Metrika manuscript No.
(will be inserted by the editor)

Information matrices

for non full rank subsystems

Pierre Druilhet[1]**, Augustyn Markiewicz**[2]

[1] CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal, 35 170 BRUZ, France
email: druilhet@ensai.fr

[2] Department of Mathematical and Statistical Methods, Agricultural University of Poznán, Wojska Polskiego 28, PL-60637 Poznán, Poland
email: amark@au.poznan.pl

Received:

Abstract

Consider the standard linear model $Y = X\theta + \varepsilon$. If the parameter of interest is a full rank subsystem $K'\theta$ of mean parameters, the associated information matrix can be defined via an extremal representation. For rank deficient subsystems, Pukelsheim (1993) introduced the notion of generalized information matrices that inherit many properties of the information matrices. However, this notion is not a direct extension of the full rank case in the sense that the definition of the generalized information matrix ap-

plied to full rank subsystems does not lead to the usual information matrix. In this paper, we propose a definition of the information matrix via an extremal representation that encompasses the full rank and the non-full rank cases. We also study its properties and show its links with the generalized information matrices.

---

1 Introduction

Information matrices play a central role in experimental design theory: optimal designs are found by maximizing some criteria based on them. Consider for example the standard linear model $Y = X\theta + \varepsilon$ with X a $(n \times k)$ matrix. In many cases, the parameter of interest is not the whole parameter $\theta \in \mathbb{R}^k$, but a subsystem $K'\theta$, for some $(k \times s)$ matrix K. Denote by $\mathcal{M}_n^+$ the set of $(n \times n)$ nonnegative symmetric information matrices and by I_n the $(n \times n)$ identity matrix. On $\mathcal{M}_n^+$, the Loewner ordering is defined as follows: let M and N be two matrices in $\mathcal{M}_n^+$; then M is said to be lower than N, denoted by $M \leq N$, if $N - M$ belongs to $\mathcal{M}_n^+$. When K is a full rank matrix, the information matrix C_K can be defined (Gaffke, 1987) by the mapping:

$$\begin{aligned} C_K : \mathcal{M}_k^+ &\longrightarrow \mathcal{M}_s^+ \\ A &\longmapsto C_K[A] = \min_{L \in \mathbb{R}^{k \times s} : L'K = I_s} L'AL, \end{aligned} \tag{1}$$

where the minimum is taken relative to the Loewner ordering. The existence, uniqueness and many properties of this extremal representation can be found in Pukelsheim (1993). However, in some situations, K is rank deficient and a full rank reparameterization is generally of no help. Pukelsheim (1993) introduced the notion of generalized information matrix to give an analogue to Formula (1) in non full rank cases. It is defined by the mapping:

$$A \longmapsto A_K = \min_{U \in \mathbb{R}^{k \times k} : U'K = K} U'AU, \tag{2}$$

where the minimum is taken relative to the Loewner ordering. This definition has many good properties but also admits some imperfections. The first one is that A_K is a $k \times k$ matrix whereas a $s \times s$ matrix is expected, as in the full rank case. The second one is that A_K is not a direct generalization of $C_K[A]$ since $A_K = K\,C_K[A]K'$ for full rank subsystems. The third one and perhaps the most important one is that A_K depends only on Range(K) and not on the exact parameterization $K'\theta$. More precisely, for any full rank $(s \times s)$ matrix M, the constraint $UKM = KM$ is equivalent to $UK = K$ and then:

$$A_K = A_{KM}. \tag{3}$$

For instance, it will be seen in Section 4.1 that this feature leads to some inconsistency when we want to use label permutations in order to derive optimal designs and especially universally optimal design (Kiefer, 1975).

In this paper, we propose another generalization of information matrices for the non full rank case. This generalization does not suffer from the three points mentioned above and encompasses the full rank case. So we will just call it information matrix. The term ”generalized information matrix” will refer to Pukelsheim's definition. In Section 2, we give the definition and some basic properties of the information matrices for rank deficient subsystems. We also give two fundamental lemmas and a general scheme useful to derive further properties of the information matrix. In Section 3, we give further properties of information matrices. In Section 4, we compare information matrices and generalized information matrices in some siutations.

