6 Moore–Penrose Pseudoinverse

Contents

In this subsection we follow Albert (1972).

6.1 Classical least squares problem

Lemma 6.1. *Let x be a vector and $\mathcal{L}$ is a linear manifold in $\mathbb{R}^n$ (that is, if $x, y \in \mathcal{L}$, then $\alpha x + \beta y \in \mathcal{L}$ for any scalars α, β). Then if*

$$x = \hat{x} + \tilde{x} \tag{6.1}$$

where $\hat{x} \in \mathcal{L}$ and $\tilde{x} \perp \mathcal{L}$, then $\hat{x}$ is "nearest" to x, or, in other words, it is the ***projection of x to the manifold $\mathcal{L}$***.

Proof. For any $y \in \mathcal{L}$ we have

$$\begin{aligned}
\|x - y\|^2 &= \|\hat{x} + \tilde{x} - y\|^2 = \|(\hat{x} - y) + \tilde{x}\|^2 \\
&= \|\hat{x} - y\|^2 + 2(\hat{x} - y, \tilde{x}) + \|\tilde{x}\|^2 = \|\hat{x} - y\|^2 + \|\tilde{x}\|^2 \\
&\geq \|\tilde{x}\|^2 = \|x - \hat{x}\|^2
\end{aligned}$$

with strict inequality holding unless $\|y - \hat{x}\|^2 = 0$. □

Theorem 6.1. *Let z be an n-dimensional real vector and $H \in \mathbb{R}^{n \times m}$.*

1. There is always a vector, in fact a unique vector $\hat{x}$ of minimal (Euclidean) norm, which minimizes

$$\|z - Hx\|^2$$

2. *The vector $\hat{x}$ is the unique vector in the range*

$$\mathcal{R}(H^{\mathsf{T}}) := \{x : x = H^{\mathsf{T}} z,\ z \in \mathbb{R}^n\}$$

which satisfies the equation

$$H\hat{x} = \hat{z}$$

where $\hat{z}$ is the projection of z on $\mathcal{R}(H)$.

Proof. By (6.1) we can write

$$z = \hat{z} + \tilde{z}$$

where $\hat{z}$ is the projection of z on the kernel (the null space)

$$\mathcal{N}(H^{\mathsf{T}}) := \{z \in \mathbb{R}^n : 0 = H^{\mathsf{T}} z\}$$

Since $Hx \in \mathcal{R}(H)$ for any $x \in \mathbb{R}^m$, it follows that

$$\hat{z} - Hx \in \mathcal{R}(H)$$

and, since $\tilde{z} \in \mathcal{R}^{\perp}(H)$,

$$\tilde{z} \perp \hat{z} - Hx$$

Therefore,

$$\begin{aligned} \|z - Hx\|^2 &= \|(\hat{z} - Hx) + \tilde{z}\|^2 \\ &= \|\hat{z} - Hx\|^2 + \|\tilde{z}\|^2 \geq \|\tilde{z}\|^2 = \|z - \hat{z}\|^2 \end{aligned}$$

This low bound is attainable since $\hat{z}$, being the range of H, is the afterimage of some x^*, that is, $\hat{z} = Hx^*$.

1. Let us show that x^* has a minimal norm. Since x^* may be decomposed into two orthogonal vectors

$$x^* = \hat{x}^* + \tilde{x}^*$$

where $\hat{x}^* \in \mathcal{R}\left(H^{\perp}\right)$ and $\tilde{x}^* \in \mathcal{N}(H)$. Thus $Hx^* = H\hat{x}$ we have

$$\|z - Hx^*\|^2 = \|z - H\hat{x}\|^2$$

and

$$\|x^*\|^2 = \|\hat{x}^*\|^2 + \|\tilde{x}^*\|^2 \geq \|\hat{x}^*\|^2$$

with strict inequality unless $x^* = \hat{x}^*$. So, x^* may be selected equal to $\hat{x}^*$.

2. Show now that $x^* = \hat{x}^*$ is unique. Suppose that $Hx^* = Hx^{**} = \hat{z}$. Then

$$(x^* - x^{**}) \in \mathcal{R}(H)$$

But $H(x^* - x^{**}) = 0$, which implies,

$$(x^* - x^{**}) \in \mathcal{N}(H) = \mathcal{R}^{\perp}\left(H^{\perp}\right)$$

Thus $(x^* - x^{**})$ is orthogonal to itself, which means that $\left\|x^* - x^{**}\right\|^2 = 0$, or equivalently, $x^* = x^{**}$. □

Corollary 6.1. $\left\|z - Hx\right\|^2$ *is minimized by* x_0 *if and only if* $Hx_0 = \hat{z}$ *where* $\hat{z}$ *is the projection of* z *on* $\mathcal{R}(H)$.

Corollary 6.2. *There is always an n-dimensional vector* y *such that*

$$\left\|z - HH^{\mathsf{T}}y\right\|^2 = \inf_x \left\|z - Hx\right\|^2$$

and if

$$\left\|z - Hx_0\right\|^2 = \inf_x \left\|z - Hx\right\|^2$$

then

$$\left\|x_0\right\|^2 \geq \left\|H^{\mathsf{T}}y\right\|^2$$

with strict inequality unless $x_0 = H^{\mathsf{T}}y$. *The vector* y *satisfies the equation*

$$HH^{\mathsf{T}}y = \hat{z}$$

Theorem 6.2. (on the system of normal equations) *Among those vectors* x, *which minimize* $\left\|z - Hx\right\|^2$, $\hat{x}$, *the one having minimal norm, is the unique vector of the form*

$$\boxed{\hat{x} = H^{\mathsf{T}}y} \tag{6.2}$$

satisfying

$$\boxed{H^{\mathsf{T}}H\hat{x} = H^{\mathsf{T}}z} \tag{6.3}$$

Proof. By direct differentiation we have

$$\frac{\partial}{\partial x}\left\|z - Hx\right\|^2 = 2H^{\mathsf{T}}(z - Hx) = 0$$

which gives (6.3). The representation (6.2) follows from the previous corollary. □

6.2 Pseudoinverse characterization

We are now in the position to exhibit an explicit representation for the minimum norm solution to a least square problem.

