Journal of Multivariate Analysis **72**, 100–119 (2000)
Article ID jmva.1999.1852, available online at http://www.idealibrary.com on IDEAL®

Canonical Analysis Applied to Multivariate Analysis of Variance

Michel Lejeune[1]

CEDRIC, Conservatoire National des Arts et Métiers, Paris, France

and

Tadeusz Caliński

Agricultural University of Poznań, Poznań, Poland

Received January 25, 1996

In this paper it is shown how a generalized form of canonical analysis can be useful to reveal which parametric functions of a MANOVA model, for instance treatment contrasts or combinations of observed variables, are responsible for rejection of a general linear hypothesis on these functions. For the decomposition in successive canonical terms the choice of a matrix norm is crucial. It is shown that the norm derived from the standardization of the least squares estimators of the parametric functions involved is equal to the Lawley–Hotelling statistic for testing the hypothesis under investigation. Thus, some useful interpretations based on canonical variates can be given in terms of the contributions of the various parametric functions to the overall test statistic or to statistics relevant to specific subhypotheses. Corresponding to these possibilities for interpretation, three different types of biplot are proposed. As an example, an agricultural block design experiment is thoroughly analyzed. © 2000 Academic Press

AMS 1991 subject classifications: 62H25, 62H15.

Key words and phrases: MANOVA, linear hypothesis, Lawley–Hotelling statistic, simultaneous test, canonical variate analysis, discriminant analysis, singular value decomposition, biplot.

1. INTRODUCTION

When applying multivariate analysis of variance (MANOVA) to experimental data one is usually interested in testing some hypotheses concerning various sources of variation. Such hypotheses are linear and can be expressed in the form H_0: $\mathbf{\Omega} = \mathbf{0}$ where $\mathbf{\Omega}$ is a matrix of linear functions of the parameters of the model. If a hypothesis is rejected, then one may wish

[1] Current address: University of Grenoble 2, Grenoble, France. E-mail: michel.lejeune@upmf-grenoble.fr.

0047-259X/00 $35.00

to find the main reasons for the rejection by investigating the data related to the hypothesis under examination. One approach, for instance, is to use a simultaneous test procedure by which the tested hypothesis can be decomposed into some more detailed component hypotheses to be tested simultaneously. In this paper we consider the analysis of the least squares estimator $\hat{\mathbf{\Omega}}$ of $\mathbf{\Omega}$ via canonical variate analysis (CVA) in order to get better insight into the structure of the data responsible for the rejection of the overall hypothesis (see Seal [23, Chap. 7], Mardia *et al.* [17, Sect. 12.5], and Seber [24, Sects. 5.8 and 10.1.4]). The analysis, in fact, is based on a singular value decomposition (SVD) with respect to an appropriate norm. It is shown that there is a natural norm to be derived from the covariance structure of $\hat{\mathbf{\Omega}}$ which is quite suitable for the problem of hypothesis testing because it yields the corresponding Lawley–Hotelling test statistic. Thus the resulting canonical decomposition is a way to exhibit the essential contributions to that test statistic. In particular the contributions of rows and columns of $\hat{\mathbf{\Omega}}$ will be of special interest for the interpretation of the data. The decomposition also allows one to produce biplots as graphical aids to reveal the main features concerning the hypothesis. Various uses of the biplot technique are proposed, each of which having a different meaningful interpretation.

The paper consists of five sections. The following section recalls the testing procedures in MANOVA and the relevant results to be used in our approach. Section 3 is devoted to the main concept of the paper, i.e., CVA applied to MANOVA as an exploratory method allowing one to measure various contributions to the test statistic. This is then illustrated by an example in Section 4. A concluding discussion follows in Section 5.

2. MULTIVARIATE ANALYSIS OF VARIANCE

The purpose of this section is to recall some basic properties of the general theory of MANOVA and to introduce the notations which will be used subsequently. In most cases the justification of the results presented here may be found in or deduced from Anderson [1], Rao [22], and Seber [24]. In other cases specific references are given.

2.1. *The Multivariate Linear Hypothesis and Related Tests*

The p-variate MANOVA model may be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{\Xi} + \mathbf{U}, \tag{2.1}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)'$ denotes the $n \times p$ matrix of observations, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)'$ denotes the $n \times q$ design matrix of rank $r \leqslant q$, $\mathbf{\Xi} = (\xi_1, \xi_2, ..., \xi_p)$

denotes the $q \times p$ matrix of parameters, and $\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n)'$ denotes the matrix of errors. Under the usual assumptions, including normality, the p-dimensional observations $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ are independent and $\mathbf{e}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ for all i.

In the analysis of multivariate experimental data, the interest may be in estimating not only some linear parametric functions of the type $\mathbf{c}'\mathbf{\Xi}$, but also in estimating functions of the type

$$\mathbf{c}'(m_1\xi_1 + m_2\xi_2 + \cdots + m_p\xi_p) = \mathbf{c}'\mathbf{\Xi}\mathbf{m},$$

where $\mathbf{m} = (m_1, m_2, ..., m_p)'$ is an arbitrary $p \times 1$ vector. This, in particular, may be the case when the characteristics (variables) observed on the experimental units are not interesting as such but rather as one or more linear compounds or comparisons among them. More generally, the experimenter may be interested in testing some hypotheses of the type

$$H_0\colon \mathbf{C\Xi M} = \mathbf{0}, \tag{2.2}$$

where the $g \times q$ matrix $\mathbf{C}$ is usually of rank g and the $p \times u$ matrix $\mathbf{M}$ is usually of rank u. Oftentimes the rows of $\mathbf{C}$ represent a set of contrasts between the q rows of $\mathbf{\Xi}$ and, therefore, the term "contrast" will be used when referring to a row of $\mathbf{C}$. As for the columns of $\mathbf{M}$, they usually define some combinations of the columns of $\mathbf{\Xi}$ which correspond to the observed variables. Therefore the term "combination of variables" will be used when referring to a column of $\mathbf{M}$. A necessary and sufficient condition for (2.2) to be testable is $\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X} = \mathbf{C}$, where $\mathbf{A}^-$ stands for a generalized inverse of $\mathbf{A}$. Then the best linear unbiased estimator for $\mathbf{\Omega} = \mathbf{C\Xi M}$, obtainable by the least squares method, is $\hat{\mathbf{\Omega}} = \mathbf{C}\hat{\mathbf{\Xi}}\mathbf{M}$, where $\hat{\mathbf{\Xi}} = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{Y}$.

The hypothesis H_0 may be tested by applying one of several known multivariate tests. Any of them involves the computation of the two matrices

$$\mathbf{E} = \mathbf{M}'\mathbf{Y}'\mathbf{Q}_E\mathbf{YM} = (n-r)\, \mathbf{M}'\mathbf{SM}, \tag{2.3}$$

the error (or residual) sum of squares of products (SSP) matrix, and

$$\mathbf{H} = \mathbf{M}'\mathbf{Y}'\mathbf{Q}_H\mathbf{YM} = \hat{\mathbf{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1}\, \hat{\mathbf{\Omega}},$$

the SSP matrix due to the deviation from hypothesis, where $\mathbf{Q}_E = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'$, $\mathbf{Q}_H = \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1}\, \mathbf{C}(\mathbf{XX})^- \mathbf{X}'$, and

$$\mathbf{S} = (n-r)^{-1}\, \mathbf{Y}'\mathbf{Q}_E\mathbf{Y}, \tag{2.4}$$

which is an unbiased estimator of $\mathbf{\Sigma}$. (See Morrison [19, Chap. 5] and Seber [24, Chap. 8 and 9]).

