ORTHOGONAL PROJECTION

Orthogonal Projection

For $\mathbf{v} \in \mathcal{V}$, let $\mathbf{v} = \mathbf{m} + \mathbf{n}$, where $\mathbf{m} \in \mathcal{M}$ and $\mathbf{n} \in \mathcal{M}^{\perp}$.

- $\mathbf{m}$ is called the ***orthogonal projection*** of $\mathbf{v}$ onto $\mathcal{M}$.
- The projector $\mathbf{P}_{\mathcal{M}}$ onto $\mathcal{M}$ along $\mathcal{M}^{\perp}$ is called the ***orthogonal projector*** onto $\mathcal{M}$.
- $\mathbf{P}_{\mathcal{M}}$ is the unique linear operator such that $\mathbf{P}_{\mathcal{M}}\mathbf{v} = \mathbf{m}$ (see p. 386).

These ideas are illustrated illustrated in Figure 5.13.1 for $\mathcal{V} = \Re^3$.

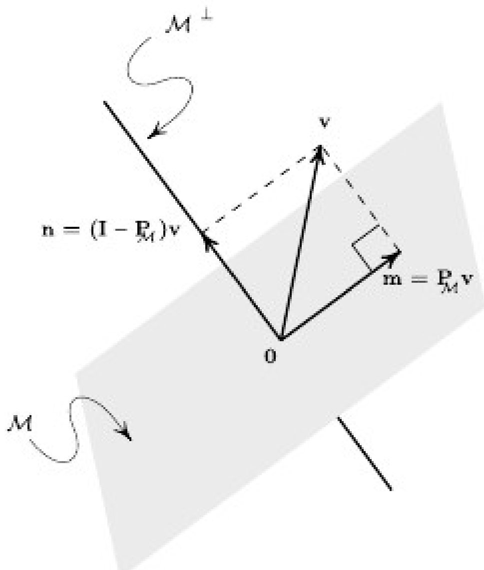

Figure 5.13.1

Constructing Orthogonal Projectors

Let $\mathcal{M}$ be an r-dimensional subspace of $\Re^n$, and let the columns of $\mathbf{M}_{n\times r}$ and $\mathbf{N}_{n\times n-r}$ be bases for $\mathcal{M}$ and $\mathcal{M}^\perp$, respectively. The orthogonal projectors onto $\mathcal{M}$ and $\mathcal{M}^\perp$ are

- $\mathbf{P}_{\mathcal{M}} = \mathbf{M}\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T$ and $\mathbf{P}_{\mathcal{M}^\perp} = \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T$. (5.13.3)

If $\mathbf{M}$ and $\mathbf{N}$ contain *orthonormal* bases for $\mathcal{M}$ and $\mathcal{M}^\perp$, then

- $\mathbf{P}_{\mathcal{M}} = \mathbf{M}\mathbf{M}^T$ and $\mathbf{P}_{\mathcal{M}^\perp} = \mathbf{N}\mathbf{N}^T$. (5.13.4)
- $\mathbf{P}_{\mathcal{M}} = \mathbf{U}\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^T$, where $\mathbf{U} = \left(\mathbf{M} \,|\, \mathbf{N}\right)$. (5.13.5)
- $\mathbf{P}_{\mathcal{M}^\perp} = \mathbf{I} - \mathbf{P}_{\mathcal{M}}$ in all cases. (5.13.6)

Problem: Let $\mathbf{u}_{n\times 1} \neq \mathbf{0}$, and consider the line $\mathcal{L} = span\{\mathbf{u}\}$. Construct the orthogonal projector onto $\mathcal{L}$, and then determine the orthogonal projection of a vector $\mathbf{x}_{n\times 1}$ onto $\mathcal{L}$.