Lemma 6.2. *For any real symmetric matrix $A \in \mathbb{R}^{n\times n}$ the limit*

$$\boxed{P_A := \lim_{\delta\to 0} (A + \delta I_{n\times n})^{-1} A} \tag{6.4}$$

always exists. For any vector $z \in \mathbb{R}^n$

$$\hat{z} = P_A z$$

*is the **projection** of x on $\mathcal{R}(A)$.*

Proof. By symmetricity of A for all $\delta > 0$ such that $0 < |\delta| < \min\limits_{j:\lambda_j(A)\neq 0} |\lambda_j(A)|$ the matrix $(A + \delta I_{n\times n})^{-1}$ exists. Any $z \in \mathbb{R}^n$ may be represented as

$$z = \hat{z} + \tilde{z}$$

where $\hat{z} \in \mathcal{R}(A)$, $\tilde{z} \in \mathcal{N}(A)$ and $Az = A\hat{z}$. There exists x_0 such that $\hat{z} = Ax_0$, so

$$(A + \delta I_{n\times n})^{-1} Az = (A + \delta I_{n\times n})^{-1} A\hat{z} = (A + \delta I_{n\times n})^{-1} A(Ax_0)$$

By the spectral theorem (4.4) for symmetric matrices it follows that

$$A = T\Lambda T^{\intercal}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $T^{\intercal} = T^{-1}$. Thus

$$\begin{aligned}
(A + \delta I_{n\times n})^{-1} Az &= (A + \delta I_{n\times n})^{-1} A^2 x_0 \\
&= (T\Lambda T^{\intercal} + \delta TT^{\intercal})^{-1} T\Lambda^2 T^{\intercal} x_0 \\
&= \left(T[\Lambda + \delta I_{n\times n}]T^{\intercal}\right)^{-1} T\Lambda^2 T^{\intercal} x_0 \\
&= T\left([\Lambda + \delta I_{n\times n}]^{-1} \Lambda^2\right) T^{\intercal} x_0
\end{aligned}$$

It is plain to see that

$$\begin{aligned}
\lim_{\delta\to 0} [\Lambda + \delta I_{n\times n}]^{-1} \Lambda^2 &= \lim_{\delta\to 0} [\Lambda + \delta I_{n\times n}]^{-1} [\Lambda + \delta I_{n\times n} - \delta I_{n\times n}] \Lambda \\
&= \lim_{\delta\to 0} \left[I_{n\times n} - \delta [\Lambda + \delta I_{n\times n}]^{-1}\right] \Lambda \\
&= \left[\lim_{\delta\to 0} \operatorname{diag}\left(1 - \frac{\delta}{\lambda_1 + \delta}, \ldots, 1 - \frac{\delta}{\lambda_n + \delta}\right)\right] \Lambda = \Lambda
\end{aligned}$$

since

$$1 - \frac{\delta}{\lambda_i + \delta} = \begin{cases} 0 & \text{if} \quad \lambda_i = 0 \\ \to 1 & \text{if} \quad \lambda_i \neq 0 \end{cases}$$

This implies

$$\lim_{\delta \to 0} (A + \delta I_{n\times n})^{-1} Az = T \Lambda T^{\mathsf{T}} x_0 = Ax_0 = \hat{z}$$

□

Theorem 6.3. *For any real $(n \times m)$-matrix H the limit*

$$\boxed{\begin{aligned} H^+ := &\lim_{\delta \to 0} \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} \\ = &\lim_{\delta \to 0} H^{\mathsf{T}} \left(H H^{\mathsf{T}} + \delta^2 I_{n\times n}\right)^{-1} \end{aligned}} \tag{6.5}$$

always exists. For any vector $z \in \mathbb{R}^n$

$$\hat{x} = H^+ z$$

is the vector of minimal norm among those which minimize

$$\left\| z - Hx \right\|^2$$

Proof. It is clear that the right sides in (6.5) are equal, if either exists, since

$$\begin{aligned} H^{\mathsf{T}} H H^{\mathsf{T}} + \delta^2 H^{\mathsf{T}} &= \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right) H^{\mathsf{T}} \\ &= H^{\mathsf{T}} \left(H H^{\mathsf{T}} + \delta^2 I_{n\times n}\right) \end{aligned}$$

and the matrices $\left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)$ and $\left(H H^{\mathsf{T}} + \delta^2 I_{n\times n}\right)$ are inverse for any $\delta^2 > 0$. By the composition

$$z = \hat{z} + \tilde{z}$$

where $\hat{z} \in \mathcal{R}(H^{\mathsf{T}})$, $\tilde{z} \in \mathcal{N}(H^{\mathsf{T}})$ and $H^{\mathsf{T}} z = H^{\mathsf{T}} \hat{z}$, there exists x_0 such that $\hat{z} = Hx_0$. So,

$$\begin{aligned} \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} z &= \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} \hat{z} \\ &= \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} H x_0 \end{aligned}$$

