Canonical analysis

Pierre Legendre
Département de sciences biologiques
Université de Montréal

© Pierre Legendre 2020
Outline of the presentation

1. Canonical analysis: definition
2. Redundancy analysis (RDA): basic principle
   2.1. Multivariate regression
   2.2. Statistics in simple RDA
   2.3. The PCA step
   2.4. Partial RDA
   2.5. Applications of partial RDA
3. Canonical correspondence analysis (CCA)
4. Other methods of canonical analysis
   4.1. Linear discriminant analysis (LDA)
   4.2. Canonical correlation analysis (CCorA)
   4.3. Co-inertia (CoIA) and Procrustes (Proc) analyses
   4.4. Multiple factor analysis (MFA)
5. Selection of explanatory variables in RDA and CCA
6. Multivariate variation partitioning
7. Multivariate anova by RDA
8. R software
9. References
1. Canonical analysis: definition

Canonical analysis is the simultaneous analysis of two, or possibly several data tables.

Methods of canonical analysis allow ecologists to perform direct comparisons of two data matrices (also called “direct gradient analysis”) about the same objects (rows).

In **symmetric methods** of canonical analysis, the two (or more) matrices play the same role in the analysis. Their order can be inverted in the calculation. Methods: CCorA, CoIA, Proc, MFA.

In **asymmetric methods**, the two matrices do not play the same role. As in linear regression, there is a response matrix $Y$ and an explanatory matrix $X$. Methods: RDA, CCA, LDA.
Canonical analysis –

An expression coined by C. R. Rao when he discovered how to solve the problem of multiple discriminant analysis (1948). He called the new method *Canonical Variate Analysis*.

This expressions refers to the canonical form of a matrix. The canonical form is the simplest and most comprehensive form to which certain functions, relations, or expressions can be reduced without loss of generality.

For example, the canonical form of a covariance matrix is the matrix of eigenvalues computed from it.

The methods of canonical analysis described in this presentation use eigen-analysis or SVD for decomposition.
2. Redundancy analysis (RDA): basic principle

RDA is an asymmetric method of canonical analysis. The analysis involves a response matrix $Y$ containing quantitative variables and an explanatory matrix $X$.

In RDA, one may be interested, for example, in the relationship between a matrix $Y$ describing species composition and a matrix $X$ containing environmental descriptors, observed at the same locations.

In partial RDA, a matrix of covariables $W$ will also be involved.
Example (artificial data, 6 species). Legendre & Legendre 2012, Table 11.3.

Artificial data set representing observations (fish abundances) at 10 sites along a tropical reef transect. The variables are further described in the text.

<table>
<thead>
<tr>
<th>Site No.</th>
<th>Sp. 1</th>
<th>Sp. 2</th>
<th>Sp. 3</th>
<th>Sp. 4</th>
<th>Sp. 5</th>
<th>Sp. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>5</td>
<td>17</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>7</td>
<td>13</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Sum</td>
<td>60</td>
<td>50</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Depth (m)</th>
<th>Coral</th>
<th>Sand</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note – The rank of explanatory matrix $X$ (right) is 3 in this example.
Relationships between
(a) ordination,
(b) regression, and
(c) two asymmetric forms of canonical analysis.

(a) Simple ordination of matrix $Y$:
- principal comp. analysis (PCA)
- correspondence analysis (CA)

(b) Ordination of $y$ (single axis) under constraint of $X$:
- multiple regression

Model: $\hat{y} = b_0 + b_1x_1 + \ldots + b_mx_m$

(c) Ordination of $Y$ under constraint of $X$:
- redundancy analysis (RDA)
- canonical correspondence analysis (CCA)
2.1. Multivariate regression

Multiple linear regression involves a vector of response values $y$ and a matrix of explanatory variables $X$.

The regression coefficients (vector $b$) are found by solving the eq.

$$ b = [X'X]^{-1} X'y $$

The vector of fitted values is found by solving: $\hat{y} = X \left[ X'X \right]^{-1} X'y$

Likewise, in RDA, the matrix of fitted values is found by solving the multivariate linear regression equation:

$$ \hat{Y} = X[X'X]^{-1}X'Y $$

where $Y = \text{multivariate response matrix}$

$X = \text{multivariate matrix of explanatory variables}$
First step: multivariate regression of $Y$ on $X$ =>

$$\hat{Y} = X[X'X]^{-1}X'Y$$

- Save the fitted values $\hat{Y}$
- Save the residuals $Y_{res} = Y - \hat{Y}$

Regress each variable $y$ on table $X$ and compute the fitted ($\hat{y}$) and residual ($y_{res}$) values
2.2. Statistics in simple RDA

\[ R^2_{Y|x} = \frac{SS(\hat{Y})}{SS(Y)} \]

\[ R_a^2 = 1 - (1 - R^2_{Y|x}) \frac{(n - 1)}{(n - m - 1)} \]

\[ AIC = n \log_e \left( \frac{(1 - R^2)}{n} \right) + 2k \]

\[ AIC_c = AIC + \frac{2k(k + 1)}{n - k - 1} \]

\[ F = \frac{R^2_{Y|x}/m}{(1 - R^2_{Y|x}) / (n - m - 1)} \]

Permutation test of \( F \) → \( p \)-value

Test of significance of the individual axes

where \( n \) = number of sites, \( m \) = rank of \( X_{\text{cent}} \)
$R^2$ or $R^2_{adj}$: which one is preferable?