Solution: The vector $\mathbf{u}$ by itself is a basis for $\mathcal{L}$, so, according to (5.13.3),

$$\mathbf{P}_{\mathcal{L}} = \mathbf{u}\left(\mathbf{u}^T\mathbf{u}\right)^{-1}\mathbf{u}^T = \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T\mathbf{u}}$$

is the orthogonal projector onto $\mathcal{L}$. The orthogonal projection of a vector $\mathbf{x}$ onto $\mathcal{L}$ is therefore given by

$$\mathbf{P}_{\mathcal{L}}\mathbf{x} = \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T\mathbf{u}}\mathbf{x} = \left(\frac{\mathbf{u}^T\mathbf{x}}{\mathbf{u}^T\mathbf{u}}\right)\mathbf{u}.$$

Note: If $\|\mathbf{u}\|_2 = 1$, then $\mathbf{P}_{\mathcal{L}} = \mathbf{u}\mathbf{u}^T$, so $\mathbf{P}_{\mathcal{L}}\mathbf{x} = \mathbf{u}\mathbf{u}^T\mathbf{x} = (\mathbf{u}^T\mathbf{x})\mathbf{u}$, and

$$\|\mathbf{P}_{\mathcal{L}}\mathbf{x}\|_2 = |\mathbf{u}^T\mathbf{x}|\;\|\mathbf{u}\|_2 = |\mathbf{u}^T\mathbf{x}|.$$

This yields a geometrical interpretation for the magnitude of the standard inner product. It says that if $\mathbf{u}$ is a vector of unit length in $\mathcal{L}$, then, as illustrated in Figure 5.13.2, $|\mathbf{u}^T\mathbf{x}|$ is the length of the orthogonal projection of $\mathbf{x}$ onto the line spanned by $\mathbf{u}$.

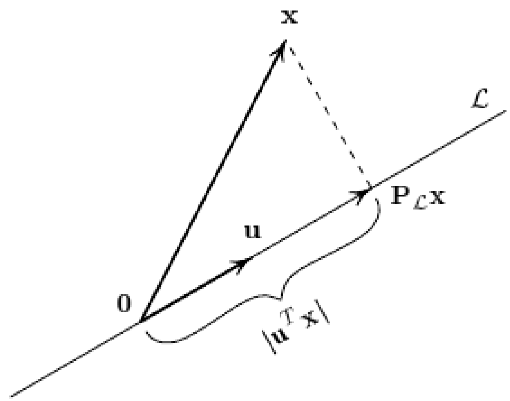

FIGURE 5.13.2

Finally, notice that since $\mathbf{P}_\mathcal{L} = \mathbf{u}\mathbf{u}^T$ is the orthogonal projector onto $\mathcal{L}$, it must be the case that $\mathbf{P}_{\mathcal{L}^\perp} = \mathbf{I} - \mathbf{P}_\mathcal{L} = \mathbf{I} - \mathbf{u}\mathbf{u}^T$ is the orthogonal projection onto $\mathcal{L}^\perp$. This was called an ***elementary orthogonal projector*** on p. 322—go back and reexamine Figure 5.6.1.

Of course, not all projectors are orthogonal projectors, so a natural question to ask is, "What characteristic features distinguish orthogonal projectors from more general oblique projectors?" Some answers are given below.

Orthogonal Projectors

Suppose that $\mathbf{P} \in \Re^{n\times n}$ is a projector—i.e., $\mathbf{P}^2 = \mathbf{P}$. The following statements are equivalent to saying that $\mathbf{P}$ is an *orthogonal* projector.

- $R(\mathbf{P}) \perp N(\mathbf{P})$. (5.13.8)
- $\mathbf{P}^T = \mathbf{P}$ (i.e., orthogonal projector $\iff \mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$). (5.13.9)
- $\|\mathbf{P}\|_2 = 1$ for the matrix 2-norm (p. 281). (5.13.10)

Proof: Every projector projects vectors onto its range along (parallel to) its nullspace, so statement (5.13.8) is essentially a restatement of the definition of an orthogonal projector. To prove (5.13.9), note that if P is an orthogonal projector, then (5.13.3) insures that P is symmetric. Conversely, if a projector P is symmetric, then it must be an orthogonal projector because (5.11.5) on p. 405 allows us to write

$$\mathbf{P} = \mathbf{P}^T \implies R(\mathbf{P}) = R\left(\mathbf{P}^T\right) \implies R(\mathbf{P}) \perp N(\mathbf{P}).$$

To see why (5.13.10) characterizes projectors that are orthogonal, refer back to Example 5.9.2 on p. 389 (or look ahead to (5.15.3)) and note that $\|\mathbf{P}\|_2 = 1/\sin\theta$, where θ is the angle between $R(\mathbf{P})$ and $N(\mathbf{P})$. This makes it clear that $\|\mathbf{P}\|_2 \geq 1$ for all projectors, and $\|\mathbf{P}\|_2 = 1$ if and only if $\theta = \pi/2$, (i.e., if and only if $R(\mathbf{P}) \perp N(\mathbf{P})$). ■

Problem: For $\mathbf{A} \in \Re^{m\times n}$ such that $rank(\mathbf{A}) = r$, describe the orthogonal projectors onto each of the four fundamental subspaces of $\mathbf{A}$.