By the previous Lemma there exists the limit

$$\lim_{\delta \to 0} \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} H = P_{H^{\mathsf{T}} H}$$

which gives

$$\lim_{\delta \to 0} \left(H^{\mathsf{T}} H + \delta^2 I_{m\times m}\right)^{-1} H^{\mathsf{T}} H x_0 = (P_{H^{\mathsf{T}} H})\, x_0 := \hat{x}_0$$

where $\hat{x}_0$ is the projection on $\mathcal{R}(H^\intercal H) = \mathcal{R}(H^\intercal)$. Thus we conclude that

$$\begin{aligned}\hat{x}_0 &= \lim_{\delta\to 0}\left(H^\intercal H + \delta^2 I_{m\times m}\right)^{-1} H^\intercal \hat{z}\\ &= \lim_{\delta\to 0}\left(H^\intercal H + \delta^2 I_{m\times m}\right)^{-1} H^\intercal z\end{aligned}$$

always exists and is an element of $\mathcal{R}(H^\intercal)$ satisfying $\hat{z} = H\hat{x}_0$. □

Definition 6.1. *The matrix limit H^+ (6.5) is called the* ***pseudoinverse*** *(the generalized inverse) of H in the* ***Moore–Penrose sense****.*

Remark 6.1. *It follows that*

- *(HH^+z) is the projection of z on $\mathcal{R}(H)$;*
- *(H^+Hx) is the projection of x on $\mathcal{R}(H^\intercal)$;*
- *$(I_{n\times n} - HH^+)z$ is the projection of z on $\mathcal{N}(H^\intercal)$;*
- *$(I_{n\times n} - H^+H)x$ is the projection of x on $\mathcal{N}(H)$.*

The following properties can be proven by the direct application of (6.5).

Corollary 6.3. *For any real $n \times m$ matrix H*

1.

$$\boxed{H^+ = (H^\intercal H)^+ H^\intercal} \tag{6.6}$$

2.

$$\boxed{(H^\intercal)^+ = \left(H^+\right)^\intercal} \tag{6.7}$$

3.

$$\boxed{H^+ = H^\intercal (HH^\intercal)^+} \tag{6.8}$$

4.

$$\boxed{H^+ = H^{-1}} \tag{6.9}$$

if H is square and nonsingular.

6.3 Criterion for pseudoinverse checking

The next theorem represents the **criterion** for a matrix B, to be the pseudoinverse H^+ of H.

Theorem 6.4. *For any real $n \times m$ matrix H the matrix $B = H^{+}$ if and only if*

1.

$$\boxed{HB \text{ and } BH \text{ are symmetric}} \tag{6.10}$$

2.

$$\boxed{HBH = H} \tag{6.11}$$

3.

$$\boxed{BHB = B} \tag{6.12}$$

Proof.

1. *Necessity.* Let $B = H^{+}$.
   (a) Since

$$HH^{+} = \lim_{\delta \to 0} HH^{\intercal} \left(HH^{\intercal} + \delta^2 I_{n \times n} \right)^{-1}$$

$$\left(HH^{+}\right)^{\intercal} = \left(H \lim_{\delta \to 0} \left(H^{\intercal} H + \delta^2 I_{m \times m} \right)^{-1} H^{\intercal} \right)^{\intercal}$$

$$= H \left[\lim_{\delta \to 0} \left(H^{\intercal} H + \delta^2 I_{m \times m} \right)^{-1} \right] H^{\intercal} = HH^{+}$$

and

$$H^{+}H = \left[\lim_{\delta \to 0} H^{\intercal} \left(HH^{\intercal} + \delta^2 I_{n \times n} \right)^{-1} \right] H$$

$$\left(H^{+}H\right)^{\intercal} = \left(\left[\lim_{\delta \to 0} H^{\intercal} \left(HH^{\intercal} + \delta^2 I_{n \times n} \right)^{-1} \right] H \right)^{\intercal}$$

$$H^{\intercal} \lim_{\delta \to 0} \left(H^{\intercal} H + \delta^2 I_{m \times m} \right)^{-1} H = H^{+}H$$

the symmetricity (6.10) takes place.

   (b) Since by (6.1) HH^{+} is a projector on $\mathcal{R}(H)$ and the projection of any vector from $\mathcal{R}(H)$ coincides with the same vector, one has for any $z \in \mathbb{R}^n$

$$HH^{+}(Hz) = Hz$$

which gives (6.11). By (6.6)

$$H^{+}H = \left(H^{\intercal} H\right)^{+} \left(H^{\intercal} H\right)$$

which, in view of (6.11) and the symmetricity property (6.10) of HH^+, implies

$$\begin{aligned} H^+ &= (H^\intercal H)^+ H^\intercal = (H^\intercal H)^+ \left(HH^+H\right)^\intercal \\ &= (H^\intercal H)^+ H^\intercal \left(HH^+\right)^\intercal = (H^\intercal H)^+ H^\intercal \left(HH^+\right) \\ &= (H^\intercal H)^+ (H^\intercal H) H^+ = H^+HH^+ \end{aligned}$$

So, (6.12) is proven.

2. *Sufficiency*. Suppose B satisfies (6.10), (6.11) and (6.12). Since

$$BH = (BH)^\intercal, \quad H = HBH$$

then

$$H = HBH = H(BH)^\intercal$$

Using this representation and since $HH^+H = H$, we derive

$$H^+H = H^+H(BH)^\intercal = \left[HH^+H\right]^\intercal B^\intercal = H^\intercal B^\intercal = BH \tag{6.13}$$

Analogously, since $B = BHB$ and HB is symmetric, we have

$$B^\intercal = HBB^\intercal$$

Pre-multiplying this identity by HH^+, we obtain

$$HH^+B^\intercal = HH^+HBB^\intercal = HBB^\intercal = B^\intercal$$

Taking transposes and in view of (6.13) we get

$$B = B\left(HH^+\right)^\intercal = B\left(HH^+\right) = BHH^+ = H^+HH^+ = H^+$$

□

The theorem above is extremely useful as a method for proving identities. If one thinks that a certain expression coincides with the pseudoinverse of a certain matrix H, a good way to decide is to run the expressions through conditions (6.10), (6.11), (6.12) and observe whether or not they hold.

