LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK

Leo Knüsel

The General Linear Model and the Generalized Singular Value Decomposition; Some Examples

Technical Report Number 049, 2009
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

The General Linear Model and the Generalized Singular Value Decomposition; Some Examples

Leo Knüsel

Department of Statistics
University of Munich
80539 Munich
knuesel@stat.uni-muenchen.de

Abstract: The general linear model $y = X\beta + \varepsilon$ with correlated error variables can be transformed by means of the generalized singular value decomposition to a very simple model (canonical form) where the least squares solution is obvious. The method works also if X and the covariance matrix of the error variables do not have full rank or are nearly rank deficient (rank-k approximation). By backtransformation one obtains the solution for the original model. In this paper we demonstrate the method with some examples.

Keywords
General linear model, canonical form, generalized singular value decomposition, CS-decomposition of an orthogonal matrix, multicollinearity, rank-k approximation.

Introduction and summary

The general linear model is given by

$$y = X\beta + \varepsilon,\ \mathrm{E}(\varepsilon) = 0,\ \mathrm{var}(\varepsilon) = \sigma^2 W,\ X = (n \times p),\ W = (n \times n),\ n > p\ .$$

$\sigma^2 W$ is the covariance matrix of ε and we assume that the matrix W is given (symmetric and positive semidefinite) while σ^2 is unknown. If $W = I_n$ we have the simple linear model with uncorrelated error variables $\varepsilon_1, \ldots, \varepsilon_n$. If $rk(W) = k$ W can be written as $W = FF^\top$ where $F = (n \times k)$. The random error ε can now be given in the form $\varepsilon = Fu$ with $u \sim (0, \sigma^2 I_k)$ i.e. with $\mathrm{E}(u) = 0$, $\mathrm{var}(u) = \sigma^2 I_k$ as $\mathrm{E}(\varepsilon) = \mathrm{E}(Fu) = 0$ and $\mathrm{var}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon^\top) = F\,\mathrm{E}(uu^\top)F^\top = \sigma^2 FF^\top = \sigma^2 W$. So the general linear model is equivalent to

$$y = X\beta + Fu,\ \text{where}\ X = (n \times p), F = (n \times k)\ \text{and}\ u \sim (0, \sigma^2 I_k)\ .$$

In Knüsel (2008, 2009) the solution of the problem by means of the simple and generalized singular value decomposition is treated and in this paper we give nine examples that deal in particular with the case of rank deficient and nearly rank deficient matrices X and W (multicollinearity, weak multicollinearity). The following table gives an overview of the examples.

Example	Title	Page
1	Simple linear model, regular case	2
2	Simple linear model with strict multicollinearity	4
3	Simple linear model with weak multicollinearity	7
4	General linear model, regular case	11
5	General linear model with strict multicollinearity	17
6	General linear model with rank deficient covariance matrix W	22
7	General linear model with rank deficient X and W	26
8	General linear model with rank deficient X, W, and $(X \mid F)$	31
9	General linear model with nearly rank deficient X, W, and $(X \mid F)$	36
10	General linear model with linear restrictions, regular case	41

The computations in the examples are done with Matlab (2008) and Maple (2006). Matlab offers a procedure *gsvd* (general singular value decomposition) that includes a subfunction *csd* (CS-decomposition), and this subfunction is used for computing the CS-decomposition of an orthogonal matrix (see Golub – Van Loan, 1996).

Example 1: Simple linear model, regular case

In this example we consider the simple linear model

(1) $y = X\beta + \varepsilon,$ where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 I_n\right)$.

Let

(2) $$X = (5 \times 3) = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 9 & 8 \\ 4 & 1 & 4 \\ 7 & 2 & 0 \\ 5 & 6 & 5 \end{pmatrix}.$$

We obtain the singular values

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} 16.560 \\ 7.612 \\ 2.795 \end{pmatrix},$$

and this means that the matrix X has full rank 3.

a) Classical solution

The classical least squares estimator of the parameter vector β is given by

(3) $$\begin{aligned} &\hat{\beta} = (X^\top X)^{-1} X^\top y \\ &\operatorname{var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1} \\ &\hat{\sigma}^2 = \frac{1}{n-p} e^\top e, \text{ where } e = y - X\hat{\beta} = (I_n - P)y \text{ with } P = X(X^\top X)^{-1} X^\top \end{aligned}$$

and we obtain

$$X^\top X = \begin{pmatrix} 100 & 63 & 58 \\ * & 126 & 112 \\ * & * & 114 \end{pmatrix}, \quad (X^\top X)^{-1} = \begin{pmatrix} 0.014\,658 & -0.005\,525 & -0.002\,030 \\ * & 0.064\,720 & -0.060\,773 \\ * & * & 0.069512 \end{pmatrix},$$

(4) $$(X^\top X)^{-1} X^\top = \begin{pmatrix} 0.026\,835 & -0.051\,302 & 0.044\,988 & 0.091\,555 & 0.029\,992 \\ -0.069\,455 & 0.090\,766 & -0.200\,474 & 0.090\,766 & 0.056\,827 \\ 0.080\,900 & 0.007\,103 & 0.209\,155 & -0.135\,754 & -0.027\,230 \end{pmatrix},$$

$$P = X(X^\top X)^{-1} X^\top = \begin{pmatrix} 0.184\,294 & 0.048\,934 & 0.361\,484 & 0.048\,934 & 0.121\,942 \\ * & 0.822\,415 & -0.086\,030 & -0.177\,585 & 0.323\,599 \\ * & * & 0.816\,101 & -0.086\,030 & 0.067\,877 \\ * & * & * & 0.822\,415 & 0.323\,599 \\ * & * & * & * & 0.354\,775 \end{pmatrix}.$$

b) Solution with singular value decomposition

The singular value decomposition of X is given by $X = UDV^\top$, where $U = (5 \times 5)$ and $V = (3 \times 3)$ are orthogonal and $D = (5 \times 3) = \operatorname{diag}(\sigma_1, \sigma_2, \sigma_3)$:

(5) $$U = \begin{pmatrix} -0.270 & 0.148 & -0.299 & -0.744 & -0.512 \\ -0.679 & -0.578 & 0.165 & 0.281 & -0.314 \\ -0.295 & 0.264 & -0.812 & 0.417 & 0.100 \\ -0.265 & 0.745 & 0.444 & 0.281 & -0.314 \\ -0.556 & 0.139 & 0.163 & -0.337 & 0.729 \end{pmatrix}, \quad V = \begin{pmatrix} -0.441 & 0.897 & -0.021 \\ -0.653 & -0.305 & 0.694 \\ -0.616 & -0.320 & -0.720 \end{pmatrix},$$

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 15.560 & 0 & 0 \\ 0 & 7.612 & 0 \\ 0 & 0 & 2.795 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

From

$$y = X\beta + \varepsilon = UDV^{\mathsf{T}}\beta + \varepsilon$$

we obtain the canonical form of the linear model (1)

(6) $\tilde{y} = D\tilde{\beta} + \tilde{\varepsilon}$, where $\tilde{y} = U^{\mathsf{T}} y,\ \tilde{\beta} = V^{\mathsf{T}}\beta,\ \tilde{\varepsilon} = U^{\mathsf{T}}\varepsilon$,

and as U is orthogonal we have $\tilde{\varepsilon} \sim (0, \sigma^2 I_n)$ i.e. the error variables $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ are again uncorrelated each with variance σ^2. The canonical model (6) explicitly written has the form

$$\begin{aligned}
&\tilde{y}_1 = \sigma_1\tilde{\beta}_1 + \tilde{\varepsilon}_1,\\
&\tilde{y}_2 = \sigma_2\tilde{\beta}_2 + \tilde{\varepsilon}_2,\\
(7)\quad &\tilde{y}_3 = \sigma_3\tilde{\beta}_3 + \tilde{\varepsilon}_3,\\
&\tilde{y}_4 = \tilde{\varepsilon}_4,\\
&\tilde{y}_5 = \tilde{\varepsilon}_5.
\end{aligned}$$

The least squares estimator of $\tilde{\beta}$ is given by $\hat{\tilde{\beta}}_i = \tilde{y}_i / \sigma_i, i = 1, \ldots, 3$ or in matrix notation

$$\begin{aligned}
&\hat{\tilde{\beta}} = D^{+}\tilde{y} \text{ where } D^{+} = (p \times n) = (3 \times 5) = \operatorname{diag}(1/\sigma_1, 1/\sigma_2, 1/\sigma_3),\\
(8)\quad &\operatorname{var}(\hat{\tilde{\beta}}) = \sigma^2 (D^{\mathsf{T}} D)^{-1},\\
&\hat{\sigma}^2 = \frac{1}{n-p}\tilde{e}^{\mathsf{T}}\tilde{e} = \tfrac{1}{2}(\tilde{u}_4^2 + \tilde{u}_5^2).
\end{aligned}$$

D^{+} is the Moore-Penrose inverse of D and we obtain

$$D^{+} = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & 0\\ 0 & 1/\sigma_2 & 0 & 0 & 0\\ 0 & 0 & 1/\sigma_3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.060\,386 & 0 & 0 & 0 & 0\\ 0 & 0.131\,366 & 0 & 0 & 0\\ 0 & 0 & 0.357\,751 & 0 & 0 \end{pmatrix},$$

(9)

$$(D^{\mathsf{T}} D)^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0\\ 0 & 1/\sigma_2^2 & 0\\ 0 & 0 & 1/\sigma_3^2 \end{pmatrix} = \begin{pmatrix} 0.003\,646 & 0 & 0\\ 0 & 0.017\,257 & 0\\ 0 & 0 & 0.127\,986 \end{pmatrix}.$$

By backtransformation we find the least squares estimators of the original parameters $\beta = V\tilde{\beta}$:

$$\begin{aligned}
&\hat{\beta} = V\hat{\tilde{\beta}} = V D^{+}\tilde{y} = X^{+} y \text{ where } X^{+} = (p \times n) = (3 \times 5) = V D^{+} U^{\mathsf{T}},\\
(10)\quad &\operatorname{var}(\hat{\tilde{\beta}}) = \sigma^2 V (D^{\mathsf{T}} D)^{-1} V^{\mathsf{T}},\\
&\hat{\sigma}^2 = \frac{1}{n-p}\tilde{e}^{\mathsf{T}}\tilde{e} = \frac{1}{n-p} e^{\mathsf{T}} e \text{ where } e = y - X\hat{\beta} = y - XX^{+}y = (I_n - XX^{+})y.
\end{aligned}$$

$X^{+} = V D^{+} U^{\mathsf{T}}$ is the Moore-Penrose inverse of *X,* and as

$$\begin{aligned}
&(X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} = X^{+}\\
(11)\quad &(X^{\mathsf{T}} X)^{-1} = V (D^{\mathsf{T}} D)^{-1} V^{\mathsf{T}}\\
&P = X (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} = X X^{+}
\end{aligned}$$

we obtain the same results as with the classical solution.

Example 2: Simple linear model with strict multicollinearity

We consider again the simple linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 I_n\right)$,

but this time

$$X = (x_1, x_2, x_3) = (5 \times 3) = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix}.$$

We have $x_3 = 2x_1 - x_2$, i.e. the third column of X is a linear combination of the first two columns, and so the matrix X has the rank $\mathrm{rk}(X) = r_X = 2$. As $X^\mathsf{T} X$ has the same rank as X the inverse $(X^\mathsf{T} X)^{-1}$ does not exist and the classical procedure breaks down.

Solution with singular value decomposition

The singular value decomposition of X is given by $X = UDV^\mathsf{T}$, where $U = (5 \times 5)$ and $V = (3 \times 3)$ are orthogonal and $D = (5 \times 3) = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$:

$$U = \begin{pmatrix} -0.276 & -0.118 & 0.790 & 0.223 & -0.486 \\ 0.227 & -0.878 & -0.238 & -0.081 & -0.340 \\ -0.432 & 0.011 & 0.027 & -0.893 & -0.123 \\ -0.746 & -0.006 & -0.503 & 0.376 & -0.219 \\ -0.358 & -0.464 & 0.256 & 0.070 & 0.765 \end{pmatrix}, \quad V = \begin{pmatrix} -0.497 & -0.294 & 0.816 \\ -0.138 & -0.902 & -0.408 \\ -0.857 & 0.315 & -0.408 \end{pmatrix},$$

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 18.802 & 0 & 0 \\ 0 & 12.103 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that $\sigma_3 = 0$ which means that $rk(X) = 2$. From

$$y = X\beta + \varepsilon = UDV^\mathsf{T}\beta + \varepsilon$$

we obtain the canonical form of the linear model (1)

(2) $\tilde{y} = D\tilde{\beta} + \tilde{\varepsilon}$, where $\tilde{y} = U^\mathsf{T} y$, $\tilde{\beta} = V^\mathsf{T}\beta$, $\tilde{\varepsilon} = U^\mathsf{T}\varepsilon$,

and as U is orthogonal we have $\tilde{\varepsilon} \sim (0, \sigma^2 I_n)$ i.e. the error variables $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ are again uncorrelated each with variance σ^2. The canonical model (2) explicitly written has the form

$$\begin{aligned} \tilde{y}_1 &= \sigma_1\tilde{\beta}_1 + \tilde{\varepsilon}_1, \\ \tilde{y}_2 &= \sigma_2\tilde{\beta}_2 + \tilde{\varepsilon}_2, \\ \tilde{y}_3 &= \tilde{\varepsilon}_3, \\ \tilde{y}_4 &= \tilde{\varepsilon}_4, \\ \tilde{y}_5 &= \tilde{\varepsilon}_5. \end{aligned} \tag{3}$$

The least squares estimator of $\tilde{\beta}$ is given by $\hat{\tilde{\beta}}_i = \tilde{y}_i / \sigma_i$, $i = 1,2$; the parameter $\tilde{\beta}_3$ is not to be found in the canonical model as $\sigma_3 = 0$, it can have arbitrary values, and it cannot be estimated. From

$X = UDV^\mathsf{T}$ we derive $XV = UD$,

and as the third column of UD is zero the same is true for the third column of XV. Let $V = (v_1, v_2, v_3)$, then $Xv_3 = 0.816x_1 - 0.408x_2 - 0.408x_3 = 0$, and this is equivalent to $2x_1 - x_2 - x_3 = 0$ as $(2, -1, -1)/\sqrt{6} = (0.816, -0.408, -0.408)$. Now there are three ways how to proceed.

A) Eliminate one of the original parameters

As $2x_1 - x_2 - x_3 = 0$ we can eliminate one of the columns x_1, x_2, x_3. We have e.g.

$$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 = \beta_1 x_1 + \beta_2 x_2 + \beta_3(2x_1 - x_2) = (\beta_1 + 2\beta_3)x_1 + (\beta_2 - \beta_3)x_2 .$$

We can introduce the new parameters

$$\bar{\beta}_1 = \beta_1 + 2\beta_3$$
$$\bar{\beta}_2 = \beta_2 - \beta_3$$

and the new matrix $\bar{X} = (x_1, x_2)$. Then model (1) is equivalent to

$$y = \bar{X}\bar{\beta} + \varepsilon, \text{ where } \bar{X} = (n \times r_x) = (5 \times 2) \text{ and } \varepsilon \sim \left(0, \sigma^2 I_n\right).$$

As $\bar{X}$ has full rank two, we can now apply the methods of example 1, and we find the covariance matrix

$$\operatorname{var}(\hat{\bar{\beta}}) = \begin{pmatrix} 0.014\,599 & -0.007\,299 \\ * & 0.011\,586 \end{pmatrix}.$$

B) Introduce the canonical parameters

We have

$$XV = (Xv_1, Xv_2, Xv_3) \text{ and } V^{\mathsf{T}}\beta = \begin{pmatrix} v_1^{\mathsf{T}}\beta \\ v_2^{\mathsf{T}}\beta \\ v_3^{\mathsf{T}}\beta \end{pmatrix} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \end{pmatrix} = \tilde{\beta}$$

and so

$$X\beta = XVV^{\mathsf{T}}\beta = XV\tilde{\beta} = \tilde{\beta}_1 Xv_1 + \tilde{\beta}_2 Xv_2 + \tilde{\beta}_3 Xv_3 .$$

As $Xv_3 = 0$ the parameter $\tilde{\beta}_3$ can possess arbitrary values and it does not appear in the canonical model (3). So we are interested only in the two remaining parameters

$$\tilde{\beta}_1 = v_1^{\mathsf{T}}\beta = -0.497\beta_1 - 0.138\beta_2 - 0.857\beta_3$$
$$\tilde{\beta}_2 = v_2^{\mathsf{T}}\beta = -0.294\beta_1 - 0.902\beta_2 + 0.315\beta_3$$

The least squares estimators are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1/\sigma_1 ,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/\sigma_2 ,$$

and if denote the vector of estimable parameters as

$$\hat{\tilde{\beta}}_{\mathrm{e}} = \left(r_X \times 1\right) = \begin{pmatrix} \hat{\tilde{\beta}}_1 \\ \hat{\tilde{\beta}}_2 \end{pmatrix}, \text{ then } \operatorname{var}\left(\hat{\tilde{\beta}}_{\mathrm{e}}\right) = \sigma^2 \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix} = \begin{pmatrix} 0.002\,829 & 0 \\ 0 & 0.006\,826 \end{pmatrix},$$

where σ_1, σ_2 are the singular values of X and σ^2 denotes the unknown variance in model (1). So the least squares estimators are uncorrelated. The unknown variance σ^2 can be estimated from (3) by

$$\hat{\sigma}^2 = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2)\text{; note that } n - r_X = 3 .$$

C) Minimum length solution

The parameter $\tilde{\beta}_3$ does not appear in the canonical model (3) and so it can have arbitrary values. We set this parameter to zero. This way it is defined such that $\tilde{\beta}^{\mathsf{T}}\tilde{\beta} = \sum \tilde{\beta}_i^2 = \min$. Now $\tilde{\beta} = V^{\mathsf{T}}\beta$ and V is orthogonal. So $\tilde{\beta}^{\mathsf{T}}\tilde{\beta} = \beta^{\mathsf{T}}\beta$, and all parameters in our linear model become identifiable by the requirement $\beta^{\mathsf{T}}\beta = \min$; this parameter definition is called the *minimum length definition*:

$$\tilde{\beta}_1 = -0.497\beta_1 - 0.138\beta_2 - 0.857\beta_3,$$
$$\tilde{\beta}_2 = -0.294\beta_1 - 0.902\beta_2 + 0.315\beta_3,$$
$$\tilde{\beta}_3 = 0.816\beta_1 - 0.408\beta_2 - 0.408\beta_3 = 0.$$

