3. Applications of Orthogonal Complements and Orthogonal Projection

3.1 Definition: For two affine spaces P and Q in $\mathbb{R}^n$, we define the **distance** between P and Q to be

$$\text{dist}(P,Q) = \min\left\{\text{dist}(a,b)\big| a \in P, b \in Q\right\}.$$

3.2 Theorem: *Let p and q be points in $\mathbb{R}^n$, let U and V be subspaces of $\mathbb{R}^n$, and let $P = p + U$ and $Q = q + V$. Then*

$$\text{dist}(P,Q) = \left\|\text{Proj}_{(U+V)^\perp}(p-q)\right\|.$$

Proof: We have

$$\begin{aligned}
\text{dist}(P,Q) &= \min\left\{\text{dist}(x,y)\big| x \in P, y \in Q\right\} \\
&= \min\left\{\text{dist}(p+u, q+v)\big| u \in U, v \in V\right\} \\
&= \min\left\{\|(q+v)-(p+u)\| \,\big| u \in U, v \in V\right\} \\
&= \min\left\{\|(q-p)-(u-v)\| \,\big| u \in U, v \in V\right\} \\
&= \min\left\{\|(q-p)-w\| \,\big| w \in U+V\right\} \\
&= \big\|(q-p) - \text{Proj}_{U+V}(q-p)\big\| \\
&= \big\|\text{Proj}_{(U+V)^\perp}(q-p)\big\|
\end{aligned}$$

where, on the second last line, we used the fact that $\text{Proj}_{U+V}(q-p)$ is the (unique) point on $U+V$ which is nearest to $q-p$.

3.3 Definition: For two subspaces $U, V \subseteq \mathbb{R}^n$, we define the **angle** between U and V, which we write as $\theta(U,V)$, as follows.

(1) If $U \subseteq V$ or $V \subseteq U$ then we define $\theta(U,V) = 0$.

(2) Otherwise, if $U \cap V = \{0\}$ then we define

$$\theta(U,V) = \min\left\{\theta(u,v)\Big| 0 \neq u \in U, 0 \neq v \in V\right\},$$

(3) and if $U \cap V = W \neq \{0\}$ then we define $\theta(U,V) = \theta(U \cap W^\perp, V \cap W^\perp)$, noting that $(U \cap W^\perp) \cap (V \cap W^\perp) = (U \cap V) \cap W^\perp = W \cap W^\perp = \{0\}$.

We define the **angle** between two affine spaces in $\mathbb{R}^n$ to be the angle between their associated vector spaces.

3.4 Theorem: *Let $\{0\} \neq U, V \subseteq \mathbb{R}^n$ be non-trivial subspaces with $U \cap V = \{0\}$. Then*
(1) In the case that $\dim(U) = 1$ with $U = \text{Span}\,\{u\}$ where $u \in \mathbb{R}^n$ with $\|u\| = 1$, we have

$$\cos\,\theta(U,V) = \big\|\text{Proj}_V(u)\big\|.$$

(2) In general, we have

$$\cos\,\theta(U,V) = \max_{u\in U, \|u\|=1} \big\|\text{Proj}_V(u)\big\|.$$

Proof: To prove part (1), suppose that $U = \text{Span}\,\{u\}$ where $u \in \mathbb{R}^n$ with $\|u\| = 1$. Since every nonzero vector in U is of the form tu for some $0 \neq t \in \mathbb{R}$, by the definition of $\theta(U,V)$ we have

$$\theta(U,V) = \min\left\{\theta(tu,w)\big|0 \neq t \in \mathbb{R}, 0 \neq w \in V\right\}.$$

