# Aspects of Gifi

Jan de Leeuw

First created July 03, 2021. Last update July 11, 2021

**Abstract**

Brouhaha

## Contents

aspect (NOUN)
*particular part or feature of something*

# 1 Intro

The *Gifi System* for descriptive multivariate analysis has a complicated history.

## 1.1 Phase One: Starters

De Leeuw (1973)

## 1.2 Phase Two: ALSOS

Young, De Leeuw, and Takane (1976)

## 1.3 Phase Three: Meet-Loss

Gifi (1990), (**michailidis_deleeuw_R_96b?**)

(within this Van der Burg and De Leeuw (1983), Van der Burg and De Leeuw (1988), Van der Burg and De Leeuw (1990))

## 1.4 Phase Four: Correlational Aspects

De Leeuw (1988a), De Leeuw (1988b)

## 1.5 Phase Five: Meet-Loss as an Aspect

De Leeuw (2004)

## 1.6 Phase Six: Meet-Loss in R

De Leeuw and Mair (2009), De Leeuw (2009)

## 1.7 Phase Seven: Gifi 2021

De Leeuw (2019), De Leeuw (2021)

# 2 Canonical Analysis

Suppose $X$ and $Y$ are $n \times r$ and $n \times s$ matrices of real numbers, with $X$ containing measurements of $n$ objects on a first set of $r$ variables, and with $Y$ measurements of the same $n$ objects on a second set of $s$ variables. Both $X$ and $Y$ are supposed to be column-centered and of full column rank. Without loss of generality we assume diag $X'X = I$ and diag $Y'Y = I$, so that $X'X$, $Y'Y$, and $X'Y$ are correlation matrices.

In canonical analysis we define the fit function(or goodness-of-fit measure) in $p$ dimensions, where $p \leq \min(r, s)$, as

$$\rho_p^\star(X, Y) := \frac{1}{p}\left\{\max_{A'X'XA=I}\ \max_{B'Y'YB=I} \text{tr}\ A'X'YB\right\} \tag{1}$$

Here matrix $A$ is $r \times p$ and matrix $B$ is $s \times p$. It is clear from this formulation that $\rho_p(X, Y) = \rho_p(XS, YT)$ for all non-singular $S$ and $T$, specifically for non-singular diagonal $S$ and $T$. Thus we assume, without loss of generality, that diag $X'X = I$ and diag $Y'Y = I$, so that $X'X$, $Y'Y$, and $X'Y$ are correlation matrices. The invariance under right multiplication shows that $\rho_p(X, Y)$ is really a characteristic of the column-spaces of $X$ and $Y$, and is independent of the choice of bases for these two spaces.

The stationary equations are

$$
\begin{aligned}
X'YB &= X'XA\Phi,\\
Y'XA &= Y'YB\Psi,\\
A'X'XA &= I,\\
B'Y'YB &= I,
\end{aligned}
\tag{2}
$$

where $\Phi$ and $\Psi$ are two symmetric matrices of Lagrange multipliers. It follows directly from these equations that $\Phi = \Psi$.

Define $\tilde{A} = (X'X)^{\frac{1}{2}}A$ and $\tilde{B} = (Y'Y)^{\frac{1}{2}}B$. Any matrix square root will do, so we can use the Cholesky factor, or the eigen factorization, or the symmetric square root. Also define $\tilde{X} = X(X'X)^{-\frac{1}{2}}$ and $\tilde{Y} = Y(Y'Y)^{-\frac{1}{2}}$. Then the stationary equations (2) become

$$
\begin{aligned}
\tilde{X}'\tilde{Y}\tilde{B} &= \tilde{A}\Phi,\\
\tilde{Y}'\tilde{X}\tilde{A} &= \tilde{B}\Psi,\\
\tilde{A}'\tilde{A} &= I,\\
\tilde{B}'\tilde{B} &= I.
\end{aligned}
\tag{3}
$$

It follows that $\Phi = \Psi = M\mathrm{P}M'$, where $M$ is an arbitrary rotation matrix with $M'M = MM' = I$, and $\mathrm{P}$ is a diagonal matrix with $p$ singular values of

$$
\tilde{C} := \tilde{X}'\tilde{Y} = (X'X)^{-\frac{1}{2}}X'Y(Y'Y)^{-\frac{1}{2}}. \tag{4}
$$

We always choose the singular values to be non-negative.