Simulations by Peres-Neto et al. (2006)

**Up:** 2 variables, $n = 100$, true $R^2 = 0.61$

Add $\{0, 1, 2, \ldots, 20\}$ explanatory variables consisting of random normal deviates. Compute $R^2$ (▼) and $R^2_{adj}$ (▲).

Repeat with 1000 independent datasets. Compute the mean of the $R^2$ and $R^2_{adj}$.

- $R^2$ (▼) increases as columns of random numbers are added to the $X$ set. Using $m = (n - 1)$ random variables in $X$ produces $R^2 = 1$.

- $R^2_{adj}$ (▲) remains constant.

**Down:** same with true $R^2 = 0.20$

Conclusion: $R^2$ is biased, $R^2_{adj}$ is not.
Example (artificial data) – Compute the statistics

Total variance in the response data, $\text{Var}(Y) = 112.8889$

$R^2 = 0.95971$

$R^2_{adj} = 0.93958$

$\text{AIC} = 24.0936$

$\text{AIC}_c = 39.0936$

$F = 47.64169$

$\Rightarrow$ p-value of the $F$-test of $R^2$ after 999 permutations: 0.001

Note: the $R^2_{adj}$ statistic can be negative. This indicates that the proportion of variance of $Y$ explained by $X$ is worse than a set of $m$ random normal deviates would do.
2.3. The PCA step

If there is a significant $Y \sim X$ relationship,
compute a PCA of the matrix $\hat{Y}$ of fitted values to reduce the
dimensionality of the solution.

The number of canonical axes obtained is $\leq \min[p, m, n – 1]$. In
other words, it cannot be larger than

- $p =$ number of response variables
- $m =$ number of explanatory variables (actually, the rank of $X$)
- $(n – 1) =$ max. dimension of unconstrained ordination of $n$ objects.

Scalings 1 and 2 are useful in RDA, as they are in PCA –

- Scaling type 1 preserves the Euclidean distances among the
  sites in matrix $\hat{Y}$.
- Scaling type 2 preserves the covariances among the species.
Two ways of representing the objects in the ordination plot:

- \( Z = \hat{Y}U \)
- \( F = YU \)

Regress each variable \( y \) on table \( X \) and compute the fitted (\( \hat{y} \)) and residual (\( y_{res} \)) values.

\[ \hat{Y} = X [X'X]^{-1} X'Y \]

\( U = \) matrix of eigenvectors (canonical)

\( Z = \hat{Y}U \) = ordination in the space of variables \( X \)

\( F = YU \) = ordination in the space of variables \( Y \)
Test the significance of the individual canonical axes.

How many axes explain significantly more variation than random data would?

Example data:
axis 1, $p = 0.005$
axis 2, $p = 0.001$
axis 3, $p = 0.003$

Regress each variable $y$ on table $X$ and compute the fitted ($\hat{y}$) and residual ($y_{res}$) values.
Represent the explanatory variables in triplots –

Compute the correlations of the explanatory variables $X$ in the canonical ordination space:

$$\text{cor}(X, Z)$$

These correlations are used to represent the positions of the explanatory variables in RDA triplots =>

- in scaling 1: modify the correlations; multiply each correlation in column [axis] $k$ by

$$\sqrt{\lambda_k} / \text{Total variance in } Y$$

- in scaling 2: use the correlations as they are to represent the explanatory variables in the triplot.
RDA of the example data – Total variance in $Y = 112.889$

<table>
<thead>
<tr>
<th>Canonical axes</th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalues</td>
<td>74.52267</td>
<td>24.94196</td>
<td>8.87611</td>
</tr>
<tr>
<td>Rel. eigenvalues</td>
<td>0.66014</td>
<td>0.22094</td>
<td>0.07863</td>
</tr>
<tr>
<td>Cumul. rel. eig.</td>
<td>0.66014</td>
<td>0.88108</td>
<td>0.95971</td>
</tr>
</tbody>
</table>

Correlations of environmental var. with the $Z$ site scores (for scaling 2)

<table>
<thead>
<tr>
<th></th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>0.42265</td>
<td>–0.55914</td>
<td>–0.71325</td>
</tr>
<tr>
<td>Coral</td>
<td>0.98850</td>
<td>0.15079</td>
<td>–0.01178</td>
</tr>
<tr>
<td>Sand</td>
<td>–0.55652</td>
<td>0.81760</td>
<td>0.14771</td>
</tr>
<tr>
<td>Other subs.</td>
<td>–0.40408</td>
<td>–0.90584</td>
<td>–0.12715</td>
</tr>
</tbody>
</table>

Biplot scores of the environmental variables (for scaling 1)

<table>
<thead>
<tr>
<th></th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>0.34340</td>
<td>–0.26282</td>
<td>–0.20000</td>
</tr>
<tr>
<td>Coral</td>
<td>0.80314</td>
<td>0.07088</td>
<td>–0.00330</td>
</tr>
<tr>
<td>Sand</td>
<td>–0.45216</td>
<td>0.38431</td>
<td>0.04142</td>
</tr>
<tr>
<td>Other subs.</td>
<td>–0.32831</td>
<td>–0.42579</td>
<td>–0.03565</td>
</tr>
</tbody>
</table>
Scaling 1 triplot for the example data.
Another way of representing qualitative variables (factors) in triplots:

For each class of the factor, compute the centroid of the sites that have that class.