The least squares estimators are now given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1/\sigma_1,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/\sigma_2,$$
$$\hat{\tilde{\beta}}_3 = 0.$$

and we have

$$\operatorname{var}\left(\hat{\tilde{\beta}}\right) = \sigma^2 \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sigma^2 \begin{pmatrix} 0.002\,829 & 0 & 0 \\ 0 & 0.006\,826 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D^+\tilde{y} \quad \text{where } D^+ = (p\times n) = (3\times 5) = \operatorname{diag}(1/\sigma_1, 1/\sigma_2, 0),$$
$$\operatorname{var}(\hat{\tilde{\beta}}) = \sigma^2 (D^\mathsf{T} D)^+,$$
$$\hat{\sigma}^2 = \frac{1}{n-2}\tilde{e}^\mathsf{T}\tilde{e} = \tfrac{1}{3}(\tilde{u}_3^2 + \tilde{u}_4^2 + \tilde{u}_5^2) = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2),$$

where $\tilde{e} = \tilde{y} - D\hat{\tilde{\beta}}$. D^+ is the Moore-Penrose inverse of D and we obtain

$$D^+ = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.053\,186 & 0 & 0 & 0 & 0 \\ 0 & 0.082\,622 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(D^\mathsf{T} D)^+ = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.002\,829 & 0 & 0 \\ 0 & 0.006\,826 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By backtransformation we find the least squares estimator of the original parameters (with minimum length definition):

$$\hat{\beta} = V\hat{\tilde{\beta}} = V\,D^+\tilde{y} = V\,D^+U^\mathsf{T} y = X^+ y \quad \text{where } X^+ = (p\times n) = (3\times 5) = V\,D^+U^\mathsf{T},$$
$$\operatorname{var}(\hat{\tilde{\beta}}) = \sigma^2 V (D^\mathsf{T} D)^+ V^\mathsf{T},$$
$$\hat{\sigma}^2 = \frac{1}{n-r_X}\tilde{e}^\mathsf{T}\tilde{e} = \frac{1}{n-r_X} e^\mathsf{T} e \text{ where } e = y - X\hat{\beta} = y - XX^+y = (I_n - XX^+)y,$$

and where $r_X = 2$ denotes the rank of X. $X^+ = V\,D^+U^\mathsf{T}$ is the Moore-Penrose inverse of $X = UDV^\mathsf{T}$. We obtain

$$X^+ = \begin{pmatrix} 0.010\,157 & 0.015\,294 & 0.011\,161 & 0.019\,890 & 0.020\,739 \\ 0.010\,794 & 0.063\,782 & 0.002\,356 & 0.005\,928 & 0.037\,249 \\ 0.009\,520 & -0.033\,194 & 0.019\,967 & 0.033\,851 & 0.004\,229 \end{pmatrix},$$

$$V(D^\mathsf{T} D)^+ V^\mathsf{T} = \begin{pmatrix} 0.001\,287 & 0.002\,002 & 0.000\,573 \\ * & 0.005\,613 & -0.001\,609 \\ * & * & 0.002\,755 \end{pmatrix},$$

$$XX^+ = \begin{pmatrix} 0.090\,140 & 0.040\,667 & 0.118\,036 & 0.206\,929 & 0.153\,632 \\ * & 0.821\,689 & -0.107\,404 & -0.163\,712 & 0.326\,382 \\ * & * & 0.186\,769 & 0.322\,442 & 0.149\,809 \\ * & * & * & 0.557\,293 & 0.270\,421 \\ * & * & * & * & 0.344\,108 \end{pmatrix}.$$

Example 3: Simple linear model with weak multicollinearity

We consider again the simple linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 I_n\right)$,

but this time

$$X_{\text{ori}} = (x_1, x_2, x_3) = (5 \times 3) = \sqrt{2}\begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix} \approx \begin{pmatrix} 4.243 & 2.828 & 5.657 \\ 1.414 & 12.728 & -9.899 \\ 5.657 & 1.414 & 9.899 \\ 9.899 & 2.828 & 16.971 \\ 7.071 & 8.485 & 5.657 \end{pmatrix} = X .$$

We have $x_3 = 2x_1 - x_2$ as in Example 2, and so the original matrix X_{ori} has rank 2 whereas for the matrix X (= X_{ori} rounded to 3 decimal places) we obtain the singular values

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} 26.590 \\ 17.116 \\ 0.000\,748 \end{pmatrix},$$

and this means that the rounded matrix X has full rank 3.

a) Classical solution

The classical least squares estimator of the parameter vector β is given by

$$\hat{\beta} = (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} y$$

$$\text{var}(\hat{\beta}) = \sigma^2 (X^{\mathsf{T}} X)^{-1}$$

$$\hat{\sigma}^2 = \frac{1}{n-p} e^{\mathsf{T}} e, \text{ where } e = y - X\hat{\beta} = (I_n - P)y \text{ with } P = X(X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}}$$

and we obtain

$$(X^{\mathsf{T}} X)^{-1} = \begin{pmatrix} 1190\,235 & -595\,091 & -595\,127 \\ * & 297\,532 & 297\,550 \\ * & * & 297\,568 \end{pmatrix},$$

(2) $$(X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} = \begin{pmatrix} 618.561 & -163.829 & 541.697 & -675\,912 & 174.596 \\ -309.256 & 81.962 & -270.831 & 337.952 & -87.261 \\ -309.275 & 81.898 & -270.835 & 337.992 & -87.289 \end{pmatrix},$$

$$P = X(X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} = \begin{pmatrix} 0.411\,600 & -0.044\,486 & 0.399\,576 & -0.144\,337 & 0.244\,361 \\ * & 0.844\,241 & -0.181\,951 & -0.070\,678 & 0.302\,354 \\ * & * & 0.433\,286 & 0.014\,814 & 0.229\,265 \\ * & * & * & 0.941\,155 & 0.171\,272 \\ * & * & * & * & 0.369\,718 \end{pmatrix}.$$

b) Solution with singular value decomposition

The singular value decomposition of X is given by $X = UDV^{\mathsf{T}}$, where $U = (5 \times 5)$ and $V = (3 \times 3)$ are orthogonal and $D = (5 \times 3) = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$:

(3)
$$U=\begin{pmatrix}-0.276 & -0.118 & -0.567 & -0.596 & -0.483\\ 0.227 & -0.878 & 0.150 & 0.201 & -0.340\\ -0.432 & 0.011 & -0.497 & 0.747 & -0.095\\ -0.747 & -0.006 & 0.620 & -0.050 & -0.237\\ -0.358 & -0.464 & -0.160 & -0.210 & 0.766\end{pmatrix},\quad V=\begin{pmatrix}-0.497 & -0.294 & -0.817\\ -0.138 & -0.902 & 0.408\\ -0.857 & 0.315 & 0.408\end{pmatrix},$$

$$D=\begin{pmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & \sigma_3\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}=\begin{pmatrix}26.590 & 0 & 0\\ 0 & 17.116 & 0\\ 0 & 0 & 0.000\,748\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.$$

From

$$y=X\beta+\varepsilon=UDV^{\mathsf{T}}\beta+\varepsilon$$

we obtain the canonical form of the linear model (1)

(4) $\tilde{y}=D\tilde{\beta}+\tilde{\varepsilon}$, where $\tilde{y}=U^{\mathsf{T}}y,\ \tilde{\beta}=V^{\mathsf{T}}\beta,\ \tilde{\varepsilon}=U^{\mathsf{T}}\varepsilon$,

and as U is orthogonal we have $\tilde{\varepsilon}\sim(0,\sigma^2 I_n)$ i.e. the error variables $\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n$ are again uncorrelated each with variance σ^2. The least squares estimator of $\tilde{\beta}$ is given by $\hat{\tilde{\beta}}_i=\tilde{y}_i/\sigma_i, i=1,\ldots,3$ or in matrix notation

$$\hat{\tilde{\beta}}=D^{+}\tilde{y}\ \text{ where }\ D^{+}=(p\times n)=(3\times 5)=\operatorname{diag}(1/\sigma_1,1/\sigma_2,1/\sigma_3),$$

(5)
$$\operatorname{var}(\hat{\tilde{\beta}})=\sigma^2(D^{\mathsf{T}}D)^{-1},$$
$$\hat{\sigma}^2=\frac{1}{n-p}\tilde{e}^{\mathsf{T}}\tilde{e}=\tfrac{1}{2}(\tilde{u}_4^2+\tilde{u}_5^2).$$

D^{+} is the Moore-Penrose inverse of *D*. We obtain

$$(D^{\mathsf{T}}D)^{-1}=\begin{pmatrix}1/\sigma_1^2 & 0 & 0\\ 0 & 1/\sigma_2^2 & 0\\ 0 & 0 & 1/\sigma_3^2\end{pmatrix}=\begin{pmatrix}0.001\,414 & 0 & 0\\ 0 & 0.003\,413 & 0\\ 0 & 0 & 1.785\times10^{6}\end{pmatrix},$$

and so the variance of the parameter $\tilde{\beta}_3$ is very large as compared with $\tilde{\beta}_1$ and $\tilde{\beta}_2$. By backtransformation we find the original least squares estimator

$$\hat{\beta}=V\hat{\tilde{\beta}}=V\,D^{+}\tilde{y}=X^{+}y,\quad\text{where}\quad X^{+}=(p\times n)=(3\times5)=V\,D^{+}U^{\mathsf{T}},$$
$$\operatorname{var}(\hat{\tilde{\beta}})=\sigma^2V(D^{\mathsf{T}}D)^{-1}V^{\mathsf{T}},$$
$$\hat{\sigma}^2=\frac{1}{n-p}\tilde{e}^{\mathsf{T}}\tilde{e}=\frac{1}{n-p}e^{\mathsf{T}}e,\quad\text{where}\quad e=y-X\hat{\beta}=y-XX^{+}y=(I_n-XX^{+})y.$$

$X^{+}=V\,D^{+}U^{\mathsf{T}}$ is the Moore-Penrose inverse of *X,* and as

$$(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}=X^{+}$$
$$(X^{\mathsf{T}}X)^{-1}=V(D^{\mathsf{T}}D)^{-1}V^{\mathsf{T}}$$
$$P=X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}=X\,X^{+}$$

we obtain the same results as with the classical solution. The variances and covariances of the original parameters $\beta=V\tilde{\beta}$ are rather large (see the matrix $(X^{\mathsf{T}}X)^{-1}$ in (2)), as all three parameters β_1,β_2,β_3 depend on $\tilde{\beta}_3$ which has an very large variance as compared with $\tilde{\beta}_1$ and $\tilde{\beta}_2$.

c) Solution with rank-k-approximation

The matrix X is given with three decimal places and so the smallest singular value of X, $\sigma_3 = 0.000\,748$, is near the maximal rounding error of X, that amounts to 0.0005. The rank-2-approximation of X will give a matrix $X_1 = (x_{ij}^{(1)}) = (5\times 3)$ with $\mathrm{rk}(X_1) = 2$ and with $\max\left|x_{ij} - x_{ij}^{(1)}\right| \le \sigma_3 = 0.000\,748$. We want to determine this matrix X_1. The singular value decomposition of X is given by $X = UDV^\top$, where $U = (5\times 5)$ and $V = (3\times 3)$ are orthogonal and where $D = (5\times 3) = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ as given in (3). Let

$$D_1 = (5\times 3) = \mathrm{diag}(\sigma_1,\sigma_2,0) = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 26.590 & 0 & 0 \\ 0 & 17.116 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Now we define X_1 as

(6) $\quad X_1 = UD_1V^\top$.

As $\mathrm{rk}(D_1) = 2$ we also have $\mathrm{rk}(X_1) = 2$, and if we compute X_1 we find $\max\left|x_{ij} - x_{ij}^{(1)}\right| = 0.000\,379$ which is smaller than $\sigma_3 = 0.000\,748$ and even smaller than the maximal rounding error of X. So we will work in the following with X_1 instead of X as the rank of X_1 is numerically stable in the sense that it cannot be made smaller just by small perturbations of the matrix elements. From the model

$$y = X_1\beta + \varepsilon,\ \text{where } X_1 = (n\times p) = UD_1V^\top \text{ and } \varepsilon \sim \left(0, \sigma^2 I_n\right)$$

we obtain the canonical model

(7) $\quad \tilde{y} = D_1\tilde{\beta} + \tilde{\varepsilon}$, where $\tilde{y} = U^\top y,\ \tilde{\beta} = V^\top\beta,\ \tilde{\varepsilon} = U^\top\varepsilon$,

and as U is orthogonal we have again $\tilde{\varepsilon} \sim (0, \sigma^2 I_n)$, which means that $\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n$ are again uncorrelated each with variance σ^2. The canonical model (7) can be written as

$$\begin{aligned} \tilde{y}_1 &= \sigma_1\tilde{\beta}_1 + \tilde{\varepsilon}_1, \\ \tilde{y}_2 &= \sigma_2\tilde{\beta}_2 + \tilde{\varepsilon}_2, \\ \tilde{y}_3 &= \tilde{\varepsilon}_3, \\ \tilde{y}_4 &= \tilde{\varepsilon}_4, \\ \tilde{y}_5 &= \tilde{\varepsilon}_5. \end{aligned} \tag{8}$$

The parameter $\tilde{\beta}_3$ does not appear in this model as the corresponding diagonal element in D_1 is zero. So $\tilde{\beta}_3$ can possess arbitrary values, it is not identifiable and not estimable. The least squares estimators for $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1/\sigma_1,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/\sigma_2.$$

As the transformed observations $\tilde{y}_i$ are uncorrelated each with variance σ^2 also the two least squares estimators are uncorrelated with variances $\mathrm{var}(\hat{\tilde{\beta}}_i) = \sigma^2/\sigma_i^2$, $i = 1,2$. The unknown variance σ^2 can be estimated by

$$\hat{\sigma}^2 = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2);\ \text{note that } n - \mathrm{rk}(X_1) = 5 - 2 = 3.$$

Minimum length solution

The parameter $\tilde{\beta}_3$ does not appear in the canonical model and so it can possess arbitrary values. We set this parameter to zero so that $\tilde{\beta}^{\mathsf{T}}\tilde{\beta} = \beta^{\mathsf{T}}\beta = \min$ (minimum length definition). The least squares estimator of $\tilde{\beta}_3$ will also be zero as well as the variance of this estimator. In matrix notation we now have

$$\hat{\tilde{\beta}} = D_1^+ y \text{, where } D_1^+ = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The covariance matrix of $\hat{\tilde{\beta}}$ is given by

$$\operatorname{var}(\hat{\tilde{\beta}}) = \sigma^2 (D_1^{\mathsf{T}} D_1)^+ = \sigma^2 \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By backtransformation we find the original least squares estimator

$$\hat{\beta} = V\hat{\tilde{\beta}} = V D_1^+ \tilde{y} = X_1^+ y \quad \text{where } X_1^+ = (p \times n) = (3 \times 5) = V D_1^+ U^{\mathsf{T}},$$

$$\operatorname{var}(\hat{\beta}) = \sigma^2 V (D_1^{\mathsf{T}} D_1)^+ V^{\mathsf{T}},$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\tilde{e}^{\mathsf{T}}\tilde{e} = \frac{1}{n-p} e^{\mathsf{T}} e \text{ where } e = y - X_1\hat{\beta} = (I_n - X_1 X_1^+) y.$$

We obtain

$$X_1^+ = \begin{pmatrix} 0.007\,182 & 0.010\,814 & 0.007\,892 & 0.014\,064 & 0.014\,665 \\ 0.007\,633 & 0.045\,102 & 0.001\,666 & 0.004\,191 & 0.026\,340 \\ 0.006\,732 & -0.023\,471 & 0.014\,118 & 0.023\,937 & 0.002\,991 \end{pmatrix},$$

$$V(D_1^{\mathsf{T}} D_1)^+ V^{\mathsf{T}} = \begin{pmatrix} 0.000\,644 & 0.001\,001 & 0.000\,286 \\ * & 0.002\,807 & -0.000\,805 \\ * & * & 0.001\,377 \end{pmatrix} = M,$$

$$P = X_1 X_1^+ = \begin{pmatrix} 0.090\,144 & 0.040\,660 & 0.118\,065 & 0.206\,935 & 0.153\,632 \\ * & 0.821\,687 & -0.107\,386 & -0.163\,722 & 0.326\,385 \\ * & * & 0.186\,757 & 0.322\,436 & 0.149\,811 \\ * & * & * & 0.557\,301 & 0.270\,416 \\ * & * & * & * & 0.344\,110 \end{pmatrix}.$$

Note that we had a much larger covariance matrix $\operatorname{var}(\hat{\beta})$ with the classical solution a). If we did the same computations with the original matrix X_{ori} instead of X_1, which is the rank-2 approximation to the rounded matrix *X*, we would find essentially the same results. The maximum difference $maxdiff = \max |a_{ij} - b_{ij}|$ between the corresponding matrices is given here:

original data	rank-2 approximation	*maxdiff*
X_{ori}	X_1	0.000 449
X_{ori}^+	X_1^+	0.154×10^{-5}
$V(D_2^{\mathsf{T}} D_2)^+ V^{\mathsf{T}}$	$V(D_2^{\mathsf{T}} D_2)^+ V^{\mathsf{T}}$	0.165×10^{-6}
$X_{\text{ori}} X_{\text{ori}}^+$	$X_1 X_1^+$	0.178×10^{-4}

Example 4: General linear model, regular case

In this example we consider the general linear model

(1) $y = X\beta + \varepsilon,$ where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$.

Here the error variables $\varepsilon_1, \ldots, \varepsilon_n$ are correlated with covariance matrix $\sigma^2 W$, where σ^2 is unknown and $W = (n \times n)$ is a known positive semidefinite matrix.

