Question 1

$$\langle K_a\alpha, K_b\beta \rangle$$
$$= \alpha^T K_a^T K_b \beta$$

Because K_a, K_b are kernel matrices (symmetric)

$$K_a^T = K_a$$
$$K_b^T = K_b$$
$$\langle K_a\alpha, K_b\beta \rangle = \alpha^T K_a K_b \beta$$

The constraints:

$$\|K_a\alpha\|_2 = 1 \Rightarrow \|K_a\alpha\|_2^2 = 1 \Rightarrow \alpha^T K_a^2 \alpha = 1$$
$$\|K_b\beta\|_2 = 1 \Rightarrow \|K_b\beta\|_2^2 = 1 \Rightarrow \beta^T K_b^2 \beta = 1$$

The Lagrangian:

$$\mathcal{L} = \alpha^T K_a K_b \beta - \frac{\rho_1}{2}(\alpha^T K_a^2 \alpha - 1) - \frac{\rho_2}{2}(\beta^T K_b^2 \beta^T - 1)$$

The partial derivative w.r.t α, β:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \alpha} = K_a K_b \beta - \rho_1 K_a^2 \alpha = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \beta} = K_b K_a \alpha - \rho_2 K_b^2 \beta = 0 \end{cases}$$

Which is equal to:

$$\begin{cases} K_a K_b \beta = \rho_1 K_a^2 \alpha \quad \dots \quad ① \\ K_b K_a \alpha = \rho_2 K_b^2 \beta \quad \dots \quad ② \end{cases}$$

$$\begin{cases} \alpha^T K_a K_b \beta = \rho_1 \alpha^T K_a^2 \alpha = \rho_1 \\ \beta^T K_b K_a \alpha = \rho_2 \beta^T K_b^2 \beta = \rho_2 \end{cases} \quad (\alpha^T K_a^2 \alpha = \beta^T K_b^2 \beta = 1)$$

$\because$ $\alpha^T K_a K_b \beta = (\beta^T K_b K_a \alpha)^T = \beta^T K_b K_a \alpha$

$$\rho_1 = \rho_2 = \rho$$

by ①, $\frac{1}{\rho}(K_a + c_a I)^{-2} K_a K_b \beta = \alpha$

so, $\frac{1}{\rho} K_b K_a (K_a + c_a I)^{-2} K_a K_b \beta = \rho_2 K_b^2 \beta$

$$(K_b + c_b I)^{-2} K_b K_a (K_a + c_a I)^{-2} K_a K_b \beta = \rho^2 \beta$$

Substitute K_a^{-2} and K_b^{-2} with $(K_a + c_a I)^{-2}$ and $(K_b + c_b I)^{-2}$ in ① and ②

So the problem can be formulated as:

$$\begin{pmatrix} 0 & K_a K_b \\ K_b K_a & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} (K_a + c_a I)^2 & 0 \\ 0 & (K_b + c_b I)^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Question 2

$$\hat{\mathrm{cov}}(\langle \phi_a(X_a), W_a\rangle, \langle \phi_b(X_b), W_b\rangle) = \frac{1}{n}\sum_{k=1}^{n} \langle \phi_a(X_a^k), W_a\rangle \langle \phi_b(X_b^k), W_b\rangle$$

$$= \frac{1}{n}\sum_{k=1}^{n} \langle \phi_a(X_a^k), W_a^{\parallel} + W_a^{\perp}\rangle \langle \phi_b(X_b^k), W_b^{\parallel} + W_b^{\perp}\rangle$$

$$= \frac{1}{n}\sum_{k=1}^{n} \langle \phi_a(X_a^k), W_a^{\parallel}\rangle \langle \phi_b(X_b^k), W_b^{\parallel}\rangle$$

$$= \frac{1}{n}\sum_{k=1}^{n} \phi_a(X_a^k)^T \sum_{i=1}^{n} \alpha_i \phi_a(X_a^i) \; \phi_b(X_b^k)^T \sum_{j=1}^{n} \beta_j \phi_b(X_b^j)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \phi_a(X_a^k)^T \alpha_i \phi_a(X_a^i) \phi_b(X_b^k)^T \beta_j \phi_b(X_b^j)$$