If the singular value decomposition is $\tilde{C} = K\mathrm{P}L'$, then the maximum in (1) is attained for $A = (X'X)^{-\frac{1}{2}}K_pM$ and $B = (Y'Y)^{-\frac{1}{2}}L_pM$, where $K_p$ and $L_p$ are singular vectors corresponding with the $p$ largest singular values, $\rho_1(X,Y) \geq \cdots \geq \rho_p(X,Y)$ and $M$ is the arbitrary rotation matrix. At the maximum

$$
\rho_p^\star(X,Y) = \frac{1}{p}\sum_{s=1}^{p}\rho_s(X,Y). \tag{5}
$$

The $\rho_s(X,Y)$ are the *canonical correlations*. Becase they are proper correlations, we have $0 \leq \rho_s(X,Y) \leq 1$. Thus $\rho_p^\star(X,Y)$ is the average of the $p$ largest canonical correlations. We also define the *canonical weights* as the maximizers $A$ and $B$, the *canonical variables* as $XA$ and $YB$. and the *canonical self-loadings* as the correlations $X'XA$ and $Y'YB$ between the original variables and the canonical varables. The *canonical cross-loadings* are $X'YB$ and $Y'XA$, but from equations (2) we see that the cross loadings are a simple rescaling of the self-loadings.

```r
x <- normy(matrix(rnorm(300), 100, 3))
y <- normy(matrix(rnorm(500), 100, 5))
h <- cancor(x, y)
r <- h$cor
a <- h$xcoef
b <- h$ycoef[,1:3]
```

```r
mprint(r)
```

```
## [1]  +0.364068  +0.245573  +0.106719
```

```r
mprint(a)
```

```
##       [,1]       [,2]       [,3]
## [1,]  +0.417988  -0.074202  -0.911973
## [2,]  +0.337779  -0.893083  +0.338824
## [3,]  -0.767089  -0.559736  -0.339876
```

```r
mprint(b)
```

```
##       [,1]       [,2]       [,3]
## [1,]  +0.912158  +0.253359  +0.151970
## [2,]  +0.033129  +0.675454  -0.780475
## [3,]  +0.074446  -0.612854  -0.175790
## [4,]  +0.079878  +0.650497  +0.251760
## [5,]  +0.404800  -0.162086  -0.722730
```

```r
mprint(crossprod(x %*% a))
```

```
##       [,1]       [,2]       [,3]
## [1,]  +1.000000  -0.000000  -0.000000
## [2,]  -0.000000  +1.000000  -0.000000
## [3,]  -0.000000  -0.000000  +1.000000
```

```r
mprint(crossprod(y %*% b))
```

```
##       [,1]       [,2]       [,3]
## [1,]  +1.000000  -0.000000  -0.000000
## [2,]  -0.000000  +1.000000  -0.000000
## [3,]  -0.000000  -0.000000  +1.000000
```

```
mprint(crossprod(x %*% a, y %*% b))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.364068  -0.000000  -0.000000
## [2,]  +0.000000  +0.245573  -0.000000
## [3,]  -0.000000  +0.000000  +0.106719
```

```
mprint(crossprod(x, x %*% a))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.486329  -0.143087  -0.861981
## [2,]  +0.478495  -0.829921  +0.286835
## [3,]  -0.827929  -0.443415  -0.343390
```

```
mprint(crossprod(y, y %*% b))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.916635  +0.073821  +0.208122
## [2,]  -0.201484  +0.487258  -0.665195
## [3,]  +0.081807  -0.467680  -0.138525
## [4,]  +0.029081  +0.507080  +0.364293
## [5,]  +0.400557  -0.220273  -0.460943
```

```
mprint(crossprod(x, x %*% a))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.486329  -0.143087  -0.861981
## [2,]  +0.478495  -0.829921  +0.286835
## [3,]  -0.827929  -0.443415  -0.343390
```

```
mprint(crossprod(x, y %*% b))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.177057  -0.035138  -0.091990
## [2,]  +0.174205  -0.203806  +0.030611
## [3,]  -0.301423  -0.108891  -0.036646
```

```
mprint(crossprod(y, x %*% a))
```

```
##          [,1]      [,2]      [,3]
## [1,]  +0.333717  +0.018129  +0.022211
## [2,]  -0.073354  +0.119658  -0.070989
## [3,]  +0.029783  -0.114850  -0.014783
## [4,]  +0.010587  +0.124525  +0.038877
## [5,]  +0.145830  -0.054093  -0.049191
```

