\boldsymbol{x} + c$ where $c = 12$, $\boldsymbol{b} = (18, 9)'$, and*

$$\boldsymbol{Q} = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$$

Back to Projections

Suppose we have a collection of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ that is linearly independent but not necessarily orthogonal. The problem of finding the projection

$$\hat{\boldsymbol{y}} = \Pi(\boldsymbol{y}|\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k))$$

amounts to finding the coefficients of the linear combination

$$\hat{\boldsymbol{y}} = \hat{\beta}_1 \boldsymbol{x}_1 + \cdots + \hat{\beta}_k \boldsymbol{x}_k,$$

that minimizes the distance from $\boldsymbol{y}$ to $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$. (Note that the notation for the coefficients is chosen to be suggestive of the linear model.)

The square of the distance from $\boldsymbol{y}$ to an arbitrary vector $\boldsymbol{x} = \beta_1 \boldsymbol{x}_1 + \cdots + \beta_k \boldsymbol{x}_k$ in $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ is

$$\| \boldsymbol{y} - (\beta_1 \boldsymbol{x}_1 + \cdots + \beta_k \boldsymbol{x}_k) \|^2$$

Define $\boldsymbol{X}$ to be the $n \times k$ matrix whose columns are $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Then we want to minimize

$$f(\boldsymbol{\beta}) = \| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \|^2 = \boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{X} \boldsymbol{\beta} - 2\boldsymbol{y}' \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{y}' \boldsymbol{y}$$

which is a quadratic in $\boldsymbol{\beta}$. Note that $\boldsymbol{X}'\boldsymbol{X}$ is positive definite, because $\boldsymbol{a}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a} = \| \boldsymbol{X}\boldsymbol{a} \|^2 > 0$ for $\boldsymbol{a} \neq 0$ if $\boldsymbol{X}$ is full rank.

The solution to the quadratic equation is found using Proposition **??**.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{y}.$$

Note that if the columns of $\boldsymbol{X}$ form an orthonormal set of vectors in $\mathbb{R}^n$, this reduces to

$$\hat{\boldsymbol{\beta}} = \boldsymbol{X}' \boldsymbol{y},$$

or, $\hat{\boldsymbol{\beta}}_i = \langle \boldsymbol{x}_i, \boldsymbol{y} \rangle$, for $i = 1, \ldots, k$. (Recall the previous result about projections onto subspaces when an orthonormal basis is available.) Here $\hat{\beta}_i$ is the coefficient of the projection of $\boldsymbol{y}$ onto $\boldsymbol{x}_i$, because $\| \boldsymbol{x}_i \|^2 = 1$.

Definition: The notation $\mathcal{C}(\boldsymbol{X})$ indicates the linear space spanned by the columns of the matrix $\boldsymbol{X}$.

The Projection Matrix: Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ be a collection of linearly independent vectors in $\mathbb{R}^n$, and let the $n \times k$ matrix $\boldsymbol{X}$ have columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Note that for any $\boldsymbol{y} \in \mathbb{R}^n$, the coefficients of the projection of $\boldsymbol{y}$ onto $C(X)$ is given by the formula for $\hat{\boldsymbol{\beta}}$, the projection $\hat{\boldsymbol{y}}$ given by

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

We define the **projection matrix** for the subspace $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ to be

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.$$

Note that $\boldsymbol{P}$ is a $n \times n$ *idempotent* matrix, that is, $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$. This makes sense because if $\boldsymbol{y}$ is in the subspace $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$, $\hat{\boldsymbol{y}} = \boldsymbol{y}$. We can state this last claim as a theorem and prove it.

Theorem 2 *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ be a collection of linearly independent vectors in $I\!R^n$, and let $V = \mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$. If $\boldsymbol{y} \in V$, then $\boldsymbol{P}_V \boldsymbol{y} = \boldsymbol{y}$, where $\boldsymbol{P}_V$ is the projection matrix associated with V.*

Proof: If $\boldsymbol{X}$ is the $n \times k$ matrix whose columns are the vectors $\boldsymbol{x}_i$, then any vector $\boldsymbol{y} \in V$ can be written as $a_1\boldsymbol{x}_1 + \cdots + a_k\boldsymbol{x}_k$ which is $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a}$, where $\boldsymbol{a} = (a_1, \ldots, a_k)$. Then

$$\begin{aligned} \boldsymbol{P}\boldsymbol{y} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a} \\ &= \boldsymbol{X}\boldsymbol{a} = \boldsymbol{y}. \end{aligned}$$

