KYBERNETIKA — VOLUME 24 (1988), NUMBER 3

ON OPTIMUM EXPERIMENTAL DESIGN FOR RIDGE ESTIMATES

EMIL HORVÁTH

In the paper a method minimizing the summary variance of a ridge estimate for an unknown vector parameter of a linear regression model is suggested. The minimization is performed under the condition that the norm of the bias divided by the norm of the unknown vector parameter is bounded from above. From the corresponding extremal problem a new optimality criterion in the regression experiment is deduced. In particular, this criterion follows the known A-optimality criterion for least-squares estimates.

1. INTRODUCTION

A. The Standard Linear Regression Model and Ridge Estimates

Let $y^{\mathrm{T}} \equiv (y_1, \ldots, y_N)$ be the vector of observed real random variables satisfying the standard linear regression model

$$\mathsf{E}y = F\beta\,, \quad \mathsf{D}y = I\,, \tag{1}$$

in which F is the regression $N \times m$ matrix, rank $F = m$, and β is the m-dimensional unknown parameter.

The least-squares estimate $\beta^* \equiv (F^{\mathrm{T}}F)^{-1} F^{\mathrm{T}}y$ of β belongs to the class $\{\beta^*(h): h \in H \equiv \langle 0, +\infty)\}$ of ridge estimates of the form

$$\beta^*(h) \equiv (F^{\mathrm{T}}F + hI)^{-1} F^{\mathrm{T}}y\,. \tag{2}$$

These estimates were introduced by Hoerl and Kennard [1]. The basic matrix characteristics of $\beta^*(h)$ are

$$\begin{aligned} \mathsf{D}\,\beta^*(h) &\equiv \mathsf{E}\,[\beta^*(h) - \mathsf{E}\,\beta^*(h)]\,[\beta^*(h) - \mathsf{E}\,\beta^*(h)]^{\mathrm{T}}\,; \\ \text{bias}\quad \beta^*(h) &\equiv \mathsf{E}\,\beta^*(h) - \beta\,; \\ \mathsf{W}\,\beta^*(h) &\equiv \mathsf{E}\,[\beta^*(h) - \beta]\,[\beta^*(h) - \beta]^{\mathrm{T}} \end{aligned}$$

and they can be written as

(3) $$\mathrm{D}\,\beta^*(h) = (M + hI)^{-1} M (M + hI)^{-1} ;$$

(4) $$\text{bias}\quad \beta^*(h) = -h(M + hI)^{-1} \beta ;$$

(5) $$\mathrm{W}\,\beta^*(h) = (M + hI)^{-1} (M + h^2 \beta\beta^{\mathrm{T}}) (M + hI)^{-1} ,$$

where $M \equiv F^{\mathrm{T}}F$ is the information matrix.

Let $t(M, h)$ denote the summary variance, $b(M, h, \beta)$ the square of the bias and $w(M, h, \beta)$ the quadratic loss function of the ridge estimate. That means

$$t(M, h) \equiv \mathrm{Tr}\,\mathrm{D}\,\beta^*(h) ;$$
$$b(M, h, \beta) \equiv \|\text{bias}\, \beta^*(h)\|^2 ;$$
$$w(M, h, \beta) \equiv \mathrm{Tr}\,\mathrm{W}\,\beta^*(h) .$$

For these basic numerical characteristics of the ridge estimate the following equalities hold:

(6) $$t(M, h) = \mathrm{Tr}\, M(M + hI)^{-2} ;$$

(7) $$b(M, h, \beta) = h^2 \beta^{\mathrm{T}}(M + hI)^{-2} \beta ;$$

(8) $$w(M, h, \beta) = t(M, h) + b(M, h, \beta) .$$

Relations (3)–(8) are very well known; see, e.g., any paper in the references.

The main argument in favour of using the ridge estimate is that for any $\beta \in \mathbb{R}^m$ there exists some $h = h(\beta) > 0$ such that $w(M, h, \beta) < w(M, 0, \beta)$, i.e. according to the value of the quadratic loss function "w", the ridge estimate is better than the least-squares estimate (cf. [3], [4]). (A considerable improvement can be expected when the minimum eigenvalue of the matrix M is sufficiently small; cf. [3].)

The most applied method of choice of the optimum "h" is also based on the function $w(M, h, \beta)$ of the variable "h"; in particular the optimum "h" is given by

$$h^* = \arg\min_{h \in H} w(M, h, \beta) .$$

However, there are many other methods of choice of the appropriate "h" (the most important of them are presented in [5]). In all of them the regression matrix F in model (1) is fixed. Usually, this cannot be assumed, and the experimenter has possibility to prepare the experiment (measurements) in different ways, and each of them leads in general to another regression matrix. Consequently, it makes sense to design the linear regression experiment for ridge estimates optimally. This means we have to find simultaneously the adequate regression matrix and the adequate ridge parameter.

Remark 1. A similar approach is applied in [2], but for the case of a fixed $h > 0$. Taking the limit $h \downarrow 0$, one general optimality criterion for ridge estimates is compared with the classical optimality criteria.

B. Notations

Consider the linear regression model

$$\mathsf{E}\, y(x) = f^{\mathrm{T}}(x)\,\beta \tag{9}$$

with uncorrelated measurements (observations) $y(x)$, $x \in X$, in which

(a) X is a compact set;

(b) $f^{\mathrm{T}}(x) \equiv (f_1(x), \ldots, f_m(x))$, where $f_1, \ldots, f_m$ are real, continuous and linearly independent functions defined on X;

(c) β is the m-dimensional unknown parameter.