To test H_0 in (2.2), it will be convenient for the present approach to use the Lawley–Hotelling trace statistic defined as

$$T_0^2 = (n-r)\,\mathrm{trace}(\mathbf{E}^{-1}\mathbf{H})$$
$$= \mathrm{trace}\{(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1}\,\hat{\boldsymbol{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\,\mathbf{C}']^{-1}\,\hat{\boldsymbol{\Omega}}\} = \sum_{k=1}^{u} \mu_k, \tag{2.5}$$

where $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_u \geqslant 0$ are the eigenvalues of $(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1}\,\mathbf{H}$. Under the adopted assumptions the rank of $\mathbf{H}$ is equal to $s = \min(g, u)$ and the s positive eigenvalues are distinct, with probability 1 (see Seber [24, Sects. 2.5.2 and A2.8]). The distribution of T_0^2 depends only on u, $m_H = g$ and $m_E = n - r$, hence the critical value at level α may be denoted by $T^2_{0,\,\alpha;\,u,\,m_H,\,m_E}$. Corresponding tables are given by Seber [24, Appendix D].

Also, McKeon [18] has given the following approximation via an F-distribution with a and b degrees of freedom,

$$\frac{\mathrm{trace}(\mathbf{E}^{-1}\mathbf{H})}{c} \sim F_{a,\,b} \qquad \text{(approximately)}, \tag{2.6}$$

where $a = u m_H$, $b = 4 + (a+2)/(B-1)$, $c = a(b-2)/[b(m_E - u - 1)]$, $m_H = g$ and $m_E = n - r$, with $B = (m_E + m_H - u - 1)(m_E - 1)/[(m_E - u - 3)(m_E - u)]$. As noticed by Seber [24, p. 39], this approximation is surprisingly accurate. The F-distribution is exact when $s = 1$. In the special case of $\mathbf{M} = \mathbf{I}_p$ in (2.2), the number u above is equal to p.

2.2. *Tests for Component Hypotheses*

If H_0 is rejected, one may be interested in finding which parametric functions are responsible for that rejection. A natural way to proceed is to test some *component hypotheses* implied by H_0, particularly those obtained when the matrix $\mathbf{M}$ is replaced by its jth column $\mathbf{m}_j$ or/and the matrix $\mathbf{C}$ is replaced by its ith row $\mathbf{c}_i'$. Tests of the hypotheses $H_{0,\,i.}$: $\mathbf{c}_i'\boldsymbol{\Xi}\mathbf{M} = \mathbf{0}'$ for $i = 1, 2, ..., g$ will allow one to find which rows of the matrix $\mathbf{C}\boldsymbol{\Xi}\mathbf{M}$ might be responsible for rejecting H_0, tests of $H_{0,\,.j}$: $\mathbf{C}\boldsymbol{\Xi}\mathbf{m}_j = \mathbf{0}$ for $j = 1, 2, ..., u$ will allow one to find responsible columns of that matrix, and tests of $H_{0,\,ij}$: $\mathbf{c}_i'\boldsymbol{\Xi}\mathbf{m}_j = 0$ for all i and j will allow one to identify responsible individual elements of $\mathbf{C}\boldsymbol{\Xi}\mathbf{M}$.

The appropriate Lawley–Hotelling statistics for testing these component hypotheses are as follows. For testing $H_{0,\,i.}$ one can use

$$T^2_{0,\,i.} = (n-r)[\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^{-}\,\mathbf{c}_i]^{-1}\,\mathbf{c}_i'\hat{\boldsymbol{\Xi}}\mathbf{M}\mathbf{E}^{-1}\mathbf{M}'\hat{\boldsymbol{\Xi}}'\mathbf{c}_i$$
$$= (n-r)[\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^{-}\,\mathbf{c}_i]^{-1}\,\hat{\boldsymbol{\Omega}}_i\mathbf{E}^{-1}\hat{\boldsymbol{\Omega}}_i', \tag{2.7}$$

where $\hat{\boldsymbol{\Omega}}_i = \mathbf{c}_i' \hat{\boldsymbol{\Xi}} \mathbf{M}$ is the ith row of the matrix $\hat{\boldsymbol{\Omega}} = \mathbf{C} \hat{\boldsymbol{\Xi}} \mathbf{M}$. For testing $H_{0,\,.j}$ one can use

$$T^2_{0,\,.j} = (n-r)\,\mathbf{m}_j'\mathbf{Y}'\mathbf{Q}_H\mathbf{Y}\mathbf{m}_j/\mathbf{m}_j'\mathbf{Y}'\mathbf{Q}_E\mathbf{Y}\mathbf{m}_j = (n-r)\,\mathbf{m}_j'\mathbf{Y}'\mathbf{Q}_H\mathbf{Y}\mathbf{m}_j/E_{jj}, \tag{2.8}$$

where E_{jj} is the jth diagonal element of the matrix $\mathbf{E}$ defined in (2.3). For testing $H_{0,\,ij}$ one can use

$$\begin{aligned} T^2_{0,\,ij} &= (n-r)[\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^- \mathbf{c}_i]^{-1} (\mathbf{c}_i' \hat{\boldsymbol{\Xi}} \mathbf{m}_j)^2/E_{jj} \\ &= (n-r)[\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^- \mathbf{c}_i]^{-1} \hat{\Omega}^2_{ij}/E_{jj}, \end{aligned} \tag{2.9}$$

where $\hat{\Omega}_{ij}$ is the (i, j)th element of the matrix $\hat{\boldsymbol{\Omega}}$. Note that $T^2_{0,\,i.}(m_E - u + 1)/(m_E u)$ has an exact $F_{u,\,m_E - u + 1}$-distribution, $T^2_{0,\,.j}/m_H$ has an exact $F_{m_H,\,m_E}$-distribution and $T^2_{0,\,ij}$ has an exact $F_{1,\,m_E}$-distribution. For this last statistic it will be convenient to define $T_{0,\,ij} = (n-r)^{1/2} [\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^- \mathbf{c}_i]^{-1/2} \hat{\Omega}_{ij}/E^{1/2}_{jj}$ which has a Student distribution with m_E degrees of freedom. Any of these distributions is central if the tested hypothesis is true.

Now, to obtain a simultaneous test procedure (STP) at level α for this family of tests one should use the critical value $T^2_{0,\,\alpha;\,u,\,m_H,\,m_E}$ for each test implied by H_0 (see Gabriel [9]). However, a disadvantage of this STP approach is that even if the overall hypothesis is rejected, it may happen that none of these implied component hypotheses is rejected by the STP, as the responsibility for rejecting H_0 may be related not only to the individual parametric functions considered in the component hypotheses but also to certain mutual relationships among them. An alternative or auxiliary method giving more thorough insight into the structure of the data responsible for rejecting the overall hypothesis (2.2) is the one based on CVA considered in the next section.

3. CANONICAL ANALYSIS FOR MANOVA

3.1. *Definition of an Appropriate Canonical Variate Analysis*

Whenever $H_0\colon \boldsymbol{\Omega} = \mathbf{0}$ is rejected on the basis of the Lawley–Hotelling statistic one would like to investigate the matrix $\hat{\boldsymbol{\Omega}}$ in order to discover which parametric functions are mostly responsible for the rejection. Suppose one would simply like to look at the elements of $\hat{\boldsymbol{\Omega}}$. Then these elements have to be standardized to become comparable among themselves. To do so, consider their joint distribution by introducing the $gu \times 1$ vector $(\hat{\boldsymbol{\Omega}}')^{\mathrm{v}}$ obtained by transposing the rows of $\hat{\boldsymbol{\Omega}}$ and then piling them up on top of one another. From the earlier normality assumptions it can be easily established that

$$(\hat{\mathbf{\Omega}}')^{\mathrm{v}} \sim N_{gu}[(\mathbf{\Omega}')^{\mathrm{v}}, \mathbf{R} \otimes \mathbf{T}],$$