2 Information matrices for non-full rank subsystem: definition and first properties

In this section, we propose an analogue to Formula (1) for a rank deficient matrix K. However, contrary to the full rank case, the condition $L'K = I_s$ cannot be satisfied. Moreover, a generalization of the information matrix cannot be achieved without a dose of arbitrariness. The idea here is to replace I_s by $\mathrm{pr}_{K'}$, where $\mathrm{pr}_{K'} = K'(KK')^+K = K^+K$ is the orthogonal projector onto $\mathrm{Range}(K')$.

Definition 1 *Let K be a $(k \times s)$ matrix. Then the information matrix for $K'\theta$ is defined by*

$$\begin{aligned} C_K : \mathcal{M}_k^+ &\longrightarrow \mathcal{M}_s^+ \\ A &\longmapsto C_K[A] = \min_{L \in \mathbb{R}^{k \times s} : L'K = \mathrm{pr}_{K'}} L'AL. \end{aligned}$$

The minimum is taken relative to the Loewner ordering.

Remarks:

1. The existence and uniqueness of $C_K[A]$ follow from the Gauss-Markov theorem: see Pukelsheim (1993)'s Theorem 1.19 with $U' = K^+ = K'(KK')^+$, $X = K$ and $V = A$.
2. For a full rank matrix K, $C_K[A]$ correspond to the usual information matrix since $\mathrm{pr}_{K'} = I_s$.

The choice of the orthogonal projector $\mathrm{pr}_{K'} = K^+K$ is somewhat arbitrary: any other non-orthogonal projector K^-K, with K^- a generalized inverse of K, could have been chosen: see Pukelsheim (1993 p. 19) for a similar discussion in the paradigm of the Gauss-Markov theorem and estimable functions. Heuristically, the choice of K^+ is motivated by the following argument: consider $c \in \ker(X)$, then the corresponding least squares estimator $c'K'\hat{\theta}$ of $c'K'\theta$ is equal to 0 and thus its variance is also equal to 0. Since the information matrix is roughly speaking the inverse of the variance matrix, the way to obtain a simple relation between them, as given in Corollary 2, is to impose that $c'\,C_K[A]\,c = 0$ for $c \in \ker(K)$ or equivalently that $\mathrm{Range}(C_K[A]) \subset \mathrm{Range}(K')$. This is achieved by choosing $K^- = K^+$, as seen in lemma 4.

The following basic properties of information matrices follow directly from the extremal definition. They are direct extensions of the full rank case (see Pukelsheim 1993, p.77).

Lemma 1

$$
\begin{array}{rcll}
C_K[\alpha\, A] & = & \alpha\, C_K[A] & \forall \alpha \geq 0, \forall A \in \mathcal{M}_k^+, \\
C_K[A+B] & \geq & C_K[A] + C_K[B] & \forall A, B \in \mathcal{M}_k^+, \\
C_K[\alpha\, A + (1-\alpha)B] & \geq & \alpha\, C_K[A] \,+\, (1-\alpha)\, C_K[B] & \forall \alpha \in]0,1[, \forall A, B \in \mathcal{M}_k^+, \\
& & A \geq B \Longrightarrow C_K[A] \geq C_K[B] & \forall A, B \in \mathcal{M}_k^+.
\end{array}
$$

We present now two lemmas and a general scheme to compute information matrices for a non full rank subsystem from information matrices for a full rank subsystem. This general scheme is also very useful to generalize properties of information matrices for full rank subsystem to the non full rank case.

Lemma 2 *If $K = (K_1|0)$ with K_1 a $(k \times s_1)$ matrix $(s_1 < s)$ and 0 the zero matrix with appropriate dimensions, then :*

$$
C_K[A] = C_{(K_1|0)}[A] = \left(\begin{array}{c|c} C_{K_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right). \tag{4}
$$

Proof For any $k \times s$ matrix L, we write $L = (L_1|L_2)$ where L_1 and L_2 are $(k \times s_1)$ and $(k \times s - s_1)$ matrices, respectively. We have, for $K = (K_1|0)$,

$$
L'K = \mathrm{pr}_{K'} \Longleftrightarrow \left(\begin{array}{c|c} L_1'K_1 & 0 \\ \hline L_2'K_2 & 0 \end{array}\right) = \left(\begin{array}{c|c} \mathrm{pr}_{K_1'} & 0 \\ \hline 0 & 0 \end{array}\right) \Longleftrightarrow \begin{cases} L_1'K_1 = \mathrm{pr}_{K_1'}, \\ L_2'K_1 = 0. \end{cases}
$$

and

$$
L'AL = \left(\begin{array}{c|c} L_1'AL_1 & L_1'AL_2 \\ \hline L_2'AL_1 & L_2'AL_2 \end{array}\right).
$$

