Question 6

(a) $E_N(V) \le E_L(V)$

Proof:

$$L(x_i^{(k)}; V) = V_k \alpha_i^{(k)}$$

Suppose we define $\alpha_i = \begin{bmatrix} \alpha_i^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

The first k rows of vector $\alpha_i$ are the same as $\alpha_i^{(k)}$ and the rest are 0. In this notation,

$L(x_i^{(k)}; V) = V\alpha_i$ where $V$ is the entire basis matrix.

$N(x_i^{(k)}; V) = V\beta_i$ where $\beta_i = \underset{c_i \text{ s.t. } c_i \text{ has atmost } k\text{-nonzero components}}{\arg\min} \|x_i - Vc_i\|^2$

By def$^n$ of $\beta_i$,

$\|x_i - V\beta_i\|^2 \le \|x_i - Vc_i\|^2 \quad \forall c_i$ such that above condition is satisfied.

Since $\alpha_i$ is a special kind of $c_i$ s.t. only the first k components are non-zero, reconstruction error using non-linear method is lower or equal.

$$\therefore \|x_i - N(x_i^{(k)}; V)\|^2 \le \|x_i - L(x_i^{(k)}; V)\|^2$$

On summing on both sides over all points,

$$E_N(V) \le E_L(V)$$

(b) Here we have $N > d$ and the covariance matrix $C$ is of size $d \times d$. Furthermore, $V$ ~~has~~ forms a basis matrix for a $d$-dimensional vector space.

$\therefore$ Let $x_i = V\alpha_i$ where $\alpha_i \in \mathbb{R}^{d \times 1}$

and $\alpha_i = V^T x_i$.

Now for each $x_i$, we want to find a $c_i$ s.t. $\|x_i - Vc_i\|^2$ is minimized.*

Putting $x_i = V\alpha_i$,

$$
\begin{aligned}
\|V\alpha_i - Vc_i\|^2 &= \|V(\alpha_i - c_i)\|^2 \\
&= (\alpha_i - c_i)^T V^T V (\alpha_i - c_i) \\
&= (\alpha_i - c_i)^T (\alpha_i - c_i) \\
&= \|\alpha_i - c_i\|^2 = \sum_{j=1}^{d} (\alpha_{ij} - c_{ij})^2
\end{aligned}
$$

To minimize this, we need to consider the $K$ largest absolute values in $\alpha_i$ and use them as our $c_i$ with the remaining positions having 0.

* Since we want to minimize the sum of squares, it will be smallest only when the largest $K$ terms do not contribute to this. That happens when $c_i$ has the $K$ largest terms in terms of magnitude of $\alpha_i$.

To do this, we need to sort each $\alpha_i$ according to magnitude. This takes $O(d \log d)$. For $N$ points $O(N d \log d)$ complexity does the job.

## (c)

It is possible to construct a W such that $E_N(W) < E_N(V)$. Consider the following scenario.

$$x1 = [-1; 10; 0.3]^T$$

$$x2 = [-0; 12; -1.3]^T$$

$$x3 = [-0.1; -3; 2.12]^T$$

$$x4 = [-0; 2.33; -3.31]^T$$

W =

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The approximation errors are on doing a degree 1 approximation i.e. k = 1:

| | |
|---|---|
| V | 12.939904 |
| W | 12.713300 |

The approximation errors are on doing a degree 2 approximation i.e. k = 2:

| | |
|---|---|
| V | 0.287779 |
| W | 0.100000 |

The associated code to test this has been placed in the `code` folder.