where $\otimes$ is the Kronecker product symbol, $\mathbf{R} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-} \mathbf{C}'$ and $\mathbf{T} = \mathbf{M}'\mathbf{\Sigma}\mathbf{M}$. Then $(\mathbf{R} \otimes \mathbf{T})^{-1/2} (\hat{\mathbf{\Omega}}')^{\mathrm{v}} \sim N_{gu}[(\mathbf{R} \otimes \mathbf{T})^{-1/2} (\mathbf{\Omega}')^{\mathrm{v}}, \mathbf{I}_{gu}]$ so that the matrix $(\mathbf{R} \otimes \mathbf{T})^{-1/2}$ provides a natural standardization for the elements of $(\hat{\mathbf{\Omega}}')^{\mathrm{v}}$. Since $(\mathbf{R} \otimes \mathbf{T})^{-1/2} (\hat{\mathbf{\Omega}}')^{\mathrm{v}}$ is equal to $(\mathbf{T}^{-1/2}\hat{\mathbf{\Omega}}'\mathbf{R}^{-1/2})^{\mathrm{v}}$ (see Mardia *et al.* [17, p. 460]), the elements of $\mathbf{R}^{-1/2}\hat{\mathbf{\Omega}}\mathbf{T}^{-1/2}$ are i.i.d. according to $N(0, 1)$ under H_0. Hence, the sum of squares of these elements has a chi-square distribution with gu degrees of freedom, which is central if H_0 is true. However, $\mathbf{\Sigma}$ is unknown and has to be replaced by its unbiased estimator $\mathbf{S}$ given in (2.4). Finally, the natural standardized form of $\hat{\mathbf{\Omega}}$ is thus

$$\hat{\mathbf{\Omega}}_{std} = [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-} \mathbf{C}']^{-1/2} \hat{\mathbf{\Omega}}(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1/2},$$

and the total variation of the deviations from H_0 can be expressed by

$$\begin{aligned} \mathrm{trace}(\hat{\mathbf{\Omega}}_{std}\hat{\mathbf{\Omega}}'_{std}) &= \mathrm{trace}\{(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1} \hat{\mathbf{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-} \mathbf{C}']^{-1} \hat{\mathbf{\Omega}}\} \\ &= (n-r)\,\mathrm{trace}(\mathbf{E}^{-1}\mathbf{H}), \end{aligned} \tag{3.1}$$

which is the Lawley–Hotelling statistic for testing H_0, given in (2.5).

This result is particularly important in order to get deeper insight into the contrasts and combinations of variables responsible for the rejection of H_0. Namely, it provides a justification for the squared norm on $\hat{\mathbf{\Omega}}$ as in (3.1) to be used for a CVA of the type proposed by Rao [21] and Seal [23] in the special case of one-way MANOVA, alias discriminant coordinate analysis (see also Seber [24, p. 270]), and later suggested by various authors for extension to general MANOVA (see for instance Chatfield and Collins [7, p. 153]). This CVA has been applied to a two-way MANOVA, e.g., by Caliński *et al.* [2, 3] in the context of agricultural crop variety research and by Camussi *et al.* [5] in connection with a study of genetic distances. The use of the appropriate norm for CVA demonstrates a direct link between the canonical decomposition of $\hat{\mathbf{\Omega}}$ and the Lawley–Hotelling statistic, thus revealing new interpretative features.

Following Lejeune [16], canonical representation can be seen as the approximation of $\hat{\mathbf{\Omega}}$ by a sequence of matrices of rank one, each of which can yield two sets of coordinates in one dimension: one for rows (contrasts) and one for columns (combinations of variables). The approximation should operate with respect to the norm defined by

$$\|\mathbf{A}\|_*^2 = \mathrm{trace}\{(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1} \mathbf{A}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-} \mathbf{C}']^{-1} \mathbf{A}\},$$

as in (3.1) for $\hat{\boldsymbol{\Omega}}$, so as to recover most of the test statistic in a few dimensions. According to the approximation theorem of Eckart and Young [8] the appropriate solution is given by the SVD of $\hat{\boldsymbol{\Omega}}_{std}$, i.e.,

$$\hat{\boldsymbol{\Omega}}_{std} = \sum_{k=1}^{s} \mu_k^{-1/2} \mathbf{v}_k \mathbf{t}'_k, \tag{3.2}$$

where the μ_k's are the nonzero eigenvalues of $\hat{\boldsymbol{\Omega}}'_{std}\hat{\boldsymbol{\Omega}}_{std}$ (or of $\hat{\boldsymbol{\Omega}}_{std}\hat{\boldsymbol{\Omega}}'_{std}$) and, from (3.1), are identical to those in (2.5), $\mathbf{t}_k$ is the eigenvector of this matrix corresponding to μ_k, normed as $\mathbf{t}'_k\mathbf{t}_k = \mu_k$, and $\mathbf{v}_k$ is the eigenvector of $\hat{\boldsymbol{\Omega}}_{std}\hat{\boldsymbol{\Omega}}'_{std}$, normed as $\mathbf{v}'_k\mathbf{v}_k = \mu_k$. Then, applying the transition equation $\hat{\boldsymbol{\Omega}} = (\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}')^{1/2} \hat{\boldsymbol{\Omega}}_{std}(\mathbf{M}'\mathbf{S}\mathbf{M})^{1/2}$, one derives the appropriate decomposition for $\hat{\boldsymbol{\Omega}}$ itself, as

$$\hat{\boldsymbol{\Omega}} = \sum_{k=1}^{s} \mu_k^{-1/2} \boldsymbol{\psi}_k \boldsymbol{\varphi}'_k, \tag{3.3}$$

where $\boldsymbol{\psi}_k = [\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{1/2} \mathbf{v}_k$ and $\boldsymbol{\varphi}_k = (\mathbf{M}'\mathbf{S}\mathbf{M})^{1/2} \mathbf{t}_k$. Thus, these vectors satisfy the equalities

$$\hat{\boldsymbol{\Omega}}(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1} \hat{\boldsymbol{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1} \boldsymbol{\psi}_k = \mu_k \boldsymbol{\psi}_k$$

and

$$\hat{\boldsymbol{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1} \hat{\boldsymbol{\Omega}}(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1} \boldsymbol{\varphi}_k = \mu_k \boldsymbol{\varphi}_k,$$

being orthonormalized in the following way,

$$\boldsymbol{\psi}'_k[\mathbf{C}(\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1} \boldsymbol{\psi}_{k'} = \begin{cases} \mu_k, & \text{if } \ k = k', \\ 0, & \text{if } \ k \neq k', \end{cases} \tag{3.4}$$

and

$$\boldsymbol{\varphi}'_k(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1} \boldsymbol{\varphi}_{k'} = \begin{cases} \mu_k, & \text{if } \ k = k', \\ 0, & \text{if } \ k \neq k'. \end{cases} \tag{3.5}$$

Since the squared norm $\|\mu_k^{-1/2} \boldsymbol{\psi}_k \boldsymbol{\varphi}'_k\|_*^2$ of the kth *canonical term* (as the kth matrix, of rank one, of the decomposition (3.3) will be called) is equal to μ_k, it follows that μ_k is the part of the Lawley–Hotelling statistic $T_0^2 = \sum_{k=1}^{s} \mu_k$ accounted for by the kth canonical term. Let μ_k be called the *contribution of the canonical term k to the test statistic* (or *to the total variation*).

One can also look at the contribution to T_0^2 of a given row of $\hat{\boldsymbol{\Omega}}$ corresponding to a contrast defined by the same row of the matrix $\mathbf{C}$. For this, write $\mathbf{v}_k = (v_{k1}, v_{k2}, ..., v_{kg})'$. Then the contribution of row i to the kth canonical term can be seen as equal to v_{ki}^2, since $T_0^2 = \sum_{k=1}^{s} \mu_k = \sum_{k=1}^{s} \mathbf{v}_k' \mathbf{v}_k$. Let v_{ki}^2/μ_k be called the *relative contribution of row i to the canonical term k*. These contributions can be used in order to identify a canonical term.

