Notes on Canonical Correlation

Suppose we have a collection of random variables in a \((q + p) \times 1\) vector \(X\) that we partition in the following form (and supposing without loss of generality that \(p \leq q\):

\[
X = \begin{pmatrix}
X_1 \\
\vdots \\
X_p \\
\vdash\vdash\vdash \\
X_{p+1} \\
\vdots \\
X_{p+q}
\end{pmatrix} = \begin{pmatrix}
X_1 \\
\vdash\vdash\vdash \\
X_2
\end{pmatrix} \sim \text{MVN}(\mu, \Sigma),
\]

where

\[
\mu = \begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}; \quad \Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix},
\]

and remembering that \(\Sigma_{21} = \Sigma_{12}'\), and

\[
\text{Cor}(a'X_1, b'X_2) = a'\Sigma_{12}b/\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}.
\]

Suppose

\[
\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'a = \lambda a,
\]

with roots \(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0\), and corresponding eigenvectors \(a_1, \ldots, a_p\). Also, let

\[
\Sigma_{22}^{-1}\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}b = \lambda b,
\]
with roots $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $\lambda_{p+1} = \lambda_q = 0$; the corresponding eigenvectors are $b_1, \ldots, b_p$.

Looking at the two linear combinations, $a'_i X_1$ (called the $i^{th}$ canonical variate in the first set), and $b'_i X_2$ (called the $i^{th}$ canonical variate in the second set), the squared correlation between them is $\lambda_i$; the $i^{th}$ canonical correlation is $\sqrt{\lambda_i}$. The maximum correlation between any two linear combinations is $\sqrt{\lambda_1}$, and is obtained for $a_1$ and $b_1$. For $a_i$ and $b_i$, these are uncorrelated with every canonical variate up to that point, and maximize the correlation subject to that restriction.

Points to make:

a) The matrices $\Sigma^{-1/2}_{11} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}'$ and $\Sigma^{-1/2}_{22} \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12}'$ are not symmetric and so the standard eigenvector/eigenvalue decompositions are not straightforward. However, the two matrices

$$\Sigma^{-1/2}_{11} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}' \Sigma_{11}^{-1/2}$$

and

$$\Sigma^{-1/2}_{22} \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2}$$

are symmetric. Also,

$$\Sigma^{-1/2}_{11} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}' \Sigma_{11}^{-1/2} e_i = \lambda_i e_i ,$$

and

$$\Sigma^{-1/2}_{22} \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} f_i = \lambda_i f_i ,$$

where the roots, i.e., the $\lambda_i$s, are the same as before. We can then obtain $a_i = \Sigma^{-1/2}_{11} e_i$, and $b_i = \Sigma^{-1/2}_{22} f_i$. Both $\Sigma^{-1/2}_{11}$ and $\Sigma^{-1/2}_{22}$ are constructed from the spectral decompositions of $\Sigma_{11} = PDP'$ and $\Sigma_{22} = QFQ'$ as $\Sigma^{-1/2}_{11} = PD^{-1/2}P'$ and $\Sigma^{-1/2}_{22} = QF^{-1/2}Q'$. Note
the normalizations of Var($a'_i X_1$) = $a'_i \Sigma_{11} a'_i = e'_i \Sigma_{11}^{-1/2} \Sigma_{11} \Sigma_{11}^{-1/2} e_i = 1$ and Var($b'_i X_2$) = 1.

b) There are three different normalizations that are commonly used for $a_i$ and $b_i$:

(i) leave as unit length so $a'_i a_i = b'_i b_i = 1$;

(ii) make the largest value 1.0 in both $a_i$ and $b_i$;

(iii) do as we did above and make $a'_i \Sigma_{11} a'_i = 1 = b'_i \Sigma_{22} b'_i$.

(c) Special cases: When $p = 1$ and $q = 1$, $\lambda_1$ is the (simple) squared correlation between two variables; when $p = 1$ and $q > 1$, $\lambda_1$ is a squared multiple correlation. In considering $a'_i X_1$ versus $X_2$, $\lambda_i$ is the squared multiple correlation of $a'_i X_1$ with $X_2$; $b_i$ gives the regression weights.

(d) When moving to the sample, all items have direct analogues. The one restriction on sample size is $n \geq p + q + 1$.

(e) Suppose the variables $X_1$ and $X_2$ are transformed by nonsingular matrices, $A_{p \times p}$ and $B_{q \times q}$, as follows:

$$Y_1 = A_{p \times p} X_1 + c_{p \times 1}$$

$$Y_2 = B_{q \times q} X_2 + d_{q \times 1}$$

The same canonical variates and correlations using $Y_1$ and $Y_2$ would be generated as from $X_1$ and $X_2$; the weights in $a_i$ and $b_i$ would be on the transformed variables, obviously. In particular, we could work with standardized variables without loss of any generality, and just use the correlation matrix.
(f) To evaluate $H_0 : \Sigma_{12} = \mathbf{0}$, a likelihood ratio test is available:

$$-(n - 1 - (1/2)(p + q + 1)) \ln \prod_{i=1}^{p} (1 - \lambda_i) \sim \chi_{pq}^2.$$  

Also, sometimes a sequential process is used to test the remaining roots until nonsignificance is reached:

$$-(n - 1 - (1/2)(p + q + 1)) \ln \prod_{i=k+1}^{p} (1 - \lambda_i) \sim \chi_{(p-k)(q-k)}^2.$$  

This latter sequential procedure is a little problematic because there is no real control over the overall significance level with this strategy.

Generally, there is some tortuous difficulty in interpreting the canonical weights substantively. I might suggest using a constrained least-squares approach (iteratively moving from one set to a second), where the weights are forced to be nonnegative.