Solution 1: Let $\mathbf{B}_{m\times r}$ and $\mathbf{N}_{n\times n-r}$ be matrices whose columns are bases for $R(\mathbf{A})$ and $N(\mathbf{A})$, respectively—e.g., $\mathbf{B}$ might contain the basic columns of $\mathbf{A}$. The orthogonal decomposition theorem on p. 405 says $R(\mathbf{A})^\perp = N(\mathbf{A}^T)$ and $N(\mathbf{A})^\perp = R(\mathbf{A}^T)$, so, by making use of (5.13.3) and (5.13.6), we can write

$$\mathbf{P}_{R(\mathbf{A})} = \mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T,$$

$$\mathbf{P}_{N(\mathbf{A}^T)} = \mathbf{P}_{R(\mathbf{A})^\perp} = \mathbf{I} - \mathbf{P}_{R(\mathbf{A})} = \mathbf{I} - \mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T,$$

$$\mathbf{P}_{N(\mathbf{A})} = \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T,$$

$$\mathbf{P}_{R(\mathbf{A}^T)} = \mathbf{P}_{N(\mathbf{A})^\perp} = \mathbf{I} - \mathbf{P}_{N(\mathbf{A})} = \mathbf{I} - \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T.$$

Note: If $rank(\mathbf{A}) = n$, then all columns of $\mathbf{A}$ are basic and

$$\mathbf{P}_{R(\mathbf{A})} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T. \tag{5.13.11}$$

so, according to (5.13.4),

$$\begin{aligned} \mathbf{P}_{R(\mathbf{A})} &= \mathbf{U}_1\mathbf{U}_1^T = \mathbf{A}\mathbf{A}^\dagger, \quad & \mathbf{P}_{N(\mathbf{A}^T)} &= \mathbf{I} - \mathbf{P}_{R(\mathbf{A})} = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger, \\ \mathbf{P}_{R(\mathbf{A}^T)} &= \mathbf{V}_1\mathbf{V}_1^T = \mathbf{A}^\dagger\mathbf{A}, \quad & \mathbf{P}_{N(\mathbf{A})} &= \mathbf{I} - \mathbf{P}_{R(\mathbf{A}^T)} = \mathbf{I} - \mathbf{A}^\dagger\mathbf{A}. \end{aligned} \tag{5.13.12}$$

The notion of orthogonal projection in higher-dimensional spaces is consistent with the visual geometry in $\Re^2$ and $\Re^3$. In particular, it is visually evident from Figure 5.13.4 that if $\mathcal{M}$ is a subspace of $\Re^3$, and if $\mathbf{b}$ is a vector outside of $\mathcal{M}$, then the point in $\mathcal{M}$ that is closest to $\mathbf{b}$ is $\mathbf{p} = \mathbf{P}_{\mathcal{M}}\mathbf{b}$, the orthogonal projection of $\mathbf{b}$ onto $\mathcal{M}$.

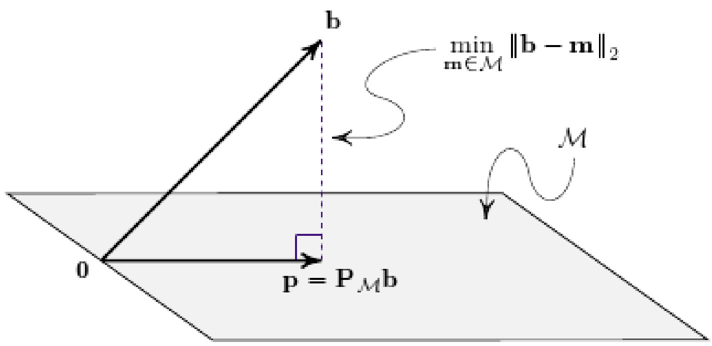

Figure 5.13.4

The situation is exactly the same in higher dimensions. But rather than using our eyes to understand why, we use mathematics—it's surprising just how easy it is to "see" such things in abstract spaces.