Example for factor “Substrate type” –

<table>
<thead>
<tr>
<th>Canonical axes</th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalues</td>
<td>74.52267</td>
<td>24.94196</td>
<td>8.87611</td>
</tr>
<tr>
<td>Rel. eigenvalues</td>
<td>0.66014</td>
<td>0.22094</td>
<td>0.07863</td>
</tr>
</tbody>
</table>

Centroids of the site coord. (in \( \mathbf{Z} \)) with code 1 for each class of the factor

- **Coral**: 12.36599, 1.09129, –0.05088
- **Sand**: –6.96197, 5.91719, 0.63774
- **Other subs.**: –4.05301, –5.25636, –0.44014

Canonical analysis
Scaling 1 triplot for the example data.
Summary of the RDA method

Regress each variable $y$ on table $X$ and compute the fitted ($\hat{y}$) and residual ($y_{res}$) values.

Data table $Y$ (centred variables)

Data table $X$ (centred var.)

Fitted values from the multiple regressions

$\hat{Y} = X [X'X]^{-1} X'Y$

Explanatory var.

Response variables

$F = YU =$ ordination in the space of variables $Y$

$F = YU =$ ordination in the space of variables $X$

Residual values from the multiple regressions

$Y_{res} = Y - \hat{Y}$

$Z = \hat{Y}U =$ ordination in the space of residuals

$Y_{res}U_{res} =$ ordination in the space of residuals

From Legendre & Legendre (2012), Fig. 11.2.
An RDA of the Doubs River fish data is presented in Borcard et al. (2018), Section 6.3.2. All steps of the analysis, which uses vegan’s rda() function, are described in detail in the script.
2.4. Partial RDA

Definition and objective of partial RDA

Partial RDA is the analysis of response matrix $Y$ by explanatory matrix $X$ in the presence of additional explanatory variables, $W$, called covariables.

In partial RDA, the linear effects of the explanatory matrix $X$ on the response matrix $Y$ are adjusted for the effects of the covariables $W$, as in partial linear regression.
Two calculation methods for partial RDA ($Y \sim X \mid W$)

[The two methods are the same as in partial regression]

0. Centre the $Y$, $X$ and $W$ matrices to column means of 0

1. Compute residuals of $Y$ on $W$: $Y_{\text{res}|W} = Y - W[W'W]^{-1} W' Y$

   Compute residuals of $X$ on $W$: $X_{\text{res}|W} = X - W[W'W]^{-1} W' X$

2a. RDA of $Y_{\text{res}|W}$ on $X_{\text{res}|W}$ produces the partial $R^2$.

2b. RDA of $Y$ on $X_{\text{res}|W}$ produces the semipartial $R^2$.

The other type of $R^2$ can be derived in each case.

$\Rightarrow$ The matrices of fitted values ($Y_{\text{fit } X\mid W}$) produced by these two multivariate regressions are the same.
Partial RDA (\(Y \sim X \mid W\))

The amount of variance associated with \(Y_{\text{fit} X \mid W}\) is fraction [a] in the Venn diagram:

Total variation in response matrix \(Y\) = [a] + [b] + [c].

Unexplained variation (residual variation) = [d]

From Legendre & Legendre (2012), Fig. 10.10.
Statistics in partial RDA ($Y \sim X \mid W$) for test of $[a]$

Permutation test of $F$-stat:
Three methods –
- Permute raw data
- Permute residuals of reduced model
- Permute residuals of full model

Test of significance of the individual axes

$p$-value

$SS(Y) = [a+b+c+d]$
$SS(Y_{fit}) = [a]$
$SS(Y_{res}) = [d]$
$SS(Y_{res|W}) = [a+d]$

$F = \frac{SS(Y_{fit})/m}{SS(Y_{res})/(n-m-q-1)}$

$R^2_{Y_{res|W}|X_{res|W}} = \frac{SS(Y_{fit})}{SS(Y_{res|W})}$

**Partial $R^2 = [a]/[a+d]$**

$R^2_{Y|X_{res|W}} = \frac{SS(Y_{fit})}{SS(Y)}$

**Semipartial $R^2 = [a]/[a+b+c+d]$**

$R^2_{adj}$ is obtained from the variation partitioning table

where $n =$ number of sites, $m =$ rank $X_{cent}$ and $q =$ rank of $W_{cent}$
2.5. Applications of partial RDA

Examples –

1. Control for well-known linear effects. — Measure (by semipartial $R^2$) the effect of $X$ on $Y$ while controlling for the linear influence of covariable matrix $W$. Test significance of that effect.

2. Isolate the effect of a single explanatory variable or factor. — Measure and test the unique contribution to $Y$ of a single explanatory variable by placing it in $X$; put all other explanatory variables in $W$. Repeat with the other explanatory variables, one at a time.
Examples of application of partial RDA (continued) –

3. Analysis of related samples. — X contains the explanatory variable of interest, W contains a factor indicating the class of the pairing variable used as covariable.

