Problem 1.a. Show that it is always possible to find an optimal solution in the span of the data, that is

$$w_x = X\alpha_x, w_y = Y\alpha_y.$$

with some coefficient vectors $\alpha_x \in \mathbb{R}^N$ and $\alpha_y \in \mathbb{R}^N$.

Solution: Take the original optimization problem and insert the definitions above

$$\begin{aligned} w_x^T XY^T w_y &= \alpha_x^T X^T XY^T Y\alpha_y & (1)\\ &= \alpha_x^T A \cdot B\alpha_y, & (2)\end{aligned}$$

where $A = X^T X$ and $B = Y^T Y$. The constraints become

$$\begin{aligned} w_x^T XX^T w_x &= \alpha_x^T X^T XX^T X\alpha_x = \alpha_x^T A^2\alpha_x = 1 & (3)\\ w_y^T YY^T w_y &= \alpha_y^T Y^T YY^T Y\alpha_y = \alpha_y^T B^2\alpha_y = 1\,. & (4)\end{aligned}$$

□

Problem 1.b. Show that the dual optimization problem is equivalent to finding the solution of the generalized eigenvalue problem

$$\begin{bmatrix} 0 & A\cdot B \\ B\cdot A & 0 \end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix} = \rho\begin{bmatrix} A^2 & 0 \\ 0 & B^2\end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix},$$

in α_x, α_y, where $A = X^T X$ and $B = Y^T Y$.

Solution: From the slides, we know that the dual optimization problem can be written as

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}\begin{bmatrix}w_x\\ w_y\end{bmatrix} = \alpha\begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy}\end{bmatrix}\begin{bmatrix}w_x\\ w_y\end{bmatrix}. \tag{5}$$

Substituting $w_x = X\alpha_x, w_y = Y\alpha_y$, $\alpha = \rho$ and using $C_{xy} = \frac{1}{N}XY^T, C_{xx} = \frac{1}{N}XX^T$ yields

$$\begin{aligned} &\begin{bmatrix} 0 & XY^T \\ YX^T & 0 \end{bmatrix}\begin{bmatrix}X\alpha_x\\ Y\alpha_y\end{bmatrix} = \rho\begin{bmatrix} XX^T & 0 \\ 0 & YY^T\end{bmatrix}\begin{bmatrix}X\alpha_x\\ Y\alpha_y\end{bmatrix} & (6)\\ \iff &\begin{bmatrix} 0 & XY^TY \\ YX^TX & 0 \end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix} = \rho\begin{bmatrix} XX^TX & 0 \\ 0 & YY^TY\end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix} & (7)\end{aligned}$$

Multiply $\begin{bmatrix} X^T & Y^T\end{bmatrix}$ on both sides from the left yields

$$\begin{bmatrix} 0 & X^TXY^TY \\ Y^TYX^TX & 0 \end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix} = \rho\begin{bmatrix} X^TXX^TX & 0 \\ 0 & Y^TYY^TY\end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix}. \tag{8}$$

Substituting $A = X^T X, B = Y^T Y$ yields the desired form

$$\begin{bmatrix} 0 & A\cdot B \\ B\cdot A & 0 \end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix} = \rho\begin{bmatrix} A^2 & 0 \\ 0 & B^2\end{bmatrix}\begin{bmatrix}\alpha_x\\ \alpha_y\end{bmatrix}. \tag{9}$$

□

Problem 1.c. Show how a solution to the original problem can be obtained from a solution of the generalized eigenvalue problem above.

Solution: A solution for the original problem can be obtained from a solution for the problem above by computing

$$w_x = X\alpha_x, \tag{10}$$
$$w_y = Y\alpha_y \tag{11}$$

□

Problem 2.a. The objective of CCA is to find linear combination $\phi_x w_x$ and $\phi_y w_y$ such that the correlation is maximized.
Note that w_x and w_y lie in span of ϕ_x and ϕ_y these can be expressed by the linear transformation: $w_x = \phi_x \alpha_x$ and $w_y = \phi_y \alpha_y$
letting $k_x = \phi_x.\phi_x^\top$ and $k_y = \phi_y.\phi_y^\top$ with k_x and k_y becomming the associated kernel functions to each space.
Recall: Letting $C_{xx} = X^\top.X, C_{yy} = Y^\top.Y, C_{xy} = X^\top.Y$
The CCA optimization problem is:

$$\rho = \max_{w_x, w_y} Corr(Xw_x, Yw_y) \tag{12}$$

subject to:

$$w_x^\top.C_{xx}w_x = w_y^\top.C_{yy}w_y = 1$$

The CCA optimization problem in (12) now becomes:

$$\max_{w_x, w_y} Corr(\phi_x w_x, \phi_y w_y) = \max_{\alpha_x, \alpha_y} Corr(K_x\alpha_x, K_y\alpha_y) \tag{13}$$

subject to:

$$w_x^\top.\phi_x^\top.\phi_x.w_x = \alpha_x^\top.K_x^2.\alpha_x = 1$$

and

$$w_y^\top.\phi_y^\top.\phi_y.w_y = \alpha_y^\top.K_y^2.\alpha_y = 1$$

so the covariance is:

$$\begin{aligned}\widehat{Cov}(<\phi_x, w_x>, <\phi_y, w_y>) &= \frac{1}{N}\sum_{k=1}^{N} <\phi_x^k, \sum_{i=1}^{N}\phi_x^i.\alpha_x^i><\phi_y^k, \sum_{j=1}^{N}\phi_y^j\alpha_y^j> \\ &= \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_x^i.K_x(x^i, x^k).K_y(y^j, y^k)\alpha_y^j \\ &= \frac{1}{N}\alpha_x^\top K_x K_y \alpha_y\end{aligned}$$

WhereK_x and k_y are the Gram-matrix associated with the data sets:$\{x^i\}$ and $\{y^i\}$. respectively, we obtain:

$$\widehat{Var}(<\phi_x, \sum \alpha_x^i \phi_x^i>) = \frac{1}{N}\alpha_x^\top K_x^2 \alpha_x$$

and:

$$\widehat{Var}(<\phi_y, \sum \alpha_y^i \phi_y^i>) = \frac{1}{N}\alpha_y^\top K_y^2 \alpha_y$$

Putting these results together, our Kernelized CCA problem becomes:

$$\widehat{\rho}(K_x, K_y) = \max_{\alpha_x, \alpha_y} \frac{\alpha_x^\top K_x K_y \alpha_y}{(\alpha_x^\top K_x^2 \alpha_x)^{1/2}(\alpha_y^\top K_y^2 \alpha_y)^{1/2}}$$

with the covariance-Matrix equal to: $\begin{bmatrix} K_x^2 & K_x K_y \\ K_x K_y & K_y^2 \end{bmatrix}$

Solution:
Now let write the lagrangian:

$$arg \max_{\alpha_x,\alpha_y}(\alpha_x^\top K_x K_y \alpha_y)$$

s.t:

$$\alpha_x^\top K_x^2 \alpha_x = 1 \quad and \quad \alpha_y^\top K_y^2 \alpha_y = 1$$

$$\mathcal{L} = \alpha_x^\top K_x K_y \alpha_y - \mu(\alpha_x^\top K_x^2 \alpha_x - 1) - \vartheta(\alpha_y^\top K_y^2 \alpha_y - 1)$$

Partial derivative:

$$\frac{\partial \mathcal{L}}{\partial \alpha_x^\top} = K_x K_y \alpha_y - \mu K_x^2 \alpha_x$$

and:

$$\frac{\partial \mathcal{L}}{\partial \alpha_y^\top} = K_y K_x \alpha_x - \vartheta K_y^2 \alpha_y$$

$$\begin{cases} K_x K_y \alpha_y = \mu K_x^2 \alpha_x \\ K_y K_x \alpha_x = \vartheta K_y^2 \alpha_y \end{cases} \tag{14}$$

multiply by $\alpha_x^\top$ and $\alpha_y^\top$ wich gives:

$$\alpha_x^\top K_x K_y \alpha_y = \mu$$

$$\alpha_x^\top K_x K_y \alpha_y = \vartheta$$

$\implies \mu = \vartheta$
Setting: $\rho = \mu = \vartheta$ Then we can write equation(14)as :

$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}. \tag{15}$$

□

Problem 2.b.

