We can use Lagrangian multiplier on CCA to find out
\[ C_x^{-1}C_{xy}C_y^{-1}C_{yx}w_x = \lambda^2 w_x \]
where \( C_x = XX^T, C_y = YY^T, C_{xy} = XY^T, C_{yx} = YX^T \)

Then we take the top eigenvector of \( C_x^{-1}C_{xy}C_y^{-1}C_{yx} \) to find the solution to CCA.

The trick is to analytically diagonalize \( C_x^{-1}C_{xy}C_y^{-1}C_{yx} \)

Define \( H = Y^T(YY^T)^{-1/2} \), also svd \( x = U \Sigma V^T = [U_1, U_2] \text{diag}(\Sigma_r, 0) [V_1, V_2]^T = U_1 \Sigma_r V_1^T \)

\[
C_x^{-1}C_{xy}C_y^{-1}C_{yx} = (XX^T)^{-1}XHHX^T
= (U_1 \Sigma_r^{-2} U_1^T) \cdot XHHX^T
= U_1 \Sigma_r^{-1} \cdot (\Sigma_r^{-1} U_1^T XH)H^T X^T
\]

Define \( A = \Sigma_r^{-1} U_1^T XH \), also svd \( A = P \Sigma_A Q^T \)

\[
C_x^{-1}C_{xy}C_y^{-1}C_{yx} = U_1 \Sigma_r^{-1} \cdot A \cdot H^T X^T UU^T
\]

Note that
\[
U \begin{bmatrix} \Sigma_r^{-1} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} P \Sigma_r & 0 \\ 0 & I \end{bmatrix} U^T = I
\]
which means \( C_x^{-1}C_{xy}C_y^{-1}C_{yx} \) is similar to
\[
\begin{pmatrix}
\Sigma_A^2 & 0 \\
0 & 0
\end{pmatrix}
\]
and \( U_1 \Sigma_r P \) are the top eigenvectors.

Now consider the linear regression problem \((X, \tilde{Y})\) where \( \tilde{Y} = H^T = (YY^T)^{-1/2}Y \)

The solution is the well-known
\[
(XX^T)^{-1}XY^T = (XX^T)^{-1}XH
= U_1 \Sigma_r^{-2}U_1^TXH
= U_1 \Sigma_r^{-1}(\Sigma_r^{-1}U_1^TXH)
= U_1 \Sigma_r^{-1}A
= U_1 \Sigma_r^{-1}P\Sigma_A Q^T
\]

It can be shown that \( \Sigma_A = I \).

Then
\[
(XX^T)^{-1}XY^T = (U_1 \Sigma_r^{-1}P)Q^T
\]

The main result: suppose we have a CCA problem \((X, Y)\) with a solution \( W_{CCA} \) and
a linear regression problem \((X, (YY^T)^{-1/2}Y)\) with a solution \( W_{LS} \), then
\( W_{LS} = W_{CCA}Q^T \).
That is, the two differ by an orthogonal matrix.