If $\tilde{L} = (\tilde{L}_1|\tilde{L}_2)$ minimizes $L'AL$ under the constraint $L'K = \mathrm{pr}_{K'}$, necessarily $\tilde{L}_2$ minimizes $L_2'AL_2$ under the constraint $L_2'K = 0$. Therefore $\tilde{L}_2'A\tilde{L}_2 = 0$ and

$$
\begin{aligned}
C_{(K_1|0)}[A] &= \min_{L\in\mathbb{R}^{k\times s}: L'K=\mathrm{pr}_{K'}} L'AL, && (5)\\
&= \left(\begin{array}{c|c} \min\limits_{L_1\in\mathbb{R}^{k\times s_1}: L_1'K_1=\mathrm{pr}_{K_1'}} L_1'AL_1 & 0 \\ \hline 0 & 0 \end{array}\right), && (6)\\
&= \left(\begin{array}{c|c} C_{K_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right). && (7)
\end{aligned}
$$

The following lemma give an equivariance property of the information matrix under orthogonal transformations.

Lemma 3 *Let T be a $(s \times s)$ orthogonal matrix (i.e. $T'T = I_s$). Then:*

$$C_{K\,T}[A] = T'\,C_K[A]\,T, \tag{8}$$

or equivalently,

$$C_K[A] = T\,C_{K\,T}[A]\,T'. \tag{9}$$

Proof First note that $\mathrm{pr}_{(KT)'} = T'\,\mathrm{pr}_{K'}T$ for any orthogonal matrix T. We have:

$$
\begin{aligned}
T\,C_{K\,T}[A]\,T' &= T\,\left(\min_{L\in\mathbb{R}^{k\times s}:L'KT=\mathrm{pr}_{(KT)'}} L'AL\right)T',\\
&= \min_{L\in\mathbb{R}^{k\times s}:L'KT=\mathrm{pr}_{(KT)'}} TL'ALT',\\
&= \min_{L\in\mathbb{R}^{k\times s}:L'KT=T'\mathrm{pr}_{K'}T} TL'ALT',\\
&= \min_{L\in\mathbb{R}^{k\times s}:TL'K=\mathrm{pr}_{K'}} TL'ALT',\\
&= \min_{U\in\mathbb{R}^{k\times s}:U'K=\mathrm{pr}_{K'}} U'AU \qquad (\text{with } U=LT'),\\
&= C_K[A].
\end{aligned}
$$

We can now use Lemmas 2 and 3 to derive the calculation and some properties of $C_K[A]$ from the full rank case as follows:

1. First, diagonalize $K'K$ as $T'K'KT = \Delta_K$ where Δ_K is the diagonal matrix with decreasingly ordered diagonal entries and T is an orthogonal matrix whose columns correspond to the eigenvectors of $K'K$.
2. Then, break up T into two parts : $T = (T_1|T_2)$. The columns of T_1 correspond to the eigenvectors associated to the non-zero eigenvalues and the columns of T_2 correspond to the kernel of K. So, $K_1 = K\,T_1$ is a full rank matrix and $K\,T_2 = 0$:

$$KT = (K\,T_1|K\,T_2) = (K\,T_1|0) = (K_1|0). \tag{10}$$

3. Then, by Lemma 3 and Lemma 2, we have

$$C_K[A] = T\,C_{KT}[A]\,T' = T\left(\begin{array}{c|c} C_{K\,T_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right)T' = T_1\,C_{KT_1}[A]\,T_1', \tag{11}$$

where $C_{K\,T_1}[A]$ is the usual information matrix for the full rank subsystem given by $K\,T_1$.

3 Further properties of information matrices

In this section, we give several properties of the information matrices for the non full rank case. The first one is a range inclusion useful for establishing the link between information matrices and generalized information matrices.

Lemma 4 *We have the following range inclusion:*

$$\mathrm{Range}(C_K[A]) \subset \mathrm{Range}(K'), \tag{12}$$

or equivalently

$$C_K[A] = \mathrm{pr}_{K'}\, C_K[A]\, \mathrm{pr}_{K'}. \tag{13}$$

Proof Define $T = (T_1|T_2)$ as in Formula 10, where $\mathrm{Range}(T_2) = \ker(K)$ and $\mathrm{Range}(T_1) = \ker(K)^{\perp} = \mathrm{Range}(K')$. By Formula 11, it is obvious that $\mathrm{Range}(C_K[A]) \subset \mathrm{Range}(T_1) = \mathrm{Range}(K')$.