Let

$$(2) \qquad X = (5 \times 3) = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 9 & 8 \\ 4 & 1 & 4 \\ 7 & 2 & 0 \\ 5 & 6 & 5 \end{pmatrix}$$

as in Example 1. We obtain the singular values

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} 16.560 \\ 7.612 \\ 2.795 \end{pmatrix},$$

and this means that the matrix X has full rank 3. Furthermore let

$$F_0 = (5 \times 5) = \begin{pmatrix} 4 & 9 & 8 & -5 & -6 \\ 1 & -7 & -9 & -5 & -7 \\ 7 & -3 & 9 & 4 & -9 \\ 4 & -9 & 5 & 9 & -8 \\ -6 & -8 & 7 & -6 & 1 \end{pmatrix}$$

and now we define

$$(3) \qquad W_0 = \tfrac{1}{17} F_0 F_0^{\mathsf{T}} = (5 \times 5) \approx \begin{pmatrix} 13.059 & -3.765 & 6.294 & -1.294 & -0.941 \\ * & 12.059 & -0.588 & 1.941 & 0.588 \\ * & * & 13.882 & 12.235 & 0.706 \\ * & * & * & 15.706 & 1.235 \\ * & * & * & * & 10.941 \end{pmatrix} = W.$$

W is the matrix W_0 rounded to three decimal places. The singular values of W are

$$\begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_5 \end{pmatrix} = \begin{pmatrix} 27.951 \\ 18.089 \\ 10.654 \\ 8.839 \\ 0.114 \end{pmatrix},$$

and so W is a symmetric and positive definite matrix with full rank 5.

a) Classical procedure, Aitken estimator

We consider the general linear model (1) with $X = (n \times p)$ and $W = (n \times n)$ given by (2) and (3).

The eigenvalue decomposition of W is given by $W = R\Lambda R^{\mathsf{T}}$, where R is orthogonal and $\Lambda = (5 \times 5) = \operatorname{diag}(\lambda_1, \ldots, \lambda_5)$, $\lambda_i = \sigma_i$ (as W is positive definite). We set

(4) $F = R\Lambda^{1/2} R^{\mathsf{T}}$, where $\Lambda^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_5})$

and obtain

$$F \approx \begin{pmatrix} 3.310 & -0.519 & 1.221 & -0.566 & -0.142 \\ * & 3.421 & -0.086 & 0.272 & 0.072 \\ * & * & 2.802 & 2.126 & 0.096 \\ * & * & * & 3.282 & 0.141 \\ * & * & * & * & 3.299 \end{pmatrix}.$$

F is symmetric and $F^2 = FF^\top = R\Lambda R^\top = W$. The random error of our model (1) can now be written in the form

$$\varepsilon = Fu \text{ with } u \sim (0, \sigma^2 I_n), \text{ i.e. with } \mathrm{E}(u) = 0 \text{ and } \mathrm{var}(u) = \sigma^2 I_n,$$

as $\mathrm{E}(\varepsilon) = \mathrm{E}(Fu) = 0$ and $\mathrm{var}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon^\top) = F\,\mathrm{E}(uu^\top)F^\top = \sigma^2 FF^\top = \sigma^2 W$. So model (1) can also be given as

(5) $\quad y = X\beta + Fu, \text{ where } X = (n \times p), F = (n \times n), u \sim \left(0, \sigma^2 I_n\right).$

As F is a regular matrix (i.e. F has full rank n) we can write (5) as

(6) $\quad \bar{y} = \bar{X}\beta + u, \text{ where } \bar{y} = F^{-1}y \text{ and } \bar{X} = F^{-1}X.$

Note that the inverse of $F = R\Lambda^{1/2}R^\top$ is given by $F^{-1} = R\Lambda^{-1/2}R^\top$ where $\Lambda^{-1/2} = \mathrm{diag}(1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_n})$. In (6) we have the simple linear model and its solution is given by

$$\hat{\beta} = (\bar{X}^\top \bar{X})^{-1}\bar{X}^\top \bar{y} = (X^\top W^{-1}X)^{-1}X^\top W^{-1}y,$$

$$\mathrm{var}(\hat{\beta}) = \sigma^2 (X^\top W^{-1}X)^{-1},$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\bar{e}^\top \bar{e} = \frac{1}{n-p}e^\top W^{-1}e \quad \text{as} \quad \bar{e} = \bar{y} - \bar{X}\hat{\beta} = F^{-1}(y - X\hat{\beta}) = F^{-1}e,$$

and we obtain

$$(X^\top W^{-1}X)^{-1} = \begin{pmatrix} 0.294\,584 & -0.313\,265 & 0.211\,856 \\ * & 0.373\,243 & -0.216\,344 \\ * & * & 0.185\,071 \end{pmatrix}$$

$$(X^\top W^{-1}X)^{-1}X^\top W^{-1} = \begin{pmatrix} 0.045\,300 & -0.067\,985 & 0.031\,067 & 0.074\,873 & 0.056\,742 \\ 0.105\,058 & 0.112\,274 & -0.266\,029 & 0.112\,274 & -0.029\,850 \\ -0.088\,830 & -0.013\,145 & 0.273\,161 & -0.156\,002 & 0.055\,801 \end{pmatrix}.$$

b) Procedure with singular value decomposition

We start with the singular value decomposition of $\bar{X} = F^{-1}X$ in model (6) that is given by $\bar{X} = UDV^\top$ where $U = (n \times n)$ and $V = (p \times p)$ are orthogonal and $D = (n \times p) = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$; $\sigma_1, \ldots, \sigma_n$ are the singular values of $\bar{X} = F^{-1}X$. Here are the three matrices:

$$U = \begin{pmatrix} 0.409 & 0.282 & 0.249 & 0.670 & 0.492 \\ 0.025 & 0.772 & -0.472 & -0.347 & 0.248 \\ -0.615 & 0.252 & 0.666 & -0.187 & 0.283 \\ 0.659 & -0.077 & 0.418 & -0.610 & 0.115 \\ 0.141 & 0.505 & 0.312 & 0.156 & -0.777 \end{pmatrix}, \quad V = \begin{pmatrix} 0.793 & 0.102 & 0.600 \\ 0.425 & 0.612 & -0.667 \\ -0.435 & 0.784 & 0.442 \end{pmatrix},$$

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 9.818 & 0 & 0 \\ 0 & 4.778 & 0 \\ 0 & 0 & 1.119 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We obtain the canonical model

(7) $\quad \tilde{y} = D\tilde{\beta} + \tilde{u} \quad \text{where} \quad \tilde{y} = U^\top F^{-1}y, \tilde{\beta} = V^\top \beta, \tilde{u} = U^\top u,$

and so

$$\tilde{y}_1 = \sigma_1 \tilde{\beta}_1 + u_1,$$
$$\tilde{y}_2 = \sigma_2 \tilde{\beta}_2 + u_2,$$
$$\tilde{y}_3 = \sigma_3 \tilde{\beta}_3 + u_3,$$
$$\tilde{y}_4 = u_4,$$
$$\tilde{y}_5 = u_5.$$

The least squares estimators of the canonical parameters are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1 / \sigma_1,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2 / \sigma_2,$$
$$\hat{\tilde{\beta}}_3 = \tilde{y}_3 / \sigma_3,$$

or in matrix notation

$$\hat{\tilde{\beta}} = D^+ \tilde{y}, \text{ where } D^+ = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 1/\sigma_3 & 0 & 0 \end{pmatrix}.$$

D^+ is the Moore-Penrose inverse of D. The covariance matrix of $\hat{\tilde{\beta}}$ is given by

$$\text{var}(\hat{\tilde{\beta}}) = D^+ \text{var}(\tilde{y})(D^+)^\mathsf{T} = \sigma^2 D^+ (D^+)^\mathsf{T}) = \sigma^2 (D^\mathsf{T} D)^{-1} = \sigma^2 \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 1/\sigma_3^2 \end{pmatrix}.$$

By backtransformation we find the least squares estimators of the original parameters $\beta = V \tilde{\beta}$:

$$\hat{\beta} = V \hat{\tilde{\beta}} = V D^+ \tilde{y} = \bar{X}^+ \bar{y} = \bar{X}^+ F^{-1} y \text{ where } \bar{X}^+ = (p \times n) = (3 \times 5) = V D^+ U^\mathsf{T},$$
$$\text{var}(\hat{\tilde{\beta}}) = \sigma^2 V (D^\mathsf{T} D)^{-1} V^\mathsf{T},$$
$$\hat{\sigma}^2 = \frac{1}{n-p} \tilde{e}^\mathsf{T} \tilde{e} = \frac{1}{n-p} e^\mathsf{T} e \text{ where } e = y - \bar{X}\hat{\beta} = y - \bar{X}\bar{X}^+ y = (I_n - XX^+) y.$$

$\bar{X}^+ = V D^+ U^\mathsf{T}$ is the Moore-Penrose inverse of $\bar{X}$, and as

$$(X^\mathsf{T} W^{-1} X)^{-1} = V (D^\mathsf{T} D)^{-1} V^\mathsf{T},$$
$$(X^\mathsf{T} W^{-1} X)^{-1} X^\mathsf{T} W^{-1} = \bar{X}^+ F^{-1},$$

we obtain the same results as with the classical procedure of Aitken.

c) Procedure with generalized singular value decomposition

We consider the general linear model as given in (5):

(8) $y = X\beta + Fu$, where $X = (n \times p)$, $F = (n \times n)$, $u \sim \left(0, \sigma^2 I_n\right)$

with X and F as given above in (2) and (4).

(i) *Singular value decomposition of the matrix* $(X \mid F)$

The singular value decomposition of $(X \mid F) = (n \times m) = (5 \times 8)$ with $m = p + n = 8$ is given by

$$(X \mid F) = P \Delta Q^\mathsf{T},$$

where $P = (n \times n) = (5 \times 5)$ and $Q = (m \times m) = (8 \times 8)$ are orthogonal and $\Delta = (n \times m) = (5 \times 8) = \text{diag}(\sigma_1, \ldots, \sigma_5)$, and where the singular values $\sigma_1, \ldots, \sigma_5$ of $(X \mid F)$ are given in the following table. We have $r_c = rk(X \mid F) = n = 5$, and this simplifies the further procedure.

Table 1: Singular values of $X, F, (X\,|\,F)$ and $\overline{X}$

	$X=(5\times 3)$	$F=(5\times 5)$	$(X\|F)=(5\times 8)$	$\overline{X}=F^{-1}X$
singular values	$\begin{pmatrix}16.560\\7.612\\2.795\end{pmatrix}$	$\begin{pmatrix}5.287\\4.253\\3.264\\2.973\\0.338\end{pmatrix}$	$\begin{pmatrix}17.008\\8.802\\4.563\\3.685\\2.120\end{pmatrix}$	$\begin{pmatrix}9.818\\4.778\\1.119\end{pmatrix}$

(ii) *CS-decomposition of Q*

As $r_c=\text{rk}(X\,|\,F)=n=5$ we have to determine the CS-decomposition of the orthogonal matrix $Q=(m\times m)=(8\times 8)$ with the format

$$(9)\qquad Q=\left(\begin{array}{c|c}Q_{11} & Q_{12}\\ \hline Q_{21} & Q_{22}\end{array}\right)=\left(\begin{array}{c|c}(p\times n) & (p\times p)\\ \hline (n\times n) & (n\times p)\end{array}\right)=\left(\begin{array}{c|c}(3\times 5) & (3\times 3)\\ \hline (5\times 5) & (5\times 3)\end{array}\right).$$

We obtain orthogonal matrices

$$U=(8\times 8)=\left(\begin{array}{c|c}U_1 & 0\\ \hline 0 & U_2\end{array}\right)=\left(\begin{array}{c|c}(3\times 3) & *\\ \hline * & (5\times 5)\end{array}\right)$$

$$V=(8\times 8)=\left(\begin{array}{c|c}V_1 & 0\\ \hline 0 & V_2\end{array}\right)=\left(\begin{array}{c|c}(5\times 5) & *\\ \hline * & (3\times 3)\end{array}\right)$$

with

$$U_1=\begin{pmatrix}0.793 & 0.102 & 0.600\\ 0.425 & 0.612 & -0.667\\ -0.435 & 0.784 & 0.442\end{pmatrix},\quad U_2=\begin{pmatrix}0.409 & 0.282 & 0.249 & 0.626 & -0.547\\ 0.025 & 0.772 & -0.472 & -0.366 & -0.218\\ -0.615 & 0.252 & 0.666 & -0.210 & -0.266\\ 0.659 & -0.077 & 0.418 & -0.617 & -0.063\\ 0.141 & 0.505 & 0.312 & 0.221 & 0.761\end{pmatrix}$$

$$V_1=\begin{pmatrix}-0.360 & -0.918 & -0.144 & 0.084 & 0\\ -0.577 & 0.351 & -0.722 & 0.117 & 0.088\\ -0.407 & 0.147 & 0.428 & 0.590 & -0.531\\ 0.235 & -0.024 & -0.003 & 0.738 & 0.632\\ 0.562 & -0.112 & -0.524 & 0.293 & -0.558\end{pmatrix},\quad V_2=\begin{pmatrix}0.468 & -0.096 & -0.878\\ -0.842 & 0.251 & -0.476\\ -0.266 & -0.963 & -0.037\end{pmatrix}$$

such that

$$U^{\mathsf{T}}QV=D=\left(\begin{array}{c|c}D_{11} & D_{12}\\ \hline D_{21} & D_{22}\end{array}\right)=\left(\begin{array}{c|c}(3\times 5) & (3\times 3)\\ \hline (5\times 5) & (5\times 3)\end{array}\right)$$

where

$$D=\left(\begin{array}{ccccc|ccc}c_1 & 0 & 0 & 0 & 0 & s_1 & 0 & 0\\ 0 & c_2 & 0 & 0 & 0 & 0 & s_2 & 0\\ 0 & 0 & c_3 & 0 & 0 & 0 & 0 & s_3\\ \hline s_1 & 0 & 0 & 0 & 0 & -c_1 & 0 & 0\\ 0 & s_2 & 0 & 0 & 0 & 0 & -c_2 & 0\\ 0 & 0 & s_3 & 0 & 0 & 0 & 0 & -c_3\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\end{array}\right).$$

The diagonal elements c_i and s_i are given in the following table. Note that $c_i^2+s_i^2=1$ for $i=1,2,3$ and so the matrix *D* is orthogonal, too.

Table 2: Diagonal elements c_i, s_i

i	c_i	s_i
1	0.994 852	0.101 334
2	0.978795	0.204 843
3	0.745 620	0.666 371

(iii) *Generalized singular value decomposition and canonical model*

We now have

$$(10)\qquad \begin{aligned} (X\mid F) &= P\Delta Q^{\mathsf{T}}, \\ Q &= UDV^{\mathsf{T}}, \end{aligned}$$

and from this we find the so-called *generalized singular value decomposition* of the pair X, F

$$(11)\qquad \begin{aligned} P^{\mathsf{T}} X U_1 &= \Delta_0 V_1 D_{11}^{\mathsf{T}}, \\ P^{\mathsf{T}} F U_2 &= \Delta_0 V_1 D_{21}^{\mathsf{T}}, \end{aligned}$$

where $\Delta_0 = (5\times 5) = \mathrm{diag}(\sigma_1,\ldots,\sigma_5)$ and where $\sigma_1,\ldots,\sigma_5$ are the singular values of $(X\mid F)$ as given above. For P we have

$$P = \begin{pmatrix} -0.264 & -0.156 & 0.725 & 0.202 & 0.583 \\ -0.667 & 0.588 & -0.121 & -0.416 & 0.150 \\ -0.314 & -0.367 & 0.440 & -0.379 & -0.655 \\ -0.290 & -0.701 & -0.476 & -0.257 & 0.362 \\ -0.551 & -0.057 & -0.201 & 0.759 & -0.278 \end{pmatrix}.$$

Our model (8) can now be written in the *canonical form*

$$(12)\qquad \tilde{y} = D_1\tilde{\beta} + D_2\tilde{u},$$

where

$$\tilde{y} = V_1^{\mathsf{T}}\Delta_0^{-1}P^{\mathsf{T}}y,\quad \tilde{\beta} = U_1^{\mathsf{T}}\beta,\quad \tilde{u} = U_2^{\mathsf{T}}u,$$

$$D_1 = D_{11}^{\mathsf{T}} = (5\times 3) = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\qquad D_2 = D_{21}^{\mathsf{T}} = (5\times 5) = \begin{pmatrix} s_1 & 0 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and so

$$(13)\qquad \begin{aligned} \tilde{y}_1 &= c_1\tilde{\beta}_1 + s_1\tilde{u}_1, \\ \tilde{y}_2 &= c_2\tilde{\beta}_2 + s_2\tilde{u}_2, \\ \tilde{y}_3 &= c_3\tilde{\beta}_3 + s_3\tilde{u}_3, \\ \tilde{y}_4 &= \tilde{u}_4, \\ \tilde{y}_5 &= \tilde{u}_5. \end{aligned}$$

The least squares estimators are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1/c_1,$$

$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/c_2,$$

$$\hat{\tilde{\beta}}_3 = \tilde{y}_3/c_3.$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^+ \tilde{y} \quad \text{as} \quad D_1^+ = \begin{pmatrix} 1/c_1 & 0 & 0 & 0 & 0 \\ 0 & 1/c_2 & 0 & 0 & 0 \\ 0 & 0 & 1/c_3 & 0 & 0 \end{pmatrix},$$

$$\text{var}(\tilde{y}) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2 \begin{pmatrix} s_1^2 & 0 & 0 & 0 & 0 \\ 0 & s_2^2 & 0 & 0 & 0 \\ 0 & 0 & s_3^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\text{var}(\hat{\tilde{\beta}}) = D_1^+ \,\text{var}(\tilde{y})(D_1^+)^{\mathsf{T}} = \sigma^2 D_1^+ D_2 D_2^{\mathsf{T}} (D_1^+)^{\mathsf{T}} = \sigma^2 \begin{pmatrix} s_1^2/c_1^2 & 0 & 0 \\ 0 & s_2^2/c_2^2 & 0 \\ 0 & 0 & s_3^2/c_3^2 \end{pmatrix} = \sigma^2 D_0 \,.$$

For the original parameters we obtain

$$\beta = U_1 \tilde{\beta}, \quad \hat{\beta} = U_1 \hat{\tilde{\beta}}, \quad \text{var}(\hat{\beta}) = U_1 \,\text{var}(\hat{\tilde{\beta}}) U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}} \,.$$

The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \tfrac{1}{2}(\tilde{y}_4^2 + \tilde{y}_5^2)\,.$$

By backtransformation we obtain again the same results as in a) and b), but the classical Aitken procedure a) works only in the regular case, i.e. if X and F have full rank, procedure b) with the simple singular value decomposition works also if multicollinearity is present i.e. if X does not have full rank, and method c) with the generalized singular value decomposition works even if both X und F are rank deficient.

