Minimum Norm Least Squares with Pseudoinverse

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Consider the SVD $A = U\Sigma V^\top$ and set $A^+ = V\Sigma^+ U^\top$, where $\Sigma^+ = \mathrm{diag}(\frac{1}{\sigma_1},\dots,\frac{1}{\sigma_r},0,\dots,0)$ is the pseudoinverse of a diagonal matrix as derived in the lecture. Show that

$$x^+ := A^+ b = \arg \min_{x\in\{x:\ A^\top Ax = A^\top b\}} \|x\|_2^2,$$

i.e., x^+ is the minimum norm least squares solution.

Hint: First consider the simple case that A is diagonal and then use the SVD for the general case.

Solution:

(1) Special Case: Diagonal matrix

Let us start with the simple case: $A \in \mathbb{R}^{m\times n}$ diagonal

$$A = \begin{pmatrix} a_{11} & & & & 0 \\ & \ddots & & & \\ & & a_{rr} & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix} \in \mathbb{R}^{m\times n},\ a_{ii} \neq 0, \quad A^T A = \begin{pmatrix} a_{11}^2 & & & & 0 \\ & \ddots & & & \\ & & a_{rr}^2 & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix} \in \mathbb{R}^{n\times n}$$

Normal equation:

$$A^T A x = A^T b = \begin{pmatrix} a_{11}b_1 \\ \vdots \\ a_{rr}b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{matrix} a_{11}^2 x_1 = a_{11} b_1 \\ \vdots \\ a_{rr}^2 x_r = a_{rr} b_r \\ 0 \cdot x_i = 0 \\ (i > r) \end{matrix} \quad \Leftarrow \quad \begin{matrix} x_1 = \frac{b_1}{a_{11}} \\ \vdots \\ x_r = \frac{b_r}{a_{rr}} \\ x_{r+1} = 0 \\ \vdots \\ x_n = 0 \end{matrix}$$

$$\Rightarrow \quad x^+ = \begin{pmatrix} \frac{1}{a_{11}} b_1 \\ \vdots \\ \frac{1}{a_{rr}} b_r \\ 0 \\ \vdots \end{pmatrix} 0 = A^+ b, \quad A^+ = \begin{pmatrix} \frac{1}{a_{11}} & & & & 0 \\ & \ddots & & & \\ & & \frac{1}{a_{rr}} & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix} \in \mathbb{R}^{n\times m}$$

Note: The x_i for $i > r$ can be chosen arbitrarily, but setting them to zero gives the smallest vector.

(2) General Case

By using the SVD $A = U\Sigma V^\top$ we find

$$A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T\Sigma V^T,$$

so that the normal equation reads as

$$(*) \quad A^T A x = A^T b \ \Leftrightarrow\ V\Sigma^T\Sigma(V^T x) = V\Sigma^T(U^T b)$$
$$\overset{V^T\cdot|}{\Leftrightarrow}\ \underbrace{\Sigma^T\Sigma(V^T x) = \Sigma^T(U^T b)}_{\text{(normal equation for}(\Sigma,\ U^T b))} \quad (\sharp)$$

Consequently, x solves $(*)$ if and only if $y := V^T x$ solves $(\sharp)$. Since V is orthogonal both solutions have the same norm, more precisely,

$$\|x\|_2^2 = x^\top x = x^\top (V^\top V) x = \|Vx\|_2^2 = \|y\|_2^2.$$

For diagonal matrices we have shown that $y^+ = \Sigma^+ U^T b$ is the smallest solution of $(\sharp)$. Thus, $x^+ := Vy^+ = \Sigma^+ U^T b$ is the smallest solution of $(*)$, i.e., the minimum norm least squares solution.

All in all: Since orthogonal matrices (here U and V) are not only invertible but also isometric and the SVD $A = U\Sigma V^\top$ always exists, we could rely on the result for diagonal matrices (here Σ).