The sum $\sum_{k=1}^{s} v_{ki}^2$ defines the contribution of row i to the total variation. Consequently, $v_{ki}^2/\sum_{t=1}^{s} v_{ti}^2$ describes the *part of contribution of row i explained by the canonical term k*. It tells how well the contribution of the corresponding contrast can be represented by the canonical term k. The same definitions apply for the contributions of the combinations of variables implied by the columns of $\mathbf{M}$, via the $\mathbf{t}_k$'s, where $\mathbf{t}_k = (t_{k1}, t_{k1}, ..., t_{ku})'$.

To see the contribution of the (i, j)th element of $\hat{\mathbf{\Omega}}$ to the test statistic T_0^2, one has to take

$$\hat{\Omega}^2_{std,\, ij} = \left(\sum_{k=1}^{s} \mu_k^{-1/2} v_{ki} t_{kj} \right)^2, \tag{3.6}$$

since $T_0^2 = \mathrm{trace}(\hat{\mathbf{\Omega}}_{std} \hat{\mathbf{\Omega}}'_{std})$, as shown in (3.1), i.e.,

$$T_0^2 = \sum_{i=1}^{g} \sum_{j=1}^{u} \hat{\Omega}^2_{std,\, ij}. \tag{3.7}$$

These specific contributions, though attributable to rows, columns or elements of the matrix $\hat{\mathbf{\Omega}}$, depend in fact on all its elements, because the matrices $[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-} \mathbf{C}']^{-1/2}$ and $(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1/2}$ are not diagonal. Their interest stems from the fact, however, that they sum up to the actual value of T_0^2.

3.2. *Geometrical Representations*

For a geometrical representation of $\hat{\mathbf{\Omega}}$, let its g rows be mapped into g points in the usual s-dimensional Euclidean space with coordinates $\boldsymbol{\psi}_k = (\psi_{k1}, \psi_{k2}, ..., \psi_{kg})'$ on the kth axis, which may be called the kth *canonical coordinates* (see Mardia *et al.* [17, p. 343]) or discriminant coordinates (see Gnanadesikan [13, p. 86]). Furthermore, let the u columns of $\hat{\mathbf{\Omega}}$ be mapped into points in the same space by giving them the coordinates $\mu_k^{-1/2} \boldsymbol{\varphi}_k = \mu_k^{-1/2} (\varphi_{k1}, \varphi_{k2}, ..., \varphi_{ku})'$ on the kth axis, which may be called the kth dual canonical coordinates. Now, to describe geometrically the magnitudes of the individual elements of $\hat{\mathbf{\Omega}}$, note that its (i, j)th element, $\hat{\Omega}_{ij} = \sum_{k=1}^{s} \psi_{ki} (\mu_k^{-1/2} \varphi_{kj})$, is equal to the scalar product of the vectors

$$\boldsymbol{\psi}_{(i)} = (\psi_{1i}, \psi_{2i}, ..., \psi_{si})'$$

and

$$\tilde{\boldsymbol{\varphi}}_{(j)} = (\mu_1^{-1/2}\varphi_{1j}, \mu_2^{-1/2}\varphi_{2j}, ..., \mu_s^{-1/2}\varphi_{sj})',$$

representing the ith row and the jth column of $\hat{\boldsymbol{\Omega}}$, respectively. Also, from (3.3) and (3.5),

$$\hat{\boldsymbol{\Omega}}(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1}\hat{\boldsymbol{\Omega}}' = \sum_{k=1}^{s} \boldsymbol{\psi}_k \boldsymbol{\psi}_k',$$

so that the Euclidean squared distance of row i to the origin, $\sum_{k=1}^{s} \psi_{ki}^2$, is equal to the ith diagonal element of $\hat{\boldsymbol{\Omega}}(\mathbf{M}'\mathbf{S}\mathbf{M})^{-1}\hat{\boldsymbol{\Omega}}'$, and thus can be interpreted as the Mahalanobis (squared) distance of this row to the null row, i.e., as its departure from H_0.

Now, in an alternative manner, $\mu_k^{-1/2}\psi_{ki}$ and φ_{kj} can be used as the coordinates of row i and column j on the kth axis, i.e., these components can be represented by

$$\tilde{\boldsymbol{\psi}}_{(i)} = (\mu_1^{-1/2}\psi_{1i}, \mu_2^{-1/2}\psi_{2i}, ..., \mu_s^{-1/2}\psi_{si})'$$

and

$$\boldsymbol{\varphi}_{(j)} = (\varphi_{1j}, \varphi_{2j}, ..., \varphi_{sj})',$$

respectively. Then, from (3.3) and (3.4),

$$\mathbf{H} = \hat{\boldsymbol{\Omega}}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}']^{-1}\hat{\boldsymbol{\Omega}} = \sum_{k=1}^{s} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k',$$

so that the squared distance of column j to the origin, $\sum_{k=1}^{s} \varphi_{kj}^2$, is then equal to the jth diagonal element of the SSP matrix $\mathbf{H}$ and can be interpreted as the sum of squares due to deviation from the null hypothesis for this column.

When reducing the representation to the first two axes only, one obtains an approximation of the matrix $\hat{\boldsymbol{\Omega}}$. It can be then displayed by a biplot, as originally proposed by Gabriel [10] in connection with principal component analysis (PCA) and later by Gabriel [11, 12] in connection with CVA (see also Gower and Hand [14, Chap. 5]).

If one would rather interpret the rows, columns and elements of $\hat{\boldsymbol{\Omega}}$ in terms of the statistics defined in (2.7), (2.8), and (2.9) for testing the component hypotheses $H_{0,\,i.}$, $H_{0,\,.j}$, and $H_{0,\,ij}$, respectively, one should use the coordinates obtained by transforming the vectors $\boldsymbol{\psi}_k$ and $\boldsymbol{\varphi}_k$ to

$$\boldsymbol{\psi}_k^* = [\operatorname{diag}\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'\}]^{-1/2}\boldsymbol{\psi}_k$$

and

$$\boldsymbol{\varphi}_k^* = [\operatorname{diag}\{\mathbf{M}'\mathbf{S}\mathbf{M}\}]^{-1/2}\, \boldsymbol{\varphi}_k,$$

respectively, for $k = 1, 2, ..., s$. In a representation analogous to the first one above, the Euclidean squared distance of row i to the origin, $\sum_{k=1}^{s} \psi_{ki}^{*2}$, will be equal to the test statistic $T_{0,\, i.}^2$, and the corresponding scalar product, $\sum_{k=1}^{s} \psi_{ki}^*(\mu_k^{-1/2} \varphi_{kj}^*)$, of the vectors representing the ith row and the jth column, will be equal to the test statistic $T_{0,\, ij}$. In the alternative manner, where the ψ_{ki}^*'s are weighted by the $\mu_k^{-1/2}$'s, the squared distance of column j to the origin, $\sum_{k=1}^{s} \varphi_{kj}^{*2}$, will yield the test statistic $T_{0,\, .j}^2$. Again, the biplot technique can be applied to indicate which contrasts, combinations of variables or individual elements of $\hat{\boldsymbol{\Omega}}$ are most likely to be responsible for rejecting H_0.

Lastly, to get a geometrical representation of the standardized matrix $\hat{\boldsymbol{\Omega}}_{std}$ one should use, according to (3.2), the vector $\mathbf{v}_k$ instead of $\boldsymbol{\psi}_k$, and the vector $\mathbf{t}_k$ instead of $\boldsymbol{\varphi}_k$ for $k = 1, 2, ..., s$. Since $\hat{\boldsymbol{\Omega}}_{std} \hat{\boldsymbol{\Omega}}'_{std} = \sum_{k=1}^{s} \mathbf{v}_k \mathbf{v}'_k$, the squared distance of row i to the origin, $\sum_{k=1}^{s} v_{ki}^2$, can be directly interpreted as its contribution to the T_0^2-statistic in the first kind of geometrical representation. In the alternative representation, since $\hat{\boldsymbol{\Omega}}'_{std} \hat{\boldsymbol{\Omega}}_{std} = \sum_{k=1}^{s} \mathbf{t}_k \mathbf{t}'_k$, the Euclidean squared distance of column j to the origin, $\sum_{k=1}^{s} t_{kj}^2$, can be interpreted as the contribution of this column. In both cases the scalar products yield the contributions of the individual elements of $\hat{\boldsymbol{\Omega}}$ to the global T_0^2-statistic, as explained in (3.6) and (3.7). As before, the biplot technique can be used to display approximately these contributions.