<table>
<thead>
<tr>
<th>Ponds</th>
<th>Response data Y</th>
<th>Variable or factor X</th>
<th>Covariable W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phytoplankton community composition</td>
<td>before before before after after after after</td>
<td>#1 #2 #3 #4 #1 #2 #3 #4</td>
<td></td>
</tr>
</tbody>
</table>

Example =>

Time: before/after some treatment, e.g. fertilization

Ponds to which treatments 1 to 4 were applied
Examples of application of partial RDA \textit{(continued)} –

4. Selection of explanatory variables in RDA — Methods: forward, backward, stepwise. See Section 5 of this presentation.

5. Multivariate anova (Manova) by RDA — See Section 7, “Manova by RDA”, and the presentation on Space-Time Interaction (STI).
2.6. Summary: RDA for community data

*Figure next slide*
Comparison of (a) classical RDA and CCA, and (b and c) alternative approaches forcing RDA to preserve other distances adapted to species composition data. L&L 2012, Fig. 11.4.
tb-RDA (transformation-based RDA, Legendre & Gallagher 2001): transform the species data with vegan’s decostand(), then use the transformed data matrix as input matrix $Y$ in RDA.

db-RDA (distance-based RDA, Legendre & Anderson 1999): compute PCoA from a pre-computed dissimilarity matrix $D$, then use the principal coordinates as input matrix $Y$ in RDA.

db-RDA can also be computed directly by function dbrda() in vegan. For a $D$ matrix with Euclidean property, the result is the same as if PCoA had been computed, followed by regular RDA.

Function dbrda() can directly handle non-Euclidean $D$ matrices, but beware of the results if the matrix is strongly non-Euclidean and make certain this is what you want.
3. Canonical correspondence analysis (CCA)

CCA is –

• a weighted form of RDA

• applied to the same matrix $\mathbf{Q}$ of contributions to the $\chi^2$ statistic as the one used in CA.

• Matrix $\overline{\mathbf{Q}}$ is computed from frequency or presence-absence data, as in correspondence analysis (CA).
In more detail –

CCA is RDA with two modifications:

- The species data are transformed into matrix $\bar{Q}$ of contributions to chi-square.

The following slides are from the teaching presentation on correspondence analysis (CA) –
Computation of the Q-bar matrix

Example: community composition data

\[ \mathbf{Y} = [f_{ij}] = \begin{bmatrix}
S_{p1} & S_{p2} & S_{p3} & S_{p4} & S_{p5} \\
Site1 & 45 & 10 & 15 & 0 & 10 \\
Site2 & 25 & 8 & 10 & 0 & 3 \\
Site3 & 7 & 15 & 20 & 14 & 12 \\
\end{bmatrix} \begin{bmatrix}
f_{i+} \\
80 \\
46 \\
68 \\
\end{bmatrix} \]

\[ [f_{+j}] = \begin{bmatrix}
77 & 33 & 45 & 14 & 25 \\
\end{bmatrix} f_{++} = 194 \]

Canonical analysis
Compute the matrix of contributions to chi-square

Matrix $\bar{Q} = \left[ q_{ij} \right] = \left[ \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right] = \frac{(O_{ij} - E_{ij})}{\sqrt{E_{ij}}} \sqrt{f_{++}}$

where $p_{ij} = f_{ij} / f_{++}$
$p_{i+} = f_{i+} / f_{++}$
$p_{+j} = f_{+j} / f_{++}$

The statistics $(O_{ij} - E_{ij})/\sqrt{E_{ij}}$ in the numerator are called “components of chi-square” in contingency table analysis. They are the square roots of the statistics that are summed to produce the Pearson chi-square statistic. Reference: Legendre & Legendre (2012, eq. 6.26).
Compute the matrix of contributions to chi-square

where \( p_{ij} = \frac{f_{ij}}{f_{++}} \)
\( p_{i+} = \frac{f_{i+}}{f_{++}} \)
\( p_{+j} = \frac{f_{+j}}{f_{++}} \)

Matrix \( \bar{Q} = [q_{ij}] = \left[ \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right] = \frac{(O_{ij} - E_{ij})}{\sqrt{E_{ij}}} \)

\[
\bar{Q} = \begin{bmatrix}
\text{Site1} & & & & \\
0.169 & -0.070 & -0.059 & -0.173 & -0.007 \\
\text{Site2} & & & & \\
0.113 & 0.004 & -0.015 & -0.131 & -0.086 \\
\text{Site3} & & & & \\
-0.276 & 0.072 & 0.076 & 0.295 & 0.079 \\
\end{bmatrix}
\]

Total inertia: \( \sum (q_{ij}^2) = 0.285 = \sum \text{of the CA eigenvalues} \)
In more detail –

CCA is RDA with two modifications:

• the species data are transformed into matrix Q-bar contributions to chi-square

• a matrix of weights is included in the multivariate regression calculation, where the weights \((p_{i+})\) are the row sums \((f_{i+})\) of \(Y\) divided by the sum total of the matrix \((f_{++})\).

\[
\begin{bmatrix}
0.41237 & 0 & 0 \\
0 & 0.23711 & 0 \\
0 & 0 & 0.35052 \\
\end{bmatrix}
\]

For the example, \(D(p_{i+})\) =
The equation to compute matrix $B$ of the regression coefficients in RDA was:

$$B = [X'X]^{-1} X'Y$$

In CCA, incorporate the matrix of row weights into that equation which becomes:

$$B = [X_{stand}' D(p_{i+}) X_{stand}]^{-1} X_{stand}' D(p_{i+})^{1/2} \bar{Q}$$

where matrix $X$ is standardized to $X_{stand}$ using weights $D(f_{i+})$.

