Recall that the design matrix $\mathbf{\Phi}$ for M base functions and N input entries is defined by:

$$\mathbf{\Phi}_{nj} = \phi_j(\boldsymbol{x}_n)$$

$$\mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{x}_1)^\top \\ \boldsymbol{\phi}(\boldsymbol{x}_2)^\top \\ \vdots \\ \boldsymbol{\phi}(\boldsymbol{x}_N)^\top \end{bmatrix}$$

which is basically the vertical stack of transformed inputs $\boldsymbol{\phi}(\boldsymbol{x}_n)^\top$.

We consider an arbitrary vector $\boldsymbol{v} \in \mathbb{R}^N$. Given that $\text{Span}(\{\boldsymbol{\phi}_j\}_{j\in\{1,\ldots,M\}}) = F$ is a closed subspace of Hilbert space (true for any finite-dimensional subspace), so that $\mathbb{R}^N = F \oplus F^\perp$, where $F^\perp$ is the orthogonal complement of F and $\oplus$ is inner direct sum.

- If $\boldsymbol{v} \in F$, which means that $\sum_{j=1}^{M} \lambda_j \boldsymbol{\phi}_j = \mathbf{\Phi}\boldsymbol{\lambda}$, then we have

$$\mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\boldsymbol{v} = \mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\mathbf{\Phi}\boldsymbol{\lambda} = \mathbf{\Phi}\boldsymbol{\lambda} = \boldsymbol{v}$$

- If $\boldsymbol{v} \in F^\perp$, so that $\boldsymbol{\phi}_j^\top\boldsymbol{v} = 0$, then

$$\mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\boldsymbol{v} = \mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{0} = \mathbf{0}$$

- If $\boldsymbol{v}$ in the subspace other than former cases, it can be decomposed into two vectors $\boldsymbol{v} = \boldsymbol{v}_F + \boldsymbol{v}_{F^\perp}$, such that $\boldsymbol{v}_F \in F, \boldsymbol{v}_{F^\perp} \in F^\perp$.

$$\mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\boldsymbol{v} = \mathbf{\Phi}(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top(\boldsymbol{v}_F + \boldsymbol{v}_{F^\perp}) = \boldsymbol{v}_F + \mathbf{0} = \boldsymbol{v}_F$$

For the least-square solution

$$\boldsymbol{w}_{\text{ML}} = (\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\mathsf{t}, \tag{3.15}$$

we know, by using the results above, that there exist a vector $\boldsymbol{\lambda}_F \in \mathbb{R}^M$ in the manifold $\mathcal{S}$ such that after the orthogonal projection by $(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top$, t becomes

$$\begin{aligned}
&(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\mathsf{t} \\
=&(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top(\mathsf{t}_F + \mathsf{t}_{F^\perp}) \\
=&(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\mathsf{t}_F \\
=&(\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top\mathbf{\Phi}\boldsymbol{\lambda}_F = \boldsymbol{\lambda}_F
\end{aligned}$$

■