Closest Point Theorem

Let $\mathcal{M}$ be a subspace of an inner-product space $\mathcal{V}$, and let $\mathbf{b}$ be a vector in $\mathcal{V}$. The unique vector in $\mathcal{M}$ that is closest to $\mathbf{b}$ is $\mathbf{p} = \mathbf{P}_{\mathcal{M}}\mathbf{b}$, the orthogonal projection of $\mathbf{b}$ onto $\mathcal{M}$. In other words,

$$\min_{\mathbf{m}\in\mathcal{M}} \|\mathbf{b}-\mathbf{m}\|_2 = \|\mathbf{b}-\mathbf{P}_{\mathcal{M}}\mathbf{b}\|_2 = dist\,(\mathbf{b},\mathcal{M}). \tag{5.13.13}$$

This is called the ***orthogonal distance*** between $\mathbf{b}$ and $\mathcal{M}$.

Proof. If $\mathbf{p} = \mathbf{P}_{\mathcal{M}}\mathbf{b}$, then $\mathbf{p}-\mathbf{m} \in \mathcal{M}$ for all $\mathbf{m} \in \mathcal{M}$, and

$$\mathbf{b}-\mathbf{p} = (\mathbf{I}-\mathbf{P}_{\mathcal{M}})\mathbf{b} \in \mathcal{M}^{\perp},$$

so $(\mathbf{p}-\mathbf{m}) \perp (\mathbf{b}-\mathbf{p})$. The Pythagorean theorem says $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ whenever $\mathbf{x} \perp \mathbf{y}$ (recall Exercise 5.4.14), and hence

$$\|\mathbf{b}-\mathbf{m}\|_2^2 = \|\mathbf{b}-\mathbf{p}+\mathbf{p}-\mathbf{m}\|_2^2 = \|\mathbf{b}-\mathbf{p}\|_2^2 + \|\mathbf{p}-\mathbf{m}\|_2^2 \geq \|\mathbf{p}-\mathbf{m}\|_2^2.$$

In other words, $\min_{\mathbf{m}\in\mathcal{M}} \|\mathbf{b}-\mathbf{m}\|_2 = \|\mathbf{b}-\mathbf{p}\|_2$. Now argue that there is not another point in $\mathcal{M}$ that is as close to $\mathbf{b}$ as $\mathbf{p}$ is. If $\widehat{\mathbf{m}} \in \mathcal{M}$ such that $\|\mathbf{b}-\widehat{\mathbf{m}}\|_2 = \|\mathbf{b}-\mathbf{p}\|_2$, then by using the Pythagorean theorem again we see

$$\|\mathbf{b}-\widehat{\mathbf{m}}\|_2^2 = \|\mathbf{b}-\mathbf{p}+\mathbf{p}-\widehat{\mathbf{m}}\|_2^2 = \|\mathbf{b}-\mathbf{p}\|_2^2 + \|\mathbf{p}-\widehat{\mathbf{m}}\|_2^2 \implies \|\mathbf{p}-\widehat{\mathbf{m}}\|_2 = 0,$$

and thus $\widehat{\mathbf{m}} = \mathbf{p}$. ■

To illustrate some of the previous ideas, consider $\Re^{n\times n}$ with the inner product $\langle \mathbf{A}|\mathbf{B}\rangle = trace\left(\mathbf{A}^T\mathbf{B}\right)$. If $\mathcal{S}_n$ is the subspace of $n\times n$ real-symmetric matrices, then each of the following statements is true.