# 3 Gifi Meet-Loss

In Gifi (1990) we define *meet-loss* for two sets as the least squares loss function (or badness-of-fit measure)

$$\sigma_p^\star(X,Y) := \frac{1}{2}\frac{1}{p}\left\{\min_{Z'Z=I}\min_A\min_B\{\mathrm{SSQ}(Z-XA)+\mathrm{SSQ}(Z-YB)\}\right\}, \tag{6}$$

where we use SSQ as shorthand for sum of squares. The name meet-loss derives from the fact that $\sigma_p^\star(X,Y)=0$ if and only if the intersection (or meet) of the column spaces of $X$ and $Y$ has dimension $d \geq p$.

In equation (6) the matrices $X$, $Y$, $A$, and $B$ have the same definitions and dimensions as before. The new component is the *target* $Z$, an orthonormal $n \times p$ matrix. Note there are no constraints on the weights $A$ and $B$ in this formulation.

The minimum over $A$ and $B$ for fixed $Z$ is attained at

$$A = (X'X)^{-1}X'Z, \tag{7}$$
$$B = (Y'Y)^{-1}Y'Z. \tag{8}$$

Thus

$$\sigma_p^\star(X,Y) = 1 - \frac{1}{p}\max_{Z'Z=I}\mathrm{tr}\ Z'\overline{P}Z, \tag{9}$$

where

$$\overline{P} := \frac{1}{2}\left\{X(X'X)^{-1}X' + Y(Y'Y)^{-1}Y'\right\} \tag{10}$$

is the *average projector*.

If $\overline{P} = V\Sigma V'$ is the eigen decomposition of $\overline{P}$, then the optimum in (9) is attained for $Z = V_pM$, where $V_p$ are the eigenvectors of $\overline{P}$ corresponding with the $p$ largest eigenvalues $\sigma_1(X,Y) \geq \cdots \geq \sigma_p(X,Y)$ and $M$ is again an arbitrary rotation matrix. Note that $0 \leq \sigma_s(X,Y) \leq 1$ for all $s$. Also

$$\sigma_p^\star(X,Y) = 1 - \frac{1}{p}\sum_{s=1}^{p}\sigma_s(X,Y). \tag{11}$$

Thus meet-loss is one minus the average of the $p$ largest eigenvalues of the average projector. We also see that $\sigma_p^\star(X,Y)=1$ if and only if the column spaces of $X$ and $Y$ are orthogonal.