At this point it should be emphasized that the first couple of geometrical representations of the elements of $\hat{\boldsymbol{\Omega}}$ given above refers to the magnitudes of these elements, the second couple refers to the single statistics for testing the nullity of the corresponding parametric functions, i.e., the hypotheses $H_{0,\, ij}$, while the third couple, that of the elements of $\hat{\boldsymbol{\Omega}}_{std}$, refers to the elementary contributions to the T_0^2-statistic used for testing the overall hypothesis H_0.

For the representation of the elements of $\hat{\boldsymbol{\Omega}}$ the biplot technique has been proposed by Gabriel [11] in a one-way MANOVA of a rainmaking experiment and, e.g., by Caliński *et al.* [3] when analyzing genotype-region interactions of wheat varieties. The other forms of geometrical representations proposed here enhance the means of investigation for any hypothesis after its rejection. Along with any plot it will be useful to specify the accuracy of the representation of each vector as the ratio of the squared norm of its projection in the two dimensions to its squared norm in the s-dimensional space. The use of the various representations and of the attached elements of interpretation will be illustrated by the following example.

4. EXAMPLE

The example is taken from Ceranka *et al.* [6]. The data come from a plant breeding research on $v=10$ strains of sunflower, which were compared in a field experiment laid out in an incomplete block design, with $b=20$ blocks of varying sizes (since only a selection of 10 from a larger set of strains has been analyzed) and with unequal replication numbers (due to the fact that two strains used as standards were replicated on 20 plots each, while the others were replicated only on 4 plots each). The experimental data from the $n=72$ plots are multivariate, as the observations were taken on $p=4$ quantitative traits (variables):

(i) the average height of plants in cm (PH),

(ii) the yield of seeds (achenes) in g per plant (SY),

(iii) the weight of 1000 seeds in g (SW),

(iv) the percentage of husk content in seeds (HC).

The experiment has been analysed under the usual model for a block design (see, e.g., Pearce [20, Sect. 3.1]), which in the multivariate case can be written, in accordance with (2.1), as

$$\mathbf{Y} = \mathbf{X}_0 \boldsymbol{\Xi}_0 + \mathbf{X}_1 \boldsymbol{\Xi}_1 + \mathbf{U}.$$

Here, $\boldsymbol{\Xi}_0$ denotes the 20×4 matrix of block parameters and $\boldsymbol{\Xi}_1$ the 10×4 matrix of strain parameters (effects), both for the four traits, while $\mathbf{X}_0$ and $\mathbf{X}_1$ are the corresponding design matrices.

Suppose we want to test the hypothesis that there are no strain effects, i.e., $H_0\colon \boldsymbol{\Xi}_1 = \mathbf{0}$ with the suitable identifiability constraint $\mathbf{r}'\boldsymbol{\Xi}_1 = \mathbf{0}'$ where $\mathbf{r}$ is the vector of the numbers of treatment (strain) replications. The null hypothesis can be written equivalently $H_0\colon \mathbf{C}_1\boldsymbol{\Xi}_1 = \mathbf{0}$ where $\mathbf{C}_1 = \mathbf{I}_v - n^{-1}\mathbf{1}_v\mathbf{r}'$ (of rank $v-1$). In the notations of Section 2.1 one would write $H_0\colon \mathbf{C}\boldsymbol{\Xi}\mathbf{M} = \mathbf{0}$ where $\mathbf{C} = (\mathbf{0} : \mathbf{C}_1)$ is a 10×30 matrix, $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_0' : \boldsymbol{\Xi}_1')'$ and $\mathbf{M} = \mathbf{I}_4$.

Thus, the matrix to be analyzed is $\mathbf{C}_1\hat{\boldsymbol{\Xi}}_1$, where $\hat{\boldsymbol{\Xi}}_1 = (\bar{\mathbf{X}}_1', \bar{\mathbf{X}}_1)^{-1} \bar{\mathbf{X}}_1'\mathbf{Y} - \mathbf{1}_v\hat{\boldsymbol{\theta}}'$ with $\bar{\mathbf{X}}_1 = [\mathbf{I}_n - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' + n^{-1}\mathbf{1}_n\mathbf{1}_n']\,\mathbf{X}_1$ and $\hat{\boldsymbol{\theta}} = n^{-1}\mathbf{Y}'\mathbf{1}_n$. Since $\mathbf{r}'\hat{\boldsymbol{\Xi}}_1 = \mathbf{0}'$, $\mathbf{C}_1\hat{\boldsymbol{\Xi}}_1$ is equal to $\hat{\boldsymbol{\Xi}}_1$. The authors (Ceranka *et al.* [6]) give the estimates of the true strain effects as shown in Table I along with the vector of overall means $\hat{\boldsymbol{\theta}}$.

The Lawley–Hotelling statistic defined in (2.5) for testing H_0 is here of the form

$$T_0^2 = \operatorname{trace}(\mathbf{S}^{-1}\hat{\boldsymbol{\Xi}}_1'\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1\hat{\boldsymbol{\Xi}}_1),$$

TABLE I

Estimates of the True Strain Effects

		Trait			
Strain	No. replic.	PH	SY	SW	HC
1	4	−0.32	4.45	−9.80	1.22
2	4	2.71	2.16	−2.20	−0.28
3	4	−0.61	−1.46	−5.85	0.28
4	4	−8.11	3.38	−1.15	−0.76
5	4	0.75	−0.81	−7.46	0.27
6	4	−8.75	−6.46	8.74	0.43
7	4	−9.70	−7.31	2.55	1.10
8	4	5.84	5.96	2.84	−1.64
9	20	−1.05	−2.34	1.41	−0.74
10	20	4.70	2.36	1.06	0.61
Overal means	72	68.21	45.19	60.76	27.85

since $\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1$ can be taken as $[\mathbf{C}_1(\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)^{-1}\mathbf{C}_1']^-$ (a generalized inverse is to be used because of the rank deficiency). The eigenvalues of $\mathbf{S}^{-1}\hat{\mathbf{\Xi}}_1'\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1\hat{\mathbf{\Xi}}_1$ are

$$\mu_1 = 99.935 \qquad (52.7\,\%)$$

$$\mu_2 = 65.558 \qquad (34.6\,\%)$$

$$\mu_3 = 18.314 \qquad (9.7\,\%)$$

$$\mu_4 = 5.758 \qquad (3.0\,\%)$$

$$\sum_{k=1}^{4} \mu_k = 189.565 \qquad (100.0\,\%).$$

Thus, the Lawley–Hotelling test statistic is $T_0^2 = 189.565$ which, according to (2.6), can be converted to $T_0^2/(m_E c) = 4.751$, to be compared with the significance points of the F-distribution with 36 and 98 degrees of freedom. This gives for T_0^2 the critical values $T_{0;\,0.05}^2 = 61.36$ and $T_{0;\,0.01}^2 = 73.19$. The P-value for T_0^2 is around $(0.5)\,10^{-9}$ and therefore H_0 is obviously rejected.

For the component hypotheses $H_{0,\,i.}$ and $H_{0,\,.j}$ the test statistics are given by

$$T_{0,\,i.}^2 = [(\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)_{(ii)}^{-1} - n^{-1}]^{-1}\,\hat{\mathbf{\Xi}}_{1i.}\mathbf{S}^{-1}\hat{\mathbf{\Xi}}_{1i.}'.$$

TABLE II

Values of the Separate Test Statistics and Contributions to T_0^2.