6.4 Some identities for pseudoinverse matrices

Lemma 6.3. $b \in \mathcal{R}(A) := \operatorname{Im}(A) \subseteq \mathbb{R}^n$ *if and only if*

$$\boxed{AA^+b = b} \tag{6.14}$$

Proof.

(a) *Necessity*. If $b \in \mathcal{R}(A)$, then there exists a vector $d \in \mathbb{R}^n$ such that $b = Ad$, and, therefore,

$$AA^+b = AA^+(Ad) = \left(AA^+A\right)d = Ad = b$$

(b) *Sufficiency*. Suppose that (6.14) is true. Any vector b can be represented as $b = Ad + b^{\perp}$ with $b^{\perp} \perp Ad$, namely, $b^{\perp} = \left(I - AA^{+}\right) v$. Then

$$AA^{+}\left(Ad + b^{\perp}\right) = Ad + b^{\perp}$$

which implies $AA^{+}b^{\perp} = b^{\perp}$, and, hence,

$$AA^{+}\left(I - AA^{+}\right) v = 0 = \left(I - AA^{+}\right) v = b^{\perp}$$

So, $b^{\perp} = 0$, and, hence, $b = Ad$, or, equivalently, $b \in \mathcal{R}(A)$. Lemma is proven. □

The following identities can be proven more easily by simple verification of (6.10), (6.11), (6.12).

Claim 6.1.

1.

$$\boxed{(O_{m\times n})^{+} = O_{n\times m}} \tag{6.15}$$

2. For any $x \in \mathbb{R}^n$ $(x \neq 0)$

$$\boxed{x^{+} = \frac{x^{\mathsf{T}}}{\|x\|^2}} \tag{6.16}$$

3.

$$\boxed{\left(H^{+}\right)^{+} = H} \tag{6.17}$$

4. In general,

$$\boxed{(AB)^{+} \neq B^{+}A^{+}} \tag{6.18}$$

The identity takes place if

$$\boxed{\begin{array}{c} A^{\mathsf{T}}A = I \quad \text{or} \\ BB^{\mathsf{T}} = I \quad \text{or} \\ B = A^{\mathsf{T}} \quad \text{or} \\ B = A^{+} \quad \text{or} \\ \text{both } A \text{ and } B \text{ are of full rank, or} \\ \operatorname{rank} A = \operatorname{rank} B \end{array}}$$

The identity in (6.18) holds ***if and only if***

$$\boxed{\mathcal{R}\left(BB^{\mathsf{T}}A^{\mathsf{T}}\right) \subseteq \mathcal{R}\left(A^{\mathsf{T}}\right)} \tag{6.19}$$

and

$$\boxed{\mathcal{R}\left(A^{\intercal}AB\right) \subseteq \mathcal{R}\left(B\right)} \tag{6.20}$$

5.

$$\boxed{(AB)^{+} = B_1^{+}A_1^{+}} \tag{6.21}$$

where

$$\boxed{\begin{aligned} B_1 &= A^{+}AB \\ A_1 &= AB_1B_1^{+} \end{aligned}}$$

6.

$$\boxed{\left(H^{\intercal}H\right)^{+} = H^{+}\left(H^{\intercal}\right)^{+}, \left(HH^{\intercal}\right)^{+} = \left(H^{\intercal}\right)^{+}H^{+}} \tag{6.22}$$

7. *If* A *is symmetric and* $\alpha > 0$, *then*

$$\boxed{\begin{aligned} (A^{\alpha})^{+} &= \left(A^{+}\right)^{\alpha} \\ A^{\alpha}(A^{\alpha})^{+} &= (A^{\alpha})^{+}A^{\alpha} = AA^{+} \\ A^{+}A^{\alpha} &= A^{\alpha}A^{+} \end{aligned}} \tag{6.23}$$

8. *If* $A = U\Lambda V^{\intercal}$ *where* U, V *are orthogonal and* Λ *is a diagonal matrix, then*

$$\boxed{A^{+} = V\Lambda^{+}U^{\intercal}} \tag{6.24}$$

9. *Greville's formula (Greville 1960): if* $C_{m+1} = \left[C_m \vdots c_{m+1}\right]$ *then*

$$\boxed{C_{m+1}^{+} = \begin{bmatrix} C_m^{+}\left[I - c_{m+1}k_{m+1}^{\intercal}\right] \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ k_{m+1}^{\intercal} \end{bmatrix}} \tag{6.25}$$

where

$$k_{m+1} = \begin{cases} \dfrac{\left[I - C_mC_m^{+}\right]c_{m+1}}{\left\|\left[I - C_mC_m^{+}\right]c_{m+1}\right\|^2} & \textit{if} \quad \left[I - C_mC_m^{+}\right]c_{m+1} \neq 0 \\ \dfrac{\left(C_m^{+}\right)^{\intercal}C_m^{+}c_{m+1}}{1 + \left\|C_m^{+}c_{m+1}\right\|^2} & \textit{otherwise} \end{cases}$$

10. *If* H *is rectangular and* S *is symmetric and nonsingular then*

$$\boxed{(SH)^{+} = H^{+}S^{-1}\left[I - \left(QS^{-1}\right)^{+}\left(QS^{-1}\right)\right]} \tag{6.26}$$

where

$$Q = I - H^{+}H$$

Example 6.1. *Simple verification of (6.10), (6.11), (6.12) shows that*

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^{+} = \begin{bmatrix} -1.333 & -0.333 & 0.667 \\ 1.8033 & 0.333 & -0.4167 \end{bmatrix}$$

6.5 Solution of least squares problem using pseudoinverse

Theorem 6.5.

(a) The vector x_0 minimizes $\|z - Hx\|^2$ if and only if it is of the form

$$\boxed{x_0 = H^{+}z + \left(I - H^{+}H\right) y} \tag{6.27}$$

for some vector y.