In RDA, the matrix of fitted values was found by solving the multivariate linear regression equation:

$$\hat{Y} = X B$$

In CCA, that equation is:

$$\hat{Y} = D(p_{i+})^{1/2} X_{stand} B$$

See Legendre & Legendre (2012, Section 11.2.1) for demonstration and details.
CCA is the eigen-decomposition of $S_{\hat{Y}^\prime \hat{Y}} = \hat{Y}^\prime \hat{Y}$ by eigen() or svd(). One obtains matrices $\Lambda$ of eigenvalues and $U$ of eigenvectors.

Canonical correspondence analysis is thus a weighted form of redundancy analysis, applied to response matrix $\bar{Q}$.

Next, one obtains matrix $\hat{U}$, as in correspondence analysis, and the matrices necessary to plot the species and sites in scalings 1 and 2, as well as the triplot scores for the explanatory (environmental) variables in these two scalings. See L&L (2012, Section 11.2.1).
Example (artificial data, 9 species). Legendre & Legendre 2012, Table 11.3.

Artificial data set representing observations (fish abundances) at 10 sites along a tropical reef transect. The variables are further described in the text.

<table>
<thead>
<tr>
<th>Site No.</th>
<th>Sp. 1</th>
<th>Sp. 2</th>
<th>Sp. 3</th>
<th>Sp. 4</th>
<th>Sp. 5</th>
<th>Sp. 6</th>
<th>Sp. 7</th>
<th>Sp. 8</th>
<th>Sp. 9</th>
<th>Depth (m)</th>
<th>Substrate type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>5</td>
<td>17</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>7</td>
<td>13</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Sum</td>
<td>60</td>
<td>50</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>45</td>
<td>35</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CCA triplots for the example data, scaling 1. `cca()` and `plot.cca()` functions of the vegan package.

Substrate types as dummy variables. As arrows, the last one not printed.

Substrate types declared as a factor. Types printed at centroids of the sites.
CCA triplots for the example data, scalings 1 and 2. CCA() and triplot.CCA() functions on Web page http://adn.biol.umontreal.ca/~numericalecology/Rcode/.
4. Other methods of canonical analysis

Asymmetric method –
Linear discriminant analysis (LDA) analyses a qualitative variable $y$ (factor or classification) with a matrix $X$ of explanatory variables.

Symmetric methods of analysis –
Canonical correlation analysis (CCorA): analyse the correlations between 2 matrices $Y$ about the same objects (rows). Inappropriate for community composition data.
Coinertia analysis (CoIA), Procrustes analysis (Proc): analyse the covariances between 2 matrices $Y$.
Multiple factor analysis (MFA): joint PCA of several matrices $Y$. 
4.1. Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is an asymmetric method, like RDA and CCA. It analyses a qualitative variable $y$ (factor or classification) using a matrix $X$ of explanatory variables.

The problem is to find a linear combination of variables that separates two or more classes of objects. The discriminant functions can be used as linear classifiers that separate the groups of observations from one another.
Example of a discriminant function for two groups, A and B.

The two groups are overlapping along both axes $x_1$ and $x_2$.

The two groups are perfectly separated along discriminant axis $z$.

From Legendre & Legendre (2012), Fig. 11.11.
R function to compute discriminant analysis:

function lda() in {MASS}

Detailed explanations and examples:

• “Numerical ecology” (2012), Section 11.3
• “Numerical ecology with R” (2018), Section 6.5.
4.2. Canonical correlation analysis (CCorA)

The first symmetric method of canonical analysis. It analyses the correlation matrix between two matrices, $Y_1$ and $Y_2$.

Method inappropriate for community composition data because the algebra of CCorA standardizes the variables, and this is inappropriate in most cases for species data.

Furthermore, CCorA can only handle matrices with fewer variables than $(n - 1)$ objects, i.e. with $p_1$ and $p_2$ smaller than $(n - 1)$.

For community composition, genetic and molecular data, $p >>> (n - 1)$ in most cases. — CCorA cannot compute results with this kind of data.

OK to compare two matrices of physical variables: the variables in each set are standardized in CCorA, so they are not required to have, originally, the same physical dimensions within each set.
4.3. Co-inertia (CoIA) and Procrustes (Proc)

CoIA and Proc are two symmetric methods that analyse the same covariance matrix between $Y_1$ and $Y_2$, centred to column means of 0. The two matrices must be about the same objects (rows).

\[
\text{Cov}_{12} = \frac{1}{n - 1} Y_{1,\text{cent}}' Y_{2,\text{cent}}
\]

The CoIA and Proc methods differ in the way they present the joint ordination of the data.

These methods are appropriate for community composition data. They can handle matrices with more variables than sites.

All variables in each set must have the same physical dimensions. Variables may differ in nature between the two sets.