	$T_{0,i.}^2$, $T_{0,.j}^2$, and $T_{0,ij}$ statistics (critical values at $\alpha = 0.05$ are in parentheses)					Elementary contributions to the T_0^2 statistic in %				
Trait	PH	SY	SW	HC		PH	SY	SW	HC	Total
Strain		$T_{0,ij}$ (± 2.02)			$T_{0,i.}^2$ (11.21)					
1	-0.15	1.21	**-3.56**	1.62	**32.70**	0.55	1.56	$^{-}$**9.42**	$^{+}$**4.85**	16.4
2	1.26	0.59	-0.80	-0.37	2.31	0.40	0.15	0.23	0.10	0.9
3	-0.28	-0.40	**-2.13**	0.37	7.64	0.28	0.02	2.98	1.12	4.4
4	**-3.77**	0.92	-0.42	-1.00	**22.06**	$^{-}$**9.37**	1.18	0.49	0.02	11.1
5	0.34	-0.22	**-2.66**	0.35	9.96	0.04	0.00	$^{-}$**4.62**	1.16	5.8
6	**-4.06**	-1.76	**3.17**	0.57	**24.77**	$^{-}$**6.16**	1.42	$^{+}$**4.80**	0.09	12.5
7	**-4.51**	-1.99	0.93	1.45	**28.41**	$^{-}$**10.1**	1.34	0.12	3.09	14.5
8	**2.71**	1.62	1.03	**-2.16**	**25.42**	3.92	0.83	0.94	$^{-}$**6.06**	11.7
9	-1.37	-1.79	1.43	**-2.73**	**17.93**	0.26	1.29	1.30	3.90	6.7
10	**6.12**	1.80	1.08	**2.27**	**42.73**	$^{+}$**14.5**	0.37	0.77	0.20	15.9
	85.35	17.04	**38.95**	**19.66**	$\leftarrow T_{0,.j}^2$ (18.96)	45.6	8.1	25.7	20.6	100.0

and

$$T_{0,.j}^2 = \hat{\boldsymbol{\Xi}}_{1.j}' \bar{\mathbf{X}}_1' \bar{\mathbf{X}}_1 \hat{\boldsymbol{\Xi}}_{1.j} / S_{jj},$$

where $\hat{\boldsymbol{\Xi}}_{1i.}$ and $\hat{\boldsymbol{\Xi}}_{1.j}$ are the ith row and the jth column of $\hat{\boldsymbol{\Xi}}_1$, respectively, and $(\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)_{(ii)}^{-1}$ stands for the ith diagonal element of the matrix $(\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)^{-1}$, S_{jj} for the jth diagonal element of the matrix $\mathbf{S}$. For the $H_{0,ij}$'s the $T_{0,ij}$'s are given by the matrix

$$[\operatorname{diag}\{(\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)^{-1}\} - n^{-1}\mathbf{I}_v]^{-1/2} \hat{\boldsymbol{\Xi}}_1 [\operatorname{diag}\{\mathbf{S}\}]^{-1/2} = \sum_{k=1}^{4} \mu_k^{-1/2} \boldsymbol{\psi}_k^* (\boldsymbol{\varphi}_k^*)'.$$

Each $T_{0,ij}$ has a Student distribution with 43 degrees of freedom, $(10/43)\,T_{0,i.}^2$ has a $F_{4,40}$-distribution and $T_{0,.j}^2/9$ has a $F_{9,43}$-distribution. These statistics are given in Table II with their corresponding critical values for separate testing (not for the STP, which would require all the T_0^2 type statistics given in Table II to be compared with a common critical value, here $T_{0;0.05}^2 = 61.36$ or $T_{0;0.05} = \pm 7.83$).

The contributions of individual parametric functions to T_0^2 are given by the squared elements of the matrix $\hat{\boldsymbol{\Xi}}_1$ standardized, i.e., $\hat{\boldsymbol{\Xi}}_{1,std} = (\bar{\mathbf{X}}_1'\bar{\mathbf{X}}_1)^{1/2} \hat{\boldsymbol{\Xi}}_1 \mathbf{S}^{-1/2}$, whereas the contributions of strains and traits are obtained by summing across rows and columns of this matrix respectively. Table II also contains these contributions (in percents for T_0^2).

TABLE III

Canonical Coordinates for Three Representations

	Strain effects: $\hat{\boldsymbol{\Xi}}_1$		Test statistics $T_{0,ij}$			Contributions to T_0^2: $\hat{\boldsymbol{\Xi}}_{1,std}$				
	Axis 1	Axis 2	Axis 1	Axis 2	cosines	Axis 1	Axis 2	cosines	[a]	[b]
Strain	$\boldsymbol{\psi}_1$	$\boldsymbol{\psi}_2$	$\boldsymbol{\psi}_1^*$	$\boldsymbol{\psi}_2^*$	*(raw)*	$\mathbf{v}_1$	$\mathbf{v}_2$	*(raw)*		
1	1.883	2.452	3.467	4.515	0.99	3.214	4.519	0.99	10.3%	31.2%
2	−0.531	0.552	−0.978	1.016	0.93	−0.776	0.856	0.89	0.6%	1.1%
3	0.976	1.067	1.797	1.965	0.96	1.935	1.994	0.96	3.7%	6.1%
4	2.076	−0.310	3.813	−0.569	0.82	3.698	−0.606	0.82	13.7%	0.6%
5	0.816	1.460	1.476	2.640	0.96	1.698	2.716	0.96	2.9%	11.3%
6	1.327	−2.233	2.436	−4.101	0.96	2.365	−4.028	0.96	5.6%	24.8%
7	2.606	−0.951	4.786	−1.747	0.96	4.747	−1.715	0.96	22.5%	4.5%
8	−2.482	−0.504	−4.558	−0.926	0.92	−4.237	−0.823	0.92	18.0%	1.0%
9	−0.237	−0.702	−1.223	−3.616	0.90	−0.837	−3.088	0.89	0.7%	14.5%
10	−1.097	0.395	−5.653	2.036	0.92	−4.681	1.816	0.92	21.9%	5.0%
Trait	$\boldsymbol{\varphi}_1$	$\boldsymbol{\varphi}_2$	$\boldsymbol{\varphi}_1^*$	$\boldsymbol{\varphi}_2^*$	*(rescaled)*	$\mathbf{t}_1$	$\mathbf{t}_2$	*(resc.)*	*(100%)*	*(100%)*
PH	−31.24	18.68	−7.899	4.724	0.98	−8.578	3.366	0.99	73.6%	17.3%
SY	−13.84	17.22	−2.049	2.549	0.38	−0.892	2.693	0.72	0.8%	11.1%
SW	−9.40	−28.19	−1.856	−5.570	0.71	−2.947	−6.002	0.96	8.7%	54.9%
HC	1.70	2.95	1.218	2.120	0.29	4.106	3.310	0.84	16.9%	16.7%

Note. Column [a], contributions of strains and traits to canonical term 1; column [b], contributions of strains and traits to canonical term 2.

Table III gives, for each of the three representations discussed in Section 3, the coordinates for the two first canonical terms.

As a starting point for a short analysis of these results, let us first look at the contributions to T_0^2 on the right part of Table II. The strains most responsible for the rejection of the hypothesis of no strain effects are strains 1, 10, 7, 6, 8, and 4, which contribute to the test statistic T_0^2 for 16.4, 15.9, 14.6, 12.5, 11.7, and 11.1%, respectively. For the traits the contribution of PH is largely predominant with 45.6% whereas that of SY is negligible. As for the relations of strains with variables one should distinguish whether they are positive or negative by considering the sign of the coordinates. Retaining contributions above 4.5% on an average of 2.5%, the most contributive relations are in decreasing order: (a) strain 10 with PH, (b) strain 1 with HC, and (c) strain 6 with SW on the positive side; (d) strain 7 with PH, (e) strain 1 with SW, (f) strain 4 with PH, (g) strain 6 with PH, (h) strain 8 with HC, and (i) strain 5 with SW on the negative side.