- $\mathcal{S}_n^{\perp} =$ the subspace $\mathcal{K}_n$ of $n\times n$ skew-symmetric matrices.
  - ▷ $\mathcal{S}_n \perp \mathcal{K}_n$ because for all $\mathbf{S}\in\mathcal{S}_n$ and $\mathbf{K}\in\mathcal{K}_n$,

$$\begin{aligned}\langle \mathbf{S}|\mathbf{K}\rangle &= trace\left(\mathbf{S}^T\mathbf{K}\right) = -trace\left(\mathbf{S}\mathbf{K}^T\right) = -trace\left(\mathbf{S}\mathbf{K}^T\right)^T \\ &= -trace\left(\mathbf{K}\mathbf{S}^T\right) = -trace\left(\mathbf{S}^T\mathbf{K}\right) = -\langle \mathbf{S}|\mathbf{K}\rangle \\ \implies &\quad \langle \mathbf{S}|\mathbf{K}\rangle = 0.\end{aligned}$$

  - ▷ $\Re^{n\times n} = \mathcal{S}_n \oplus \mathcal{K}_n$ because every $\mathbf{A}\in\Re^{n\times n}$ can be uniquely expressed as the sum of a symmetric and a skew-symmetric matrix by writing

$$\mathbf{A} = \frac{\mathbf{A}+\mathbf{A}^T}{2} + \frac{\mathbf{A}-\mathbf{A}^T}{2} \qquad \text{(recall (5.9.3) and Exercise 3.2.6).}$$

- The orthogonal projection of $\mathbf{A}\in\Re^{n\times n}$ onto $\mathcal{S}_n$ is $\mathbf{P}(\mathbf{A}) = (\mathbf{A}+\mathbf{A}^T)/2$.
- The closest symmetric matrix to $\mathbf{A}\in\Re^{n\times n}$ is $\mathbf{P}(\mathbf{A}) = (\mathbf{A}+\mathbf{A}^T)/2$.
- The distance from $\mathbf{A}\in\Re^{n\times n}$ to $\mathcal{S}_n$ (the deviation from symmetry) is

$$dist(\mathbf{A},\mathcal{S}_n) = \|\mathbf{A}-\mathbf{P}(\mathbf{A})\|_F = \left\|(\mathbf{A}-\mathbf{A}^T)/2\right\|_F = \sqrt{\frac{trace\left(\mathbf{A}^T\mathbf{A}\right)-trace\left(\mathbf{A}^2\right)}{2}}.$$

Affine Projections. If $\mathbf{v} \neq \mathbf{0}$ is a vector in a space $\mathcal{V}$, and if $\mathcal{M}$ is a subspace of $\mathcal{V}$, then the set of points $\mathcal{A} = \mathbf{v} + \mathcal{M}$ is called an ***affine space*** in $\mathcal{V}$. Strictly speaking, $\mathcal{A}$ is not a subspace (e.g., it doesn't contain $\mathbf{0}$), but, as depicted in Figure 5.13.5, $\mathcal{A}$ is the translate of a subspace—i.e., $\mathcal{A}$ is just a copy of $\mathcal{M}$ that has been translated away from the origin through $\mathbf{v}$. Consequently, notions such as projection onto $\mathcal{A}$ and points closest to $\mathcal{A}$ are analogous to the corresponding concepts for subspaces.

Problem: For $\mathbf{b} \in \mathcal{V}$, determine the point $\mathbf{p}$ in $\mathcal{A} = \mathbf{v} + \mathcal{M}$ that is closest to $\mathbf{b}$. In other words, explain how to project $\mathbf{b}$ orthogonally onto $\mathcal{A}$.

Solution: The trick is to subtract $\mathbf{v}$ from $\mathbf{b}$ as well as from everything in $\mathcal{A}$ to put things back into the context of subspaces where we already know the answers. As illustrated in Figure 5.13.5, this moves $\mathcal{A}$ back down to $\mathcal{M}$, and it translates $\mathbf{v} \to \mathbf{0}$, $\mathbf{b} \to (\mathbf{b} - \mathbf{v})$, and $\mathbf{p} \to (\mathbf{p} - \mathbf{v})$.