And,

$$\frac{1}{n}\alpha^T K_a K_b \beta = \frac{1}{n}(\alpha^T K_a)\cdot(\beta^T K_b^T)^T = \frac{1}{n}(\alpha^T K_a)\cdot(\beta^T K_b)^T$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} \alpha_i K_{ai} \cdot \sum_{j=1}^{n} \beta_j K_{bj}^T \quad (K_i \text{ is the ith row of } K)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \beta_j K_{ai}\cdot K_{bj}^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \beta_j \sum_{k=1}^{n} K_{ai,k}\cdot K_{bj,k}$$

by $K_a(X_a^i, X_a^k) = \langle \phi_a(X_a^i), \phi_a(X_a^k)\rangle$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \beta_j \sum_{k=1}^{n} \phi_a(X_a^k)^T \cdot \phi_a(X_a^i)\cdot \phi_b(X_b^k)^T \cdot \phi_b(X_b^j)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \beta_j \phi_a(X_a^k)^T \phi_a(X_a^i)\cdot \phi_b(X_b^k)^T \phi_b(X_b^j)$$

So,

$$\begin{cases} \hat{\mathrm{cov}}(\langle \phi_a(X_a), W_a\rangle, \langle \phi_b(X_b), W_b\rangle) = \frac{1}{n}\sum_{i,j,k=1}^{n} \alpha_i\beta_j \phi_a(X_a^k)^T \phi_a(X_a^i) \phi_b(X_b^k)^T \phi_b(X_b^j) \\ \frac{1}{n}\sum_{i,j,k=1}^{n} \alpha_i\beta_j \phi_a(X_a^k)^T \phi_a(X_a^i) \phi_b(X_b^k)^T \phi_b(X_b^j) = \frac{1}{n}\alpha^T K_a K_b \beta \end{cases}$$

$$\Downarrow$$

$$\hat{\mathrm{cov}}(\langle \phi_a(X_a), W_a\rangle, \langle \phi_b(X_b), W_b\rangle) = \frac{1}{n}\alpha^T K_a K_b \beta$$

$\mathrm{cov}(\langle \phi_a(X_a), W_a \rangle, \langle \phi(X_b), W_b \rangle)$

$= \frac{1}{n} \sum \phi_a(X_a^k)^T W_a \cdot \phi_b(X_b^k)^T W_b$

$= \frac{1}{n} \sum_{k=1}^{n} \phi_a(X_a^k)^T (W_a^{\parallel} + W^{\perp}) \cdot \phi_b(X_b^k)^T (W_b^{\parallel} + W^{\perp})$

$= \frac{1}{n} \sum_{k=1}^{n} \phi_a(X_a^k)^T \sum_{i=1}^{n} \alpha_i \phi_a(X_a^i) \phi_b(X_b^k)^T \sum_{j=1}^{n} \beta_j \phi_b(X_b^j)$

$\alpha^T K_a = [\alpha_1, \ldots \alpha_i \ldots \alpha_n]$

$[\alpha \ldots]$

$\sum (X_a^k)^T X_a^T \alpha \quad n\times 1 \, (X_b^k)^T X_b^T \beta$

$\phi_a \quad \alpha^T X_a \qquad X_a^T \alpha \; X_b^T \alpha \sum$

$[\sum \alpha_i k_{i,i}, \ldots, \sum \alpha_i k_{i,j} \ldots \sum \alpha_i k_{i,n}]$

$= \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_a(X_a^k)^T \alpha_i^T \phi_a(X_a^i) \quad \phi_b(X_b^j)$

$=$

$\frac{1}{n} \alpha^T K_a K_b \beta = \sum_{i=1}^{n} \alpha_i k_i \cdot \sum \beta_j k_j^T \ldots$

$1\times n \; n\times n \; n\times n \; n\times 1$

$= \sum \sum \alpha_i \beta_j k_{ai} k_{bj}^T$

$= \sum_{i,j,k} \sum \sum \alpha_i \beta_j \, k_{ai,k} \cdot k_{bj,k}$

$= \sum_i \sum_j \sum_k \alpha_i \beta_j \cdot \phi_a(X_a^k)^T \cdot \phi_b(X_a^i) \, \phi_b(X_b^k)^T \phi_b(X_b^j)$