Remark

By multiplying the canonical model (12) with D_2^{-1} we obtain a model that is equivalent to the canonical model (7) as we have $D_2^{-1} D_1 = D$, where $D = (5 \times 3) = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$ and where $\sigma_1, \sigma_2, \sigma_3$ are the singular values of $\bar{X} = F^{-1} X$.

Example 5: General linear model with strict multicollinearity

In this example we consider the general linear model

(1) $y = X\beta + \varepsilon,$ where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$

as in Example 4, but this time

$$X = (x_1, x_2, x_3) = (5 \times 3) = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix},$$

whereas the symmetric matrix $W = (5 \times 5)$ is the same as there:

$$W = \begin{pmatrix} 13.059 & -3.765 & 6.294 & -1.294 & -0.941 \\ * & 12.059 & -0.588 & 1.941 & 0.588 \\ * & * & 13.882 & 12.235 & 0.706 \\ * & * & * & 15.706 & 1.235 \\ * & * & * & * & 10.941 \end{pmatrix}.$$

We have $x_3 = 2x_1 - x_2$, i.e. the third column of X is a linear combination of the first two columns, and so the matrix X has rank 2; so we face the problem of strict multicollinearity. As $X^\top W^{-1} X$ has the same rank as X the inverse $(X^\top W^{-1} X)^{-1}$ does not exist and the classical procedure of Aitken breaks down.

a) Procedure with singular value decomposition

As in Example 4 the general linear model can be written in the form

(2) $y = X\beta + Fu,$ where $X = (n \times p),$ $F = (n \times n),$ $u \sim \left(0, \sigma^2 I_n\right),$

and where $FF^\top = W$. As in Example 4 we start with the singular value decomposition of $\overline{X} = F^{-1}X$ that is given by $\overline{X} = UDV^\top$ where $U = (n \times n)$ and $V = (p \times p)$ are orthogonal and $D = (n \times p) = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$; $\sigma_1, \ldots, \sigma_n$ are the singular values of $\overline{X} = F^{-1}X$. The following table gives the singular values of the different matrices involved.

Table 1: Singular values of $X, F, \overline{X} = F^{-1}X, (X \mid F)$

	$X = (5 \times 3)$	$F = (5 \times 5)$	$\overline{X} = (5 \times 5)$	$(X \mid F)$
singular values	$\begin{pmatrix} 18.802 \\ 12.103 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 5.287 \\ 4.253 \\ 3.264 \\ 2.973 \\ 0.338 \end{pmatrix}$	$\begin{pmatrix} 14.548 \\ 3.802 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 19.428 \\ 12.572 \\ 3.945 \\ 3.619 \\ 1.213 \end{pmatrix}$

And here are the three matrices U, V, D:

$$U = \begin{pmatrix} -0.427 & -0.200 & -0.842 & -0.257 & -0.060 \\ 0.061 & -0.892 & 0.126 & 0.258 & -0.345 \\ 0.535 & 0.065 & -0.498 & 0.642 & 0.223 \\ -0.709 & 0.175 & 0.115 & 0.673 & -0.032 \\ -0.161 & -0.361 & 0.119 & -0.053 & 0.909 \end{pmatrix}, \quad V = \begin{pmatrix} -0.536 & -0.216 & 0.816 \\ -0.271 & -0.872 & -0.408 \\ -0.800 & 0.440 & -0.408 \end{pmatrix},$$

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 14.548 & 0 & 0 \\ 0 & 3.802 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We obtain the canonical model

(3) $\tilde{y} = D\tilde{\beta} + \tilde{u}$, where $\tilde{y} = U^{\mathsf{T}} F^{-1} y, \tilde{\beta} = V^{\mathsf{T}}\beta, \tilde{u} = U^{\mathsf{T}} u$,

and so

$$\tilde{y}_1 = \sigma_1 \tilde{\beta}_1 + \tilde{u}_1,$$
$$\tilde{y}_2 = \sigma_2 \tilde{\beta}_2 + \tilde{u}_2,$$
$$\tilde{y}_3 = \tilde{u}_3,$$
$$\tilde{y}_4 = \tilde{u}_4,$$
$$\tilde{y}_5 = \tilde{u}_5.$$

The least squares estimators of the canonical parameters are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1/\sigma_1,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/\sigma_2.$$

The parameter $\tilde{\beta}_3$ can have arbitrary values, it is not identifiable and not estimable without further assumptions. We set this parameter to zero and so its value is defined such that $\tilde{\beta}^{\mathsf{T}}\tilde{\beta} = \tilde{\beta}_1^2 + \tilde{\beta}_2^2 + \tilde{\beta}_3^2 = \min$. For the original parameters we have $\beta = V\tilde{\beta}$, and as $\beta^{\mathsf{T}}\beta = \tilde{\beta}^{\mathsf{T}}\tilde{\beta}$ our parameters are made identifiable and estimable by *the minimum length requirement* $\beta^{\mathsf{T}}\beta = \tilde{\beta}^{\mathsf{T}}\tilde{\beta} = \min$. As now $\tilde{\beta}_3 = 0$ also the least squares estimate of $\tilde{\beta}_3$ is zero. In matrix notation we can write

$$\hat{\tilde{\beta}} = D^{+}\tilde{y}, \quad \hat{\beta} = V\hat{\tilde{\beta}} = VD^{+}\tilde{y} = VD^{+}U^{\mathsf{T}}F^{-1}y,$$
$$\mathrm{var}(\hat{\tilde{\beta}}) = \sigma^2 D^{+}D^{+\mathsf{T}} = \sigma^2 (D^{\mathsf{T}}D)^{+}, \quad \mathrm{var}(\hat{\beta}) = \sigma^2 V(D^{\mathsf{T}}D)^{+}V^{\mathsf{T}},$$
$$\hat{\sigma}^2 = \frac{1}{n - r_X}\tilde{e}^{\mathsf{T}}\tilde{e} = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2) \quad \text{with} \quad r_X = rk(X) = 2.$$

We obtain

$$(D^{\mathsf{T}}D)^{+} = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.004\,725 & 0 & 0 \\ 0 & 0.069\,175 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$V(D^{\mathsf{T}}D)^{+}V^{\mathsf{T}} = \begin{pmatrix} 0.004\,576 & 0.013\,698 & -0.004\,545 \\ * & 0.052\,905 & -0.025\,509 \\ * & * & 0.016\,418 \end{pmatrix},$$

$$VD^{+}U^{\mathsf{T}}F^{-1} = \begin{pmatrix} 0.049\,401 & 0.014\,657 & -0.076\,112 & 0.061\,121 & 0.009\,415 \\ 0.050\,009 & 0.063\,111 & -0.049\,466 & 0.026\,076 & 0.027\,105 \\ 0.048\,792 & -0.033\,797 & -0.102\,757 & 0.096\,166 & -0.008\,275 \end{pmatrix}.$$

b) Procedure with generalized singular value decomposition

We consider the general linear model as given in (2) with *X* and *F* as given there.

(i) *Singular value decomposition of the matrix* $(X \mid F)$

The singular value decomposition of $(X \mid F)$ is given by

$$(X \mid F) = P\Delta Q^{\mathsf{T}},$$

where $P = (5 \times 5)$ and $Q = (8 \times 8)$ are orthogonal and $\Delta = (5 \times 8) = \mathrm{diag}(\sigma_1, \ldots, \sigma_5)$, and where the singular values $\sigma_1, \ldots, \sigma_5$ of $(X \mid F)$ are given above in Table 1. For *P* we obtain

$$P = \begin{pmatrix} 0.275 & -0.090 & 0.847 & -0.273 & 0.353 \\ -0.217 & -0.882 & -0.167 & -0.380 & 0.051 \\ 0.451 & 0.011 & 0.052 & -0.411 & -0.790 \\ 0.744 & -0.014 & -0.472 & -0.123 & 0.457 \\ 0.349 & -0.462 & 0.169 & 0.773 & -0.199 \end{pmatrix}.$$

(ii) *CS-decomposition of Q*

As $r_c = \text{rk}(X \mid F) = n = 5$ we have to determine the CS-decomposition of the orthogonal matrix $Q = (m \times m) = (8 \times 8)$ with the format

$$Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right) = \left(\begin{array}{c|c} (p \times n) & (p \times p) \\ \hline (n \times n) & (n \times p) \end{array}\right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 3) \\ \hline (5 \times 5) & (5 \times 3) \end{array}\right).$$

We obtain orthogonal matrices

$$U = (8 \times 8) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right) = \left(\begin{array}{c|c} (3 \times 3) & * \\ \hline * & (5 \times 5) \end{array}\right)$$

$$V = (8 \times 8) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right) = \left(\begin{array}{c|c} (5 \times 5) & * \\ \hline * & (3 \times 3) \end{array}\right)$$

with

$$U_1 = \begin{pmatrix} -0.536 & 0.216 & -0.816 \\ -0.271 & 0.872 & 0.408 \\ -0.800 & -0.440 & 0.408 \end{pmatrix}, \quad U_2 = \begin{pmatrix} -0.427 & 0.200 & 0.388 & -0.778 & -0.150 \\ 0.061 & 0.892 & -0.448 & -0.027 & -0.007 \\ 0.535 & -0.065 & -0.035 & -0.169 & -0.825 \\ -0.709 & -0.175 & -0.450 & 0.210 & -0.469 \\ -0.161 & 0.361 & 0.667 & 0.569 & -0.277 \end{pmatrix}$$

$$V_1 = \begin{pmatrix} -0.959 & -0.148 & 0 & 0 & -0.241 \\ 0.139 & -0.985 & 0.041 & -0.084 & 0.049 \\ -0.024 & 0.087 & 0.724 & -0.683 & 0.043 \\ -0.092 & -0.026 & 0.619 & 0.680 & 0.381 \\ -0.227 & 0.021 & -0.302 & -0.253 & 0.890 \end{pmatrix}, \quad V_2 = \begin{pmatrix} -0.747 & 0.566 & 0.350 \\ 0.659 & 0.557 & 0.506 \\ -0.091 & -0.608 & 0.788 \end{pmatrix}$$

such that

$$U^{\mathsf{T}} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 3) \\ \hline (5 \times 5) & (5 \times 3) \end{array}\right)$$

where

$$D = \left(\begin{array}{ccccc|ccc} c_1 & 0 & 0 & 0 & 0 & s_1 & 0 & 0 \\ 0 & c_2 & 0 & 0 & 0 & 0 & s_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline s_1 & 0 & 0 & 0 & 0 & -c_1 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 & 0 & -c_2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right).$$

The diagonal elements c_i, s_i are given in Table 2. Note that $c_i^2 + s_i^2 = 1$ for $i = 1, 2$, and so the matrix D is orthogonal, too.

Table 2: Diagonal elements c_i, s_i

i	c_i	s_i
1	0.997 646	0.068 578
2	0.967 110	0.254 360

(iii) *Generalized singular decomposition and canonical model*

Now we have

$$(X\,|\,F) = P\Delta Q^{\mathsf{T}},$$
$$Q = UDV^{\mathsf{T}},$$

and from this we find the so-called *generalized singular value decomposition* of the pair X, F

$$\text{(4)}\qquad \begin{aligned} P^{\mathsf{T}} X U_1 &= \Delta_0 V_1 D_{11}^{\mathsf{T}}, \\ P^{\mathsf{T}} F U_2 &= \Delta_0 V_1 D_{21}^{\mathsf{T}}, \end{aligned}$$

where $\Delta_0 = (5\times 5) = \mathrm{diag}(\sigma_1,\ldots,\sigma_5)$ and where $\sigma_1,\ldots,\sigma_5$ are the singular values of $(X\,|\,F)$ as given in Table 1. Our model (2) can now be written in the canonical form

$$\text{(5)}\qquad \tilde{y} = D_1\tilde{\beta} + D_2\tilde{u}\,,$$

where

$$\tilde{y} = V_1^{\mathsf{T}}\Delta_0^{-1}P^{\mathsf{T}}y,\quad \tilde{\beta} = U_1^{\mathsf{T}}\beta,\quad \tilde{u} = U_2^{\mathsf{T}}u,$$

$$D_1 = D_{11}^{\mathsf{T}} = (5\times 3) = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\qquad D_2 = D_{21}^{\mathsf{T}} = (5\times 5) = \begin{pmatrix} s_1 & 0 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The generalized singular value decomposition (4) shows that $\mathrm{rk}(X) = \mathrm{rk}(D_{11}) = \mathrm{rk}(D_1) = 2$ and $\mathrm{rk}(F) = \mathrm{rk}(D_{21}) = \mathrm{rk}(D_2) = 5$. Now the canonical model (5) explicitly written has the form

$$\begin{aligned} \tilde{y}_1 &= c_1\tilde{\beta}_1 + s_1\tilde{u}_1, \\ \tilde{y}_2 &= c_2\tilde{\beta}_2 + s_2\tilde{u}_2, \\ \tilde{y}_3 &= \tilde{u}_3, \\ \tilde{y}_4 &= \tilde{u}_4, \\ \tilde{y}_5 &= \tilde{u}_5. \end{aligned}$$

The parameter $\tilde{\beta}_3$ can possess arbitrary values as it is not met in the canonical model and we set this parameter to zero (minimum length definition). The least squares estimators are given by

$$\begin{aligned} \hat{\tilde{\beta}}_1 &= \tilde{y}_1/c_1\,, \\ \hat{\tilde{\beta}}_2 &= \tilde{y}_2/c_2\,, \\ \hat{\tilde{\beta}}_3 &= 0. \end{aligned}$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^{+}\tilde{y}\quad \text{as}\quad D_1^{+} = \begin{pmatrix} 1/c_1 & 0 & 0 & 0 & 0 \\ 0 & 1/c_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathrm{var}(\tilde{y}) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2 \begin{pmatrix} s_1^2 & 0 & 0 & 0 & 0 \\ 0 & s_2^2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathrm{var}(\hat{\tilde{\beta}}) = D_1^{+}\,\mathrm{var}(\tilde{y})(D_1^{+})^{\mathsf{T}} = \sigma^2 D_1^{+} D_2 D_2^{\mathsf{T}} (D_1^{+})^{\mathsf{T}} = \sigma^2 \begin{pmatrix} s_1^2/c_1^2 & 0 & 0 \\ 0 & s_2^2/c_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sigma^2 D_0\,.$$

For the original parameters we obtain

$$\beta = U_1\tilde{\beta}, \quad \hat{\beta} = U_1\hat{\tilde{\beta}}, \quad \mathrm{var}(\hat{\beta}) = U_1 \,\mathrm{var}(\hat{\tilde{\beta}})U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}} .$$

The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2) .$$

We obtain the same results as in a) with the simple singular value decomposition, but that procedure works only if the matrix *W* (and so the matrix *F*) has full rank, whereas the method with the generalized singular value decomposition also works if both *X* und *F* are rank deficient as we will see in the next example.

By multiplying the canonical model (5) with D_2^{-1} we obtain a model that is equivalent to (3) and we have $D_2^{-1}D_1 = D$, where $D = (5\times 3) = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ and where $\sigma_1, \sigma_2, \sigma_3$ are the singular values of $\overline{X} = F^{-1}X$.

Example 6: General linear model with rank deficient covariance matrix

In this example we consider again the general linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$.

Let

$$X = (5 \times 3) = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 9 & 8 \\ 4 & 1 & 4 \\ 7 & 2 & 0 \\ 5 & 6 & 5 \end{pmatrix}$$

as in Example 4. The matrix X has full rank 3. But in this example the matrix W will be rank deficient. Let

$$F_0 = (5 \times 4) = \begin{pmatrix} 4 & 9 & 8 & -5 \\ 1 & -7 & -9 & -5 \\ 7 & -3 & 9 & 4 \\ 4 & -9 & 5 & 9 \\ -6 & -8 & 7 & -6 \end{pmatrix}$$

and now we define

$$W = \tfrac{1}{20} F_0 F_0^{\mathsf{T}} = (5 \times 5) = \begin{pmatrix} 9.3 & -5.3 & 2.65 & -3.5 & -0.5 \\ * & 7.8 & -3.65 & -1.15 & 0.85 \\ * & * & 7.75 & 6.8 & 1.05 \\ * & * & * & 10.15 & 1.45 \\ * & * & * & * & 9.25 \end{pmatrix}.$$

The singular values of W are

$$\begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_5 \end{pmatrix} = \begin{pmatrix} 17.462 \\ 15.181 \\ 8.812 \\ 2.796 \\ 0 \end{pmatrix},$$

and so W is a symmetric and positive semidefinite matrix with rank 4. Now we consider the general linear model (1) with $X = (n \times p)$ and $W = (n \times n)$ as given above. As $rk(W) = 4$ the inverse W^{-1} does not exist and so the classical Aitken procedure as well as the procedure with the simple singular value decomposition cannot work. Therefore we apply the procedure with the generalized singular value decomposition.

(i) *Factorization of W*

We want to find a matrix $F = (n \times k)$ such that $W = FF^{\mathsf{T}}$ where $k = rk(W) = 4$. The eigenvalue decomposition of W is given by $W = R\Lambda R^{\mathsf{T}}$, where R is orthogonal and $\Lambda = (5 \times 5) = \mathrm{diag}(\lambda_1, \ldots, \lambda_5)$, $\lambda_i = \sigma_i$ (as W is positive semidefinite). As $\sigma_5 = \lambda_5 = 0$ we set

(2) $D = (5 \times 4) = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_4})$ and $F = (5 \times 4) = RD$.