Since $\theta(tu,w) = \theta(u,\pm w)$ $\big($indeed when $t > 0$ we have $\theta(tu,w) = \theta(u,w)$ and when $t < 0$ we have $\theta(tu,w) = \theta(u,-w)$ $\big)$ it follows that

$$\theta(U,V) = \min\left\{\theta(u,w)\big|0 \neq w \in V\right\}.$$

If $u \in V^\perp$ then $\text{Proj}_V(u) = 0$ and for all $0 \neq w \in V$ we have $u \cdot w = 0$ so that $\theta(u,w) = \frac{\pi}{2}$, and so $\theta(U,V) = \min\left\{\theta(u,w)\big|0 \neq w \in V\right\} = \frac{\pi}{2}$, and hence $\cos\,\theta(U,V) = 0 = \big\|\text{Proj}_V(u)\big\|$. Suppose that $u \notin V^\perp$ and let

$$v = \text{Proj}_V(u).$$

Note that $v \neq 0$ since $u \notin V^\perp$. By Trigonometric Ratios, we have

$$\cos\theta(u,v) = \tfrac{\|v\|}{\|u\|} = \|v\|.$$

Since $\cos\theta(u,v) \geq 0$ we have $\theta(u,v) \in \left[0, \frac{\pi}{2}\right]$. Let $0 \neq w \in V$ and let

$$y = \text{Proj}_w(U) = \tfrac{u \cdot w}{\|w\|^2}\, w.$$

If $y = 0$ then $u \cdot w = 0$ and so $\theta(u,w) = \frac{\pi}{2} \geq \theta(u,v)$. Suppose that $y \neq 0$. By Trigonometric Ratios, we have $\cos(u,y) = \frac{\|y\|}{\|u\|} = \|y\|$. Since $\theta(u,y) \geq 0$ we have $\theta(u,y) \in \left[0, \frac{\pi}{2}\right]$. If $u \cdot w < 0$ so that $\theta(u,w) = \pi - \theta(u,y) \in \left[\frac{\pi}{2}, \pi\right]$, then we have $\theta(u,w) \geq \theta(u,v)$. If $u \cdot w > 0$ so that $\theta(u,w) = \theta(u,y)$, then by Trigonometric Ratios, and since v is the point in V nearest to u, we have

$$\sin\theta(u,w) = \sin\theta(u,y) = \tfrac{\|u-y\|}{\|u\|} = \|u-y\| \geq \|u-v\| = \tfrac{\|u-v\|}{\|u\|} = \sin\theta(u,v)$$

and hence $\theta(u,w) \geq \theta(u,v)$. Thus for all $0 \neq w \in V$ we have $\theta(u,w) \geq \theta(u,v)$, where $v = \text{Proj}_V(u)$. It follows that $\theta(U,V) = \min\left\{\theta(u,w)\big|0 \neq w \in V\right\} = \theta(u,v)$ and hence that $\cos\,\theta(U,V) = \cos\theta(u,v) = \|v\| = \big\|\text{Proj}_V(u)\big\|$. This completes the proof of Part (1).

To prove Part (2), we no longer assume that U is 1-dimensional. Note that

$$\begin{aligned}\theta(U,V) &= \min_{0\neq u\in U}\ \min_{0\neq v\in V}\ \theta(u,v)\\ &= \min_{u\in U,\|u\|=1}\ \min_{0\neq w\in \text{Span}\,\{u\}}\ \min_{0\neq v\in V}\ \theta(w,v)\\ &= \min_{u\in U,\|u\|=1}\ \theta\big(\text{Span}\,\{u\},V\big),\end{aligned}$$

and so, by Part (1) we have

$$\cos\,\theta(U,V) = \max_{u\in U,\|u\|=1} \cos\,\theta\big(\text{Span}\,\{u\},V\big) = \max_{u\in U,\|u\|=1} \big\|\text{Proj}_V(u)\big\|.$$

3.5 Definition: Let $a, b \in \mathbb{R}^n$ with $a \neq b$. The **perpendicular bisector** of $[a,b]$ is the hyperplane in $\mathbb{R}^n$ through the midpoint $\frac{a+b}{2}$ which is perpendicular to the vector $b-a$, in other words it is the hyperplane in $\mathbb{R}^n$ given by the equation $\left(x - \frac{a+b}{2}\right) \bullet (b-a) = 0$.