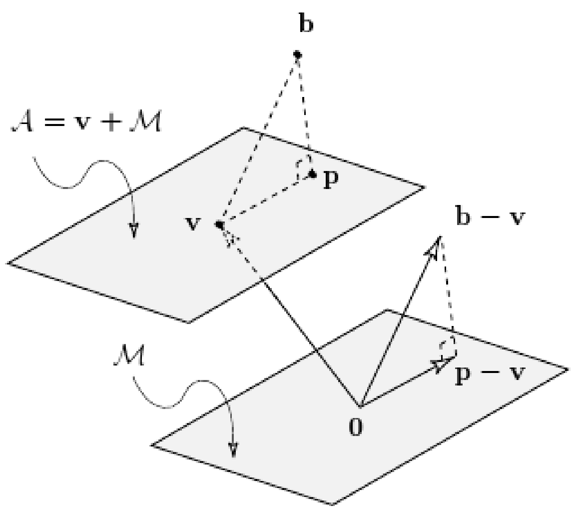

FIGURE 5.13.5

If $\mathbf{p}$ is to be the orthogonal projection of $\mathbf{b}$ onto $\mathcal{A}$, then $\mathbf{p} - \mathbf{v}$ must be the orthogonal projection of $\mathbf{b} - \mathbf{v}$ onto $\mathcal{M}$, so

$$\mathbf{p} - \mathbf{v} = \mathbf{P}_{\mathcal{M}}(\mathbf{b} - \mathbf{v}) \implies \mathbf{p} = \mathbf{v} + \mathbf{P}_{\mathcal{M}}(\mathbf{b} - \mathbf{v}), \tag{5.13.14}$$

and thus $\mathbf{p}$ is the point in $\mathcal{A}$ that is closest to $\mathbf{b}$. Applications to the solution of linear systems are developed in Exercises 5.13.17–5.13.22.

We are now in a position to replace the classical calculus-based theory of least squares presented in §4.6 with a more modern vector space

development.
In addition to being straightforward, the modern geometrical approach puts the entire least squares picture in much sharper focus. Viewing concepts from more than one perspective generally produces deeper understanding, and this is particularly true for the theory of least squares.

Recall from p. 226 that for an inconsistent system $\mathbf{A}_{m\times n}\mathbf{x} = \mathbf{b}$, the object of the least squares problem is to find vectors $\mathbf{x}$ that minimize the quantity

$$(\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2 . \tag{5.13.15}$$

The classical development in §4.6 relies on calculus to argue that the set of vectors $\mathbf{x}$ that minimize (5.13.15) is exactly the set that solves the (always consistent) system of normal equations $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$. In the context of the closest point theorem the least squares problem asks for vectors $\mathbf{x}$ such that $\mathbf{Ax}$ is as close to $\mathbf{b}$ as possible. But $\mathbf{Ax}$ is always a vector in $R(\mathbf{A})$, and the closest point theorem says that the vector in $R(\mathbf{A})$ that is closest to $\mathbf{b}$ is $\mathbf{P}_{R(\mathbf{A})}\mathbf{b}$, the orthogonal projection of $\mathbf{b}$ onto $R(\mathbf{A})$. Figure 5.13.6 illustrates the situation in $\Re^3$.

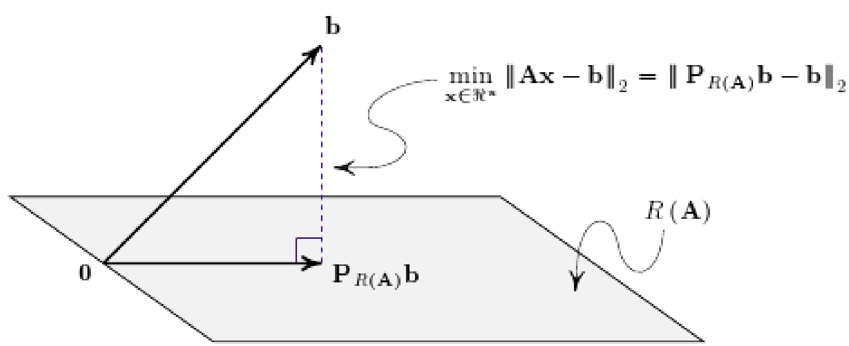

FIGURE 5.13.6

So the least squares problem boils down to finding vectors $\mathbf{x}$ such that

$$\mathbf{Ax} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b}.$$

But this system is equivalent to the system of normal equations because

$$\begin{aligned}\mathbf{Ax} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b} &\iff \mathbf{P}_{R(\mathbf{A})}\mathbf{Ax} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b} \\ &\iff \mathbf{P}_{R(\mathbf{A})}(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}\end{aligned}$$

$$\begin{aligned}
&\iff (\mathbf{Ax}-\mathbf{b}) \in N\left(\mathbf{P}_{R(\mathbf{A})}\right) = R\left(\mathbf{A}\right)^{\perp} = N\left(\mathbf{A}^T\right)\\
&\iff \mathbf{A}^T(\mathbf{Ax}-\mathbf{b}) = \mathbf{0}\\
&\iff \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}.
\end{aligned}$$

