Let $m, n \in \mathbb{N}^*$ be fixed integers.

Definition 0.1 : Moore-Penrose pseudo-inverse

Let $A \in \mathbb{R}^{m\times n}$ be a matrix. We say that a matrix $B \in \mathbb{R}^{n\times m}$ is a *Moore-Penrose pseudo-inverse* of A if it satisfies all following conditions:

- $ABA = A$;
- $BAB = B$;
- AB is symmetric ;
- BA is symmetric.

Proposition 0.2

Any matrix A has a unique Moore-Penrose pseudo-inverse. We usually denote it $A^\dagger$.

Proof. Let $A \in \mathbb{R}^{m\times n}$ be fixed. We denote $a : \mathbb{R}^n \to \mathbb{R}^m$ the linear map whose representer matrix (in the canonical basis) is A.

Let us denote $\tilde{a} : \mathrm{Ker}(a)^\perp \to \mathrm{Range}(a)$ the restriction of a to $\mathrm{Ker}(a)^\perp$. It is injective (as $\mathrm{Ker}(a)^\perp$ contains no non-zero element of $\mathrm{Ker}(a)$). As, from the rank theorem,

$$\dim(\mathrm{Range}(a)) = n - \dim(\mathrm{Ker}(a)) = \dim(\mathrm{Ker}(a)^\perp),$$

it is actually bijective. We denote $\tilde{a}^{-1} : \mathrm{Range}(a) \to \mathrm{Ker}(a)^\perp$ its inverse.

We define $b : \mathbb{R}^m \to \mathbb{R}^n$ as the only linear map such that

$$\begin{aligned} b(x) &= \tilde{a}^{-1}(x), \forall x \in \mathrm{Range}(a); \\ &= 0, \forall x \in \mathrm{Range}(a)^\perp. \end{aligned}$$

The linear map $a \circ b$ is the orthogonal projector onto $\mathrm{Range}(a)$: from its definition, it is the identity on $\mathrm{Range}(a)$ and zero on $\mathrm{Range}(a)^\perp$. This implies that $a \circ b$ is symmetric, and that $(a \circ b) \circ a = a$.

Similarly, $b \circ a$ is the orthogonal projector onto $\mathrm{Ker}(a)^\perp = \mathrm{Range}(b)$. This implies that $b \circ a$ is symmetric and $(b \circ a) \circ b = b$.

Consequently, if we define B as the matrix representing b in the canonical basis, B satisfies all properties required by the definition of the Moore-Penrose pseudo-inverse. Consequently, we have shown existence.

Let us now show uniqueness. Let C be another matrix satisfying the same properties. Let c be the associated linear map. Let us show that $c = b$.

Since $(c \circ a) \circ c = c$, we must have $\text{Range}(c) \subset \text{Range}(c \circ a) \subset \text{Range}(c)$, hence $\text{Range}(c) = \text{Range}(c \circ a)$. Similarly, $\text{Ker}(c) = \text{Ker}(a \circ c)$.

As $a \circ c$ is self-adjoint, $\text{Ker}(a \circ c) = \text{Range}(a \circ c)^\perp \supset \text{Range}(a)^\perp$, from which we deduce

$$\text{Range}(a)^\perp \subset \text{Ker}(c).$$

In particular, for any $x \in \text{Range}(a)^\perp$, $c(x) = 0 = b(x)$.

As $c \circ a$ is self-adjoint, $\text{Range}(c \circ a) = \text{Ker}(c \circ a)^\perp \subset \text{Ker}(a)^\perp$. From this, we deduce that

$$\text{Range}(c) \subset \text{Ker}(a)^\perp.$$

In particular, for any $x = a(y) \in \text{Range}(a)$, $c(x)$ is an element of $\text{Ker}(a)^\perp$ such that $\tilde{a}(c(x)) = a(c(a(y))) = a(y) = x$. Therefore, $c(x) = \tilde{a}^{-1}(x) = b(x)$.

We have shown that $b = c$ on $\text{Range}(a)$ and $\text{Range}(a)^\perp$. The equality follows on all $\mathbb{R}^m$ by linearity. □

Proposition 0.3

For any matrix A,

- if A is invertible, then $A^\dagger = A^{-1}$;
- if $A^T A$ is invertible, then $A^\dagger = (A^T A)^{-1} A^T$;
- if AA^T is invertible, then $A^\dagger = A^T (AA^T)^{-1}$.

Proof. It suffices to check, in each case, that the four properties of the definition hold. □

Proposition 0.4

Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Let $b \in \mathbb{R}^m$ be a vector. Then $A^\dagger b$ is a minimizer of the map

$$f : x \in \mathbb{R}^n \to \frac{1}{2}||Ax - b||_2^2.$$

(Even more, it is the minimizer with the smallest norm.)

Proof. Let us consider an arbitrary $x \in \mathbb{R}^n$. We write $x = z + A^\dagger b$. Then, from the properties of $A^\dagger$,

$$\begin{aligned} f(x) &= \frac{1}{2}||Az + AA^\dagger b - b||_2^2 \\ &= \frac{1}{2}||Az||_2^2 + \left\langle Az, AA^\dagger b \right\rangle - \langle Az, b\rangle + f(A^\dagger b) \\ &= \frac{1}{2}||Az||_2^2 + \left\langle AA^\dagger Az, b \right\rangle - \langle Az, b\rangle + f(A^\dagger b) \\ &= \frac{1}{2}||Az||_2^2 + \langle Az, b\rangle - \langle Az, b\rangle + f(A^\dagger b) \\ &= \frac{1}{2}||Az||_2^2 + f(A^\dagger b) \\ &\geq f(A^\dagger b). \end{aligned}$$

Therefore, $f(A^\dagger b) = \min f$.

From the previous inequalities, we also see that $f(x) = f(A^\dagger b) = \min f$ if and only if $Az = 0$. In this case,

$$\begin{aligned} ||x||_2^2 &= ||z||_2^2 + 2\left\langle z, A^\dagger b \right\rangle + ||A^\dagger b||_2^2 \\ &= ||z||_2^2 + 2\left\langle z, A^\dagger AA^\dagger b \right\rangle + ||A^\dagger b||_2^2 \\ &= ||z||_2^2 + 2\left\langle A^\dagger Az, A^\dagger b \right\rangle + ||A^\dagger b||_2^2 \\ &= ||z||_2^2 + 2\left\langle A^\dagger 0, A^\dagger b \right\rangle + ||A^\dagger b||_2^2 \\ &= ||z||_2^2 + ||A^\dagger b||_2^2 \\ &\geq ||A^\dagger b||_2^2, \end{aligned}$$

with equality if and only if $z = 0$. Consequently, $A^\dagger b$ has minimal norm among all minimizers of f.

□