(b) Among all solutions x_0 (6.27) the vector

$$\boxed{\bar{x}_0 = H^{+}z} \tag{6.28}$$

has the minimal Euclidean norm.

Proof. By theorem (6.3) we know that $H^{+}z$ minimizes $\|z - Hx\|^2$ and by (6.1), any x_0 minimizes $\|z - Hx\|^2$ if and only if $Hx_0 = \hat{z}$ where $\hat{z}$ is the projection of z on $\mathcal{R}(H)$. In view of that

$$Hx_0 = H\left(H^{+}z\right)$$

This means that $x_0 - H^{+}z$ is a null vector of H that is true if and only if

$$x_0 - H^{+}z = \left(I - H^{+}H\right) y$$

for some y. So, (a) (6.27) is proven. To prove (b) (6.28) it is sufficient to notice that

$$\begin{aligned}
\|x_0\|^2 &= \left\|H^{+}z + \left(I - H^{+}H\right) y\right\|^2 = \left\|H^{+}z\right\|^2 \\
&\quad + \left(H^{+}z, \left(I - H^{+}H\right) y\right) + \left\|\left(I - H^{+}H\right) y\right\|^2 = \left\|H^{+}z\right\|^2 \\
&\quad + \left(\left(I - H^{+}H\right)^{\intercal} H^{+}z, y\right) + \left\|\left(I - H^{+}H\right) y\right\|^2 = \left\|H^{+}z\right\|^2 \\
&\quad + \left(\left(I - H^{+}H\right) H^{+}z, y\right) + \left\|\left(I - H^{+}H\right) y\right\|^2 \\
&= \left\|H^{+}z\right\|^2 + \left\|\left(I - H^{+}H\right) y\right\|^2 \geq \left\|H^{+}z\right\|^2 = \left\|\bar{x}_0\right\|^2
\end{aligned}$$

$\square$

Corollary 6.4. (LS problem with constraints) *Suppose the set*

$$\mathcal{J} = \{x : Gx = u\}$$

is not empty. Then the vector x_0 minimizes $\|z - Hx\|^2$ over $\mathcal{J}$ if and only if

$$\boxed{\begin{aligned} x_0 &= G^{+}u + \bar{H}^{+}z + \left(I - G^{+}G\right)\left(I - \bar{H}^{+}\bar{H}\right)y \\ \bar{H} &:= H\left(I - G^{+}G\right) \end{aligned}} \tag{6.29}$$

and among all solutions

$$\boxed{\bar{x}_0 = G^{+}u + \bar{H}^{+}z} \tag{6.30}$$

has the minimal Euclidean norm.

Proof. Notice that by the Lagrange multipliers method x_0 solves the problem if it minimizes the Lagrange function

$$\|z - Hx\|^2 + (\lambda, Gx - u)$$

for some λ. This λ and x_0 satisfy the equation

$$\frac{\partial}{\partial x}\left[\|z - Hx\|^2 + (\lambda, Gx - u)\right] = -2H^{\intercal}(z - Hx_0) + G^{\intercal}\lambda = 0$$

or, equivalently,

$$H^{\intercal}Hx_0 = \left[H^{\intercal}z - \frac{1}{2}G^{\intercal}\lambda\right] := \tilde{z}$$

which in view of (6.27) implies

$$\begin{aligned} x_0 &= (H^{\intercal}H)^{+}\tilde{z} + \left[I - (H^{\intercal}H)^{+}(H^{\intercal}H)\right]y \\ &= (H^{\intercal}H)^{+}\left[H^{\intercal}z - \frac{1}{2}G^{\intercal}\lambda\right] + \left[I - (H^{\intercal}H)^{+}(H^{\intercal}H)\right]y \end{aligned} \tag{6.31}$$

But this x_0 should satisfy $Gx_0 = u$ which leads to the following equality

$$Gx_0 = G\left[(H^{\intercal}H)^{+}\left(H^{\intercal}z - \frac{1}{2}G^{\intercal}\lambda\right) + \left[I - (H^{\intercal}H)^{+}(H^{\intercal}H)\right]y\right] = u$$

or, equivalently,

$$\begin{aligned} \left[G(H^{\intercal}H)^{+}\right]\left(\frac{1}{2}G^{\intercal}\lambda\right) &= G(H^{\intercal}H)^{+}H^{\intercal}z \\ &\quad + G\left[I - (H^{\intercal}H)^{+}(H^{\intercal}H)\right]y - u \end{aligned}$$

or

$$\begin{aligned}\frac{1}{2}G^{\mathsf T}\lambda = {} & \left[G\left(H^{\mathsf T}H\right)^{+}\right]^{+}\left[G\left(H^{\mathsf T}H\right)^{+}H^{\mathsf T}z\right. \\ & \left.+\,G\left[I-\left(H^{\mathsf T}H\right)^{+}\left(H^{\mathsf T}H\right)\right]y-u\right] \\ & +\left[I-\left[G\left(H^{\mathsf T}H\right)^{+}\right]^{+}\left[G\left(H^{\mathsf T}H\right)^{+}\right]\right]\tilde{y}\end{aligned} \tag{6.32}$$

Substitution of (6.32) into (6.31) and using the properties of the pseudoinverse implies (6.29). The statement (6.30) is evident. □

6.6 Cline's formulas

In fact, the direct verification leads to the following identities (see Cline (1964, 1965)).