Characterizing the set of least squares solutions as the solutions to $\mathbf{Ax} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b}$ makes it obvious that $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$ is a particular least squares solution because (5.13.12) insures $\mathbf{AA}^{\dagger} = \mathbf{P}_{R(\mathbf{A})}$, and thus

$$\mathbf{A}(\mathbf{A}^{\dagger}\mathbf{b}) = \mathbf{P}_{R(\mathbf{A})}\mathbf{b}.$$

Furthermore, since $\mathbf{A}^{\dagger}\mathbf{b}$ is a particular solution of $\mathbf{Ax} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b}$, the general solution—i.e., the set of all least squares solutions—must be the affine space $\mathcal{S} = \mathbf{A}^{\dagger}\mathbf{b} + N\left(\mathbf{A}\right)$. Finally, the fact that $\mathbf{A}^{\dagger}\mathbf{b}$ is the least squares solution of minimal norm follows from Example 5.13.5 together with

$$R\left(\mathbf{A}^{\dagger}\right) = R\left(\mathbf{A}^T\right) = N\left(\mathbf{A}\right)^{\perp} \quad \text{(see part (g) of Exercise 5.12.16)}$$

because (5.13.14) insures that the point in $\mathcal{S}$ that is closest to the origin is

$$\mathbf{p} = \mathbf{A}^{\dagger}\mathbf{b} + \mathbf{P}_{N(\mathbf{A})}(\mathbf{0} - \mathbf{A}^{\dagger}\mathbf{b}) = \mathbf{A}^{\dagger}\mathbf{b}.$$

The classical development in §4.6 based on partial differentiation is not easily generalized to cover the case of complex matrices, but the vector space approach given in this example trivially extends to complex matrices by simply replacing $(\star)^T$ by $(\star)^*$.

Below is a summary of some of the major points concerning the theory of least squares.

Least Squares Solutions

Each of the following four statements is equivalent to saying that $\widehat{\mathbf{x}}$ is a least squares solution for a possibly inconsistent linear system $\mathbf{Ax} = \mathbf{b}$.

- $\left\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{b}\right\|_2 = \min_{\mathbf{x}\in\Re^n} \left\|\mathbf{Ax} - \mathbf{b}\right\|_2.$ (5.13.16)
- $\mathbf{A}\widehat{\mathbf{x}} = \mathbf{P}_{R(\mathbf{A})}\mathbf{b}.$ (5.13.17)
- $\mathbf{A}^T\mathbf{A}\widehat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$ ($\mathbf{A}^*\mathbf{A}\widehat{\mathbf{x}} = \mathbf{A}^*\mathbf{b}$ when $\mathbf{A} \in \mathcal{C}^{m\times n}$). (5.13.18)
- $\widehat{\mathbf{x}} \in \mathbf{A}^{\dagger}\mathbf{b} + N\left(\mathbf{A}\right)$ ($\mathbf{A}^{\dagger}\mathbf{b}$ is the minimal 2-norm LSS). (5.13.19)

Orthogonal Projections

- Some Properties:
  - Let $\text{rank}(\mathbf{A}) = r$ and partition $\mathbf{U}$ and $\mathbf{V}$

$$\mathbf{U} = [\mathbf{U}_r, \mathbf{U}_{m-r}], \qquad \mathbf{V} = [\mathbf{V}_r, \mathbf{V}_{n-r}] \tag{9}$$

    * $\mathbf{U}_r \in \Re^{m\times r}$, $\mathbf{U}_{m-r} \in \Re^{m\times(m-r)}$
    * $\mathbf{V}_r \in \Re^{n\times r}$, $\mathbf{V}_{n-r} \in \Re^{n\times(n-r)}$
  - $\mathbf{U}_r\mathbf{U}_r^T$ is a projection onto $\text{range}(\mathbf{A})$
  - $\mathbf{U}_{m-r}\mathbf{U}_{m-r}^T$ is a projection onto $\text{null}(\mathbf{A}^T)$
    * the orthogonal complement of $\text{range}(\mathbf{A})$
  - $\mathbf{V}_r\mathbf{V}_r^T$ is a projection onto $\text{range}(\mathbf{A}^T)$
    * the orthogonal complement of $\text{null}(\mathbf{A})$
  - $\mathbf{V}_{n-r}\mathbf{V}_{n-r}^T$ is a projection onto $\text{null}(\mathbf{A})$

