Exercise 2

a) Plugging $w_x = X\alpha_x$ and $w_y = Y\alpha_y$ in (1) gives us

$$w_x^T C_{xy} w_y = (X\alpha_x)^T C_{xy} Y\alpha_y = \frac{1}{N}\alpha_x^T X^T X Y^T Y \alpha_y,$$

$$w_x^T C_{xx} w_x = (X\alpha_x)^T C_{xx} X\alpha_x = \frac{1}{N}\alpha_x^T X^T X X^T X \alpha_x$$

and

$$w_y^T C_{yy} w_y = (Y\alpha_y)^T C_{yy} Y\alpha_y = \frac{1}{N}\alpha_y^T Y^T Y Y^T Y \alpha_y.$$

Now let $K_x = X^T X$ and $K_y = Y^T Y$ be the linear kernels. We then obtain the kCCA problem

$$\max_{\alpha_x,\alpha_y} \alpha_x^T K_x K_y \alpha_y$$

with

$$\alpha_x^T K_x^2 \alpha_x = 1 \quad \text{and} \quad \alpha_y^T K_y^2 \alpha_y = 1 \qquad (*)$$

The corresponding Lagrangian is

$$L = \alpha_x^T K_x K_y \alpha_y - \frac{\lambda_{\alpha_x}}{2}(\alpha_x^T K_x^2 \alpha_x - 1) - \frac{\lambda_{\alpha_y}}{2}(\alpha_y^T K_y^2 \alpha_y - 1)$$

It holds by taking partial derivatives w.r.t α_x, α_y that

$$\frac{\partial L}{\partial \alpha_x} = K_x K_y \alpha_y - \lambda_{\alpha_x} K_x^2 \alpha_x$$

and

$$\frac{\partial L}{\partial \alpha_x} = K_y K_x \alpha_x - \lambda_{\alpha_y} K_y^2 \alpha_y$$

We set the partial derivatives to 0 and multiply with α_x^T, α_y^T

$$\alpha_x^T K_x K_y \alpha_y - \lambda_{\alpha_x} \alpha_x^T K_x^2 \alpha_x = 0 \ (**) \qquad \text{and} \qquad \alpha_y^T K_y K_x \alpha_x - \lambda_{\alpha_y} \alpha_y^T K_y^2 \alpha_y = 0 \ (***)$$

Subtrachting (***) from (**) gives

$$\lambda_{\alpha_y} \alpha_y^T K_y^2 \alpha_y - \lambda_{\alpha_x} \alpha_x^T K_x^2 \alpha_x = 0$$

Due to (*) we have $\lambda_{\alpha_x} = \lambda_{\alpha_y}$. Let $\lambda = \lambda_{\alpha_x} = \lambda_{\alpha_y}$, it holds

$$\alpha_y = \frac{K_y^{-1} K_x \alpha_x}{\lambda}$$

Plugging in (**) gives us

$$K_x^2 \alpha_x - \lambda^2 K_x^2 \alpha_x = 0 \Leftrightarrow \alpha_x = (I - \lambda^2)^{-1}.$$

Thus, we always find an optimal solution in the span of the data.