Problem 1 (Problem 3.11)

Let $A = U\Sigma V^T$ be the SVD of A, let $\hat{X} = V^T X U$, we claim that $\hat{X} = \Sigma^+$ minimizes

$$||AX - I||_2^2$$

To prove this, note that

$$\begin{aligned}
||AX - I||_F^2 &= ||U\Sigma V^T V\hat{X}U^T - UU^T||_F^2 \\
&= ||\Sigma\hat{X} - I||_F^2 \\
&= \left\| \begin{bmatrix} \sigma_1 x_{11} - 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \sigma_2 x_{21} & \sigma_2 x_{22} - 1 & 0 & \cdots & 0 & \cdots & 0 \\ \sigma_3 x_{31} & \sigma_3 x_{32} & \sigma_3 x_{33} - 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_r x_{r1} & \sigma_r x_{r2} & \sigma_r x_{r3} & \cdots & \sigma_r x_{rr} - 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \right\|_F \\
&= \sum_{i=1}^{r} \left((\sigma_1 x_{ii} - 1)^2 + \sum_{i \neq j} \sigma_i^2 x_{ij}^2 \right) + (n - r)
\end{aligned}$$

To minimize the above line, we choose $\hat{X}$ such that

$$\begin{cases} \sigma_i x_{ii} - 1 = 0 \implies x_{ii} = \frac{1}{\sigma_i} \\ x_{ij} = 0 \end{cases}$$

This suggests that $\hat{X} = \Sigma^+$ minimizes the above expression, with minimum value of $\sqrt{n - r}$. Since $\hat{X} = \Sigma^+$, the minimizer

$$X = V\Sigma^+ U^T$$

Which is the Moore-Penrose pseudoinverse.

Problem 2 (Question 3.12)

Let $A \in \mathbb{R}^{w\times x}, B \in \mathbb{R}^{y\times z}, C \in \mathbb{R}^{w\times y}$. Let $A = U_1\Sigma_1 V_1, B = U_2\Sigma_2 V_2$ be the SVD of A, B, letting $X = V_1\hat{X}U_2$, we have

$$||AXB - C||_F^2 = ||\Sigma_1\hat{X}\Sigma_2 - U_1^T CV_2||_F^2$$

Note that after performing the multiplication, the nonzero entries $\Sigma_1\hat{X}\Sigma_2$ range from 1 to r_1 (in rows) and 1 to r_2 (in columns), where r_1, r_2 denote the rank of A, B respectively. Hence, to minimize the expression, it suffice minimizing

$$||\Sigma_1\hat{X}\Sigma_2 - U_1^T CV_2||_F^2$$

Constrained on 1 to r_1 (in rows) and 1 to r_2. To minimize this expression, it suffice solving

$$\Sigma_1\hat{X}\Sigma_2 = U_1^T CV_2 \tag{*}$$

We prove that $X_0 = A^+CB^+$ solves $(*)$. Since $X = V_1\hat{X}U_2$, it follows that $\hat{X} = V_1^T A^+CB^+U_2^T$

$$\begin{aligned}
\Sigma_1\hat{X}\Sigma_2 &= \Sigma_1 V_1^T A^+CB^+U_2^T\Sigma_2 \\
&= \Sigma_1 V_1^T V_1\Sigma_1^+ U_1^T CV_2\Sigma_2^+ U_2^T U_2^T \Sigma_2 \\
&= \Sigma_1\Sigma_1^+ U_1^T CV_2\Sigma_2^+\Sigma_2 \\
&= \begin{bmatrix} I_{r_1} & 0 \\ 0 & 0 \end{bmatrix} U_1^T CV_2 \begin{bmatrix} I_{r_2} & 0 \\ 0 & 0 \end{bmatrix} \\
&= U_1^T CV_2
\end{aligned}$$

Hence, $X_0 = A^+CB^+$ is a minimizer of $||AXB - C||_F^2$.

Problem 3 (Question 3.13)

Let $A = U\Sigma V^T$, then

1.

$$AA^+A = U\Sigma V^T V\Sigma^+ U^T U\Sigma V^T = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A$$

2.

$$A^+AA^+ = V\Sigma^+ U^T U\Sigma V^T V\Sigma^+ U^T = V\Sigma^+\Sigma\Sigma^+ U^T = V\Sigma^+ U^T = A^+$$

3.

$$A^+A = V\Sigma^+ U^T U\Sigma V^T = V\Sigma^+\Sigma V^T = V\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^T = (V\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^T)^T = (A^+A)^T$$

4.

$$AA^+ = U\Sigma V^T V\Sigma^+ U^T = U\Sigma\Sigma^+ U^T = U\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T = (U\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T)^T = (AA^+)^T$$

Problem 4 (Problem 3.14)

Suppose $A \in \mathbb{R}^{m\times n}$ is rectangular. Let $A = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$, where r is the rank of A. We have

$$H = \begin{bmatrix} 0 & \sum_{i=1}^r \sigma_i v_i u_i^T \\ \sum_{i=1}^r \sigma_i u_i v_i^T & 0 \end{bmatrix}$$

Note that for $k = 1, ..., r$, we have

$$H \begin{bmatrix} v_k \\ u_k \end{bmatrix} = \begin{bmatrix} 0 & \sum_{i=1}^r \sigma_i v_i u_i^T \\ \sum_{i=1}^r \sigma_i u_i v_i^T & 0 \end{bmatrix} \begin{bmatrix} v_k \\ u_k \end{bmatrix} = \sigma_k \begin{bmatrix} v_k \\ u_k \end{bmatrix}$$

$$H \begin{bmatrix} v_k \\ -u_k \end{bmatrix} = \begin{bmatrix} 0 & \sum_{i=1}^r \sigma_i v_i u_i^T \\ \sum_{i=1}^r \sigma_i u_i v_i^T & 0 \end{bmatrix} \begin{bmatrix} v_k \\ -u_k \end{bmatrix} = -\sigma_k \begin{bmatrix} v_k \\ -u_k \end{bmatrix}$$

This suggests that $\pm\sigma_k$ are H's eigenvalues (there are $2r$ of them), and the corresponding eigenvectors are by

$$\frac{1}{\sqrt{2}} \begin{bmatrix} v_k \\ \pm u_k \end{bmatrix}$$

In the case where A is square and has full rank, we are done (plug in $r = n$ gives us the result).
However, when A is rectangular with rank r, note that there are $2r$ eigenvalues $\pm\sigma_k$. The remaining $m - n - 2r$ eigenvalues are all 0, and the corresponding eigenvector can be composed by the basis vectors of the null space of A and A^T.
Since there are $n - r$ basis x_k for Null(A) and $m - r$ basis y_k for Null(A^T), it follows that the eigenvectors (of eigenvalue 0) takes in the form

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

Problem 5

Recall that in Ridge regression, we aim to minimize

$$||Ax - b|| + 2^2 + \lambda ||x||_2^2 = || \begin{bmatrix} A \\ \lambda I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} ||_2^2$$

Suppose we want to try different λ values. Consider first apply a sequence of Given's rotation to transform the matrix A into an upper triangular matrix R

$$Q_1 Q_2 ... Q_k \begin{bmatrix} A \\ \lambda I \end{bmatrix} \to \begin{bmatrix} R \\ \lambda I \end{bmatrix}$$

If we store such representation, then for each λ, it remaings using n Given's rotation to clear the n diagonal entries λ. This takes an additional $O(n^3)$ time per lambda. The system then can be solved using back-substitution.