As $DD^{\mathsf{T}} = (5 \times 5) = \mathrm{diag}(\lambda_1, \ldots, \lambda_4, 0) = \Lambda$ we have $FF^{\mathsf{T}} = RDD^{\mathsf{T}} R^{\mathsf{T}} = R\Lambda R^{\mathsf{T}} = W$. The random error of our model (1) can now be written in the form

$$\varepsilon = Fu \text{ with } u \sim (0, \sigma^2 I_k), \text{ i.e. with } \mathrm{E}(u) = 0 \text{ and } \mathrm{var}(u) = \sigma^2 I_k,$$

as $\mathrm{E}(\varepsilon) = \mathrm{E}(Fu) = 0$ and $\mathrm{var}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon^{\mathsf{T}}) = F\,\mathrm{E}(uu^{\mathsf{T}})F^{\mathsf{T}} = \sigma^2 FF^{\mathsf{T}} = \sigma^2 W$. So model (1) can also be given as

(3) $y = X\beta + Fu$, where $X = (n \times p), F = (n \times k), u \sim (0, \sigma^2 I_k)$.

(ii) *Singular value decomposition of* $(X \mid F)$

We have $X = (n \times p) = (5 \times 3)$ and $F = (n \times k) = (5 \times 4)$ and so $(X \mid F) = (n \times m) = (5 \times 7)$, where $m = p + k = 7$. Now we compute the singular value decomposition of $(X \mid F)$:

$$(X \mid F) = P \Delta Q^{\mathsf{T}},$$

where $P = (5 \times 5)$ and $Q = (7 \times 7)$ are orthogonal and $\Delta = (5 \times 7) = \mathrm{diag}(\sigma_1, \ldots, \sigma_5)$. We obtain

$$P = \begin{pmatrix} -0.258 & -0.147 & 0.718 & 0.226 & 0.588 \\ -0.671 & 0.590 & -0.196 & -0.338 & 0.222 \\ -0.299 & -0.342 & 0.428 & -0.590 & -0.513 \\ -0.275 & -0.708 & -0.507 & -0.149 & 0.379 \\ -0.565 & -0.109 & -0.075 & 0.682 & -0.445 \end{pmatrix},$$

and the singular values of $(X \mid F)$ are given in the following table.

Table 1: Singular values of $X, F, (X \mid F)$

	$X = (5 \times 3)$	$F = (5 \times 4)$	$(X \mid F) = (5 \times 7)$
singular values	$\begin{pmatrix} 16.560 \\ 7.612 \\ 2.795 \end{pmatrix}$	$\begin{pmatrix} 4.179 \\ 3.896 \\ 2.969 \\ 1.672 \end{pmatrix}$	$\begin{pmatrix} 16.771 \\ 8.540 \\ 4.246 \\ 2.883 \\ 1.924 \end{pmatrix}$

(iii) *CS-decomposition of* Q

As $r_c = \mathrm{rk}(X \mid F) = n = 5$ we have to determine the CS-decomposition of $Q = (m \times m) = (7 \times 7)$ with the format

$$Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array} \right) = \left(\begin{array}{c|c} (p \times n) & (p \times (m-n)) \\ \hline (k \times n) & (k \times (m-n)) \end{array} \right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 2) \\ \hline (4 \times 5) & (4 \times 2) \end{array} \right).$$

We obtain orthogonal matrices

$$U = (7 \times 7) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array} \right) = \left(\begin{array}{c|c} (3 \times 3) & * \\ \hline * & (4 \times 4) \end{array} \right)$$

$$V = (8 \times 8) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array} \right) = \left(\begin{array}{c|c} (5 \times 5) & * \\ \hline * & (2 \times 2) \end{array} \right)$$

with

$$U_1 = \begin{pmatrix} 0.785 & -0.214 & -0.581 \\ 0.616 & 0.369 & 0.696 \\ -0.066 & 0.905 & -0.421 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0.055 & 0.990 & 0 & 0.127 \\ 0.148 & 0.097 & 0.540 & -0.823 \\ -0.343 & 0.089 & -0.758 & -0.548 \\ -0.926 & 0.041 & 0.367 & 0.080 \end{pmatrix},$$

$$V_1 = \begin{pmatrix} -0.702 & -0.699 & 0.081 & -0.100 & 0.035 \\ -0.448 & 0.538 & 0.709 & -0.015 & 0.078 \\ -0.275 & 0.308 & -0.488 & -0.387 & 0.665 \\ 0.250 & -0.255 & 0.283 & 0.514 & 0.727 \\ 0.410 & -0.248 & 0.415 & -0.759 & 0.147 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0.483 & 0.876 \\ 0.876 & -0.483 \end{pmatrix},$$

such that

$$U^{\mathsf{T}} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array} \right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 2) \\ \hline (4 \times 5) & (4 \times 2) \end{array} \right),$$

where

$$D=\left(\begin{array}{ccccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & 0 & s_1 & 0 \\ 0 & 0 & c_2 & 0 & 0 & 0 & s_2 \\ \hline 0 & s_1 & 0 & 0 & 0 & -c_1 & 0 \\ 0 & 0 & s_2 & 0 & 0 & 0 & -c_2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right).$$

The diagonal elements c_i and s_i are given in Table 2. Note that $c_i^2 + s_i^2 = 1$ for $i = 1, 2$, and so the matrix D is orthogonal, too. Also note D_{21} and D_{12} are not classical diagonal matrices as the diagonal starts in the lower right corner and not in the upper left one.

Table 2: Diagonal elements c_i, s_i

i	c_i	s_i
1	0.988 084	0.153 915
2	0.779 148	0.626 840

(iv) *Generalized singular value decomposition and canonical model*
Now we have

$$(X \mid F) = P\Delta Q^{\mathsf{T}},$$
$$Q = UDV^{\mathsf{T}},$$

and from this we find the so-called *generalized singular value decomposition* of the pair X, F

(4)
$$P^{\mathsf{T}} X U_1 = \Delta_1 V_1 D_{11}^{\mathsf{T}},$$
$$P^{\mathsf{T}} F U_2 = \Delta_1 V_1 D_{21}^{\mathsf{T}},$$

where $\Delta_0 = (r_c \times r_c) = (5 \times 5) = \mathrm{diag}(\sigma_1, \ldots, \sigma_5)$ and where $\sigma_1, \ldots, \sigma_5$ are the singular values of $(X \mid F)$ as given above. Our model (2) can now be written in the canonical form

(5) $\tilde{y} = D_1 \tilde{\beta} + D_2 \tilde{u}$,

where

$$\tilde{y} = V_1^{\mathsf{T}} \Delta_0^{-1} P^{\mathsf{T}} y,$$
$$\tilde{\beta} = U_1^{\mathsf{T}} \beta,$$
$$\tilde{u} = U_2^{\mathsf{T}} u,$$

$$D_1 = D_{11}^{\mathsf{T}} = (5 \times 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & c_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$D_2 = D_{21}^{\mathsf{T}} = (5 \times 4) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The generalized singular value decomposition (4) shows that $\mathrm{rk}(X) = \mathrm{rk}(D_{11}) = \mathrm{rk}(D_1) = 3$ and $\mathrm{rk}(F) = \mathrm{rk}(D_{21}) = \mathrm{rk}(D_2) = 4$. The canonical model (5) explicitly written has the form

$$\tilde{y}_1 = \tilde{\beta}_1,$$
$$\tilde{y}_2 = c_1 \tilde{\beta}_2 + s_1 \tilde{u}_1,$$
$$\tilde{y}_3 = c_2 \tilde{\beta}_3 + s_2 \tilde{u}_2,$$
$$\tilde{y}_4 = \tilde{u}_3,$$
$$\tilde{y}_5 = \tilde{u}_4.$$

The observation $\tilde{y}_1$ is identical to the parameter $\tilde{\beta}_1$, this observation has no random error. As the covariance matrix $W = (5 \times 5)$ has rank 4 there exists a linear combination $\tilde{y}_1$ of the original observations ($y_1, \ldots, y_n$) with no random error, and there exists a linear combination $\tilde{\beta}_1$ of the original parameters ($\beta_1, \ldots, \beta_p$) such that $\tilde{y}_1 = \tilde{\beta}_1$. The least squares estimators are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1,$$
$$\hat{\tilde{\beta}}_2 = \tilde{y}_2 / c_1,$$
$$\hat{\tilde{\beta}}_3 = \tilde{y}_3 / c_2.$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^+ \tilde{y} \quad \text{as} \quad D_1^+ = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/c_1 & 0 & 0 & 0 \\ 0 & 0 & 1/c_2 & 0 & 0 \end{pmatrix},$$

$$\text{var}(\tilde{y}) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 & 0 & \\ 0 & s_1^2 & 0 & 0 & 0 \\ 0 & 0 & s_2^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\text{var}(\hat{\tilde{\beta}}) = D_1^+ \text{ var}(\tilde{y})(D_1^+)^{\mathsf{T}} = \sigma^2 D_1^+ D_2 D_2^{\mathsf{T}} (D_1^+)^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & s_1^2/c_1^2 & 0 \\ 0 & 0 & s_2^2/c_2^2 \end{pmatrix} = \sigma^2 D_0.$$

For the original parameters we obtain

$$\beta = U_1 \tilde{\beta}, \quad \hat{\beta} = U_1 \hat{\tilde{\beta}}, \quad \text{var}(\hat{\beta}) = U_1 \text{ var}(\hat{\tilde{\beta}}) U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}}.$$

The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \tfrac{1}{2}(\tilde{y}_4^2 + \tilde{y}_5^2).$$

Example 7: General linear model with rank deficient *X* and *W*

In this example we consider again the general linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$,

but now both matrices *X* and *W* are rank deficient, i.e. $\text{rk}(X) < p$ and $\text{rk}(W) < n$. Let

$$X = (x_1, x_2, x_3) = (5 \times 3) = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix}$$

as in Example 5. We have $x_3 = 2x_1 - x_2$, i.e. the third column of *X* is a linear combination of the first two columns, and so the matrix *X* has rank 2. Let

$$W = (5 \times 5) = \begin{pmatrix} 9.3 & -5.3 & 2.65 & -3.5 & -0.5 \\ * & 7.8 & -3.65 & -1.15 & 0.85 \\ * & * & 7.75 & 6.8 & 1.05 \\ * & * & * & 10.15 & 1.45 \\ * & * & * & * & 9.25 \end{pmatrix}$$

as in Example 6. The singular values of *W* are

$$\begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_5 \end{pmatrix} = \begin{pmatrix} 17.462 \\ 15.181 \\ 8.812 \\ 2.796 \\ 0 \end{pmatrix},$$

and so *W* is a symmetric and positive semidefinite matrix with rank 4. Now we consider the general linear model (1) with $X = (n \times p)$ and $W = (n \times n)$ as given above. As $rk(W) = 4$ the inverse W^{-1} does not exist and so the classical Aitken procedure as well as the procedure with the simple singular value decomposition cannot work. So we apply the procedure with the generalized singular value decomposition.

(i) *Factorization of W*

We want to find a matrix $F = (n \times k)$ such that $W = FF^\top$ where $k = rk(W) = 4$. The eigenvalue decomposition of *W* is given by $W = R\Lambda R^\top$, where *R* is orthogonal and $\Lambda = (5 \times 5) = \text{diag}(\lambda_1, \ldots, \lambda_5)$, $\lambda_i = \sigma_i$ (as *W* is positive semidefinite). As $\sigma_5 = \lambda_5 = 0$ we set

$$D = \left(5 \times 4\right) = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_4}) \quad \text{and} \quad F = (5 \times 4) = RD.$$

As $DD^\top = (5 \times 5) = \text{diag}(\lambda_1, \ldots, \lambda_4, 0) = \Lambda$ we have $FF^\top = RDD^\top R^\top = R\Lambda R^\top = W$. The random error of our model (1) can now be written in the form

$$\varepsilon = Fu \ \text{ with } \ u \sim (0, \sigma^2 I_k), \ \text{ i.e. with } \ \text{E}(u) = 0 \text{ and } \text{var}(u) = \sigma^2 I_k,$$

as $\text{E}(\varepsilon) = \text{E}(Fu) = 0$ and $\text{var}(\varepsilon) = \text{E}(\varepsilon\varepsilon^\top) = F\,\text{E}(uu^\top)F^\top = \sigma^2 FF^\top = \sigma^2 W$. So model (1) can also be given as

(2) $y = X\beta + Fu$, where $X = (n \times p), F = (n \times k), u \sim \left(0, \sigma^2 I_k\right)$.

(ii) *Singular value decomposition of* $(X \mid F)$

We have $X = (n \times p) = (5 \times 3)$ and $F = (n \times k) = (5 \times 4)$ and so $(X \mid F) = (n \times m) = (5 \times 7)$, where $m = p + k = 7$. Now we compute the singular value decomposition of $(X \mid F)$:

$$(X \mid F) = P \Delta Q^{\mathsf{T}},$$

where $P = (5 \times 5)$ and $Q = (7 \times 7)$ are orthogonal and $\Delta = (5 \times 7) = \mathrm{diag}(\sigma_1, \ldots, \sigma_5)$ We obtain

$$P = \begin{pmatrix} 0.270 & -0.092 & 0.839 & -0.330 & 0.326 \\ -0.236 & -0.870 & -0.181 & -0.392 & 0.019 \\ 0.442 & 0.015 & 0.039 & -0.356 & -0.822 \\ 0.743 & -0.022 & -0.475 & -0.157 & 0.444 \\ 0.353 & -0.483 & 0.189 & 0.766 & -0.141 \end{pmatrix},$$

and the singular values are given in the following table.

Table 1: Singular values of $X, F, (X \mid F)$

	$X = (5 \times 3)$	$F = (5 \times 4)$	$(X \mid F)$
singular values	$\begin{pmatrix} 18.802 \\ 12.103 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 4.179 \\ 3.896 \\ 2.969 \\ 1.672 \end{pmatrix}$	$\begin{pmatrix} 19.214 \\ 12.424 \\ 3.639 \\ 2.503 \\ 1.093 \end{pmatrix}$

(iii) *CS-decomposition of Q*

As $r_c = \mathrm{rk}(X \mid F) = n = 5$ we have to determine the CS-decomposition of $Q = (m \times m) = (7 \times 7)$ with the format

$$Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right) = \left(\begin{array}{c|c} (p \times n) & (p \times (m-n)) \\ \hline (k \times n) & (k \times (m-n)) \end{array}\right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 2) \\ \hline (4 \times 5) & (4 \times 2) \end{array}\right).$$

We obtain orthogonal matrices

$$U = (7 \times 7) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right) = \left(\begin{array}{c|c} (3 \times 3) & * \\ \hline * & (4 \times 4) \end{array}\right)$$

$$V = (8 \times 8) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right) = \left(\begin{array}{c|c} (5 \times 5) & * \\ \hline * & (2 \times 2) \end{array}\right)$$

with

$$U_1 = \begin{pmatrix} 0.569 & -0.100 & 0.816 \\ 0.446 & -0.797 & -0.408 \\ 0.691 & 0.596 & -0.408 \end{pmatrix}, \quad U_2 = \begin{pmatrix} -0.431 & 0 & 0 & 0.902 \\ -0.173 & -0.460 & 0.867 & -0.083 \\ 0.283 & -0.866 & -0.390 & 0.135 \\ 0.839 & 0.197 & 0.311 & 0.401 \end{pmatrix}$$

$$V_1 = \begin{pmatrix} 0.914 & 0.357 & 0.010 & 0.044 & -0.190 \\ -0.358 & 0.922 & -0.025 & -0.143 & -0.023 \\ 0.041 & -0.072 & 0.721 & -0.685 & -0.061 \\ -0.033 & 0.117 & 0.679 & 0.678 & 0.253 \\ 0.186 & 0.058 & -0.134 & -0.220 & 0.946 \end{pmatrix}, \quad V_2 = \begin{pmatrix} -0.641 & 0.768 \\ -0.768 & -0.641 \end{pmatrix}$$

such that

$$U^{\mathsf{T}} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right) = \left(\begin{array}{c|c} (3 \times 5) & (3 \times 2) \\ \hline (4 \times 5) & (4 \times 2) \end{array}\right)$$

where

$$D = \left(\begin{array}{ccccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & 0 & s_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & s_1 & 0 & 0 & 0 & -c_1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right).$$

The diagonal elements c_1 and s_1 are given in Table 2. Note that $c_1^2 + s_1^2 = 1$, and so the matrix *D* is orthogonal, too. Also note D_{21} and D_{12} are not classical diagonal matrices as the diagonal starts in the lower right corner and not in the upper left one.