The following proposition gives a very simple link between the generalized information matrix A_K and the information matrix $C_K[A]$. It is a generalization to the non full rank case of a result by Pukelsheim (1993, p.90, ii).

Proposition 1 *We have:*

$$A_K = K\, C_K[A]\, K', \tag{14}$$

where A_K is defined by Formula (2).

Proof The full rank case has been proved by Pukelsheim (1993, p.90). For the general case, define $T = (T_1|T_2)$ as in Formula 10. Since $T_1 K$ is of full rank, $A_{KT_1} = KT_1\, C_{KT_1}[A]\, T_1' K'$. By Formula 3, $A_{KT_1} = A_K$. By Formula 11, $T_1\, C_{KT_1}[A]\, T_1' = C_K[A]$. The result follows.

Corollary 1 *The information matrix can be obtained from the generalized information matrix as follows:*

$$C_K[A] = K^+\, A_K\, K^{+'} = K'(KK')^+\, A_K\, (KK')^+ K.$$

Proof By Formula (13) and (14), $C_K[A] = K'(KK')^+ K\; C_K[A]\; K'(KK')^+ K = K'(KK')^+\, A_K\, (KK')^+ K$.

The following proposition and its corollary give the main motivation of the definition of the information matrix: the link between the information matrix and the variance matrix for an estimable subsystem is given through the Moore-Penrose inverse. Another definition of the information matrix would have led to a more complicated generalized inverse.

Proposition 2 *If $K'\theta$ is estimable, i.e. if* $\mathrm{Range}(K) \subset \mathrm{Range}(A)$*, then:*

$$C_K[A] = (K' A^- K)^+,$$

for any generalized inverse A^- of A.

Proof Assume first that $K = (K_1|0)$ with K_1 a full rank matrix and $\mathrm{Range}(K_1) \subset \mathrm{Range}(A)$. We have:

$$
\begin{aligned}
(K'A^-K)^+ &= \left(\begin{array}{c|c} K_1'A^-K_1 & 0 \\ \hline 0 & 0 \end{array}\right)^+, \\
&= \left(\begin{array}{c|c} (K_1'A^-K_1)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right), \\
&= \left(\begin{array}{c|c} C_{K_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{(by Pukelsheim, 1993 p. 64)}, \\
&= C_K[A] \quad \text{(by Lemma 2)}.
\end{aligned}
$$

In the general case, we just assume that $\mathrm{Range}(K) \subset \mathrm{Range}(A)$. Let T be an orthogonal matrix such that $KT = (K_1|0)$ with K_1 of full rank.

$$
\begin{aligned}
(K'A^-K)^+ &= T(T'K'A^-KT)^+T', \\
&= TC_{KT}[A]T', \\
&= C_K[A].
\end{aligned}
$$

The second equality comes from the first case with K replaced by KT. The third equality comes from Lemma 3.

Corollary 2 *Consider the standard linear model $Y = X\theta + \varepsilon$ with $\mathbb{E}(\varepsilon) = 0$, $\mathrm{var}(\varepsilon) = \sigma^2 I_n$, and denote $M = X'X$. Then for any estimable function $K'\theta$, we have :*

$$
\mathrm{Var}(K'\hat{\theta}) = \sigma^2\, K'\, M^+\, K = \sigma^2\, (C_K[M])^+, \tag{15}
$$

where $\hat{\theta}$ is an OLS estimator.

We now give a chain rule for the information matrix of a subsystem of the subsystem $K'\theta$.

Proposition 3 (iterated information matrices) *Let K be an $(k \times s)$ matrix, H be an $(s \times r)$ matrix and $A \in \mathcal{M}_k^+$, then :*

$$C_{KH}[A] = C_{\mathrm{pr}_{K'}\,H}[C_K[A]]. \tag{16}$$

Proof The proof is a bit long and is given in the appendix.

Remark: Formula (16) is more complicated than for the full rank case. However, if K is of full rank (but not necessarily H), the formulae for the full-rank and non full-rank cases match and we have

$$C_{KH}[A] = C_H[C_K[A]]. \tag{17}$$

We now establish some relationships between the linear subspace of the estimable functions associated to a subsystem and the range of the information matrices. Then, we derive a rank property of the information matrix.