# 4 Relationships

Consider the partioned matrix

$$U := \left[X(X'X)^{-\frac{1}{2}} \quad | \quad Y(Y'Y)^{-\frac{1}{2}}\right] \tag{12}$$

then

$$UU' = 2\overline{P},$$

and

$$U'U = \begin{bmatrix} I & \tilde{X}'\tilde{Y} \\ \tilde{Y}'\tilde{X} & I \end{bmatrix}.$$

```
u <- cbind(orthy(x), orthy(y))
mprint(eigen(crossprod(u))$values)
```

```
## [1]  +1.364068  +1.245573  +1.106719  +1.000000  +1.000000  +0.893281  +0.754427
## [8]  +0.635932
```

```
mprint(eigen(tcrossprod(u))$values[1:8])
```

```
## [1]  +1.364068  +1.245573  +1.106719  +1.000000  +1.000000  +0.893281  +0.754427
## [8]  +0.635932
```

For any matrix $U$ the non-zero eigenvalues of $U'U$ are the same as the non-zero eigenvalues of $UU'$. The non-zero eigenvalues of $UU'$ are $2\sigma_s(X,Y)$ and those of $U'U$ are $1+\rho_s(X,Y)$ and $1-\rho_s(X,Y)$. Thus $1+\rho_s(X,Y) = 2\sigma_s(X,Y)$ and

$$\sigma_p^\star(X,Y) = 1 - \frac{1}{p}\sum_{s=1}^{p}(\rho_s - 1)/2 =$$

# 5 Optimal Scaling

# 6 Aspect Loss

The *aspect* approach to optimal scaling is due to De Leeuw (1988b), with further elaborations in De Leeuw (2004). See also Mair and De Leeuw (2010) for the *aspect* package, which provides a partial implementation in R.

In the aspect approach we minimize a concave function $\phi$ of the correlation matrix $R$ of the variables in the data.

Suppose that the standardized variables are collected in a matrix $Q$, so that $R = Q'Q$.

Because $\phi$ is concave on the space of correlation matrices we have for any two correlation matrices $R$ and $\tilde{R}$

$$\phi(R) \leq \phi(\tilde{R}) + \text{tr } G(\tilde{R})(R - \tilde{R}),$$

where $G(\tilde{R})$ is the matrix of partial derivatives of $\phi$ at $\tilde{R}$ (or, more generally, any subgradient of $\phi$ at $\tilde{R}$). Note $G$ is both symmetric and hollow (??).

If $\tilde{R}$ is our previous best solution, then we find a better solution by minimizing

$$\text{tr } G(\tilde{R})R = \text{tr } QG(\tilde{R})Q'$$

over $Q \in \mathcal{K}$.

It is shown in De Leeuw (1988b) that the squared multiple correlation of one variable with the others, the log-determinant of the $R$, the negative of the sum of the $r$ largest eigenvalues of $R$, the sum of the correlation coefficients, the negative of any norm of the correlation matrix, and any function of the form

$$\phi(R) := \min_{\Gamma \in \mathcal{R}} \log \Gamma + \text{tr } \Gamma^{-1} R$$

are concave in $R$. Thus the aspect approach covers the optimal scaling versions of multiple regression, path analysis, principal component analysis, and multinormal maximum likelihood.

In section 8 of De Leeuw (1988b) on limitations it was noticed that the canonical correlations are not concave in the joint correlation matrix of $X$ and $Y$, so aspect theory does not apply. This implied that there was no firm theoretical basis for the alternative apprach to canonical analysis discussed in Tijssen and De Leeuw (1989). But then, in De Leeuw (2004), it was discovered that if we use the joint correlation matrix of $X, Y$ and $Z$ from Gifi's meet-loss we are back in the realm concavity, and thus we can use the MM aspect algorithm.

Define

$$Q = \begin{bmatrix} Z & X & Y \end{bmatrix}$$

$$G = \begin{bmatrix} I & I \\ -A & 0 \\ 0 & -B \end{bmatrix}$$

then

$$\sigma_p^\star(X, Y) = \min_{Z'Z=I} \min_{G} \text{tr } G'RG$$

# 7 Partials

```
z<-normy(matrix(rnorm(400), 100, 4))
h <- cancor(cbind(x,z), cbind(y,z))
zx<-lsfit(z,x,intercept=FALSE)$residuals
zy<-lsfit(z,y,intercept=FALSE)$residuals
g <- cancor(zx,zy)
print(h$cor)
```

```
## [1] 1.00000000000 1.00000000000 1.00000000000 1.00000000000 0.37661694169
## [6] 0.24926259554 0.07652584154
```

```
print(g$cor)
```

```
## [1] 0.37661694169 0.24926259554 0.07652584154
```

```
xvar1<-cbind(x,z)%*%h$xcoef[,-(1:3)]
xvar2<-zx%*%g$xcoef
```