Claim 6.2. (Pseudoinverse of a partitioned matrix)

$$\boxed{\left[U \;\vdots\; V\right]^{+} = \begin{bmatrix} U^{+} - U^{+}VJ \\ \cdots\cdots\cdots\cdots \\ J \end{bmatrix}} \tag{6.33}$$

where

$$\boxed{\begin{aligned} & J = C^{+} + \left(I - C^{+}C\right)KV^{\mathsf T}\left(U^{+}\right)^{\mathsf T}U^{+}\left(I - VC^{+}\right) \\ & C = \left(I - UU^{+}\right)V \\ & K = \left(I + \left[U^{+}V\left(I - C^{+}C\right)\right]^{\mathsf T}\left[U^{+}V\left(I - C^{+}C\right)\right]\right)^{-1} \end{aligned}} \tag{6.34}$$

Claim 6.3. (Pseudoinverse of sums of matrices)

$$\boxed{\begin{gathered} \left(UU^{\mathsf T} + VV^{\mathsf T}\right)^{+} = \left(CC^{\mathsf T}\right)^{+} + \left[I - \left(VC^{+}\right)^{\mathsf T}\right] \\ \times\left[\left(UU^{\mathsf T}\right)^{+} - \left(UU^{\mathsf T}\right)^{+}V\left(I - C^{+}C\right)KV^{\mathsf T}\left(UU^{\mathsf T}\right)^{+}\right] \\ \times\left[I - \left(VC^{+}\right)^{\mathsf T}\right] \end{gathered}} \tag{6.35}$$

where C and K are defined in (6.34).

6.7 Pseudo-ellipsoids

6.7.1 Definition and basic properties

Definition 6.2. *We say that the set $\varepsilon(\mathring{x}, A) \in \mathbb{R}^n$ is **the pseudo-ellipsoid** (or **elliptic cylinder**) in $\mathbb{R}^n$ with the center at the point $\mathring{x} \in \mathbb{R}^n$ and with the matrix $0 \le A = A^{\mathsf T} \in \mathbb{R}^{n\times n}$ if it is defined by*

$$\boxed{\varepsilon(\mathring{x}, A) := \left\{x \in \mathbb{R}^n \mid \|x - \mathring{x}\|_A^2 = (x - \mathring{x}, A(x - \mathring{x})) \le 1\right\}} \tag{6.36}$$

If $A > 0$ *the set* $\varepsilon(\mathring{x}, A)$ *is an* ***ordinary ellipsoid*** *with the semi-axis equal to* $\lambda_i^{-1}(A)$ $(i = 1, \ldots, n)$.

Remark 6.2. *If*

(a) $A > 0$, *then* $\varepsilon(\mathring{x}, A)$ *is a bounded set;*
(b) $A \geq 0$, *then* $\varepsilon(\mathring{x}, A)$ *is an unbounded set.*

Lemma 6.4. *If* $0 < A = A^{\intercal} \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ *and* $\alpha < 1 - \|b\|_{A^{-1}}^2$, *then the set given by*

$$(x, Ax) - 2(b, x) + \alpha \leq 1$$

is the ellipsoid $\varepsilon\left(A^{-1}b, \dfrac{1}{1 - \alpha + \|b\|_{A^{-1}}^2} A\right)$.

Proof. It follows from the identity

$$\left\|x - A^{-1}b\right\|_A^2 - \|b\|_{A^{-1}}^2 = \|x\|_A^2 - 2(b, x)$$

□

Lemma 6.5. *If* $0 \leq A = A^{\intercal} \in \mathbb{R}^{n\times n}$, $b \in \mathcal{R}(A) \subseteq \mathbb{R}^n$ *and* $\alpha < 1 - \|b\|_{A^+}^2$, *then the set given by*

$$(x, Ax) - 2(b, x) + \alpha \leq 1$$

is the pseudo-ellipsoid $\varepsilon\left(A^{+}b, \dfrac{1}{1 - \alpha + \left\|b\right\|_{A^{+}}^2} A\right)$.

Proof. It follows from the identity

$$\left\|x - A^{+}b\right\|_A^2 - \|b\|_{A^{+}}^2 = \|x\|_A^2 - 2(b, x)$$

□

Lemma 6.6.

$$\boxed{\varepsilon\left(A^{+}A\mathring{x}, A\right) = \varepsilon(\mathring{x}, A)} \tag{6.37}$$

Proof. Indeed,

$$\begin{aligned}
&\left(x - A^{+}A\mathring{x}, A\left(x - A^{+}A\mathring{x}\right)\right) = \left(x - A^{+}A\mathring{x}, Ax - A\mathring{x}\right) \\
&(x, Ax) - \left(A^{+}A\mathring{x}, Ax\right) - (x, A\mathring{x}) + \left(A^{+}A\mathring{x}, A\mathring{x}\right) \\
&\quad = (x, Ax) - 2(x, A\mathring{x}) + (\mathring{x}, A\mathring{x}) = (x - \mathring{x}, A(x - \mathring{x}))
\end{aligned}$$

□

6.7.2 Support function

Definition 6.3. *The function* $f_S : \mathbb{R}^n \to \mathbb{R}$ *defined by*

$$\boxed{f_S(y) := \max_{x \in S} (y, x)} \tag{6.38}$$

is called the ***support*** *(or Legendre)* ***function*** *(SF) of the convex closed set* $S \subseteq \mathbb{R}^n$.

Lemma 6.7. *If* S *is the pseudo-ellipsoid* $\varepsilon(\mathring{x}, A)$ *(6.36), that is,*

$$S = \varepsilon(\mathring{x}, A) = \left\{x \in \mathbb{R}^n \mid \|x - \mathring{x}\|_A^2 = (x - \mathring{x}, A(x - \mathring{x})) \le 1\right\}$$

then

$$\boxed{f_S(y) = y^{\mathsf{T}} \mathring{x} + \sqrt{y^{\mathsf{T}} A^+ y}} \tag{6.39}$$