Definition 2 *Let K be a $(k \times s)$ matrix and $A \in \mathcal{M}_k^+$. We denote by $\mathcal{E}_K(A)$ the linear subspace of estimable functions associated to the parameter subsystem K:*

$$\mathcal{E}_K(A) = \{c \in \mathbb{R}^s \;/\; K\,c \in \mathrm{Range}(A)\}\,, \tag{18}$$

$$= \{c \in \mathbb{R}^s \;/\; K\,c \in \mathrm{Range}(A) \cap \mathrm{Range}(K)\}\,. \tag{19}$$

When K is of full rank, each element in $\mathrm{Range}(K)$ has a unique antecedent. Thus, for any generalized inverse K^-:

$$\mathcal{E}_K(A) = K^-(\mathrm{Range}(A) \cap \mathrm{Range}(K)).$$

In the general case, $K^-(\mathrm{Range}(A)\cap\mathrm{Range}(K))$ is a linear subspace of $\mathcal{E}_K(A)$ that depends on the choice of the generalized inverse K^-. To get the whole space, we have to add the kernel of K:

$$\mathcal{E}_K(A) = K^-(\mathrm{Range}(A)\cap\mathrm{Range}(K)) \oplus \ker(K). \tag{20}$$

Proposition 4 *Using the same notations as in Definition 2:*

$$\mathrm{Range}(C_K(A)) = K^+(\mathrm{Range}(A)\cap\mathrm{Range}(K)) = \mathrm{pr}_{K'}\left(\mathcal{E}_K(A)\right). \tag{21}$$

Proof This result is well known for the full rank case (see, e.g. Pukelsheim, 1993 p. 96 or Heiligers, 1991). Note that, in that case, $\mathrm{pr}_{K'} = I_s$ and that K^+ can be replaced by any generalized inverse of K. If K is rank deficient, we first consider the case $K = (K_1|0)$, with K_1 a $(k \times s_1)$ full rank matrix. We have:

$$K^+ = \begin{pmatrix} K_1^+ \\ \hline 0 \end{pmatrix}$$

and $\mathrm{Range}(K) = \mathrm{Range}(K_1)$. Denote $c = \begin{pmatrix} c_1 \\ \hline c_2 \end{pmatrix}$ with $c_1 \in \mathbb{R}^{s_1}$ and $c_2 \in \mathbb{R}^{s-s_2}$, then

$$\begin{aligned} c \in \mathrm{Range}(C_K[A]) &\iff c_2 = 0, \quad c_1 \in C_{K_1}[A] \quad \text{(by Formula (4))}, \\ &\iff c_2 = 0, \quad c_1 \in K_1^+\left(\mathrm{Range}\, A \cap \mathrm{Range}(K_1)\right), \\ &\iff c_2 = 0, \quad c_1 \in K_1^+\left(\mathrm{Range}\, A \cap \mathrm{Range}(K)\right), \\ &\iff c \in K^+(\mathrm{Range}\, A \cap \mathrm{Range}(K)). \end{aligned}$$

The general case follows from Lemmas 3 and Lemma 2. The second equality of the proposition comes from Equation 20.

Corollary 3 *We have the following rank equality*

$$\operatorname{Rank} C_K(A) = \dim(\operatorname{Range} A \cap \operatorname{Range} K) = \dim(\mathcal{E}_K(A)) - \dim(\ker(K)).$$

The last result of this section is a variant of Lemma 2.

Proposition 5 *Let* $A \in \mathcal{M}_k^+$. *Let* $K = (K_1|K_2)$ *be a* $k \times (s_1 + s_2)$ *matrix such that* $\operatorname{Range}(K_1) \subset \operatorname{Range} A$ *and* $\operatorname{Range}(K_2) \cap \operatorname{Range}(A) = \{0\}$. *Then,*

$$C_K[A] = \left(\begin{array}{c|c} C_{K_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right).$$

Remark: Note that the condition $\operatorname{Range}(K_1) \subset \operatorname{Range} A$ cannot be removed.