# 8 Partals

$$SSQ(Y-(X-ZA)B)+\alpha\ SSQ(X-ZA)$$

# 9 Canals

$$\tilde{\sigma}_p^\star(X,Y)=\min_{A'X'XA=I}\min_B \mathrm{SSQ}(XA-YB)=\min_{B'Y'Y'B=I}\min_A \mathrm{SSQ}(XA-YB) \tag{13}$$

# 10 Criminals

# 11 Morals

# 12 Redundals

Van der Burg and De Leeuw (1990)

$$SSQ(Z-XB)+SSQ(Y-ZA)$$
$$\mathrm{tr}\ Z'Q_XZ+\mathrm{tr}\ Y'(I-ZZ')Y=p+\mathrm{tr}\ Y'Y-\mathrm{tr}\ Z'P_XZ-\mathrm{tr}\ Z'YY'Z$$

$$\begin{bmatrix} X(X'X)^{-\frac{1}{2}} & \mid & Y \end{bmatrix}$$

$$\begin{bmatrix} I & \tilde{X}'Y \\ Y'\tilde{X} & Y'Y \end{bmatrix}$$

$$A + \tilde{X}'YB = AM$$

$$Y'\tilde{X}A + Y'YB = BM$$

$$SSQ(Y - XAB') = SSQ(Y - P_X Y - X(AB' - D)) = \text{tr } Y'Q_X Y + \text{tr } (D - AB')X'X(D - AB')$$

$D = (X'X)^{-1}X'Y$

$$\text{tr } (D - AB')X'X(D - AB') = K - 2\text{tr } A'X'YB + \text{tr } A'X'XAB'B$$

$$X'YB = X'XAM$$

$$Y'XA = B$$

$$X'YY'XA = X'XAM$$

$$\tilde{X}'YY'\tilde{X}\tilde{A} = \tilde{A}M$$

# 13 Totals

Total Least Squares version

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

$$Y + E = (X + D)B$$

$$\text{SSQ}(E) + \alpha \text{ SSQ}(D)$$

$$\text{SSQ}(Y - (X + Z)B) + \alpha \text{ SSQ}(Z)$$

# 14 Dynamals

# 15 Multiple Sets

$$H_k = W_k^+ \sum_{j \in \mathcal{J}_k} W_j(Z - y_j a_j^T) = Z - W_k^+ \sum_{j \in \mathcal{J}_k} W_j y_j a_j^T.$$

Let $v_j = W_k^+ W_j y_j$.

$$H_k = Z - V_k A_k$$

$$H_k'W_kH_k = Z^TW_kZ - 2Z^TW_kV_kA_k + A_k^TV_k^TW_kV_kA_k$$
$$A_k = (V_k^TW_kV_k)^{-1}V_k^TW_kZ$$
$$Z^T(W_k - W_kV_k(V_k^TW_kV_k)^{-1}V_k^TW_k)Z$$
$$Z^T(W_\bullet - \sum_{k=1}^K W_kV_k(V_k^TW_kV_k)^{-1}V_k^TW_k)Z$$

$Z^TW_\bullet Z = I$ $\tilde{Z} = W_\bullet^{\frac{1}{2}}Z$

$$\tilde{Z}^T(I - W_\bullet^{-\frac{1}{2}}\sum_{k=1}^K W_kV_k(V_k^TW_kV_k)^{-1}V_k^TW_kW_\bullet^{-\frac{1}{2}})\tilde{Z}$$
$$U_k = W_\bullet^{-\frac{1}{2}}W_kV_k(V_k^TW_kV_k)^{-\frac{1}{2}}$$

# References


De Leeuw, J. 1973. "Canonical Analysis of Categorical Data." PhD thesis, University of Leiden, The Netherlands.