3.6 Theorem: *a point $x \in \mathbb{R}^n$ lies on the perpendicular bisector of $[a,b]$ if and only if x is equidistant from a and b.*

Proof: Let P be the perpendicular bisector of $[a,b]$. Then

$$\begin{aligned} x \in P &\iff \left(x - \tfrac{a+b}{2}\right) \bullet (b-a) = 0 \iff \big(2x - (a+b)\big) \bullet (b-a) = 0 \\ &\iff 2x \bullet (b-a) = (a+b) \bullet (b-a) \\ &\iff 2x \bullet b - 2x \bullet a = a \bullet b - a \bullet a + b \bullet b - b \bullet a = \|b\|^2 - \|a\|^2 \\ &\iff -2x \bullet a + \|a\|^2 = -2x \bullet b + \|b\|^2 \\ &\iff \|x\|^2 - 2x \bullet a + \|a\|^2 = \|x\|^2 - 2x \bullet b + \|b\|^2 \\ &\iff \|x-a\|^2 = \|x-b\|^2 \iff \|x-a\| = \|x-b\| \,. \end{aligned}$$

3.7 Theorem: *Let $[a_0, a_1, \cdots, a_l]$ be an l-simplex in $\mathbb{R}^n$. For $0 \leq j < k \leq n$, let B_{jk} be the perpendicular bisector of $[a_j, a_k]$. Then there is a unique point o in the affine span $\langle a_0, a_1, \cdots, a_l \rangle$ which lies in the intersection of all of the perpendicular bisectors B_{jk}. This point o is called the circumcentre of the simplex.*

Proof: For $1 \leq i \leq l$ let $u_i = a_i - a_0$. For $x \in \langle a_0, a_1, \cdots, a_l \rangle = a_0 + \text{Span}\{u_1, u_2, \cdots, u_l\}$, we can write x uniquely as $x = a_0 + \sum_{i=1}^{l} t_i u_i = a_0 + At$ where $A = (u_1, u_2, \cdots, u_l) \in M_{n \times l}(\mathbb{R})$. For $x \in \langle a_0, a_1, \cdots, a_l \rangle$ with $x = a_0 + At$ we have

$$\begin{aligned} x \in \bigcap_{k=1}^{l} B_{0k} &\iff \left(x - \frac{a_0 + a_k}{2}\right) \bullet (a_k - a_0) = 0 \text{ for all } 1 \leq k \leq l \\ &\iff \left((a_0 + At) - \left(a_0 + \tfrac{a_k - a_0}{2}\right)\right) \bullet (a_k - a_0) = 0 \text{ for all } 1 \leq k \leq l \\ &\iff \left(At - \tfrac{1}{2}\, u_k\right) \bullet u_k = 0 \text{ for all } 1 \leq k \leq l \\ &\iff (At) \bullet u_k = \tfrac{1}{2}\|u_k\|^2 \text{ for all } 1 \leq k \leq l \\ &\iff \begin{pmatrix} (At) \bullet u_1 \\ \vdots \\ (At) \bullet u_l \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} \|u_1\|^2 \\ \vdots \|u_l\|^2 \end{pmatrix} \\ &\iff A^T A\, t = \tfrac{1}{2}\, u \text{ , where } u = \left(\|u_1\|^2, \|u_2\|^2, \cdots, \|u_l\|^2\right)^T. \end{aligned}$$

Since $\{a_0, a_1, \cdots, a_l\}$ is affinely independent, the set $\{u_1, u_2, \cdots, u_l\}$ is linearly independent so we have $\text{rank}(A^T A) = \text{rank}(A) = l$, and hence $A^T A$ is invertible. Thus there is a unique point $o \in \langle a_0, a_1, \cdots, a_l \rangle$ which lies in each bisector B_{0k} for $1 \leq k \leq l$, namely the point

$$o = a_0 + At = a_0 + \tfrac{1}{2} A (A^T A)^{-1} u \,.$$

Finally, note that for $1 \leq j < k \leq l$, by the previous theorem, since $o \in B_{0j}$ and $o \in B_{0k}$ we have $\|o - a_0\| = \|a - a_j\|$ and $\|o - a_0\| = \|o - a_k\|$ so that $\|o - a_j\| = \|o - a_k\|$, then by another application of the previous theorem, since $\|o - a_j\| = \|o - a_k\|$ it follows that $o \in B_{jk}$.