Proof We first consider the case where $(K_1|K_2)$ is a full rank matrix. Denote by P_A a projector onto $\operatorname{Range}(A)$ along a supplementary subspace of $\operatorname{Range}(A)$ containing $\operatorname{Range} K_2$. Write $Q_A = I_s - P_A$ where $s = s_1 + s_2$. Let $\tilde{L} = (\tilde{L}_1|\tilde{L}_2)$ be a $(s \times k)$ matrix that minimizes $L'AL$ under the constraint $L'K = I_s$ i.e. such that $L_1'K_1 = I_{s_1}$, $L_2'K_1' = 0$, $L_1'K_2' = 0$ and $L_2'K_2 = I_{s_2}$. Consider $L^* = (L_1^*|L_2^*) = (\tilde{L}_1, Q_A\tilde{L}_2)$. It is easy to check that $\tilde{L}^{*'}K = I_s$ and $L_2^{*'}AL_2^{*'} = 0$. Since $\tilde{L}A\tilde{L}$ is a minimum w.r.t. the Loewner ordering, necessarily, $\tilde{L}_2'A\tilde{L}_2 = 0$ and then $\tilde{L}_2'A\tilde{L}_1 = \tilde{L}_1'A\tilde{L}_2 = 0$. So,

$$C_K[A] = \tilde{L}'\,A\,\tilde{L} = \left(\begin{array}{c|c} \tilde{L}_1\,A\,\tilde{L}_1 & 0 \\ \hline 0 & 0 \end{array}\right).$$

It remains to show that $\tilde{L}_1A\tilde{L}_1 = C_K[A]$. Define $\bar{L}_1$ such that:

$$C_{K_1}[A] = \min_{L_1'K_1 = I_{s_1}} L_1'\,A\,L_1 = \bar{L}_1\,A\,\bar{L}_1.$$

We have:

$$\bar{L}_1\, A\, \bar{L}_1 \leq \min_{\substack{L_1' K_1 = I_{s_1} \\ L_1' K_2 = 0}} L_1'\, A\, L_1 = \tilde{L}_1\, A\, \tilde{L}_1.$$

Denote $\bar{L}_1^* = P_A' \bar{L}_1$. It is straightforward to check that $\bar{L}_1^{*'} K_1 = I_{s_1}$, $\bar{L}_1^{*'} K_2 = 0$ and $\bar{L}_1^{*'} A \bar{L}_1^* = \bar{L}_1 A \bar{L}_1 = C_{K_1}[A]$. The result follows.

We now consider the general case, that is, K_1 and K_2 are not necessarily of full rank. Consider the orthogonal matrices $T_1 = (T_a|T_b)$ and $T_2 = (T_c|T_d)$ such that $K_1 T_a$ and $K_2 T_c$ are full rank matrices and such that $KT_c = 0$ and $KT_d = 0$. Consider the orthogonal matrix

$$T = \left(\begin{array}{c|c|c|c} T_a & 0 & T_b & 0 \\ \hline 0 & T_c & 0 & T_d \end{array}\right).$$

We have $KT = (K_1 T_a | K_2 T_c | 0)$ with $(K_1 T_a | K_2 T_c)$ of full rank since $\mathrm{Range}\, K_1 \cap \mathrm{Range}\, K_2 = \{0\}$. The result follows easily from Lemmas 2 and 3 and the full rank case.

4 Some examples and applications

In this section, we compare the behavior of information matrices and generalized information matrices in the one-way and two-way analysis of variance under a relabelisation of the treatments.

4.1 One way ANOVA

We show that, in one-way analysis of variance, it is not necessary to center the parameter of interest. We also compare the behavior of information and generalized information matrices under a permutation group action.

Consider the standard one-way model $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ for $i = 1, ..., I$ and $j = 1, ..., n_i$, where $n_i \geq 1$. Denote $n = \sum_{i=1}^{I} n_i$. In vector notation, we have,

$$Y = \mu \, \mathbf{1}_n + A \, \alpha + \varepsilon$$

where $\mathbf{1}_n$ is the n vector of ones, A the incidence matrix of the factor α. It is well known that $M = A' \left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right) A$ is the information matrix for the parameter α. On the other hand, consider the (centered) parameterization

$$Q\alpha = \left(\alpha_1 - \frac{1}{I}\sum_i \alpha_i, ..., \alpha_I - \frac{1}{I}\sum_i \alpha_i\right)'$$

where $Q = I_I - \frac{1}{I}\mathbf{1}_I\mathbf{1}_I'$. Note that $M\mathbf{1}_I = M^+\mathbf{1}_I = 0$. Moreover, $\operatorname{Range} Q = \operatorname{Range} M$ which means that $Q\alpha$ is estimable. By Proposition 2, we have

