Problem 1 :

Given the data is zero mean

$\sum_{n=1}^{N} x^{(n)} = 0$ & $b, c = 0$ as they can be dropped for simplicity.

Now: $\hat{x}^{(n)} = V(Wx^{(n)})$

Loss function : $\sum_{n=1}^{N} \| x^{(n)} - \hat{x}^{(n)} \|^2$

$$= \sum \| x^{(n)} - V(Wx^{(n)}) \|^2$$

Now:

$$\mathcal{L}(v) = \| X - VWX \|^2$$

Do $\frac{\partial}{\partial V}$ for $\mathcal{L}$

$$\frac{\partial}{\partial V} \| X - VWX \|^2 = -2(X - VWX) X^T W^T \Rightarrow 0$$

$$(X - VWX)(X^T W^T) = 0$$

$VWXX^T W^T = XX^T W^T \Rightarrow VWW^T = W^T$ as XX^T can be cancelled

Given XX^T is non-degenerate in subspace of W,

$V = W^T$ $WW^T = I$ since W is projection matrix

Problem 2 :

$\Sigma = \frac{1}{N} \sum_{n=1}^{N} x^{(n)} (x^{(n)})^T \in \mathbb{R}^{d \times d}$ This is sample covariance matrix

$\|x - W^T W x\|^2 = \|x\|^2 - 2 \, \text{trace} (x^T W^T W x) + \|W^T W x\|^2$

Miniming the LHS $\approx$ Maximising RHS.

($W^T W = I$) orthogonal

Let $\Sigma = U \Lambda U^T$ is the eigen-decomposition of Σ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ with $\lambda_1 \geq \lambda_2 \geq \dots \lambda_d \geq 0$ and $U = [u_1, u_2, \dots u_d]$. The top k principal components are $u_1, \dots, u_k$. Thus optimal W has rows $\{u_1^T, \dots u_k^T\}$

$W^T = [u_1, u_2, \dots u_k]$

$V = W^T = [u_1, u_2, \dots u_k]$

Hence the columns of W^T are an orthonormal basis for top k- principal components.