Simple linear regression example again: Let $\boldsymbol{v}_1$ be the one-vector $(1, \ldots, 1)'$ and $\boldsymbol{v}_2 = (x_1, \ldots, x_n)'$. The vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the columns of the $n \times 2$ design matrix $\boldsymbol{X}$, and we wish to project a random "response" vector $\boldsymbol{y}$ onto the space V spanned by the columns of $\boldsymbol{X}$.

Suppose we want to find the projection matrix onto V. We could write $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, but the calculations might be easier if we do the following: Let $\boldsymbol{u}_1 = \boldsymbol{x}_1$ and $\boldsymbol{u}_2 = \boldsymbol{x}_2 - \bar{x}\boldsymbol{u}_1$, and let $\boldsymbol{U}$ be the $n \times 2$ matrix with columns $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Now the elements of $\boldsymbol{u}$ are $u_i = x_i - \bar{x}$ and $\boldsymbol{u}_1 \perp \boldsymbol{u}_2$. Because the space spanned by the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is the same as the space spanned by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$,

$$\boldsymbol{P} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}' = \begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum(x_i - \bar{x})^2} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 - \bar{x} & \cdots & x_n - \bar{x} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n} + \frac{(x_1-\bar{x})^2}{\sum(x_i-\bar{x})^2} & \frac{1}{n} + \frac{(x_1-\bar{x})(x_2-\bar{x})}{\sum(x_i-\bar{x})^2} & \cdots & \frac{1}{n} + \frac{(x_1-\bar{x})(x_n-\bar{x})}{\sum(x_i-\bar{x})^2} \\ \frac{1}{n} + \frac{(x_1-\bar{x})(x_2-\bar{x})}{\sum(x_i-\bar{x})^2} & \frac{1}{n} + \frac{(x_2-\bar{x})^2}{\sum(x_i-\bar{x})^2} & \cdots & \frac{1}{n} + \frac{(x_2-\bar{x})(x_n-\bar{x})}{\sum(x_i-\bar{x})^2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{n} + \frac{(x_1-\bar{x})(x_n-\bar{x})}{\sum(x_i-\bar{x})^2} & \frac{1}{n} + \frac{(x_2-\bar{x})(x_n-\bar{x})}{\sum(x_i-\bar{x})^2} & \cdots & \frac{1}{n} + \frac{(x_n-\bar{x})^2}{\sum(x_i-\bar{x})^2} \end{pmatrix}.$$

If we were to go through the trouble of computing $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, we would get the exact same matrix $\boldsymbol{P}$ (after a lot of simplifying of terms).

Exercise 18 *(continuation of simple linear regression example) Show directly that if $\boldsymbol{y}$ is a multiple of the one-vector, then $\boldsymbol{P}\boldsymbol{y} = \boldsymbol{y}$.*

Exercise 19 *For the one-way ANOVA problem with four levels and five observations per level, what is the projection matrix?*

The residual vector: The vector $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ is called the "residual vector."

Theorem 3 *Let $\hat{\boldsymbol{y}}$ be the projection of $\boldsymbol{y}$ onto the space V spanned by the columns of $\boldsymbol{X}$, and $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$. Then $\boldsymbol{e}$ is perpendicular to V.*

Proof: Any $\boldsymbol{v} \in V$ can be written as $\boldsymbol{X}\boldsymbol{a}$ for some vector $\boldsymbol{a} \in \mathbb{R}^k$. Then

$$\begin{aligned} \boldsymbol{e}'\boldsymbol{v} &= (\boldsymbol{y} - \hat{\boldsymbol{y}})'\boldsymbol{X}\boldsymbol{a} \\ &= \boldsymbol{y}'\boldsymbol{X}\boldsymbol{a} - \boldsymbol{y}'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{a} = 0. \end{aligned}$$

♣

Orthogonal Complements: Let V be a linear subspace of $\mathbb{R}^n$. The orthogonal complement of V, written $V^\perp$, is the set of all vectors in $\mathbb{R}^n$ that are orthogonal to V. Another name for $V^\perp$, is "perpendicular space."

