

3.14 ψ_j is a linear combination of the basis functions $\phi_0 \dots \phi_{M-1}$, giving us:

$$\psi_j(\mathbf{x}) = \sum_{i=0}^{M-1} \alpha_{i,j} \phi_i(\mathbf{x})$$

Let us define a matrix \mathbf{A} such that:

$$\mathbf{A} = \begin{bmatrix} \alpha_{0,0} & \alpha_{1,0} & \dots & \alpha_{(M-1),0} \\ \alpha_{0,1} & \alpha_{1,1} & \dots & \alpha_{(M-1),1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{0,j} & \alpha_{1,j} & \dots & \alpha_{(M-1),j} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{0,(M-1)} & \alpha_{1,(M-1)} & \dots & \alpha_{(M-1),(M-1)} \end{bmatrix}$$

This implies that:

$$\begin{aligned} \mathbf{A}\Phi^T &= \begin{bmatrix} \alpha_{0,0} & \alpha_{1,0} & \dots & \alpha_{(M-1),0} \\ \alpha_{0,1} & \alpha_{1,1} & \dots & \alpha_{(M-1),1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{0,j} & \alpha_{1,j} & \dots & \alpha_{(M-1),j} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{0,(M-1)} & \alpha_{1,(M-1)} & \dots & \alpha_{(M-1),(M-1)} \end{bmatrix} \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \phi(\mathbf{x}_1) & \phi(\mathbf{x}_2) & \dots & \phi(\mathbf{x}_N) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \\ &= \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \psi(\mathbf{x}_1) & \psi(\mathbf{x}_2) & \dots & \psi(\mathbf{x}_N) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \Psi^T \end{aligned}$$

and $\mathbf{A}\phi(\mathbf{x}_i) = \psi(\mathbf{x}_i)$

$$\implies \Psi = (\mathbf{A}\Phi^T)^T = \Phi\mathbf{A}^T$$

$$\implies \Phi = \Psi\mathbf{A}^{-T}$$

Note: \mathbf{A}^{-1} exists because change of basis matrices are invertible. Source.

Using 3.54,

$$\begin{aligned}
\mathbf{S}_N &= (\alpha \mathbf{I} + \beta \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \\
&= (\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \quad \text{Since } \alpha = 0 \\
&= (\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \\
&= (\beta (\boldsymbol{\Psi} \mathbf{A}^{-T})^T (\boldsymbol{\Psi} \mathbf{A}^{-T}))^{-1} \\
&= (\beta \mathbf{A}^{-1} \boldsymbol{\Psi}^T \boldsymbol{\Psi} \mathbf{A}^{-T})^{-1}
\end{aligned}$$

Since $\boldsymbol{\Psi}$ is orthonormal basis, $\boldsymbol{\Psi}^T \boldsymbol{\Psi} = \mathbf{I}$, giving us:

$$\begin{aligned}
\mathbf{S}_N &= (\beta \mathbf{A}^{-1} \mathbf{I} \mathbf{A}^{-T})^{-1} \\
&= \frac{1}{\beta} (\mathbf{A}^{-1} \mathbf{A}^{-T})^{-1} \\
&= \frac{1}{\beta} (\mathbf{A}^T \mathbf{A})
\end{aligned}$$

Applying this to 3.62, we get

$$\begin{aligned}
k(\mathbf{x}, \mathbf{x}') &= \beta \boldsymbol{\phi}(\mathbf{x})^T \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}') \\
&= \beta \boldsymbol{\phi}(\mathbf{x})^T \left(\frac{1}{\beta} (\mathbf{A}^T \mathbf{A}) \right) \boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\phi}(\mathbf{x})^T \mathbf{A}^T \mathbf{A} \boldsymbol{\phi}(\mathbf{x}') \\
&= (\mathbf{A} \boldsymbol{\phi}(\mathbf{x}))^T \mathbf{A} \boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\psi}(\mathbf{x})^T \boldsymbol{\psi}(\mathbf{x}')
\end{aligned}$$

Now to prove that the kernel satisfies the summation constraint:

$$\sum_{n=1}^N k(\mathbf{x}, \mathbf{x}_n) = \sum_{n=1}^N \boldsymbol{\psi}(\mathbf{x})^T \boldsymbol{\psi}(\mathbf{x}_n)$$

$$\begin{aligned}
&= \sum_{n=1}^N \sum_{m=0}^{M-1} \psi_m(\mathbf{x}) \psi_m(\mathbf{x}_n) \\
&= \sum_{m=0}^{M-1} \sum_{n=1}^N \psi_m(\mathbf{x}) \psi_m(\mathbf{x}_n) \\
&= \sum_{m=0}^{M-1} \psi_m(\mathbf{x}) \sum_{n=1}^N \psi_m(\mathbf{x}_n) \\
&= \sum_{m=0}^{M-1} \psi_m(\mathbf{x}) \sum_{n=1}^N \psi_m(\mathbf{x}_n) \psi_0(\mathbf{x}_n) \\
&= \psi_0(\mathbf{x}) \sum_{n=1}^N \psi_0(\mathbf{x}_n) \psi_0(\mathbf{x}_n) + \sum_{m=1}^{M-1} \psi_m(\mathbf{x}) \sum_{n=1}^N \psi_m(\mathbf{x}_n) \psi_0(\mathbf{x}_n) \\
&= 1.1 + \sum_{m=1}^{M-1} \psi_m(\mathbf{x}).0 \\
&= 1.
\end{aligned}$$