Proof. Using the Lagrange principle (see Theorem 21.12), for any $y \in \mathbb{R}^n$ we have

$$\arg\max_{x \in S} (y, x) = \arg\min_{\lambda \ge 0} \max_{x \in \mathbb{R}^n} L(x, \lambda \mid y)$$
$$L(x, \lambda \mid y) := (y, x) + \lambda\left[(x - \mathring{x}, A(x - \mathring{x})) - 1\right]$$

and, therefore, the extremal point (x^*, λ^*) satisfies

$$0 = \frac{\partial}{\partial x} L(x^*, \lambda^* \mid y) = y + \lambda^* A(x^* - \mathring{x})$$
$$\lambda^*\left[(x^* - \mathring{x}, A(x^* - \mathring{x})) - 1\right] = 0$$

The last identity is referred to as the *complementary slackness condition*. The x satisfying the first equation can be represented as follows

$$\arg_{x \in \mathbb{R}^n} \{y + \lambda A(x - \mathring{x}) = 0\} = \arg\min_{x \in \mathbb{R}^n} \|y + \lambda A(x - \mathring{x})\|^2$$

If $\lambda = 0$, it follows that $y = 0$. But $L(x, \lambda \mid y)$ is defined for any $y \in \mathbb{R}^n$. So, $\lambda > 0$, and hence, by (6.27),

$$x^* - \mathring{x} = \frac{1}{\lambda^*} A^+ y + \left(I - A^+ A\right) v, \quad v \in \mathbb{R}^n$$

Substitution of this expression in the complementary slackness condition and taking into account that $A^+ = \left(A^+\right)^{\mathsf{T}}$ implies

$$\begin{aligned} 1 &= \left(\frac{1}{\lambda^*} A^+ y + \left(I - A^+ A\right) v, \frac{1}{\lambda^*} A A^+ y\right) \\ &= \frac{1}{(\lambda^*)^2} \left(y, A^+ A A^+ y\right) + \left(A^+ A \left(I - A^+ A\right) v, \frac{1}{\lambda^*} y\right) \\ &= \frac{1}{(\lambda^*)^2} \left(y, A^+ y\right) \end{aligned}$$

or, equivalently, $\lambda^* = \sqrt{(y, A^+ y)}$, which finally gives

$$\begin{aligned}
f_S(y) &= \max_{x \in S}(y, x) = L(x^*, \lambda^* \mid y) = (y, x^*) \\
&= \left(y, \mathring{x} + \frac{1}{\lambda^*} A^+ y + \left(I - A^+ A\right) v\right) \\
&= (y, \mathring{x}) + \frac{1}{\lambda^*}\left(y, A^+ y\right) + \left(y, \left(I - A^+ A\right) v\right) \\
&= (y, \mathring{x}) + \frac{\left(y, A^+ y\right)}{\sqrt{(y, A^+ y)}} - \left(\lambda A\left(x - \mathring{x}\right), \left(I - A^+ A\right) v\right) \\
&= (y, \mathring{x}) + \frac{\left(y, A^+ y\right)}{\sqrt{(y, A^+ y)}} - \left(\lambda\left(x - \mathring{x}\right), A\left(I - A^+ A\right) v\right) \\
&= (y, \mathring{x}) + \frac{\left(y, A^+ y\right)}{\sqrt{(y, A^+ y)}}
\end{aligned}$$

Lemma is proven. □

6.7.3 *Pseudo-ellipsoids containing vector sum of two pseudo-ellipsoids*

The support function $f_S(y)$ (6.38) is particularly useful since the vector sum of convex closed sets and a linear transformation $\bar{A}$ of S have their simple counterparts in the support function description (see Appendix in Schlaepfer & Schweppe (1972)).

Lemma 6.8. (on SF for the vector sum of convex sets) *Let*

$$S_1 \oplus S_2 := \{x \in \mathbb{R}^n \mid x = x_1 + x_2, x_1 \in S_1, x_2 \in S_2\}$$

where S_1, S_2 are convex closed sets. Then

$$\boxed{f_{S_1 \oplus S_2}(y) = f_{S_1}(y) + f_{S_2}(y)} \tag{6.40}$$

Proof. By (6.38), it follows

$$\begin{aligned}
f_{S_1 \oplus S_2}(y) &= \max_{x \in S}(y, x) = \max_{x_1 \in S_1, x_2 \in S_2}\left((y, x_1) + (y, x_2)\right) \\
&= \max_{x_1 \in S_1}(y, x_1) + \max_{x_2 \in S_2}(y, x_2) = f_{S_1}(y) + f_{S_2}(y)
\end{aligned}$$

which completes the proof. □

Lemma 6.9. (on SF for a linear transformation) *Let*

$$\mathcal{B}_S := \{x \in \mathbb{R}^n \mid x = Bz, z \in S\}$$

where $B \in \mathbb{R}^{n \times n}$ is an $(n \times n)$ matrix. Then

$$\boxed{f_{\mathcal{B}_S}(y) = f_S\left(B^\top y\right)} \tag{6.41}$$

Proof. By (6.38), we have

$$f_{\mathcal{B}_S}(y) = \max_{x \in \mathcal{B}_S} (y, x) = \max_{z \in S} (y, Bz) = \max_{z \in S} (B^{\mathsf{T}} y, z) = f_S(B^{\mathsf{T}} y)$$

which proves the lemma. □

Lemma 6.10. (on SF for closed sets) *If two convex closed sets are related as $S_1 \supseteq S_2$, then for all $y \in S$*

$$\boxed{f_{S_1}(y) \geq f_{S_2}(y)} \tag{6.42}$$

Proof. It follows directly from the definition (6.38):

$$f_{S_1}(y) = \max_{x \in S_1} (y, x) \geq \max_{x \in S_2} (y, x) = f_{S_2}(y)$$

□

Lemma 6.11. (on SF for the vector sum of two ellipsoids) *Let S_1 and S_2 be two pseudo-ellipsoids, that is, $S_i = \varepsilon(\mathring{x}_i, A_i)$ $(i = 1, 2)$. Then*

$$\boxed{f_{S_1 \oplus S_2}(y) = f_S(y) = y^{\mathsf{T}}(\mathring{x}_1 + \mathring{x}_2) + \sqrt{y^{\mathsf{T}} A_1^{+} y} + \sqrt{y^{\mathsf{T}} A_2^{+} y}} \tag{6.43}$$