3.8 Theorem: *Let F be a field. Let* $A = \begin{pmatrix} 1 & a_0 & {a_0}^2 & \cdots & {a_0}^n \\ 1 & a_1 & {a_1}^2 & \cdots & {a_1}^n \\ & & \vdots & & \\ 1 & a_n & {a_n}^2 & \cdots & {a_n}^n \end{pmatrix}$ *with each $a_i \in F$.*

Then $\det A = \prod_{0 \le i < j \le n} (a_j - a_i)$.

Proof: Let $A_n = \begin{pmatrix} 1 & a_0 & {a_0}^2 & \cdots & {a_0}^n \\ 1 & a_1 & {a_1}^2 & \cdots & {a_1}^n \\ & & \vdots & & \\ 1 & a_n & {a_n}^2 & \cdots & {a_n}^n \end{pmatrix}$. Note that $\det A_1 = \det \begin{pmatrix} 1 & a_0 \\ 1 & a_1 \end{pmatrix} = a_1 - a_0$. Suppose, inductively, that $\det A_{n-1} = \prod_{0 \le i < j < n} (a_j - a_i)$. Note that if $a_i = a_j$ for some $i \ne j$, then A_n has two equal rows, and so in this case we have $\det A_n = 0 = \prod_{0 \le i < j \le n} (a_j - a_i)$. Suppose that $a_0, a_1, \cdots, a_n$ are all distinct. Replace a_n by x and let

$$f(x) = \det \begin{pmatrix} 1 & a_0 & {a_0}^2 & \cdots & {a_0}^n \\ 1 & a_1 & {a_1}^2 & \cdots & {a_1}^n \\ & & \vdots & & \\ 1 & a_{n-1} & {a_{n-1}}^2 & \cdots & {a_{n-1}}^n \\ 1 & x & x^2 & \cdots & x^n \end{pmatrix}.$$

By expanding along the last row we see that $f(x)$ is a polynomial of degree n with leading coefficient equal to $C = \det A_{n-1} = \prod_{0 \le i < j < n} (a_j - a_i)$. On the other hand, for each value of i with $0 \le i < n$, by subtracting the i^{th} row from the last row we see that

$$\begin{aligned} f(x) &= \det \begin{pmatrix} 1 & a_0 & {a_0}^2 & \cdots & {a_0}^n \\ 1 & a_1 & {a_1}^2 & \cdots & {a_1}^n \\ & & \vdots & & \\ 1 & a_{n-1} & {a_{n-1}}^2 & \cdots & {a_{n-1}}^n \\ 0 & x - a_i & x^2 - {a_i}^2 & \cdots & x^n - {a_i}^n \end{pmatrix} \\ &= (x - a_i) \det \begin{pmatrix} 1 & a_0 & {a_0}^2 & \cdots & {a_0}^n \\ 1 & a_1 & {a_1}^2 & \cdots & {a_1}^n \\ & & \vdots & & \\ 1 & a_{n-1} & {a_{n-1}}^2 & \cdots & {a_{n-1}}^n \\ 0 & 1 & x + a_i & \cdots & x^{n-1} + \cdots + x\,{a_i}^{n-2} + {a_i}^{n-1} \end{pmatrix} \end{aligned}$$

and so $(x - a_i)$ divides $f(x)$. Thus we must have

$$f(x) = C(x - a_0)(x - a_1) \cdots (x - a_{n-1}) = \prod_{0 \le i < j < n} (a_j - a_i) \prod_{0 \le i < n} (x - a_i)\,.$$

Replacing x by a_n gives $\det A_n = f(a_n) = \prod_{0 \le i < j \le n} (a_j - a_i)$, as required.