Table 2: Diagonal elements c_1, s_1

i	c_i	s_i
1	0.981 975	0.189 010

(iv) *Generalized singular value decomposition and canonical model*

Now we have

$$(X\,|\,F) = P\Delta Q^{\mathsf{T}} \quad \text{and} \quad Q = UDV^{\mathsf{T}},$$

and from this we find the so-called *generalized singular value decomposition* of the pair X, F

$$\text{(3)} \qquad \begin{aligned} P^{\mathsf{T}} X U_1 &= \Delta_0 V_1 D_{11}^{\mathsf{T}}, \\ P^{\mathsf{T}} F U_2 &= \Delta_0 V_1 D_{21}^{\mathsf{T}}, \end{aligned}$$

where $\Delta_0 = (5 \times 5) = \text{diag}(\sigma_1, \ldots, \sigma_5)$ and where $\sigma_1, \ldots, \sigma_5$ are the singular values of $(X\,|\,F)$ as given above. Our model (2) can now be written in the canonical form

(4) $\quad \tilde{y} = D_1 \tilde{\beta} + D_2 \tilde{u}$,

where

$$\tilde{y} = V_1^{\mathsf{T}} \Delta_0^{-1} P^{\mathsf{T}} y,$$
$$\tilde{\beta} = U_1^{\mathsf{T}} \beta,$$
$$\tilde{u} = U_2^{\mathsf{T}} u,$$
$$D_1 = D_{11}^{\mathsf{T}} = (5 \times 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$D_2 = D_{21}^{\mathsf{T}} = (5 \times 4) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ s_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

The generalized singular value decomposition (3) shows that $\text{rk}(X) = \text{rk}(D_{11}) = \text{rk}(D_1) = 2$ and $\text{rk}(F) = \text{rk}(D_{21}) = \text{rk}(D_2) = 4$. The canonical model (4) explicitly written has the form

$$(5)\qquad \begin{aligned}\tilde{y}_1 &= \tilde{\beta}_1,\\ \tilde{y}_2 &= c_1\tilde{\beta}_2 + s_1\tilde{u}_1,\\ \tilde{y}_3 &= \tilde{u}_2,\\ \tilde{y}_4 &= \tilde{u}_3,\\ \tilde{y}_5 &= \tilde{u}_4.\end{aligned}$$

The observation $\tilde{y}_1$ is identical to the parameter $\tilde{\beta}_1$, this observation has no random error. As the covariance matrix $W = (5\times 5)$ has rank 4 there exists a linear combination $\tilde{y}_1$ of the original observations ($y_1,\ldots,y_n$) with no random error, and there exists a linear combination $\tilde{\beta}_1$ of the original parameters ($\beta_1,\ldots,\beta_p$) such that $\tilde{y}_1 = \tilde{\beta}_1$. The parameter $\tilde{\beta}_3$ can have arbitrary values as it does not show up in the canonical model, and we set this parameter to zero (minimum length definition). The least squares estimators are given by

$$\begin{aligned}\hat{\tilde{\beta}}_1 &= \tilde{y}_1,\\ \hat{\tilde{\beta}}_2 &= \tilde{y}_2/c_1,\\ \hat{\tilde{\beta}}_3 &= 0.\end{aligned}$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^+\tilde{y} \quad \text{as} \quad D_1^+ = \begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1/c_1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{pmatrix},$$

$$\operatorname{var}(\tilde{y}) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2\begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 0 & s_1^2 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

$$\operatorname{var}(\hat{\tilde{\beta}}) = D_1^+\operatorname{var}(\tilde{y})(D_1^+)^{\mathsf{T}} = \sigma^2 D_1^+ D_2 D_2^{\mathsf{T}}(D_1^+)^{\mathsf{T}} = \sigma^2\begin{pmatrix}0 & 0 & 0\\ 0 & s_1^2/c_1^2 & 0\\ 0 & 0 & 0\end{pmatrix} = \sigma^2 D_0 .$$

For the original parameters we obtain

$$\beta = U_1\tilde{\beta},\quad \hat{\beta} = U_1\hat{\tilde{\beta}},\quad \operatorname{var}(\hat{\beta}) = U_1\operatorname{var}(\hat{\tilde{\beta}})U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}} .$$

The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \tfrac{1}{3}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2).$$

Remark

In our example we have $n = 5$, $p = 3$ and

$$\begin{aligned}r_X &= \operatorname{rk}(X) = 2 && < p = 3\\ r_F &= \operatorname{rk}(F) = \operatorname{rk}(W) = k = 4 && < n = 5\\ r_c &= \operatorname{rk}(X\,|\,F) = 5 && = n\\ r &= r_X + r_F - r_c = 1 && > 0\end{aligned}$$

and this is the most general case of a general linear model with $r_c = n$. In the canonical model (5) we have three categories of observations:

(a) observations with no random error, that are identical to a parameter ($\tilde{y}_1$ in the example); number of these observations: $r_X - r = n - r_F = n - r_W = 5 - 4 = 1$;

(b) "classical" observations, that depend on the parameters and possess a random error ($\tilde{y}_2$ in the example); number of these observations: $r = 1$;

(c) observations, that do not depend on the parameters and possess a random error ($\tilde{y}_3$, $\tilde{y}_4$, and $\tilde{y}_5$ in the example); number of these observations: $r_F - r = 4 - 1 = 3$.

Furthermore we have three categories of parameters:

(α) parameters, that are completely fixed by the observations ($\tilde{\beta}_1$ in the example); number of these parameters: $r_X - r = 2 - 1 = 1$;

(β) "classical" parameters, that can be estimated with a random error ($\tilde{\beta}_2$ in the example); number of these parameters: $r = 1$;

(γ) parameters that do not show up in the canonical model ($\tilde{\beta}_3$ in the example); these parameters can have arbitrary values and they can be set to zero in order to make all parameters identifiable (minimum length definition); number of these parameters: $p - r_X = 3 - 2 = 1$.

Final remark

If the matrices X and W are nearly rank deficient the rank-k approximation should be applied as described in Example 9 so that the ranks become numerically stable in the sense that small perturbations of the matrix elements cannot reduce the rank further.

Example 8: General linear model with rank deficient *X*, *W*, and (*X* | *F*)

In this example we consider again the general linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$,

but now both matrices *X* and *W* are rank deficient, i.e. $\text{rk}(X) < p$ and $\text{rk}(W) < n$, and in addition $\text{rk}(X \mid F) < n$. Let

$$X = (x_1, x_2, x_3) = (5 \times 3) = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix}$$

as in Example 5. We have $x_3 = 2x_1 - x_2$, i.e. the third column of *X* is a linear combination of the first two columns, and so the matrix *X* has rank 2. Let

$$F_0 = (5 \times 3) = \begin{pmatrix} 2 & 4.5 & 4 \\ 0.5 & -3.5 & -7 \\ 3.5 & -1.5 & 7 \\ 2 & -4.5 & 12 \\ -3 & -4 & 4 \end{pmatrix};$$

note that the last column of F_0 is identical to the last column of *X*. The singular values of *X*, *F*, and $(X \mid F)$ are given in the following table.

Table 1: Singular values of $X, F_0, (X \mid F_0), W$

	$X = (5 \times 3)$	$F_0 = (5 \times 5)$	$(X \mid F_0)$	W
singular values	$\begin{pmatrix} 18.802 \\ 12.103 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 16.939 \\ 8.211 \\ 4.487 \end{pmatrix}$	$\begin{pmatrix} 25.216 \\ 13.519 \\ 6.153 \\ 4.248 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 286.942 \\ 67.425 \\ 20.133 \\ 0 \\ 0 \end{pmatrix}$

Now we define

$$W = F_0 F_0^{\mathsf{T}} = (5 \times 5) = \begin{pmatrix} 40.25 & -42.75 & 28.25 & 31.75 & -8 \\ * & 61.5 & -42 & -67.25 & -15.5 \\ * & * & 63.5 & 97.75 & 23.5 \\ * & * & * & 168.25 & 60 \\ * & * & * & * & 41 \end{pmatrix}.$$

W_0 is a symmetric and positive semidefinite matrix with rank 3 as $\text{rk}(W) = \text{rk}(F) = 3$. The singular values of *W* are given in Table 1. We now consider the general linear model (1) with *X* and *W* as given above. As $rk(W) = 3 < n$ the inverse W^{-1} does not exist and so the classical Aitken procedure as well as the procedure with the simple singular value decomposition cannot work. So we apply the procedure with the generalized singular value decomposition.

(i) *Factorization of W*

We want to find a matrix $F = (n \times k)$ such that $W = FF^{\mathsf{T}}$ where $k = rk(W) = 3$. The eigenvalue decomposition of *W* is given by $W = R\Lambda R^{\mathsf{T}}$, where *R* is orthogonal and $\Lambda = (5 \times 5) = \text{diag}(\lambda_1, \ldots, \lambda_5)$, $\lambda_i = \sigma_i$ (as *W* is positive semidefinite). As the two smallest eigenvalues are zero we set

(2) $D = (5 \times 3) = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3})$ and $F = (5 \times 3) = RD$.

We obtain

$$F = RD \approx \begin{pmatrix} -3.444 & 5.327 & 0.093 \\ 6.155 & -4.091 & -2.623 \\ -7.602 & 0.347 & -2.365 \\ -12.748 & -2.294 & -0.695 \\ -4.112 & -4.114 & 2.677 \end{pmatrix}.$$

Note that F is different from F_0, but the singular values of F and F_0 are the same as the squared values are the singular values of W (compare Table 1 and Table 2). As $DD^{\mathsf{T}} = (5 \times 5) = \Lambda$ we have $FF^{\mathsf{T}} = RDD^{\mathsf{T}}R^{\mathsf{T}} = R\Lambda R^{\mathsf{T}} = W$. The random error of our model (1) can now be written in the form

$$\varepsilon = Fu \text{ with } u \sim (0, \sigma^2 I_k), \text{ i.e. with } \mathrm{E}(u) = 0 \text{ and } \mathrm{var}(u) = \sigma^2 I_k,$$

as $\mathrm{E}(\varepsilon) = \mathrm{E}(Fu) = 0$ and $\mathrm{var}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon^{\mathsf{T}}) = F\,\mathrm{E}(uu^{\mathsf{T}})F^{\mathsf{T}} = \sigma^2 FF^{\mathsf{T}} = \sigma^2 W$. So model (1) can also be given as

(3) $\quad y = X\beta + Fu, \text{ where } X = (n \times p), F = (n \times k), u \sim \left(0, \sigma^2 I_k\right).$

with $X = (5 \times 3)$ and $F = (5 \times 3)$ as given above.

Table 2: Singular values of $X, F, (X \mid F)$

	$X = (5 \times 3)$	$F = (5 \times 5)$	$(X \mid F)$
singular values	$\begin{pmatrix} 18.802 \\ 12.103 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 16.939 \\ 8.211 \\ 4.487 \end{pmatrix}$	$\begin{pmatrix} 25.216 \\ 13.519 \\ 6.153 \\ 4.248 \\ 0 \end{pmatrix}$

(ii) *Singular value decomposition of* $(X \mid F)$

We have $X = (n \times p) = (5 \times 3)$ and $F = (n \times k) = (5 \times 3)$ and so $(X \mid F) = (n \times m) = (5 \times 6)$, where $m = p + k = 6$. Now we compute the singular value decomposition of $(X \mid F)$:

$$(X \mid F) = P\Delta Q^{\mathsf{T}},$$

where $P = (5 \times 5)$ and $Q = (7 \times 7)$ are orthogonal and $\Delta = (5 \times 7) = \mathrm{diag}(\sigma_1, \ldots, \sigma_5)$. We obtain

$$P = \begin{pmatrix} 0.239 & 0.061 & 0.889 & -0.358 & -0.144 \\ -0.308 & -0.835 & 0.256 & 0.347 & -0.147 \\ 0.441 & -0.021 & 0.200 & 0.498 & 0.719 \\ 0.752 & -0.140 & -0.192 & 0.230 & -0.570 \\ 0.298 & -0.528 & -0.260 & -0.671 & 0.339 \end{pmatrix}$$

and the singular values $\sigma_1, \ldots, \sigma_5$ are given in Table 2.

(iii) *CS-decomposition of* Q

Now $r_c = \mathrm{rk}(X \mid F) = 4 < n$ and therefore we will find in our canonical model an additional category of observations. First we have to determine the CS-decomposition of $Q = (m \times m) = (6 \times 6)$ with the format

$$Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right) = \left(\begin{array}{c|c} (p \times r_c) & (p \times (m - r_c)) \\ \hline (k \times r_c) & (k \times (m - r_c)) \end{array}\right) = \left(\begin{array}{c|c} (3 \times 4) & (3 \times 2) \\ \hline (3 \times 4) & (3 \times 2) \end{array}\right).$$

We obtain orthogonal matrices

$$U = (6\times 6) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right) = \left(\begin{array}{c|c} (3\times 3) & * \\ \hline * & (3\times 3) \end{array}\right)$$

$$V = (6\times 6) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right) = \left(\begin{array}{c|c} (4\times 4) & * \\ \hline * & (2\times 2) \end{array}\right)$$

with

$$U_1 = \begin{pmatrix} 0.447 & -0.365 & -0.816 \\ -0.894 & 0.183 & 0.408 \\ 0 & -0.913 & 0.408 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0.974 & -0.226 & 0 \\ -0.124 & -0.535 & -0.836 \\ -0.189 & -0.814 & 0.549 \end{pmatrix}$$

$$V_1 = \begin{pmatrix} -0.209 & -0.965 & 0.156 & 0.009 \\ 0.863 & -0.236 & -0.281 & -0.347 \\ -0.419 & 0.044 & -0.240 & -0.875 \\ 0.191 & 0.104 & 0.916 & -0.338 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0.690 & 0.724 \\ 0.724 & -0.690 \end{pmatrix}$$

such that

$$U^{\mathsf T} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right) = \left(\begin{array}{c|c} (3\times 4) & (3\times 2) \\ \hline (3\times 4) & (3\times 2) \end{array}\right)$$

where

$$D = \left(\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & s_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & s_1 & 0 & 0 & -c_1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right).$$

The diagonal elements c_1 and s_1 are given in Table 3. Note that $c_1^2 + s_1^2 = 1$, and so the matrix D is orthogonal, too. Also note D_{21} and D_{12} are not classical diagonal matrices as the diagonal starts in the lower right corner and not in the upper left one.

Table 3: Diagonal elements c_1, s_1

i	c_i	s_i
1	0.738 549	0.674 200

(iv) *Generalized singular value decomposition and canonical model*

Now we have

$$(X \mid F) = P\Delta Q^{\mathsf T} \quad \text{and} \quad Q = UDV^{\mathsf T},$$

and from this we find the so-called generalized singular value decomposition of the pair X, F

(4)
$$P^{\mathsf T} X U_1 = \left(\frac{\Delta_0}{0}\right) V_1 D_{11}^{\mathsf T},$$
$$P^{\mathsf T} F U_2 = \left(\frac{\Delta_0}{0}\right) V_1 D_{21}^{\mathsf T},$$

where $\Delta_0 = (4\times 4) = \mathrm{diag}(\sigma_1, \ldots, \sigma_4)$; $\sigma_1, \ldots, \sigma_4$ are the positive singular values of $(X \mid F)$ as given above. Our model (3) can now be written in the canonical form

(5) $\tilde{y} = \left(\frac{\tilde{y}_0}{0}\right) = \left(\frac{(4\times 1)}{(1\times 1)}\right)$, where $\tilde{y}_0 = D_1\tilde{\beta} + D_2\tilde{u}$

with

$$\tilde{y}_0 = V_1^{\mathsf{T}} \Delta_0^{-1} \bar{y}_0,\ \ \bar{y} = P^{\mathsf{T}} y = \begin{pmatrix} \bar{y}_0 \\ 0 \end{pmatrix} = \begin{pmatrix} (4\times 1) \\ (1\times 1) \end{pmatrix},$$

$$\tilde{\beta} = U_1^{\mathsf{T}} \beta,$$

$$\tilde{u} = U_2^{\mathsf{T}} u,$$

$$D_1 = D_{11}^{\mathsf{T}} = (4\times 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad D_2 = D_{21}^{\mathsf{T}} = (4\times 3) = \begin{pmatrix} 0 & 0 & 0 \\ s_1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The generalized singular value decomposition (4) shows that $\operatorname{rk}(X) = \operatorname{rk}(D_{11}) = \operatorname{rk}(D_1) = 2$ and $\operatorname{rk}(F) = \operatorname{rk}(D_{21}) = \operatorname{rk}(D_2) = 3$. The canonical model (5) explicitly written has the form

$$\begin{aligned} &\tilde{y}_1 = \tilde{\beta}_1, \\ &\tilde{y}_2 = c_1 \tilde{\beta}_2 + s_1 \tilde{u}_1, \\ (6) \qquad &\tilde{y}_3 = \tilde{u}_2, \\ &\tilde{y}_4 = \tilde{u}_3, \\ &\tilde{y}_5 = 0. \end{aligned}$$

We have $\tilde{y}_1 = \tilde{\beta}_1$ and $\tilde{y}_5 = 0$; these two observations have no random error. As the covariance matrix $W_1 = (5\times 5)$ has rank 3 there exists two independent linear combinations $\tilde{y}_1$ and $\tilde{y}_5$ of the original observations ($y_1, \ldots, y_n$) with no random error. The parameter $\tilde{\beta}_3$ can have arbitrary values as it does not show up in the canonical model, and we set this parameter to zero (minimum length definition). The least squares estimators are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1,$$

$$\hat{\tilde{\beta}}_2 = \tilde{y}_2 / c_1,$$

$$\hat{\tilde{\beta}}_3 = 0.$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^{+} \tilde{y}_0 \quad \text{as} \quad D_1^{+} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/c_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\operatorname{var}(\tilde{y}_0) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & s_1^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\operatorname{var}(\hat{\tilde{\beta}}) = D_1^{+} \operatorname{var}(\tilde{y}_0) (D_1^{+})^{\mathsf{T}} = \sigma^2 D_1^{+} D_2 D_2^{\mathsf{T}} (D_1^{+})^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & s_1^2 / c_1^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sigma^2 D_0 .$$

For the original parameters we obtain

$$\beta = U_1 \tilde{\beta}, \quad \hat{\beta} = U_1 \hat{\tilde{\beta}}, \quad \operatorname{var}(\hat{\beta}) = U_1 \operatorname{var}(\hat{\tilde{\beta}}) U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}} .$$

The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \tfrac{1}{2}(\tilde{u}_2^2 + \tilde{u}_3^2) = \tfrac{1}{2}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2) .$$

Remark

In our example we have $n=5$, $p=3$ and

$$\begin{array}{ll} r_X = \text{rk}(X) = 2 & < p = 3 \\ r_F = \text{rk}(F) = \text{rk}(W) = k = 3 & < n = 5 \\ r_c = \text{rk}(X \mid F) = 4 & < n = 5 \\ r = r_X + r_F - r_c = 1 & > 0 \end{array}$$

and this is the most general case of the general linear model.

In the canonical model (6) we have four categories of observations:

(a) observations with no random error, that are identical to a parameter ($\tilde{y}_1$ in the example); number of these observations: $r_X - r = 2 - 1 = 1$;

(b) "classical" observations, that depend on the parameters and possess a random error ($\tilde{y}_2$ in the example); number of these observations: $r = 1$;

(c) observations, that do not depend on the parameters and possess a random error ($\tilde{y}_3$ and $\tilde{y}_4$ in the example); number of these observations: $r_F - r = 3 - 1 = 2$;

(d) observations, that are identical to zero ($\tilde{y}_5$ in the example); number of these observations: $n - r_c = 5 - 4 = 1$.

The number of observations in categories (a) and (d) that possess no random error is

$$(r_X - r) + (n - r_c) = n - r_F = n - \text{rk}(W) = 5 - 3 = 2.$$

Furthermore we have three categories of parameters:

(α) parameters, that are completely fixed by the observations ($\tilde{\beta}_1$ in the example); number of these parameters: $r_X - r = 2 - 1 = 1$;

(β) "classical" parameters, that can be estimated with a random error ($\tilde{\beta}_2$ in the example); number of these parameters: $r = 1$;

(γ) parameters that do not show up in the canonical model ($\tilde{\beta}_3$ in the example); these parameters can have arbitrary values and they can be set to zero in order to make all parameters identifiable (minimum length definition); number of these parameters: $p - r_X = 3 - 2 = 1$.