Physical variables that vary in dimensions within a set must be standardized.
R code: example of co-inertia analysis (CoIA)

```r
# Two matrices of environmental data of the Doubs River.
# Example from Section 6.9.2 of Numerical ecology with R (2018),
# with small corrections to the data.
# The corrected data are available in the course Practicals folder.

# Files envchem2 (water chem, 7 var), envtopo2 (physiography, 3 var)

# In ade4, the data sets must first be processed by a “dudi.xxx”
# ordination function. Here: dudi.pca(), which does a PCA.
# The data within each set are standardized by the dudi.pca() calls:
dudi.chem <- dudi.pca(envchem2, scale = TRUE, scan=FALSE)
dudi.topo <- dudi.pca(envtopo2, scale = TRUE, scan=FALSE)

# Co-inertia analysis, using function coinertia() of ade4 –
# The function recuperates the standardized data from the dudi.xxx
# output files. The PCA results are not used in co-inertia analysis.
coia.chem.topo <- coinertia(dudi.chem, dudi.topo, scan=FALSE, nf=3)
summary(coia.chem.topo)

# Plot the results
plot(coia.chem.topo)
```
Co-inertia analysis results, chemical and physiographic variables, Doubs River. In the graphs, X refers to water chemistry, Y to physiographic data.

**Upper plot** (normed site scores): position of the sites on the co-inertia axes using the chemistry (origins of the arrows) and physiography (arrowheads). Short arrows: sites represented in the same way in X and Y.

**Lower plots**: contributions of the groups of variables to the canonical space. Vectors pointing in the same direction are correlated; longer vectors contribute more to the structure.
Co-inertia analysis: the $RV$ coefficient

Test of significance of the correlation between two data matrices: the $RV$ coefficient

$$RV = \frac{\text{inertia}(\text{Cov}_{12})}{\sqrt{\text{inertia}(\text{Cov}_{11}) \times \text{inertia}(\text{Cov}_{22})}}$$

where

$$\text{inertia}(\text{Cov}_{12}) = \sum (\text{Cov}^{(2)}_{12})$$

is the sum of the squared values in the covariance matrix between $Y_1$ and $Y_2$ — Same for $\text{Cov}_{11} = \text{cov}(Y_1)$ and $\text{Cov}_{22} = \text{cov}(Y_2)$.

This statistic is a generalized form of $R^2$ for the comparison of two multivariate quantitative data matrices.

$RV$ coefficient for the example (previous slides):

$$RV = 0.5538, \text{ p-value (999 perm)} = 0.001$$

$^1$The abbreviation $RV$ stands for “Correlation [R] of (multivariate) Vectors” (Escoufier 1973).
**RV coefficient —**

For two simple variables $y_1$ and $y_2$, $RV$ is the square of their Pearson correlation coefficient.

R functions implementing tests of the $RV$ coefficient:

- permutational test: in RV.rtest() of ade4,
- parametric test: in coeffRV() of FactoMineR.
Procrustes analysis (Proc)

In Procrustes analysis, each data matrix is first standardized to have a total variance of 1 (Gower standardization, 1971).

\[ \text{Cov}_{12} = \frac{1}{n-1} Y'_{1,Gower} Y_{2,Gower} \]

Then, the decomposition of the \( \text{Cov}_{12} \) matrix is the same as in CoIA. Main difference with CoIA: presentation of the ordination graphs. For details, see Legendre & Legendre (2012), Section 11.5.2.

R functions for Procrustes analysis —

• function procrustes() in vegan,
• function procuste() in ade4 (French spelling for Προκρούστης)

Permutation test of the Procrustes statistic to test the significance of the correlation between two data matrices:

• function protest() in vegan.
Procrustes, son of Poseidon, was a rogue who physically attacked people by stretching them on an iron bed or cutting off their legs, so as to force them to fit the size of the bed.

The Greek hero **Theseus** killing the villain **Procrustes** on his ill-famed iron bed. Greek mythology. Picture available on: https://fr.wikipedia.org/wiki/Procuste
4.4. Multiple factor analysis (MFA)

MFA performs a symmetric analysis of several data matrices, \( Y_1 \) to \( Y_k \). Usually in MFA, \( k > 2 \).

This is a correlative form of analysis. It does not involve any hypothesis of causal influence of a data matrix on another matrix.

The variables must have the same mathematical type (quantitative or qualitative) within each matrix.

If all variables are quantitative, MFA is basically a PCA applied to the whole set of variables, in which each subset is weighted by the amount of variance it contributes to the MFA.

For details and example, see Borcard et al. (2018), Section 6.10.
Different graphical output files can be produced —

- The axes of the PCA of each data matrix are projected on the common MFA ordination plane.
- The variables from all subsets can be projected together on the ordination plane.
- A graph can be produced with the site scores of the separate PCAs projected in the MFA ordination, plus the centroids of these scores.

The pairwise similarity between the geometric representations derived from the $k$ groups of variables is measured by $RV$ coefficients; see previous section on co-inertia analysis.

R functions for MFA —

- function mfa() of ade4
- function MFA() of FactoMineR
  (easier to use than mfa() in ade4; offers more options).
5. Selection of explanatory variables

Selection of environmental variables in RDA and CCA.