We expect these features to be recovered from a biplot since 87% of T_0^2 is accounted for by the first two canonical terms. In Fig. 1, because the experiment is intended to compare strains, the strains are represented by their raw vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, the lengths of which are thus directly meaningful in terms of contributions to T_0^2, whereas the vectors for variables (traits) $\mathbf{t}_1$ and $\mathbf{t}_2$ are rescaled via the eigenvalues. For strains as for traits the cosines due to the projection on the two dimensions provide indications for careful interpretation (see Table III). Here strain 4 (cosine equal to 0.82), trait HC (cosine 0.84) and trait SY (cosine 0.72) are not accurately represented. Thus relation (f) is somewhat underestimated because of strain 4 and the relations (b) and (h) involving HC are slightly distorted.

With the help of the contributions of strains and traits to the first canonical term (column [a] of Table III) and taking into account the signs of coordinates, one sees that this term is essentially featuring a strong contrast between strains 8 and 10, positively linked to PH, and strain 7 as it is negatively linked to PH. This might receive some interpretations from the specialist (canonical term 1 is also characterized in a lesser extent by trait HC which tends to oppose PH on these strains). The second canonical term is clearly linked to SW and to its positive association with strain 6 and negative with strain 1. Of course these interpretations can be read off Fig. 1 but the relative importance of a strain or a trait for each axis is more precisely quantified by columns [a] and [b] of Table III. Moreover it should be noticed here that the usefulness of CVA should not be limited to biplots, as CVA extracts the successive linear combinations of treatments which are mostly responsible for the rejection of H_0 with respect to some combination of variables. For instance, further results would show that the

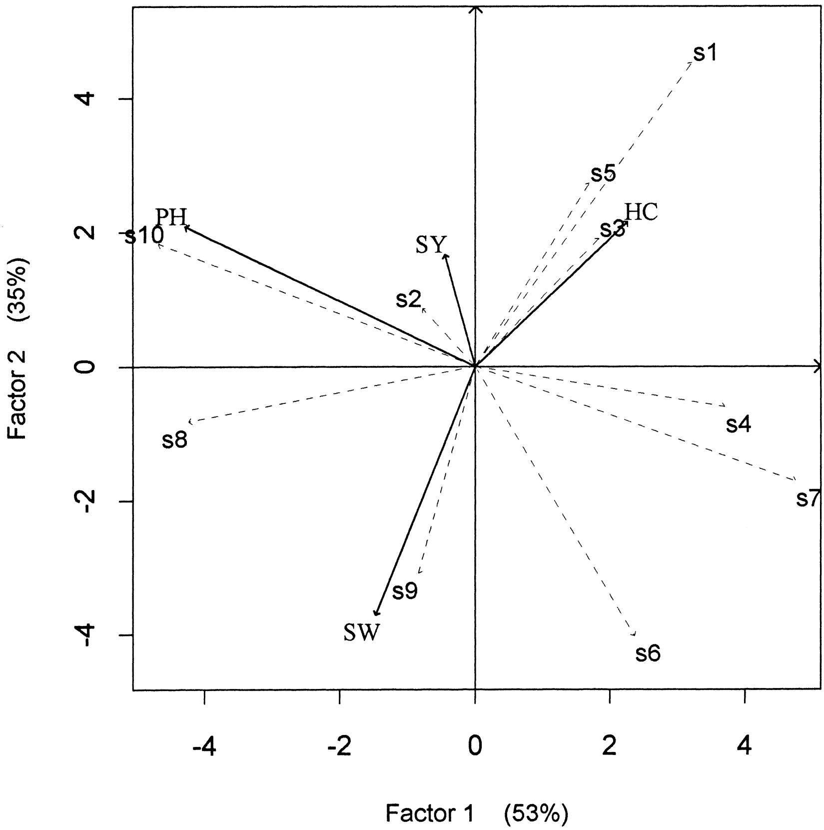

FIG. 1. Biplot for the contributions of strain means to the Lawley–Hotelling statistic in relation to the four traits, based on Table III (trait coordinates rescaled and multiplied by 5).

third canonical term is mainly specific of trait HC (contributing for 59% to it) while opposing strain 10 versus strains 4 and 8.

The left part of Table II may complement the features drawn above by examination of statistical significance in separate tests. Thus, strain 9 which was left out above because of a lack of contribution shows nonetheless a significant effect. On the contrary relation (b) does not appear to be significant (note, however, that if the relevant STP was applied instead of the above separate inferences approach, then even the effect of strain 10 in relation to PH could not be declared significant at level 0.05). A biplot could also be drawn here using coordinates from Table III, but much greater care should be taken now with respect to HC because of its poor level of representation (cosine equal to 0.29). For the representation where the lengths of the strain vectors are meaningful, the points falling out of the circle of radius $(11.21)^{1/2} = 3.35$ would correspond to

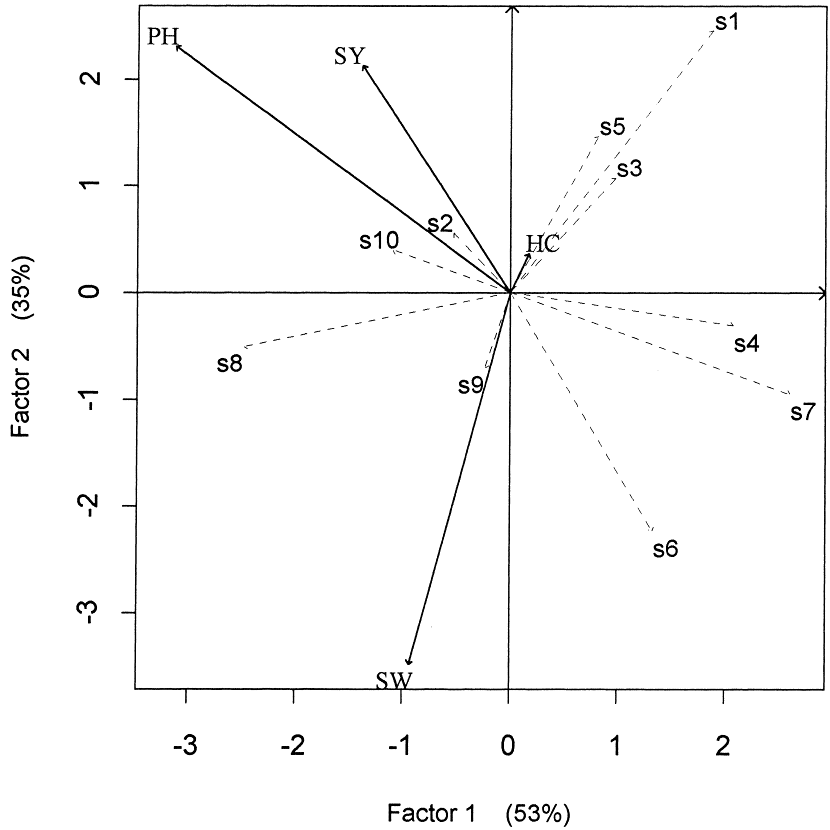

FIG. 2. Biplot for the estimates of strain means in relation to the four traits, based on Table III (trait coordinates rescaled).

strains, the effects of which are surely (separately) significant at level 0.05. Table III also gives coordinates to be used for the representation of the magnitudes of the elements of the matrix of strain effects $\hat{\boldsymbol{\Xi}}_1$. Cosines are not indicated since they are identical to those of the former biplot. These coordinates are those traditionally proposed in the literature (e.g., by Gabriel [11] or Chatfield and Collins [7]). The corresponding biplot is given in Fig. 2.

It is found that the three biplots (one for the magnitudes of strain effects as in Fig. 2, one for separate test statistics and one for elementary contributions to T_0^2 as in Fig. 1) are not fundamentally different, even though each of them has its proper logic. It is true that oftentimes the conclusions will appear to be identical; however, it may also not be necessarily so, depending in particular on the structure of the variance-covariance matrix **S**.