Example 9: General linear model with nearly rank deficient *X, W,* and (*X* | *F*)

In this example we consider again the general linear model

(1) $y = X\beta + \varepsilon$, where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$,

but now both matrices *X* and *W* are nearly rank deficient, and in addition $(X \mid F)$ is nearly rank deficient. Let

$$X_0 = (x_1, x_2, x_3) = (5 \times 3) = \begin{pmatrix} 3 & 2 & 4 \\ 1 & 9 & -7 \\ 4 & 1 & 7 \\ 7 & 2 & 12 \\ 5 & 6 & 4 \end{pmatrix}$$

X_0 is the same as *X* in Example 8. We have $x_3 = 2x_1 - x_2$, i.e. the third column of X_0 is a linear combination of the first two columns, and so the matrix X_0 has rank 2. Now let

$$X = (5 \times 3) = \begin{pmatrix} 3.0002 & 1.9996 & 4.0002 \\ 1.0005 & 8.9998 & -6.9996 \\ 3.9999 & 0.9999 & 7.0002 \\ 6.9995 & 1.9995 & 11.9998 \\ 5.0003 & 5.9997 & 3.9999 \end{pmatrix}.$$

If $X = (x_{ij})$ and $X_0 = (x_{ij}^{(0)})$ we have $\max\left|x_{ij} - x_{ij}^{(0)}\right| = 0.0005$ and so $X \approx X_0$, but *X* has full rank 3 whereas X_0 has rank 2 (see the list of singular values below). So *X* is nearly rank deficient. Let

$$F_0 = (5 \times 3) = \begin{pmatrix} 2 & 4.5 & 4 \\ 0.5 & -3.5 & -7 \\ 3.5 & -1.5 & 7 \\ 2 & -4.5 & 12 \\ -3 & -4 & 4 \end{pmatrix}.$$

F_0 is the same as *F* in Example 8. Note that the last column of F_0 is identical to the last column of X_0. Now let

$$W_0 = F_0 F_0^{\mathsf{T}} = (5 \times 5) = \begin{pmatrix} 40.25 & -42.75 & 28.25 & 31.75 & -8 \\ * & 61.5 & -42 & -67.25 & -15.5 \\ * & * & 63.5 & 97.75 & 23.5 \\ * & * & * & 168.25 & 60 \\ * & * & * & * & 41 \end{pmatrix}$$

$$W = (5 \times 5) = \begin{pmatrix} 40.251 & -42.751 & 28.249 & 31.747 & -7.997 \\ * & 61.502 & -42.001 & -67.248 & -15.501 \\ * & * & 63.499 & 97.751 & 23.497 \\ * & * & * & 168.253 & 59.999 \\ * & * & * & * & 40.995 \end{pmatrix}$$

If $W = (w_{ij})$ and $W_0 = (w_{ij}^{(0)})$ we have $\max\left|w_{ij} - w_{ij}^{(0)}\right| = 0.005$ and so $W \approx W_0$, but *W* has full rank 5 whereas W_0 has rank 3, the same as F_0 (see the list of singular values below). So also *W* is nearly rank deficient. The singular values of X_0, X, F_0, W_0, W are given in Table 1. The singular values of $W_0 = F_0 F_0^{\mathsf{T}}$ are the square of the corresponding singular values of F_0. Table 2 shows that the small perturbations added to *W* make the matrix indefinite as one of the five eigenvalues becomes negative.

Table 1: Singular values of X_0, X, F_0, W_0, W

	$X_0 = (5\times 3)$	$X = (5\times 3)$	$F_0 = (5\times 3)$	$W_0 = (5\times 5)$	$W = (5\times 5)$
Singular values	$\begin{pmatrix} 18.8018 \\ 12.1034 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 18.8014 \\ 12.1031 \\ 0.000\,328 \end{pmatrix}$	$\begin{pmatrix} 16.9394 \\ 8.2113 \\ 4.4870 \end{pmatrix}$	$\begin{pmatrix} 286.9422 \\ 67.4248 \\ 20.1330 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 286.9421 \\ 67.4244 \\ 20.1343 \\ 0.0024 \\ 0.0017 \end{pmatrix}$

Table 2: Eigenvalues of W_0, W

	$W_0 = (5\times 5)$	$W = (5\times 5)$
Eigenvalues	$\begin{pmatrix} 286.9422 \\ 67.4248 \\ 20.1330 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 286.9421 \\ 67.4244 \\ 20.1343 \\ 0.0017 \\ -0.0024 \end{pmatrix}$

(i) *Rank-k approximation and factorization of W*

Table 2 shows that all the eigenvalues of W_0 are non-negative and so W_0 is a positive semidefinite matrix, but this is not true for the matrix *W*; due to the small perturbations added to W_0 the rank of *W* has been enhanced from three to five and *W* has become indefinite. We want to find an approximation to *W* with a numerically stable rank. As the elements of *W* are given with three decimal places the two smallest singular values 0.0024 and 0.0017 are near the rounding error of *W*, we determine the rank-3 approximation of *W*. This approximation will give a matrix $W_1 = (w_{ij}^{(1)})$ with $\mathrm{rk}(W_1) = 3$ and $\max|w_{ij} - w_{ij}^{(1)}| \le \sigma_4 = 0.0024$. The singular value decomposition of *W* is given by $W = UDV^\top$ where $U = (5\times 5)$ and $V = (5\times 5)$ are orthogonal and $D = (5\times 5) = \mathrm{diag}(\sigma_1,\ldots,\sigma_5)$; $\sigma_1,\ldots,\sigma_5$ are the singular values of *W*. Now define W_1 as

(2) $W_1 = UD_1V^\top \ (= UD_1U^\top)$ where $D_1 = (5\times 5) = \mathrm{diag}(\sigma_1,\sigma_2,\sigma_3,0,0)$.

Obviously $\mathrm{rk}(W_1) = 3$ and we obtain $\max|w_{ij} - w_{ij}^{(1)}| = 0.000\,928 \le \sigma_4 = 0.0024$; so $W_1 \approx W$. The rank of W_1 is numerically stable in the sense that it cannot be reduced by small perturbations of the matrix elements $w_{ij}^{(1)}$. In a second step we have to factorize W_1 such that $W_1 = F_1F_1^\top$ with $F_1 = (5\times 3)$. Let $D_2 = (5\times 3) = \mathrm{diag}(\sqrt{\sigma_1},\sqrt{\sigma_2},\sqrt{\sigma_3})$ and $F_1 = UD_2$. Then $F_1F_1^\top = UD_2D_2^\top U^\top = UD_1U^\top = W_1$. So $W_1 = F_1F_1^\top$ and W_1 is positive semidefinite.

(ii) *Rank-k approximation of* $X = (n\times p)$

The elements of *X* are given with four decimal places and the smallest singular value 0.000 328 is near the rounding error of *X*, and so we determine the rank-2 approximation of *X*. This approximation will give a matrix $X_1 = (x_{ij}^{(1)})$ with $\mathrm{rk}(X_1) = 2$ and $\max|x_{ij} - x_{ij}^{(1)}| \le \sigma_3 = 0.000\,328$. The singular value decomposition of $X = (5\times 3)$ is given by $X = UDV^\top$ where $U = (5\times 5)$ and $V = (3\times 3)$ are orthogonal and $D = (5\times 3) = \mathrm{diag}(\sigma_1,\sigma_2,\sigma_3)$; $\sigma_1,\sigma_2,\sigma_3$ are the singular values of *X*. Now define X_1 as

(3) $X_1 = UD_1V^\top$ where $D_1 = (5\times 3) = \text{diag}(\sigma_1, \sigma_2, 0)$.

Obviously $\text{rk}(X_1) = 2$ and we obtain $\max|x_{ij} - x_{ij}^{(1)}| = 0.000\,152 \le \sigma_3 = 0.000\,328$; so $X_1 \approx X$.

(iii) *Rank-k approximation of* $(X_1 \mid F_1)$

Table 3 gives the singular values of $(X_1 \mid F_1)$. The smallest singular value of $(X_1 \mid F_1)$ is near the rounding error of *X*, and so we determine the rank-4 approximation of $(X_1 \mid F_1)$. The singular value

Table 3: Singular values of $X_1, F_1, (X_1 \mid F_1)$

	$X_1 = (5\times 3)$	$F_1 = (5\times 3)$	$(X_1 \mid F_1) = (5\times 6)$
singular values	$\begin{pmatrix} 18.8014 \\ 12.1031 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 16.9394 \\ 8.2112 \\ 4.4871 \end{pmatrix}$	$\begin{pmatrix} 25.2152 \\ 13.5189 \\ 6.1533 \\ 4.2484 \\ 0.000\,107 \end{pmatrix}$

decomposition of $(X_1 \mid F_1)$ is given by

(4) $(X_1 \mid F_1) = P\Delta Q^\top$

where $P = (5\times 5)$ and $Q = (6\times 6)$ are orthogonal and $\Delta = (5\times 6) = \text{diag}(\sigma_1, \ldots, \sigma_5)$; $\sigma_1, \ldots, \sigma_5$ are the singular values of $(X_1 \mid F_1)$. Now we define $(X_2 \mid F_2)$ as

(5) $(X_2 \mid F_2) = P\Delta_1 Q^\top$ where $\Delta_1 = (5\times 6) = \text{diag}(\sigma_1, \sigma_2, \sigma_3, \sigma_4, 0)$.

We have $\max\left|x_{ij}^{(1)} - x_{ij}^{(2)}\right| = 0.000\,047$ and $\max\left|f_{ij}^{(1)} - f_{ij}^{(2)}\right| = 0.000\,055$; so $X_1 \approx X_2$ and $F_1 \approx F_2$. The singular values of $(X_2 \mid F_2)$ are given in Table 4. Note that X_2 has the same rank as X_1; it is remarkable that a rank-k approximation of $(X_1 \mid F_1)$ cannot raise the rank of X_1.

Now we have $\max\left|x_{ij} - x_{ij}^{(2)}\right| = 0.000\,152$ and $\max\left|w_{ij} - w_{ij}^{(2)}\right| = 0.001\,725$, where $W_2 = F_2F_2^\top$. Instead of the original linear model (1) we now consider the approximate model

(6) $y = X_2\beta + \varepsilon$, where $X_2 = (n\times p) \approx X$ and $\varepsilon \sim \left(0, \sigma^2 W_2\right), W_2 = F_2F_2^\top \approx W$,

and this model is equivalent to

(7) $y = X_2\beta + F_2u$, where $X_2 = (n\times p), F_2 = (n\times k)$, and $u \sim (0, \sigma^2 I_k)$

with $n = 5, p = 3, k = 3$. The singular values of X_2, F_2 and $(X_2 \mid F_2)$ are given in Table 4; we are now in exactly the same situation as in Example 8 (see Table 2 there), and we now proceed as there.

Table 4: Singular values of $X_2, F_2, (X_2 \mid F_2)$

	$X_2 = (5\times 3)$	$F_2 = (5\times 3)$	$(X_2 \mid F_2) = (5\times 6)$
singular values	$\begin{pmatrix} 18.8014 \\ 12.1031 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 16.9394 \\ 8.2112 \\ 4.4871 \end{pmatrix}$	$\begin{pmatrix} 25.2152 \\ 13.5189 \\ 6.1533 \\ 4.2484 \\ 0 \end{pmatrix}$

(iv) *CS-decomposition of Q*

The singular value decomposition of $(X_2 \mid F_2)$ is given in (5), and we have $r_c = \mathrm{rk}(X_2 \mid F_2) = 4$. Now we have to determine the CS-decomposition of $Q = (m \times m) = (6 \times 6)$ with the format

$$\left(\begin{array}{c|c} (p \times r_c) & (p \times (m-r_c)) \\ \hline (k \times r_c) & (k \times (m-r_c)) \end{array}\right) = \left(\begin{array}{c|c} (3 \times 4) & (3 \times 2) \\ \hline (3 \times 4) & (3 \times 2) \end{array}\right).$$

We obtain orthogonal matrices

$$U = (6 \times 6) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right) = \left(\begin{array}{c|c} (3 \times 3) & * \\ \hline * & (3 \times 3) \end{array}\right)$$

$$V = (6 \times 6) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right) = \left(\begin{array}{c|c} (4 \times 4) & * \\ \hline * & (2 \times 2) \end{array}\right)$$

with

$$U_1 = \begin{pmatrix} -0.447 & -0.365 & -0.816 \\ -0.894 & 0.183 & 0.408 \\ -0.000 & -0.913 & 0.408 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0.974 & 0 & -0.226 \\ -0.124 & -0.835 & -0.535 \\ -0.189 & 0.550 & -0.814 \end{pmatrix}$$

$$V_1 = \begin{pmatrix} -0.209 & -0.965 & 0.009 & 0.157 \\ -0.863 & 0.236 & 0.347 & 0.281 \\ 0.419 & -0.043 & 0.875 & 0.240 \\ 0.191 & 0.104 & -0.338 & 0.916 \end{pmatrix}, \quad V_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

such that

(8) $$U^{\mathsf{T}} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right) = \left(\begin{array}{c|c} (3 \times 4) & (3 \times 2) \\ \hline (3 \times 4) & (3 \times 2) \end{array}\right)$$

where

(9) $$D = \left(\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & s_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & s_1 & 0 & 0 & -c_1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right);$$

the diagonal elements c_1 and s_1 are given in Table 5. Note that $c_1^2 + s_1^2 = 1$ for $i = 1,2$, and so the matrix D is orthogonal, too.

Table 5: Diagonal elements c_1, s_1

i	c_i	s_i
1	0.738 539	0.674 211

(v) *Generalized singular value decomposition and canonical model*

Now we have

(10) $(X_2 \mid F_2) = P \Delta_1 Q^{\mathsf{T}}$ and $Q = UDV^{\mathsf{T}}$,

and from this we find the so-called *generalized singular value decomposition* of X_2, F_2

$$P^{\mathsf{T}} X_2 U_1 = \begin{pmatrix} \Delta_0 \\ \hline 0 \end{pmatrix} V_1 D_{11}^{\mathsf{T}},$$

$$P^{\mathsf{T}} F_2 U_2 = \begin{pmatrix} \Delta_0 \\ \hline 0 \end{pmatrix} V_1 D_{21}^{\mathsf{T}},$$

where $\Delta_0 = (4\times 4) = \mathrm{diag}(\sigma_1,\ldots,\sigma_4)$ is the reduced form of $\Delta_1 = (5\times 6) = \mathrm{diag}(\sigma_1,\ldots,\sigma_4,0)$. Our model (7) can now be written in the canonical form

$$(11)\quad \tilde{y} = \begin{pmatrix}\tilde{y}_0\\0\end{pmatrix} = \begin{pmatrix}(4\times 1)\\(1\times 1)\end{pmatrix},\ \text{where } \tilde{y}_0 = D_1\tilde{\beta} + D_2\tilde{u},$$

with

$$\tilde{y}_0 = V_1^{\mathsf{T}}\Delta_0^{-1}\,\overline{y}_0,\ \overline{y} = P^{\mathsf{T}} y = \begin{pmatrix}\overline{y}_0\\0\end{pmatrix}$$

$$\tilde{\beta} = U_1^{\mathsf{T}}\beta,$$

$$\tilde{u} = U_2^{\mathsf{T}}u,$$

$$D_1 = D_{11}^{\mathsf{T}} = (4\times 3) = \begin{pmatrix}1&0&0\\0&c_1&0\\0&0&0\\0&0&0\end{pmatrix},\quad D_2 = D_{21}^{\mathsf{T}} = (4\times 3) = \begin{pmatrix}0&0&0\\s_1&0&0\\0&1&0\\0&0&1\end{pmatrix},$$

$$P = \begin{pmatrix}0.239&-0.061&-0.889&-0.358&-0.144\\-0.308&0.835&-0.256&0.347&-0.147\\0.441&0.021&-0.200&0.498&0.719\\0.752&0.140&0.192&0.230&-0.570\\0.298&0.528&0.260&-0.671&0.339\end{pmatrix},$$

and so

$$\begin{aligned}&\tilde{y}_1 = \tilde{\beta}_1,\\&\tilde{y}_2 = c_1\tilde{\beta}_2 + s_1\tilde{u}_1,\\(12)\quad&\tilde{y}_3 = \tilde{u}_2,\\&\tilde{y}_4 = \tilde{u}_3,\\&\tilde{y}_5 = 0.\end{aligned}$$

We have $\tilde{y}_1 = \tilde{\beta}_1$ and $\tilde{y}_5 = 0$; these two observations have no random error. As the covariance matrix $W_1 = (5\times 5)$ has rank 3 there exists two independent linear combinations $\tilde{y}_1$ and $\tilde{y}_5$ of the original observations ($y_1,\ldots,y_n$) with no random error. The parameter $\tilde{\beta}_3$ can have arbitrary values as it does not show up in the canonical model, and we set this parameter to zero (minimum length definition). The least squares estimators are given by

$$\hat{\tilde{\beta}}_1 = \tilde{y}_1,$$

$$\hat{\tilde{\beta}}_2 = \tilde{y}_2/c_1,$$

$$\hat{\tilde{\beta}}_3 = 0.$$

In matrix notation we can write

$$\hat{\tilde{\beta}} = D_1^{+}\tilde{y}_0 \quad\text{as}\quad D_1^{+} = \begin{pmatrix}1&0&0&0\\0&1/c_1&0&0\\0&0&0&0\end{pmatrix},$$

$$\mathrm{var}(\tilde{y}_0) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2\begin{pmatrix}0&0&0&0\\0&s_1^2&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix},$$

$$\mathrm{var}(\hat{\tilde{\beta}}) = D_1^{+}\,\mathrm{var}(\tilde{y}_{\mathrm{red}})(D_1^{+})^{\mathsf{T}} = \sigma^2 D_1^{+}D_2D_2^{\mathsf{T}}(D_1^{+})^{\mathsf{T}} = \sigma^2\begin{pmatrix}0&0&0\\0&s_1^2/c_1^2&0\\0&0&0\end{pmatrix} = \sigma^2 D_0.$$

The unknown variance σ^2 is estimated by $\hat{\sigma}^2 = \frac{1}{2}(\tilde{u}_2^2 + \tilde{u}_3^2) = \frac{1}{2}(\tilde{y}_3^2 + \tilde{y}_4^2 + \tilde{y}_5^2)$. Note that the matrices P, U_1, U_2 and the diagonal elements c_1 and s_1 are essentially the same as in Example 8, and so the rank-k approximation allows to eliminate small perturbations in the original matrices *X* and *W* so that we can find matrices with numerically stable ranks in the sense that the ranks cannot be reduced further by small alterations of the matrix elements.