Methods:

• Forward selection
• Backward elimination
• Stepwise selection

R functions:

• ordistep() and ordiR2step() in vegan, for RDA and CCA,
• forward.sel() in adespatial, for multiple regression and RDA.
Canonical variation partitioning allows users to partition the variation of a multivariate response data matrix $Y$ with respect to two or several explanatory matrices $X_1, X_2, X_3$, etc.

Variation partitioning also works for a single response vector $y$.

Function varpart() in vegan implements this method for 2, 3 or 4 explanatory matrices $X_1$ to $X_4$.

The significance of the unique fractions ([a] and [c] in the case of two explanatory matrices) can be tested using the test available in partial RDA.

The common fractions (e.g. [b]) are not adjusted components of variance. They cannot be estimated and tested by linear methods.
Variation partitioning results

Venn diagram illustrating the results of variation partitioning of the Doubs River fish assemblage data (29 sites) among three sets of explanatory variables: Topography, Chemistry and Geography.

The fractions of variation are identified by letters [a] to [h]. The value next to each identifier is the $R^2_{\text{adj}}$.

From Legendre & Legendre 2012, Fig. 11.6.

R functions used: varpart() and plot.varpart() in vegan.
7. Multivariate Anova by RDA

A look at **parametric multivariate analysis of variance (Manova)**

Conditions of application of parametric Manova:

- multivariate normality of data in each group,
- homogeneity among groups of the variance-covariance matrices,
- number of response variables ($p$) smaller than ($n – 1 – k$), where $k$ is the number of variables coding for the factors and their interaction(s).

The condition of **multivariate normality** is never met by community composition data

The third condition is often not met by community data, which may harbour large numbers of species.
For Manova, RDA offers an elegant alternative to parametric analysis,
while adding the versatility of permutation tests and the possibility
of representation of the results in triplots.
Manova for two crossed factors, balanced design –

The two crossed factors are represented by Helmert contrasts:

<table>
<thead>
<tr>
<th>2 groups: 1 variable</th>
<th>3 groups: 2 variables</th>
<th>4 groups: 3 variables</th>
<th>5 groups: 4 variables</th>
<th>etc.</th>
</tr>
</thead>
</table>

\[
\begin{bmatrix}
-1 \\
+1
\end{bmatrix}
\begin{bmatrix}
-1 & -1 \\
+1 & -1 \\
0 & +2
\end{bmatrix}
\begin{bmatrix}
-1 & -1 & -1 \\
+1 & -1 & -1 \\
0 & +2 & -1 \\
0 & 0 & +3
\end{bmatrix}
\begin{bmatrix}
-1 & -1 & -1 & -1 \\
+1 & -1 & -1 & -1 \\
0 & +2 & -1 & -1 \\
0 & 0 & +3 & -1 \\
0 & 0 & 0 & +4
\end{bmatrix}
\]

Their interaction is represented by variables that are the product of the Helmert variables coding for the main factors.
A univariate example; data from Sokal & Rohlf (2012) – 12 rats, males and females, were fed fresh and rancid lard (pig fat). The response variable is: how much fat (in g) each rat had eaten. The main factors are represented by Helmert contrasts; interaction variables are the product of the contrasts coding for the main factors.

\[
\begin{align*}
\text{Consumption} = & \begin{bmatrix}
709 \\
679 \\
699 \\
592 \\
538 \\
476 \\
657 \\
594 \\
677 \\
508 \\
505 \\
539
\end{bmatrix} & \text{Sex} = & \begin{bmatrix}
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
-1 \\
-1 \\
-1 \\
-1 \\
-1 \\
-1
\end{bmatrix} & \text{Lard} = & \begin{bmatrix}
+1 \\
+1 \\
-1 \\
-1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1
\end{bmatrix} & \text{Interaction} = & \begin{bmatrix}
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1 \\
+1
\end{bmatrix}
\end{align*}
\]
Properties –

1. The sum of values is zero in each Helmert-coded and interaction variable.

2. The variables coding for each factor and their interaction are orthogonal to one another within each group, i.e. their scalar products are zero.

3. The groups of variables coding for the main factors and their interaction are also orthogonal to one another (among the groups).

If the main factors were coded using ordinary (binary) contrasts, properties 1 and 3 would not occur.

=> In particular (property 3), the groups of variables would not be orthogonal among the groups coding for the main factors and their interaction.

Canonical analysis
R code for Manova by RDA for this example.

```r
# Rats data
Consumption <- c(709,679,699,592,538,476,657,594,677,508,505,539)
# Function gl() generates factors. Sex has 2 groups of 6 replicates
Sex <- gl(2,6)
Lard <- gl(2, 3, length=12)
Sex
# [1] 1 1 1 1 1 1 2 2 2 2 2 2
# Levels: 1 2
Lard
# [1] 1 1 1 2 2 2 1 1 1 2 2 2
# Levels: 1 2

# Construct matrix of contrasts. Remove the first column (intercept)
helmert <- model.matrix(~ Sex*Lard,
                      contrasts=list(Sex="contr.helmert", Lard="contr.helmert"))[,-1]

# Examine the matrix of contrasts. Which columns represent the main factors Sex and Lard? Which column represent the interaction?

# Check the orthogonality property of the coding variables
crossprod(helmert)
```
# Test interaction by RDA. Sex + Lard form the matrix of covariables

```r
test_interaction <- rda(Consumption, helmert[,3], helmert[,1:2])
anova(test_interaction, permutations = how(nperm = 9999))
```