Solution: Linear KCCA is the same as CCA.We obtain $\rho_{cca} = \rho_{kcca}$ .the solution of the kernel CCA problem is given by the eigenfunctions corresponding to the largest eigenvalue. the kernelized CCA problem does not provide a useful estimate of the population canonical correlation, in general Subsequent Directions are found by imposing the additional constraints.
$(w_x^i)^\top C_{xx} w_x^i = (w_y^i)^\top C_{yy} w_y^i = (w_x^i)^\top C_{xy} w_y^j = 0$
for $i \neq j$ and $(w_1^i)^\top C_{11} w_1^i = (w_2^i)^\top C_{22} w_2^i = 1$
i=j=1,....,p, $p = min(d_1, d_2)$

□

Problem 3.a. Explain how the unconstrained objective above relates to the original CCA objective.

Solution:
Let's assume that our goal is only to find the:

$$\max_{\theta_x,\theta_y,w_x,w_y} w_x^T C_{xy} w_y.$$

There's no constraints, thus the ingredients of the objective would rise in an unbounded manner in order to maximize the value of the term above. Having that in mind, let's analyze:

$$\min(0, 1 - w_x^T C_{xx} w_x)$$

□

If the ingredients of the first equation would rise in an unbounded manner, the value of:

$$1 - w_x^T C_{xx} w_x$$

would fall in value in an unbounded manner resulting in the term $\min(0, 1-w_x^T C_{xx} w_x)$ being a negative number. The same logic applies to:

$$\min(0, 1 - w_y^T C_{yy} w_y).$$

The unconstrained objective stated in this problem is stated as:

$$\max_{\theta_x,\theta_y,w_x,w_y} w_x^T C_{xy} w_y + \alpha[\min(0, 1 - w_x^T C_{xx} w_x) + \min(0, 1 - w_y^T C_{yy} w_y)]$$

Addition of the sum of *min* functions enforces such optimization that these *min* values won't fall in value in a way explained above in order to find better solution to the problem. Ideally we want to have:

$$1 - w_y^T C_{yy} w_y = 1 - w_x^T C_{xx} w_x = 0$$

so we can maximize the $w_x^T C_{xy} w_y$ term and have no "penalty" for high values of the variables - which is how unconstrained objective relates to the original CCA objective.

Problem 3.b. Express the gradient of the new objective with respect to x as a function of the Jacobian matrix $\frac{\partial \phi_x}{\partial \theta_x}$.

Solution:
Let's name our objective function as:

$$f(w_x, w_y, \theta_x, \theta_y) = w_x^T C_{xy} w_y + \alpha[\min(0, 1 - w_x^T C_{xx} w_x) + \min(0, 1 - w_y^T C_{yy} w_y)]$$

The derivative of f in regards to θ_x is:

$$\frac{\partial f}{\partial \theta_x} f(w_x, w_y, \theta_x, \theta_y) = \frac{\partial}{\partial \theta_x} w_x^T C_{xy} w_y + \alpha \frac{\partial}{\partial \theta_x} \min(0, 1 - w_x^T C_{xx} w_x)$$

Writing down the derivatives in sum:

$$\frac{\partial}{\partial \theta_x} w_x^T C_{xy} w_y = \frac{1}{N} w_x^T \left[\frac{\partial \phi_x}{\partial \theta_x}\right] \phi_y w_y$$

$$\frac{\partial}{\partial \theta_x} \min(0, 1 - w_x^T C_{xx} w_x) = \begin{cases} 0 \ if \ 0 < \min(0, 1 - w_x^T C_{xx} w_x), \\ \frac{2}{N} w_x^T \left[\frac{\partial \phi_x}{\partial \theta_x}\right] w_x, \ else. \end{cases}$$

Finally resulting in:

$$\frac{\partial f}{\partial \theta_x} f(w_x, w_y, \theta_x, \theta_y) = \begin{cases} \frac{1}{N} w_x^T \left[\frac{\partial \phi_x}{\partial \theta_x}\right] \phi_y w_x, \ if \ 0 < \min(0, 1 - w_x^T C_{xx} w_x) \\ \frac{1}{N} w_x^T \left[\frac{\partial \phi_x}{\partial \theta_x}\right] \phi_y w_x + \alpha \frac{2}{N} w_x^T \left[\frac{\partial \phi_x}{\partial \theta_x}\right] w_x, \ else. \end{cases}$$

$\square$