Example 10: General linear model with linear restrictions, regular case

In this example we consider the general linear model

(1) $\quad y = X\beta + \varepsilon,$ where $X = (n \times p)$ and $\varepsilon \sim \left(0, \sigma^2 W\right)$ with linear restrictions $L\beta = c$.

Here the error variables $\varepsilon_1, \ldots, \varepsilon_n$ are correlated with covariance matrix $\sigma^2 W$, where σ^2 is unknown and $W = (n \times n)$ is a known positive semidefinite matrix. $L = (r \times p)$ and $c = (r \times 1)$ are given, too. Let

(2) $$X = (5 \times 3) = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 9 & 8 \\ 4 & 1 & 4 \\ 7 & 2 & 0 \\ 5 & 6 & 5 \end{pmatrix}$$

as in Example 4. From Table 1 we see that all singular values of *X* are positive and so $\text{rk}(X) = 3$. Furthermore let

$$F_0 = (5 \times 5) = \begin{pmatrix} 4 & 9 & 8 & -5 & -6 \\ 1 & -7 & -9 & -5 & -7 \\ 7 & -3 & 9 & 4 & -9 \\ 4 & -9 & 5 & 9 & -8 \\ -6 & -8 & 7 & -6 & 1 \end{pmatrix}$$

and now we define

(3) $$W_0 = \tfrac{1}{17} F_0 F_0^{\mathsf{T}} = (5 \times 5) \approx \begin{pmatrix} 13.059 & -3.765 & 6.294 & -1.294 & -0.941 \\ * & 12.059 & -0.588 & 1.941 & 0.588 \\ * & * & 13.882 & 12.235 & 0.706 \\ * & * & * & 15.706 & 1.235 \\ * & * & * & * & 10.941 \end{pmatrix} = W.$$

W is the matrix W_0 rounded to three decimal places. The singular values of *W* are given in Table 1, they are all positive, and so the symmetric matrix *W* has full rank 5 and is positive definite. In addition we assume that the linear restriction

(4) $\quad \beta_1 - \beta_2 + \beta_3 = 17$

must be fulfilled. (4) can be written as

$$L\beta = c \text{ with } L = (1, -1, 1) \text{ and } c = (17).$$

The linear model (1) including the restriction (4) can be written as an extended general linear model

(5) $$y_e = X_e \beta_e + \varepsilon_e, \text{ where } y_e = \left(\frac{y}{c}\right),\ X_e = \left(\frac{X}{L}\right) \text{ and } \varepsilon_e = \left(\frac{\varepsilon}{0}\right) \sim \left(0, \sigma^2 W_e\right),\ W_e = \left(\begin{array}{c|c} W & 0 \\ \hline 0 & 0 \end{array}\right).$$

In the extended model we have

$$y_e = \begin{pmatrix} y \\ \hline c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ \hline 17 \end{pmatrix}, \quad X_e = \begin{pmatrix} X \\ \hline L \end{pmatrix} = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 9 & 8 \\ 4 & 1 & 4 \\ 7 & 2 & 0 \\ 5 & 6 & 5 \\ \hline 1 & -1 & 1 \end{pmatrix}, \quad \varepsilon_e = \begin{pmatrix} \varepsilon \\ \hline 0 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \hline 0 \end{pmatrix},$$

(6)

$$W_e = \left(\begin{array}{c|c} W & 0 \\ \hline 0 & 0 \end{array}\right) = \left(\begin{array}{ccccc|c} 13.059 & -3.765 & 6.294 & -1.294 & -0.941 & 0 \\ * & 12.059 & -0.588 & 1.941 & 0.588 & 0 \\ * & * & 13.882 & 12.235 & 0.706 & 0 \\ * & * & * & 15.706 & 1.235 & 0 \\ * & * & * & * & 10.941 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Note that $X_e = (n_e \times p) = (6 \times 3)$ and $W_e = (n_e \times n_e) = (6 \times 6)$ as $n_e = n + r = 5 + 1 = 6$.

Table 1: Singular values of X, X_e, W, W_e

	$X = (5\times3)$	$X_e = (6\times3)$	$W = (5\times5)$	$W_e = (6\times6)$
singular values	$\begin{pmatrix} 16.560 \\ 7.612 \\ 2.795 \end{pmatrix}$	$\begin{pmatrix} 16.565 \\ 7.665 \\ 3.136 \end{pmatrix}$	$\begin{pmatrix} 27.951 \\ 18.089 \\ 10.654 \\ 8.839 \\ 0.114 \end{pmatrix}$	$\begin{pmatrix} 27.951 \\ 18.089 \\ 10.654 \\ 8.839 \\ 0.114 \\ 0 \end{pmatrix}$

(i) *Factorization of* W_e

We want to find a matrix $F_e = (n_e \times k)$ such that $W_e = F_e F_e^\mathsf{T}$ where $k = rk(W_e) = 5$. The eigenvalue decomposition of W_e is given by $W_e = R\Lambda R^\mathsf{T}$, where R is orthogonal and $\Lambda = (6\times6) = \mathrm{diag}(\lambda_1,\dots,\lambda_6)$, $\lambda_i = \sigma_i$ (as W_e is positive semidefinite). As $\lambda_6 = \sigma_6 = 0$ we set

$$D = (6\times5) = \mathrm{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_5}) \quad \text{and} \quad F_e = (6\times5) = RD.$$

As $DD^\mathsf{T} = (6\times6) = \mathrm{diag}(\lambda_1,\dots,\lambda_5,0) = \Lambda$ we have $F_e F_e^\mathsf{T} = RDD^\mathsf{T} R^\mathsf{T} = R\Lambda R^\mathsf{T} = W_e$. The random error of our model can now be written in the form

$$\varepsilon_e = F_e u \;\text{ with }\; u \sim (0, \sigma^2 I_k), \;\text{ i.e. with }\; \mathrm{E}(u) = 0 \text{ and } \mathrm{var}(u) = \sigma^2 I_k,$$

as $\mathrm{E}(\varepsilon_e) = \mathrm{E}(F_e u) = 0$ and $\mathrm{var}(\varepsilon_e) = \mathrm{E}(\varepsilon_e \varepsilon_e^\mathsf{T}) = F_e\,\mathrm{E}(uu^\mathsf{T}) F_e^\mathsf{T} = \sigma^2 F_e F_e^\mathsf{T} = \sigma^2 W_e$. So model (5) is equivalent to

(7) $\quad y_e = X_e\beta + F_e u, \;$ where $X_e = (6\times3), F_e = (6\times5), u \sim \left(0, \sigma^2 I_5\right)$.

For F_e we obtain

$$F_e = (6\times5) = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

(ii) *Singular value decomposition of the matrix* $(X_e \mid F_e)$

The singular value decomposition of $(X_e \mid F_e) = (n_e \times m) = (6 \times 8)$ with $m = p + k = 3 + 5 = 8$ is given by

$$(X_e \mid F_e) = P \Delta Q^{\mathsf{T}},$$

where $P = (n_e \times n_e) = (6 \times 6)$ and $Q = (m \times m) = (8 \times 8)$ are orthogonal and $\Delta = (n_e \times m) = (6 \times 8) = \mathrm{diag}(\sigma_1, \ldots, \sigma_6)$, and where the singular values $\sigma_1, \ldots, \sigma_6$ of $(X_e \mid F_e)$ are given in the following table. We have $r_c = rk(X_e \mid F_e) = n_e = 6$, and this simplifies the further procedure.

Table 2: Singular values of X_e, F_e, $(X_e \mid F_e)$

	$X_e = (6 \times 3)$	$F_e = (6 \times 5)$	$(X_e \mid F_e) = (6 \times 8)$
singular values	$\begin{pmatrix} 16.565 \\ 7.665 \\ 3.136 \end{pmatrix}$	$\begin{pmatrix} 5.287 \\ 4.253 \\ 3.264 \\ 2.973 \\ 0.338 \end{pmatrix}$	$\begin{pmatrix} 17.013 \\ 8.838 \\ 4.607 \\ 3.685 \\ 3.277 \\ 1.047 \end{pmatrix}$

(iii) *CS-decomposition of Q*

As $r_c = \mathrm{rk}(X_e \mid F_e) = n_e = 6$ we have to determine the CS-decomposition of the orthogonal matrix $Q = (m \times m) = (8 \times 8)$ with the format

$$(8) \qquad Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right) = \left(\begin{array}{c|c} p \times n_e & p \times (m - n_e) \\ \hline k \times n_e & k \times (m - n_e) \end{array}\right) = \left(\begin{array}{c|c} 3 \times 6 & 3 \times 2 \\ \hline 5 \times 6 & 5 \times 2 \end{array}\right).$$

We obtain orthogonal matrices

$$U = (8 \times 8) = \left(\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right) = \left(\begin{array}{c|c} 3 \times 3 & * \\ \hline * & 5 \times 5 \end{array}\right)$$

$$V = (8 \times 8) = \left(\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right) = \left(\begin{array}{c|c} 6 \times 6 & * \\ \hline * & 2 \times 2 \end{array}\right)$$

with

$$U_1 = \begin{pmatrix} 0.577 & 0.816 & 0.005 \\ -0.577 & 0.404 & 0.709 \\ 0.577 & -0.412 & 0.705 \end{pmatrix}, \quad U_2 = \begin{pmatrix} -0.127 & -0.194 & -0.973 & 0 & 0 \\ 0.049 & 0.299 & -0.066 & -0.127 & 0.942 \\ 0.110 & 0.294 & -0.073 & 0.946 & 0.023 \\ 0.035 & 0.881 & -0.180 & -0.283 & -0.332 \\ 0.984 & -0.105 & -0.108 & -0.090 & -0.038 \end{pmatrix}$$

$$V_1 = \begin{pmatrix} -0.237 & -0.371 & -0.888 & -0.123 & -0.017 & -0.047 \\ -0.469 & -0.587 & 0.436 & -0.480 & -0.113 & 0.053 \\ 0.384 & -0.415 & 0.095 & 0.143 & -0.061 & -0.804 \\ 0.016 & -0.235 & 0.024 & 0.508 & -0.779 & 0.281 \\ -0.531 & 0.521 & -0.027 & -0.100 & -0.419 & -0.511 \\ 0.543 & 0.141 & -0.105 & -0.683 & -0.448 & 0.087 \end{pmatrix}, \quad V_2 = \begin{pmatrix} -0.081 & -0.997 \\ -0.997 & 0.081 \end{pmatrix}$$

such that

$$U^{\mathsf{T}} Q V = D = \left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right) = \left(\begin{array}{c|c} 3 \times 6 & 3 \times 2 \\ \hline 5 \times 6 & 5 \times 2 \end{array}\right)$$

where

$$D=\left(\begin{array}{cccccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & 0 & 0 & s_1 & 0 \\ 0 & 0 & c_2 & 0 & 0 & 0 & 0 & s_2 \\ \hline 0 & s_1 & 0 & 0 & 0 & 0 & -c_1 & 0 \\ 0 & 0 & s_2 & 0 & 0 & 0 & 0 & -c_2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right).$$

The diagonal elements c_i and s_i are given in the following table. Note that $c_i^2 + s_i^2 = 1$ for $i = 1, 2$, and so the matrix *D* is orthogonal, too.

Table 3: Diagonal elements c_i, s_i

i	c_i	s_i
1	0.994 856	0.101 301
2	0.978 298	0.207 204

(iv) *Generalized singular value decomposition and canonical model*

We now have

$$(9)\qquad \begin{aligned} (X_e \mid F_e) &= P\Delta Q^{\mathsf T}, \\ Q &= UDV^{\mathsf T}, \end{aligned}$$

and from this we find the so-called *generalized singular value decomposition* of the pair X_e, F_e

$$(10)\qquad \begin{aligned} P^{\mathsf T} X_e U_1 &= \Delta_0 V_1 D_{11}^{\mathsf T}, \\ P^{\mathsf T} F_e U_2 &= \Delta_0 V_1 D_{21}^{\mathsf T}, \end{aligned}$$

where $\Delta_0 = (6\times 6) = \mathrm{diag}(\sigma_1,\ldots,\sigma_6)$ and where $\sigma_1,\ldots,\sigma_6$ are the singular values of $(X_e \mid F_e)$ as given in Table 2. For $P = (6\times 6)$ we have

$$P=\begin{pmatrix} -0.264 & -0.157 & 0.705 & -0.205 & 0.591 & 0.131 \\ -0.666 & 0.588 & -0.113 & 0.417 & 0.113 & 0.108 \\ -0.314 & -0.368 & 0.437 & 0.378 & -0.546 & -0.365 \\ -0.291 & -0.695 & -0.489 & 0.257 & 0.326 & 0.144 \\ -0.550 & -0.055 & -0.199 & -0.758 & -0.264 & -0.100 \\ -0.024 & -0.092 & 0.144 & 0.008 & -0.404 & 0.898 \end{pmatrix}.$$

Our model (7) can now be written in the *canonical form*

$$(11)\qquad \tilde y = D_1\tilde\beta + D_2\tilde u,$$

where

$$\tilde y = V_1^{\mathsf T}\Delta_0^{-1}P^{\mathsf T} y_e,\quad \tilde\beta = U_1^{\mathsf T}\beta,\quad \tilde u = U_2^{\mathsf T} u,$$

$$D_1 = D_{11}^{\mathsf T} = (6\times 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & c_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad D_2 = D_{21}^{\mathsf T} = (6\times 5) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ s_1 & 0 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and so the canonical model explicitly written has the form

$$
(12)\quad
\begin{aligned}
\tilde{y}_1 &= \tilde{\beta}_1, \\
\tilde{y}_2 &= c_1\,\tilde{\beta}_2 + s_1\,\tilde{u}_1, \\
\tilde{y}_3 &= c_2\,\tilde{\beta}_3 + s_2\,\tilde{u}_2, \\
\tilde{y}_4 &= \tilde{u}_3, \\
\tilde{y}_5 &= \tilde{u}_4, \\
\tilde{y}_6 &= \tilde{u}_5.
\end{aligned}
$$

The least squares estimators are given by

$$
\begin{aligned}
\hat{\tilde{\beta}}_1 &= \tilde{y}_1, \\
\hat{\tilde{\beta}}_2 &= \tilde{y}_2 / c_1, \\
\hat{\tilde{\beta}}_3 &= \tilde{y}_3 / c_2.
\end{aligned}
$$

In matrix notation we can write

$$
\hat{\tilde{\beta}} = D_1^{+}\tilde{y} \quad \text{as} \quad D_1^{+} = (3\times 6) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/c_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/c_2 & 0 & 0 & 0 \end{pmatrix},
$$

$$
\mathrm{var}(\tilde{y}) = \sigma^2 D_2 D_2^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_2^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},
$$

$$
\mathrm{var}(\hat{\tilde{\beta}}) = D_1^{+}\,\mathrm{var}(\tilde{y})(D_1^{+})^{\mathsf{T}} = \sigma^2 D_1^{+} D_2 D_2^{\mathsf{T}} (D_1^{+})^{\mathsf{T}} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & s_1^2/c_1^2 & 0 \\ 0 & 0 & s_2^2/c_2^2 \end{pmatrix} = \sigma^2 D_0 .
$$

For the original parameters we obtain

$$
\beta = U_1\tilde{\beta}, \quad \hat{\beta} = U_1\hat{\tilde{\beta}}, \quad \mathrm{var}(\hat{\beta}) = U_1\,\mathrm{var}(\hat{\tilde{\beta}})U_1^{\mathsf{T}} = \sigma^2 U_1 D_0 U_1^{\mathsf{T}} .
$$

The unknown variance σ^2 is estimated by

$$
\hat{\sigma}^2 = \tfrac{1}{3}(\tilde{y}_4^2 + \tilde{y}_5^2 + \tilde{y}_6^2).
$$

Final Remarks

a) Note that $\tilde{\beta}_1 = \tilde{y}_1$ and according to (11) we have

$$
\tilde{y} = M_1 y_e \quad \text{where} \quad M_1 = V_1^{\mathsf{T}}\Delta_0^{-1}P^{\mathsf{T}} \quad \text{and} \quad y_e^{\mathsf{T}} = (y_1,\ldots,y_5,17).
$$

We obtain

$$
M_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \sqrt{3}/3 \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix},
$$

and thus $\tilde{\beta}_1 = \tilde{y}_1 = 17\times\sqrt{3}/3 = 9.815$. $\tilde{\beta}_1$ is completely determined by the linear restriction $\beta_1 - \beta_2 + \beta_3 = 17$ i.e. $L\beta = c$ with $L = (1,-1,1)$ and $c = 17$ as $\|L\| = \sqrt{3}$.

b) The linear restriction $L\beta = c$ holds true also for the least squares estimator $\hat{\beta}$ as

$$
L\hat{\beta} = LU_1\hat{\tilde{\beta}} = LU_1D_1^{+}\tilde{y} = LU_1D_1^{+}M_1y_e = M_2y_e
$$

and as $M_2 = (0,0,0,0,0,1)$ and $y_e^{\mathsf{T}} = (y_1,\ldots,y_5,17)$ we have $L\hat{\beta} = M_2y_e = 17$.

References

Golub, G. H. and Van Loan, C. F. (1996). Matrix computations. Third edition, 694 pp. The John Hopkins University Press.

Knüsel, L. (2008). Singulärwert-Zerlegung und Methode der kleinsten Quadrate. Technical Report Number 31, 2008, Department of Statistics, www.stat.uni-muenchen.de.

Knüsel, L. (2009). The General Linear Model and the Generalized Singular Value Decomposition. Technical Report Number 48, 2009, Department of Statistics, www.stat.uni-muenchen.de.

Matlab (2008). Technical Software for Engineering and Science. www.mathworks.com.

Maple (2006). Advanced Mathematical Software. www.maplesoft.com.