# -----
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 9999

Model: `rda(X = Consumption, Y = helmert[, 3], Z = helmert[, 1:2])`

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Variance</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>83.52</td>
<td>0.63</td>
<td>0.4406</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>1060.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# -----

# Test factor Sex. Lard + interaction form the matrix of covariables

```r
test_factor <- rda(Consumption, helmert[,1], helmert[,2:3])
anova(test_factor, permutations = how(nperm = 9999))
```

# -----
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 9999

Model: `rda(X = Consumption, Y = helmert[, 1], Z = helmert[, 2:3])`

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Variance</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>343.7</td>
<td>2.5925</td>
<td>0.1495</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>1060.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# -----
# Test factor Lard. Sex + interaction form the matrix of covariables

Lard.rda <- rda(Consumption, helmert[,2], helmert[,c(1,3)])

anova(Lard.rda, permutations = how(nperm = 9999))

# -----  
Permutation test for rda under reduced model  
Permutation: free  
Number of permutations: 9999  

Model: rda(X = Consumption, Y = helmert[, 2], Z = helmert[, c(1, 3)])

<table>
<thead>
<tr>
<th>Df</th>
<th>Variance</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>5564.0</td>
<td>41.968</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>1060.6</td>
<td></td>
</tr>
</tbody>
</table>

# -----  

# For comparison, compute a 2-way anova with aov() of {stats}  
aov.res4 = aov(Consumption ~ Sex*Lard)  
summary(aov.res4)

# -----  

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1</td>
<td>3781</td>
<td>2.593</td>
<td>0.146036</td>
</tr>
<tr>
<td>Lard</td>
<td>1</td>
<td>61204</td>
<td>41.969</td>
<td>0.000192 ***</td>
</tr>
<tr>
<td>Sex:Lard</td>
<td>1</td>
<td>919</td>
<td>0.630</td>
<td>0.450255</td>
</tr>
<tr>
<td>Residuals</td>
<td>8</td>
<td>11667</td>
<td>1458</td>
<td></td>
</tr>
</tbody>
</table>

# -----  

### The F-statistics are identical, the p-values are very close.
Manova by RDA –

This example shows that, for univariate response data $y$, RDA produces the same $F$-statistics as a parametric analysis of variance.

For response data that are not multivariate normal, the p-values of the parametric tests would be biased. Permutation testing represents a better choice.

With multivariate response data $Y$, one should check the homogeneity of the within-group dispersion matrices before Manova. Function `betadisper()` in vegan is available to do so.

We can now use Manova by RDA for multivariate data for which parametric Manova would not be applicable.

$=>$ With multivariate data, one can also compute an RDA with the main factors as the explanatory variables $X$ and produce a triplot of the Manova solution. The triplot would allow a deeper functional interpretation of the results than a simple p-value would.
Manova by RDA for multivariate response data is readily available in function `adonis2()` of vegan.

Example (the rat data are univariate)

```r
res <- adonis2(Consumption ~ Sex*Lard, met="eucl", by="term")
res

# -----
Permutation test for adonis under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 999

adonis2(formula = Consumption ~ Sex * Lard, method = "eucl", by = "term")

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SumOfSqs</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1</td>
<td>3781</td>
<td>2.5925</td>
<td>0.152</td>
</tr>
<tr>
<td>Lard</td>
<td>1</td>
<td>61204</td>
<td>41.9685</td>
<td>0.002 **</td>
</tr>
<tr>
<td>Sex:Lard</td>
<td>1</td>
<td>919</td>
<td>0.6300</td>
<td>0.416</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>11667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Try `adonis2()` with multivariate response data, e.g. community composition data, properly transformed. This function also produces correct results for unbalanced designs.
Example 2 with multivariate response data. Example prepared by Daniel Borcard for the 2018 edition of “Numerical ecology with R”. Doubs River fish data, 27 sites. Hellinger transformation. Two crossed factors are created by dividing two of the environmental variables (Elevation and pH) into 3 classes each. Balanced design, 3 observations per group.

Results –

- Interaction: not significant.
- factor Elevation: significant
- Factor pH: not significant.

A RDA triplot was produced from the analysis of the Hellinger-transformed fish data constrained by factor Elevation.
In the triplot, the species that are associated with the 3 altitude classes can easily be identified. This information was not available from the analysis-of-variance table and its p-values.
8. R software

RDA, partial RDA, CCA, partial CCA: `rda()`, `cca()`, `capscale()` in vegan.

RDA (demonstration function) in package rdaTest available on http://adn.biol.umontreal.ca/~numericalecology/Rcode/

Manova by RDA: `adonis2()` in vegan.

Selection of environmental variables in RDA: `ordistep()` and `ordiR2step()` in vegan, `forward.sel()` in adespatial.

Variation partitioning in RDA: `varpart()` and `plot.varpart()` in vegan.


Coinertia (CoIA) and Procrustes (Proc) analyses: `coinertia()` in ade4; `procrustes()` in vegan, `procuste()` in ade4.

Co-correspondence analysis (CoCA): `cocorresp()` in cocorresp.

Multiple factor analysis (MFA): `mfa()` in ade4, `MFA()` in FactoMineR.
9. References


End of the presentation