$$C_Q[M] = (QM^+Q)^+ = (M^+)^+ = M. \tag{22}$$

So, the information matrix for the centered parameter is the same as for the initial parameter. The generalized information matrix for $Q\alpha$ based on M is also :

$$M_Q = M. \tag{23}$$

Consider now a permutation, say σ, of the levels of α. Denote by P_σ the corresponding permutation matrix. This permutation can be applied at two different levels: either on the matrix M by considering $P_\sigma \, M \, P_\sigma'$ or on the parameter $Q \, \alpha$ by considering $P_\sigma \, Q \, \alpha$. At both levels, the corresponding information matrix gives the same answer. More precisely, by Formula (22) we have:

$$C_Q[P_\sigma\, M\, P'_\sigma] = P_\sigma\, M\, P'_\sigma,$$

and by Formula (8) we have:

$$C_{Q\, P'_\sigma}[M] = P_\sigma\, M\, P'_\sigma.$$

We can see that information matrices have some equivariance property. On the other hand, the corresponding generalized information matrices are different. Formula (23) applied to $P_\sigma\, M\, P'_\sigma$ gives:

$$(P_\sigma\, M\, P'_\sigma)_Q = P_\sigma\, M\, P'_\sigma$$

whereas, by Formula (3), the one obtained for $P_\sigma\, Q\, \alpha$ is

$$M_{Q\, P'_\sigma} = M_Q = M.$$

We see here that the generalized information matrix is either equivariant or invariant, depending on the way the permutation is applied.

4.2 Two way ANOVA

Consider now a connected block design with t treatments and b blocks. The corresponding model is

$$Y = \mu\, \mathbf{1}_n + A\, \alpha + B\beta + \varepsilon$$

where $\alpha \in \mathbb{R}^I$ is the vector treatment effects with incidence matrix A of size $(n \times t)$ and $\beta \in \mathbb{R}^b$ is the vector of block effects with incidence matrix B of size $(n \times b)$. The moment matrix M for the full parameter $\theta = \left(\begin{array}{c} \alpha \\ \hline \beta \end{array}\right)$ is

$$M = \left(\begin{array}{c|c} \Delta_t & W \\ \hline W' & \Delta_b \end{array}\right)$$

where $\Delta_t = A'A$, $\Delta_b = B'B$ and $W = A'B$. Denote $K_\alpha = \left(\begin{array}{c} I_t \\ \hline 0 \end{array}\right)$. Is is well known that the information matrix for $\alpha = K'_\alpha \theta$ is the Schur complement

$$C_{K_\alpha}[M] = \Delta_t - W\Delta_b^{-1}W'.$$

By Formula (14), the corresponding generalized information matrix is

$$M_{K_\alpha} = K_\alpha \, C_{K_\alpha}[M] \, K'_\alpha = \left(\begin{array}{c|c} C_{K_\alpha}[M] & 0 \\ \hline 0 & 0 \end{array}\right)$$

We see here that the information matrix does not involve unnecessary zero matrices. As in section 4.1, consider the centered parameterization $Q\,\alpha = Q\,K_\alpha\theta$. Note that $C_{K_\alpha}[M]\mathbf{1}_t = 0$ and, because the design is connected, $\mathrm{Range}(C_{K_\alpha}[M]) = \mathrm{Range}(Q)$. The information matrix for $Q\,\alpha$ can be derived either from the moment matrix M for θ, or the information matrix $C_{K_\alpha}[M]$ for α. By Formulae (17) and (22), we have

$$C_{K_\alpha Q}[M] = C_Q[C_{K_\alpha}[M]] = C_{K_\alpha}[M].$$

On the other hand, the generalized information matrices obtained either from M or from $C_{K_\alpha}[M]$ give two different answers and, in fact, two matrices of different sizes. We see here that information matrices are more consistent than generalized information matrices.

5 Appendix

We give here the proof of Proposition 3. Let us start by establishing two technical Lemmas.