3.9 Definition: The matrix A in the above theorem is called the **Vandermonde matrix** on $a_0, a_1, \cdots, a_n$.

3.10 Corollary: *Let F be any field. Let $(a_0, b_0), (a_1, b_1), \cdots, (a_n, b_n)$ be ordered pairs of elements in F with the a_i all distinct. Then there exists a unique polynomial $f \in P_n(F)$ with $f(a_i) = b_i$ for all i.*

Proof: Suppose that $a_0, a_1, \cdots, a_n$ are all distinct, and let $b_0, b_1, \cdots, b_n$ be arbitrary. Let $f \in P_n(F)$, say $f(x) = c_0 + c_1 x + \cdots c_n x^n$. Then we have

$$\begin{aligned} f(a_i) = b_i \text{ for all } i &\iff c_0 + c_1 a_i + c_2 {a_i}^2 + \cdots + c_n {a_i}^n = b_i \text{ for all } i \\ &\iff Ac = b \end{aligned}$$

where $b = (b_0, b_1, \cdots, b_n)^T$, $c = (c_0, c_1, \cdots, c_n)^T$, and A is the Vandermonde matrix on $a_0, \cdots, a_n$. By the above theorem, we have $\det A = \prod(a_j - a_i)$. Since $a_0, a_1, \cdots, a_n$ are all distinct, $\det A \neq 0$, so A is invertible and the equation $Ac = b$ has a unique solution c.

3.11 Theorem: *Let $n, l \in \mathbb{Z}^+$. Given n ordered pairs $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n) \in \mathbb{R}^2$ such that at least $l+1$ of the a_i are distinct, there exists a unique polynomial $f \in P_l(\mathbb{R})$ which minimizes the sum $\sum_{i=1}^{n} \big(f(a_i) - b_i\big)^2$. This polynomial f is called the* **least-squares best fit polynomial** *for the data points (a_i, b_i).*

Proof: For $f(x) = c_0 + c_1 x + \cdots + c_l x^l$, we have

$$\begin{pmatrix} f(a_1) \\ \vdots \\ f(a_n) \end{pmatrix} = \begin{pmatrix} c_0 + c_1 a_1 + c_2\, {a_1}^2 + \cdots + c_l\, {a_1}^l \\ \vdots \\ c_0 + c_1 a_n + c_2\, {a_n}^2 + \cdots + c_l\, {a_n}^l \end{pmatrix} = Ac$$

where

$$A = \begin{pmatrix} 1 & a_1 & {a_1}^2 & \cdots & {a_1}^l \\ 1 & a_2 & {a_2}^2 & \cdots & {a_2}^l \\ \vdots & & & & \\ 1 & a_n & {a_n}^2 & \cdots & {a_n}^l \end{pmatrix} \in M_{n\times(l+1)}(\mathbb{R}) \text{ and } c = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_l \end{pmatrix}.$$

Note that the sum $\sum_{i=1}^{n} \big(f(a_i) - b_i\big)^2$ is the square of the distance between $b = (b_1, b_2, \cdots, b_n)^T$ and $f(a) = \big(f(a_1), f(a_2), \cdots, f(a_n)\big)^T = Ac$, so to minimize the sum we need to choose c to minimize the diastance $\|b - Ac\|$. To do this Ac must be the (unique) point in $\mathrm{Col}A$ which is nearest to b, that is

$$Ac = \mathrm{Proj}_{\mathrm{Col}A}(b).$$

Since $l+1$ of the a_i are distinct, it follows that the corresponding rows of A form a Vandermonde matrix on $l+1$ distinct points. This $(l+1) \times (l+1)$ Vandermonde matrix is invertible by Theorem 3.8, so these $l+1$ rows are linearly independent. It follows that $\mathrm{rank}A = l+1$ and that the $l+1$ columns of A are linearly independent. Thus A is injective, and so there is a unique vector c with $Ac = \mathrm{Proj}_{\mathrm{Col}A}(b)$. Indeed from our formula for the orthogonal projection given in Theorem 2.15, we have

$$c = (A^T A)^{-1} A^T b.$$