———. 1988a. "Multivariate Analysis with Linearizable Regressions." *Psychometrika* 53: 437–54. http://deleeuwpdx.net/janspubs/1988/articles/deleeuw_A_88a.pdf.

———. 1988b. "Multivariate Analysis with Optimal Scaling." In *Proceedings of the International Conference on Advances in Multivariate Statistical Analysis*, edited by S. Das Gupta and J. K. Ghosh, 127–60. Calcutta, India: Indian Statistical Institute. http://deleeuwpdx.net/janspubs/1988/chapters/deleeuw_C_88b.pdf.

———. 2004. "Least Squares Optimal Scaling of Partially Observed Linear Systems." In *Recent Developments in Structural Equation Models*, edited by K. van Montfort, J. Oud, and A. Satorra. Dordrecht, Netherlands: Kluwer Academic Publishers. http://deleeuwpdx.net/janspubs/2004/chapters/deleeuw_C_04a.pdf.

———. 2009. "Regression, Discriminant Analysis, and Canonical Analysis with homals." Preprint Series 562. Los Angeles, CA: UCLA Department of Statistics. http://deleeuwpdx.net/janspubs/2009/reports/deleeuw_R_09c.pdf.

———. 2019. "Gifi Update Notes." 2019. http://deleeuwpdx.net/pubfolders/gifi/Gifi_New/gifi.pdf.

———. 2021. *Multivariate Analysis with Optimal Scaling*. Bookdown. https://github.com/deleeuw/gifi/blob/main/_book/_main.pdf.

De Leeuw, J., and P. Mair. 2009. "Homogeneity Analysis in r: The Package Homals." *Journal of Statistical Software* 31 (4): 1–21. http://deleeuwpdx.net/janspubs/2009/articles/deleeuw_mair_A_09a.pdf.

Gifi, A. 1990. *Nonlinear Multivariate Analysis*. New York, N.Y.: Wiley.

Mair, P., and J. De Leeuw. 2010. “A General Framework for Multivariate Analysis with Optimal Scaling: The r Package Aspect.” *Journal of Statistical Software* 32 (9): 1–23. http://deleeuwpdx.net/janspubs/2010/articles/mair_deleeuw_A_10.pdf.

Tijssen, R. J. W., and J. De Leeuw. 1989. “Multi-Set Nonlinear Canonical Analysis via the Burt Matrix.” In *Multiway Data Analysis*, edited by R. Coppi and S. Bolasko. Amsterdam; New York: North Holland Publishing Company. http://deleeuwpdx.net/janspubs/1989/chapters/tijssen_deleeuw_C_89.pdf.

Van der Burg, E., and J. De Leeuw. 1983. “Non-Linear Canonical Correlation.” *British Journal of Mathematical and Statistical Psychology* 36: 54–80. http://deleeuwpdx.net/janspubs/1983/articles/vanderburg_deleeuw_A_83.pdf.

———. 1988. “Use of the Multinomial Jackknife and Bootstrap in Generalized Nonlinear Canonical Correlation Analysis.” *Applied Stochastic Models and Data Analysis* 4 (159–172). http://deleeuwpdx.net/janspubs/1988/articles/vanderburg_deleeuw_A_88.pdf.

———. 1990. “Nonlinear Redundancy Analysis.” *British Journal of Mathematical and Statistical Psychology* 43: 217–30. http://deleeuwpdx.net/janspubs/1990/articles/vanderburg_deleeuw_A_90.pdf.

Young, F. W., J. De Leeuw, and Y. Takane. 1976. “Regression with Qualitative and Quantitative Data: An Alternating Least Squares Approach with Optimal Scaling Features.” *Psychometrika* 41: 505–29. http://deleeuwpdx.net/janspubs/1976/articles/young_deleeuw_takane_A_76.pdf.