Orthogonal Projections

- *Example 1.* The singular value decomposition of

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 0 & 5 \\ 3 & 1 & 8 \end{bmatrix}$$

is

$$\mathbf{U} = \begin{bmatrix} 0.3393 & -0.7427 & -0.5774 \\ 0.4735 & 0.6652 & -0.5774 \\ 0.8128 & -0.0775 & 0.5774 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 10.5826 & 0 & 0 \\ 0 & 1.4174 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 0.3393 & -0.7427 & 0.5774 \\ 0.1089 & -0.5786 & -0.8083 \\ 0.9344 & 0.3371 & -0.1155 \end{bmatrix}$$

- With $\sigma_3 = 0$ we have $r = 2$
- Verify that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

Orthogonal Projections

- Partition $\mathbf{U}$ and $\mathbf{V}$

$$\mathbf{U}_r = \begin{bmatrix} 0.3393 & -0.7427 \\ 0.4735 & 0.6652 \\ 0.8128 & -0.0775 \end{bmatrix}, \qquad \mathbf{U}_{m-r} = \begin{bmatrix} -0.5774 \\ -0.5774 \\ 0.5774 \end{bmatrix}$$

$$\mathbf{V}_r = \begin{bmatrix} 0.3393 & -0.7427 \\ 0.1089 & -0.5786 \\ 0.9344 & 0.3371 \end{bmatrix}, \qquad \mathbf{V}_{n-r} = \begin{bmatrix} 0.5774 \\ -0.8083 \\ -0.1155 \end{bmatrix}$$

- The projections

$$\mathbf{U}_r\mathbf{U}_r^T = \begin{bmatrix} 0.6667 & -0.3333 & 0.3333 \\ -0.3333 & 0.6667 & 0.3333 \\ 0.3333 & 0.3333 & 0.6667 \end{bmatrix}$$

$$\mathbf{U}_{m-r}\mathbf{U}_{m-r}^T = \begin{bmatrix} 0.3333 & 0.3333 & -0.3333 \\ 0.3333 & 0.3333 & -0.3333 \\ -0.3333 & -0.3333 & 0.3333 \end{bmatrix}$$

$$\mathbf{V}_r\mathbf{V}_r^T = \begin{bmatrix} 0.6667 & 0.4667 & 0.0667 \\ 0.4667 & 0.3467 & -0.0933 \\ 0.0667 & -0.0933 & 0.9867 \end{bmatrix}$$

$$\mathbf{V}_{n-r}\mathbf{V}_{n-r}^T = \begin{bmatrix} 0.3333 & -0.4667 & -0.0667 \\ -0.4667 & 0.6533 & 0.0933 \\ -0.0667 & 0.0933 & 0.0133 \end{bmatrix}$$

Orthogonal Projections

- Verify that $\mathbf{U}_r^T \mathbf{U}_r = \mathbf{I}$, etc
- Verify that
  - $\mathbf{U}_r \mathbf{U}_r^T + \mathbf{U}_{n-r} \mathbf{U}_{n-r}^T = \mathbf{I}$
  - $\mathbf{V}_r \mathbf{V}_r^T + \mathbf{V}_{n-r} \mathbf{V}_{n-r}^T = \mathbf{I}$
- Choose $\mathbf{x} = [1, 0, 0]^T$ and calculate

$$\mathbf{y} = \mathbf{V}_{n-r} \mathbf{V}_{n-r}^T \mathbf{x} = \begin{bmatrix} 0.3333 \\ -0.4667 \\ -0.0667 \end{bmatrix}$$

  - Verify that $\mathbf{A}\mathbf{y} = 0$
  - Verify that $\mathbf{A}\mathbf{V}_{n-r} \mathbf{V}_{n-r}^T = \mathbf{0}$
- Similarly, verify that $\mathbf{A}^T \mathbf{U}_{m-r} \mathbf{U}_{m-r}^T = \mathbf{0}$