Lemma 5 *For any full rank square matrix Q,*

$$C_{QK}[A] = C_K[Q^{-1}A\,Q^{'-1}].$$

Proof We have $\mathrm{pr}_{K'Q'} = \mathrm{pr}_{K'}$. So,

$$\begin{aligned} C_{QK}[A] &= \min_{L\in\mathbb{R}^{k\times s}:L'QK=\mathrm{pr}_{K'Q'}} L'QQ^{-1}A\,Q^{'-1}Q'L, \\ &= \min_{U\in\mathbb{R}^{k\times s}:U'K=\mathrm{pr}_{K'}} U'Q^{-1}A\,Q^{'-1}U \quad (\text{with } U = Q'L), \\ &= C_K[Q^{-1}A\,Q^{'-1}]. \end{aligned}$$

Lemma 6 *Let* $A = \left(\begin{array}{c|c} A_1 & 0 \\ \hline 0 & 0 \end{array}\right) \in \mathcal{M}_k^+$ *with* $A_1 \in \mathcal{M}_{k_1}^+$ *and* $K = \left(\begin{array}{c} K_1 \\ \hline 0 \end{array}\right)$ *a* $(k\times s)$ *matrix with* K_1 *of size* $k_1 \times s$*. Then,*

$$C_K[A] = C_{K_1}[A_1].$$

Proof The result follows from the definition of the information matrix and the fact that $\mathrm{pr}_{K'} = \mathrm{pr}_{K_1'}$, $L'K = L_1'K_1$ and $L'AL = L_1'A_1L_1$ where $L' = (L_1'|L_2')$.

Proof **(of Proposition 3)** First, note that the full rank case can be found in Pukelsheim (1993, theorem 3.19). To get the proof more clear, we split it into four cases.

Case 1: K is of full rank and $H = (H_1|0)$ with H_1 of full rank. Hence, $KH = (K\,H_1|0)$ with $K\,H_1$ of full rank. We have:

$$\begin{aligned}
C_{KH}[A] &= \left(\begin{array}{c|c} C_{KH_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{(by lemma 2)}, \\
&= \left(\begin{array}{c|c} C_{H_1}[C_K[A]] & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{(from the full rank case)}, \\
&= C_H[C_K[A]] \quad \text{(by lemma 2)}.
\end{aligned}$$

Case 2: K is of full rank. Denote by U an $(r \times r)$ orthogonal matrix such that $HU = (H_1|0)$. By Lemma 3 and Case 1, we have $C_{KH}[A] = UC_{KHU}U' = UC_{HU}[C_K[A]]U' = C_H[C_K[A]]$.

Case 3: Assume $K = (K_1|0)$ with K_1 of full rank and $H' = (H_1'|H_2')$. Note that $\mathrm{pr}_{K'}H = (H_1'|0)'$. Consequently,

$$\begin{aligned}
C_{P_{K'}H}[C_K[A]] &= C_{P_{K'}H}\left[\left(\begin{array}{c|c} C_{K_1}[A] & 0 \\ \hline 0 & 0 \end{array}\right)\right] \quad \text{(by Lemma 2)}, \\
&= \left(\begin{array}{c|c} C_{H_1}[C_{K_1}[A]] & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{(by Case 2)}, \\
&= C_{(K_1H_1|0)}[A] \quad \text{(by Lemma 6)}, \\
&= C_{KH}[A] \quad \text{(because } KH = K_1H_1\text{))}.
\end{aligned}$$

Case 4 *(general case)* : Let T be an orthogonal matrix such that $KT = (K_1|0)$ with K_1 of full rank:

$$\begin{aligned}
C_{KH}[A] &= C_{KTT'H}[A], \\
&= C_{\mathrm{pr}_{T'K'}T'H}\left[C_K T[A]\right] \quad \text{(from Case 3)}, \\
&= C_{T'\mathrm{pr}_{K'}H}\left[C_{KT}[A]\right] \quad (\text{because } \mathrm{pr}_{T'K'} = T'\mathrm{pr}_{K'}T), \\
&= C_{\mathrm{pr}_{K'}H}\left[TC_{KT}[A]\right] \quad \text{(by Lemma 5)}, \\
&= C_{\mathrm{pr}_{K'}H}\left[C_K[A]\right] \quad \text{(by Lemma 3)}.
\end{aligned}$$

References

1. Gaffke, N. (1987) Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. The Annals of Statistics 15: 942–957
2. Heiligers , B. (1991) E-optimal Polynomial Regression Designs. Habilitationsschrift, Rheinisch-Westflische Technische Hochschule Aachen.
3. Kiefer, J. (1975) Construction and optimality of generalized Youden designs. In: Srivastava, J. N. (ed) A survey of Statistical Design and Linear Models. North Holland, Amsterdam, pp 333-354
4. Pukelsheim, F. (1993), Optimal design of experiments. (John Wiley & Sons)