1) Given reconstruction loss function

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

where

$$\hat{x}^{(n)} = V(Wx^{(n)} + b) + c$$

Since we assume data is mean-centered

$$\sum_{n=1}^{N} x^{(n)} = 0$$

Simplifying this to

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

$$L = \sum_{n=1}^{N} \| x^{(n)} - VWx^{(n)} \|^2$$

$$\frac{dL}{dV} = -2 \sum_{n=1}^{N} (x^{(n)} - VWx^{(n)}) x^{(n)T}.$$

As $\frac{dL}{dV} = 0$, we get

$$\sum_{n=1}^{N} x^{(n)} x^{(n)T} = \sum_{n=1}^{N} VWx^{(n)} x^{(n)T}$$

Having $\Sigma = \frac{1}{N} \sum_{n=1}^{N} x^{(n)} x^{(n)T}$

$$\Sigma = VW\Sigma$$

$\therefore$ If W is chosen optimally, the best choice is

$$V = W^T$$

2) Need to show that optimal weight matrix W^* forms an orthonormal basis for the top k eigenvector of Σ

$$L = \sum_{n=1}^{N} \| x^{(n)} - W^T W x^{(n)} \|^2$$

$$E[\| x - W^T W x \|^2] =$$

$$tr(\Sigma - W^T W \Sigma - \Sigma W^T W + W^T W \Sigma W^T W).$$

Since Σ is symmetric

$$E(\| x - W^T W x \|^2) = tr(\Sigma) - tr(W^T W \Sigma)$$

Minimizing this is same as $tr(W^T W \Sigma)$

The optimal choice of W is to pick the eigenvectors of Σ corresponding to the largest k eigenvalues as they capture maximum variance in the data.
Let $\Sigma = U \Lambda U^T$ be the eigen-decomposition where

- $U = [\mu_1, \mu_2 \ldots \mu_d]$ are eigenvectors of Σ
- $\Lambda = diag(\lambda_1, \lambda_2 \ldots \lambda_d)$ are corresponding eigenvalues

$\therefore$ Choosing W^* as top k eigenvectors

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

This spans the same subspace as the principal eigenvectors of Σ, completing the proof.