5. DISCUSSION AND CONCLUSIONS

It has been shown how useful the canonical decomposition of $\hat{\mathbf{\Omega}} = \mathbf{C}\hat{\mathbf{\Xi}}\mathbf{M}$ can be in analyzing and visualizing the empirical departure from any null hypothesis of the form H_0: $\mathbf{C\Xi M} = \mathbf{0}$ in MANOVA. However, to achieve such a decomposition the proper choice of a matrix norm is necessary, first to measure the total variation in $\hat{\mathbf{\Omega}}$ and second to derive the corresponding canonical approximations. Here a norm has been exhibited which is meaningful in the sense that, when applied to $\hat{\mathbf{\Omega}}$, it coincides with the Lawley–Hotelling statistic for H_0. Thus, such canonical analysis attempts to recover most of the value of the test statistic in a low dimensional approximation. Of course one can think of another norm aiming, for instance, at Roy's statistic, Wilks' likelihood ratio or Pillai's trace. In practice, however, one is constrained to finding a feasible solution to the problem of approximations by rank one matrices as shown in Subsection 3.1. When the criterion is of the form trace($\hat{\mathbf{\Omega}}'\mathbf{P}\hat{\mathbf{\Omega}}\mathbf{Q}$) for some positive definite matrices $\mathbf{P}$ and $\mathbf{Q}$, then the decomposition is straightforward via SVD. Otherwise the solution is generally not tractable, which is the case for Wilks' likelihood ratio in particular.

Now, besides the choice of the matrix norm, there are several possibilities for choosing coordinates for rows of $\hat{\mathbf{\Omega}}$, i.e., contrasts (in a wider sense of linear combinations) of treatments, or for columns of $\hat{\mathbf{\Omega}}$, i.e., combinations of variables, in two-dimensional biplots. In each case these contrasts and combinations are represented by vectors such that the scalar product of the vector representing a contrast with that representing a combination of variables is, up to the approximation in two dimensions, meaningful in some sense for the association between these elements. Also the lengths of the vectors may have a specific interpretation. The most straightforward plot simply aims at depicting each element of the departure matrix $\hat{\mathbf{\Omega}}$, i.e., the magnitude of the effect of a contrast for some combination of the variables. A second plot considered in Subsection 3.2 is related to the separate inferences on component hypotheses. It is then primarily a descriptive tool to be used to explore which treatments or constrasts of them and which variables or combinations of them must be picked up to proceed, in a second step, to a confirmatory analysis by testing the appropriate parametric functions simultaneously.

As for the third plot proposed here it rests on a standardized form of the matrix $\hat{\mathbf{\Omega}}$ whose squared elements add up to the T_0^2-statistic. This is particularly convenient in order to interpret the lengths of the vectors as contributions of either treatment contrasts or variate combinations to the test statistic. However these contributions are due to the sole elements of a contrast, or respectively of a combination of variables, in the strict sense, if and only if the $\mathbf{P}$ and $\mathbf{Q}$ matrices are diagonal.

In the analysis of a biplot the degree of approximation induced by the projection must be taken into account for each vector in order to prevent from abusive conclusions. Also, the approach proposed here allows to possibly interpret each canonical term (i.e., each axis in a biplot) by looking at the relative contributions of either treatment contrasts or combinations of variates to it, thus enriching the search for causes of rejection of the hypothesis.

Another interesting aspect is the number of canonical terms to be retained in the analysis. For this problem a testing procedure has been proposed by Caliński and Lejeune [4] which allows to guard against overestimating the number of significant terms.

The example in Section 4 is an application to the classical test for main effects. However, canonical analysis can be applied to other kinds of data matrix, like a matrix of interactions, a matrix of residuals, matrices indicating the difference between two models, departure from specified values, etc. This opens a wide scope of investigations beyond the traditional use of canonical analysis. In fact biplots are frequently used in univariate analysis of variance to represent interactions between two factors (see, e.g., Kempton [15]). These biplots, however, are often derived from the usual principal component analysis, thus overlooking the appropriate norm and ignoring links with test statistics.

Finally, note that no assumption is made on the structure of the design matrix **X**. Consequently, the canonical approach can be applied to any linear model and, in particular, to multivariate analysis of covariance where **X** entails some quantitative regressors.

ACKNOWLEDGMENTS

The resarch was initiated when the first author visited the French Ecole Nationale de la Statistique et de l'Analyse de l'Information, Malakoff, supported by a grant of CREST. The authors thank the editors and a referee for their helpful comments and suggestions. The technical help of Dr. Stanisław Czajka and Mr. Jarosław Lira has also been very much appreciated.

REFERENCES

1. T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," 2nd ed., Wiley, New York, 1984.
2. T. Caliński, S. Czajka, and Z. Kaczmarek, A model for the analysis of a series of experiments repeated at several places over a period of years, *Biuletyn Oceny Odmian Cultivar Testing Bull.* **12**, Nos. 17–18 (1987), 7–71.

3. T. Caliński, S. Czajka, and Z. Kaczmarek, A model for the analysis of a series of variety trials repeated at places subject to grouping, *Biuletyn Oceny Odmian Cultivar Testing Bull.* **21–22** (1989), 27–64.
4. T. Caliński and M. Lejeune, Dimensionality in MANOVA tested by a closed testing procedure, *J. Multivariate Anal.* **65** (1998), 181–194.
5. A. Camussi, E. Ottaviano, T. Caliński, and Z. Kaczmarek, Genetic distances based on quantitative traits, *Genetics* **111** (1985), 945–962.
6. B. Ceranka, H. Chudzik, S. Czajka, and Z. Kaczmarek, Multivariate analysis of variance for the incomplete block design, *Roczniki Akad. Rolniczej Poznan. Algorytmy Biometr. Statyst.* **5** (1976), 39–58. [In Polish]
7. C. Chatfield and A. J. Collins, "Introduction to Multivariate Analysis," Chapman & Hall, London, 1980.
8. C. Eckart and G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* **1** (1936), 211–218.
9. K. R. Gabriel, Simultaneous test procedures in multivariate analysis of variance, *Biometrika* **55** (1968), 489–504.
10. K. R. Gabriel, The biplot graphic display of matrices with application to principal component analysis, *Biometrika* **58** (1971), 453–467.
11. K. R. Gabriel, Analysis of meteorological data by means of canonical decomposition and biplots, *J. Appl. Meteorol.* **11** (1972), 1071–1077.
12. K. R. Gabriel, Biplot display of multivariate matrices for inspection of data and diagnosis, *in* "Interpreting Multivariate Data" (V. Barnett, Ed.), pp. 147–173, Wiley, London, 1981.
13. R. Gnanadesikan, "Methods for Statistical Data Analysis of Multivariate Observations," Wiley, New York, 1977.
14. J. C. Gower and D. J. Hand, "Biplots," Chapman & Hall, London, 1996.
15. R. A. Kempton, The use of biplots in interpreting variety by environment interactions, *J. Agricultural Sci. Cambridge* **103** (1984), 123–135.
16. M. Lejeune, A generic look at factor analysis, *in* "Proceedings of the International Conference on Linear Statistical Inference LINSTAT'93" (T. Caliński and R. Kala, Eds.), pp. 285–291, Kluwer Academic, Dordrecht, 1994.
17. K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate Analysis," Academic Press, London, 1979.
18. J. J. McKeon, F approximations to the distribution of Hotelling's T_0^2, *Biometrika* **61** (1974), 381–383.
19. D. F. Morrison, "Multivariate Statistical Methods," 2nd ed., McGraw–Hill, New York, 1976.
20. S. C. Pearce, "The Agricultural Field Experiment," Wiley, Chichester/New York, 1983.
21. C. R. Rao, "Advanced Statistical Methods in Biometric Research," Wiley, New York, 1952.
22. C. R. Rao, "Linear Statistical Inference and Its Applications," 2nd ed., Wiley, New York, 1973.
23. H. L. Seal, "Multivariate Statistical Analysis for Biologists," Methuen, London, 1964.
24. G. A. F. Seber, "Multivariate Observations," Wiley, New York, 1984.