By definition,

$$\forall i\ \forall c_i \quad \| x_i - V\alpha_i \|^2 \le \| x_i - Vc_i \|^2$$

c_i contains at most k non-zero elements.

$$\cancel{\text{Choosing } c_i = \alpha_i^{(k)} = V_k^t x_i}$$

$$\cancel{\Rightarrow \forall i \quad \| x_i - N(x_i^{(k)}, V) \|^2 \le \| x_i - V\alpha_i^{(k)} \|}$$

$$\text{Choosing } c_i = \begin{bmatrix} \alpha_i^{(k)} \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \Big\} \ d-k \text{ zeros.}$$

$$Vc_i = V_k \, \alpha_i^{(k)}$$

$$\forall i \quad \| x_i - V\alpha_i \|^2 \le \| x_i - Vc_i \|^2$$

$$= \| x_i - V^k \alpha_i^{(k)} \|^2$$

$$\Rightarrow \forall i \quad \| x_i - N(x_i^{(k)}; V) \|^2 \le \| x_i - L(x_i^{(k)}; V) \|^2$$

$$\Rightarrow E_N(V) = \sum_{i=1}^{N} \| x_i - N(x_i^{(k)}; V) \|^2 \le \sum_{i=1}^{N} \| x_i - L(x_i^{(k)}; V) \|^2 = E_L(V)$$

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

Algorithm for finding $N(x_i^{(k)}; V)$:

① We shall find α_i as:

$$\alpha_i' = \left(V^t x_i\right)$$

Set the $(d-k)$ smallest elements in α_i' to zero to obtain α_i.

Why this works: (and finds α_i exactly)

$$V \alpha_i' = V V^t x_i = x_i$$

~~$x_i - N(x_i'$~~

$$\| x_i - Vc_i \|^2 = \| V\alpha_i' - Vc_i \|^2$$

$$= \| V(\alpha_i' - c_i) \|^2 = (\alpha_i' - c_i)^t V^t V (\alpha_i' - c_i)$$

$$= (\alpha_i' - c_i)^t (\alpha_i' - c_i)$$

$$= \sum_{j=1}^{d} (\alpha_{ij}' - c_{ij})^2 \geq \underbrace{\sum_{j : c_{ij} = 0} (\alpha_{ij}')^2}_{\text{At least } d-k \text{ terms}}$$

$\geq$ Sum of square of $(d-k)$ smallest elements of α_i' (in magnitude)

Further, this lower limit can be achieved by choosing a c_i s.t.

$$\alpha_{ij} = c_{ij} = \begin{cases} \alpha_{ij}' & , \text{ if } \alpha_{ij}' \text{ is one of the } k \text{ largest elements of } \alpha_i' \\ 0 & , \text{ otherwise.} \end{cases}$$

For this c_i, $\| x_i - Vc_i \|^2 = \| \alpha_i' - c_i \|^2 = \sum_{j=1}^{d} (\alpha_{ij}' - c_{ij})^2$

= Sum of square of $(d-k)$ smallest elements of α_i'. (in magnitude)

Complexity of finding order k non linear approximation of x_i given V is:

$$O\begin{pmatrix} d^2 \\ + d \end{pmatrix} \quad \begin{matrix} (\text{finding } V^t x_i) \\ (\text{finding } k \text{ largest elements}) \end{matrix}$$

$$= O(d^2)$$

Due to what we have learnt from PCA, we know that

$$E_L(W) \geq E_L(V) \quad \forall W \quad \forall \text{orders } k$$

i.e. V, which is a reslt from PCA is the set of orthogonal directions which best reconstructs the vectors $x_1, x_2 \ldots$ from their projections, for all orders k.

$$\therefore \quad E_L(W) \geq E_L(V) \geq E_N(V)$$