Exercise 20 *Prove that $V^\perp$ is a subspace of $\mathbb{R}^n$.*

Example: Let $\boldsymbol{v}_1 = (1, 1, 0, 0)'$ and $\boldsymbol{v}_2 = (0, 0, 1, 1)'$, and let $V = \mathcal{L}(\boldsymbol{v}_1, \boldsymbol{v}_2)$. Clearly, a vector is in $V^\perp$ if and only if it is of the form $(a, -a, b, -b)'$. In fact, we see that the vectors $(1, -1, 0, 0)$ and $(0, 0, 1, -1)$ form a basis for $V^\perp$.

Exercise 21 *Let each of the following sets of vectors span a linear space V. Find a basis for $V^\perp$.*

(a) $\boldsymbol{v}_1 = (1, 0, 0, 0, 0)'$, $\boldsymbol{v}_2 = (0, 1, 0, 0, 0)'$, *and* $\boldsymbol{v}_3 = (0, 0, 0, 0, 1)'$

(b) $\boldsymbol{v}_1 = (1, 1, 0, 0, 0)'$, $\boldsymbol{v}_2 = (0, 0, 1, 1, 0)'$, *and* $\boldsymbol{v}_3 = (0, 0, 0, 0, 1)'$

(c) $\boldsymbol{v}_1 = (1, 1, 1, 0, 0, 0)'$ *and* $\boldsymbol{v}_2 = (0, 0, 0, 1, 1, 1)'$.

Theorem 4 *Let V be a linear subspace of $\mathbb{R}^n$, and let $V^\perp$ be its orthogonal complement. If the dimension of V is k, then the dimension of $V^\perp$ is $n - k$.*

Proof: Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be an orthonormal basis for $\mathbb{R}^n$, such that $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ is an orthonormal basis for V. Let U be the space spanned by $\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n$. We show that $U = V^\perp$. Clearly, all linear combinations of $\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n$ are orthogonal to V, because the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ form an orthogonal set. Therefore, $U \subseteq V^\perp$. Any vector $\boldsymbol{v} \in \mathbb{R}^n$ can be written as $a_1\boldsymbol{u}_1 + \cdots + a_n\boldsymbol{u}_n$ for some constants $a_1, \ldots, a_n$. Note that $\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle = a_i$ Therefore, if $\boldsymbol{v} \in V^\perp$, $a_1, \ldots, a_k = 0$, and we have $V^\perp \subseteq U$.

The trick of writing an orthonormal basis for $\mathbb{R}^n$, with the first k vectors in the basis spanning V is used to prove other results, such as the next two theorems about projections onto orthogonal subspaces. The first shows that the residual vector $\boldsymbol{y} - \hat{\boldsymbol{y}}$ where $\hat{\boldsymbol{y}}$ is the projection of $\boldsymbol{y}$ onto V, is the projection of $\boldsymbol{y}$ onto $V^\perp$.

Theorem 5 $\Pi(\boldsymbol{y}|V) + \Pi(\boldsymbol{y}|V^\perp) = \boldsymbol{y}$.

Proof: Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be an orthonormal basis for $\mathbb{R}^n$, where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ span V. Then we can write

$$\boldsymbol{y} = a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k + a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n,$$

and we know that each term is the projection of $\boldsymbol{y}$ onto the corresponding $\boldsymbol{u}_i$. We can use the result that the sum of projections onto mutually orthogonal vectors is the projection onto the space spanned by the vectors, to determine that $a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k$ is $\Pi(\boldsymbol{y}|V)$, and similarly that $a_{k+1}\boldsymbol{u}_{k+1} + \cdots + a_n\boldsymbol{u}_n$ is $\Pi(\boldsymbol{y}|V^\perp)$.

Theorem 6 *Let V_1 be a subspace of $\mathbb{R}^n$ spanned by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, and let V_2 be a subspace of $\mathbb{R}^n$ spanned by $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_{k+l}$. If V_1 and V_2 are orthogonal, so that any vector $\boldsymbol{v} \in V_1$ is orthogonal to the space V_2, and $V = \mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k+l})$, then $\Pi(\boldsymbol{y}|V_1) + \Pi(\boldsymbol{y}|V_2) = \Pi(\boldsymbol{y}|V)$.*

Proof: Homework assignment!