Proof. It results from (6.39) and (6.40). □

To bound $S_1 \oplus S_2$ by some pseudo-ellipsoid $S^*_{S_1 \oplus S_2}$ means to find $(\mathring{x}^*, A^*)$ such that (see Lemma 6.11) for all $y \in \mathbb{R}^n$

$$\boxed{y^{\mathsf{T}}(\mathring{x}_1 + \mathring{x}_2) + \sqrt{y^{\mathsf{T}} A_1^{+} y} + \sqrt{y^{\mathsf{T}} A_2^{+} y} \leq y^{\mathsf{T}} \mathring{x}^* + \sqrt{y^{\mathsf{T}} (A^*)^{+} y}} \tag{6.44}$$

Lemma 6.12. *The choice*

$$\boxed{\begin{array}{l} \mathring{x}^* = \mathring{x}_1 + \mathring{x}_2 \\ A^* = \left(\gamma^{-1} A_1^{+} + (1 - \gamma)^{-1} A_2^{+}\right)^{+}, \quad \gamma \in (0, 1) \end{array}} \tag{6.45}$$

is sufficient to satisfy (6.44).

Proof. Taking $\mathring{x}^* = \mathring{x}_1 + \mathring{x}_2$, we should to prove that

$$\sqrt{y^{\mathsf{T}} A_1^{+} y} + \sqrt{y^{\mathsf{T}} A_2^{+} y} \leq \sqrt{y^{\mathsf{T}} (A^*)^{+} y}$$

or, equivalently,

$$\left(\sqrt{y^{\mathsf{T}} A_1^{+} y} + \sqrt{y^{\mathsf{T}} A_2^{+} y}\right)^2 \leq y^{\mathsf{T}} (A^*)^{+} y$$

for some A^*. Applying the inequality (12.2), for any $\varepsilon > 0$ we have

$$\left(\sqrt{y^\intercal A_1^+ y} + \sqrt{y^\intercal A_2^+ y}\right)^2 \le (1+\varepsilon)\, y^\intercal A_1^+ y + \left(1+\varepsilon^{-1}\right) y^\intercal A_2^+ y$$
$$= y^\intercal \left[(1+\varepsilon) A_1^+ + \left(1+\varepsilon^{-1}\right) A_2^+\right] y$$

Denoting $\gamma^{-1} := (1+\varepsilon)$ and taking into account the identity (6.17) we get (6.45). □

6.7.4 Pseudo-ellipsoids containing intersection of two pseudo-ellipsoids

If $S_i = \varepsilon\left(\mathring{x}_i, A_i\right)$ $(i = 1, 2)$ are two pseudo-ellipsoids, then $S_1 \cap S_2$ is not a pseudo-ellipsoid. Sure, there exists a lot of pseudo-ellipsoids $S^*_{S_1\cap S_2}$ (in fact, a set) containing $S_1 \cap S_2$. To bound $S_1 \cap S_2$ by $S^*_{S_1\cap S_2}$ means to find $(\mathring{x}^*, A^*)$ such that

$$S_1 \cap S_2 \subseteq S^*_{S_1\cap S_2}$$

where

$$\begin{aligned} &S_1 \cap S_2 := \{x \in \mathbb{R}^n \mid (x - \mathring{x}_1, A_1 (x - \mathring{x}_1)) \le 1 \\ &\text{and} \quad (x - \mathring{x}_2, A_2 (x - \mathring{x}_2)) \le 1\} \\ &S^*_{S_1\cap S_2} := \{x \in \mathbb{R}^n \mid (x - \mathring{x}^*, A^* (x - \mathring{x}^*)) \le 1\} \end{aligned} \tag{6.46}$$

Lemma 6.13. *Let $S_1 \cap S_2 \neq \varnothing$. Then $(\mathring{x}^*, A^*)$ can be selected as follows*

$$\boxed{\begin{aligned} \mathring{x}^* &= A_\gamma^+ b_\gamma \\ b_\gamma &= \gamma A_1 \mathring{x}_1 + (1-\gamma) A_2 \mathring{x}_2 \\ A_\gamma &= \gamma A_1 + (1-\gamma) A_2, \ \gamma \in (0, 1) \end{aligned}} \tag{6.47}$$

and

$$\boxed{\begin{aligned} A^* &= \frac{1}{\beta_\gamma} A_\gamma \\ \beta_\gamma &= 1 - \alpha_\gamma + \left\|b_\gamma\right\|^2_{A_\gamma^+} \\ \alpha_\gamma &= \gamma\left(\mathring{x}_1, A_1 \mathring{x}_1\right) + (1-\gamma)\left(\mathring{x}_2, A_2 \mathring{x}_2\right) \end{aligned}} \tag{6.48}$$

Proof. Notice that $S^*_{S_1\cap S_2}$ can be selected as

$$S_1 \cap S_2 := \{x \in \mathbb{R}^n \mid \gamma (x - \mathring{x}_1, A_1 (x - \mathring{x}_1))$$
$$+ (1-\gamma)(x - \mathring{x}_2, A_2 (x - \mathring{x}_2)) \le 1\}, \ \gamma \in (0, 1)$$

Straightforward calculations imply

$$\gamma (x - \mathring{x}_1, A_1 (x - \mathring{x}_1)) + (1-\gamma)(x - \mathring{x}_2, A_2 (x - \mathring{x}_2))$$
$$= \left(x, A_\gamma x\right) - 2\left(b_\gamma, x\right) + \alpha_\gamma \le 1$$

Applying Lemma 6.5 we get (6.48). Lemma is proven. □