In addition without restrictions on generality assume that $\mathsf{D}\, y(x) \equiv 1$ (cf. [2]).

The model (9) – called the standard linear regression model of experiment – is sufficiently comprising in the sense that it covers a wide range of situations for designing an experiment (compare with [2]). In this model we use the following notations:

N ... a fixed natural number; $N \geqq m$

η ... $\equiv (x_1, \ldots, x_N)$, $x_1, \ldots, x_N \in X$; the exact design of the regression experiment of size "N"

$y(\eta)$ ... N-dimensional vector with components $y(x_1), \ldots, y(x_N)$

$F(\eta)$ ... the regression $N \times m$ matrix with the rows $f^{\mathrm{T}}(x_i)$, $i = 1, \ldots, N$

$M(\eta)$ ... the information $m \times m$ matrix equal to $F^{\mathrm{T}}(\eta)\, F(\eta)$

$\mathfrak{P}_N$ ... the set of all exact designs of size "N"

$\mathfrak{M}_N$ ... $\equiv \{M(\eta): \eta \in \mathfrak{P}_N\}$

$\mathfrak{M}_N^+$ ... the set of all nonsingular matrices belonging to $\mathfrak{M}_N$

$\mathfrak{P}_N^+$ ... $\equiv \{\eta \in \mathfrak{P}_N: M(\eta) \in \mathfrak{M}_N^+\}$

ξ ... the discrete probability measure on X supported by a finite set; the asymptotical design of the experiment

$M(\xi)$ ... $\equiv \sum_{x \in X} f(x) f^{\mathrm{T}}(x)\, \xi(x)$

$\mathfrak{S}$ ... the set of all asymptotical designs

$\mathfrak{M}$ ... $\equiv \{M(\xi): \xi \in \mathfrak{S}\}$

$\mathfrak{M}^+$ ... the set of all nonsingular matrices from $\mathfrak{M}$

$R^0_{m \times m}$ ... the set of all positive semidefinite $m \times m$ matrices

$R^{0,+}_{m \times m}$ ... the set of all positive definite $m \times m$ matrices

The set $\mathfrak{P}_N^+$ is nonempty ([2], [6]). Consequently, for $\eta \in \mathfrak{P}_N^+$ and $h \in H$ we can define the ridge estimate $\beta^*(\eta, h)$ by

$$\beta^*(\eta, h) \equiv [M(\eta) + hI]^{-1}\, F^{\mathrm{T}}(\eta)\, y(\eta) \tag{10}$$

(compare with (2)).

It is evident that the equalities (3)–(8) will be used also in the case of $M = M(\eta) \in \mathfrak{M}_N^+$, i.e. for the ridge estimate (10) with the same interpretation.

2. THE A_d-OPTIMALITY CRITERION

Natural attempts to consider an optimality criterion as a function of the matrix

$$\mathrm{W}\,\beta^*(\eta, h) = [M(\eta) + hI]^{-1}\,[M(\eta) + h^2\beta\beta^{\mathrm{T}}]\,[M(\eta) + hI]^{-1}$$

(compare with (5)) meet with the following two difficulties:

(a) such a function will depend on the unknown vector β;

(b) we cannot use asymptotic designs for such criteria function.

Therefore we shall use another approach and consider the function

$$\Psi_d(A) \equiv \mathrm{Tr}\, A\left[A + \frac{d}{1-d}\,\lambda(A)\, I\right]^{-2} \quad \text{if} \quad A \in R^{0,+}_{m\times m}$$

$$\equiv +\infty \qquad\qquad \text{if} \quad A \notin R^{0,+}_{m\times m},$$

where $d \in \langle 0, 1)$ and $\lambda(A)$ is the minimal eigenvalue of the matrix A.

The statistical justification of this function will be given in Proposition 1. In Proposition 3 it will be shown further that there is a matrix $M_d^* \in \mathfrak{M}_N^+$ such that $\Psi_d(M_d^*) \leqq \Psi_d(M)$ for all M from $\mathfrak{M}_N$. This matrix will be called A_d-optimal in $\mathfrak{M}_N$ and the exact design η_d^* such that $M_d^* = M(\eta_d^*)$ will be called A_d-optimal in $\mathfrak{P}_N$.

Evidently, in the case of $d = 0$ we obtain the usual A-optimality criterion for least-squares estimates.

Let in the model (9) $m = 1$. It is easy to verify that for each $d \in \langle 0, 1)$ the A_d-optimum exact design $\eta_d^* \in \mathfrak{P}_N$ is concentrated at points $x_1^*, \ldots, x_N^*$, each of them is the argument of the minimum of the function $f^{-2}(x)$, $x \in X$. Therefore, for $m = 1$ the A_d-optimality criterion does not present anything new.

Further, we shall assume that the dimension of the model (9) is at least equal to 2.

A. Justification and First Properties of the A_d-optimality Criterion

For $d \in \langle 0, 1)$ denote by $\Gamma_{d,N}$ the set

$$\Gamma_{d,N} \equiv \{(M, h) \in \mathfrak{M}_N^+ \times H\colon b(M, h, \beta) \leqq d^2\beta^{\mathrm{T}}\beta \text{ for all } \beta \in \mathbb{R}^m\}$$

and consider the minimization problem

$$\mathrm{INF}\,(d, N) \equiv \inf\{t(M, h)\colon (M, h) \in \Gamma_{d,N}\}\,. \tag{11}$$

The interpretation of (11) is based on the fact that $(M(\eta), h) \in \Gamma_{d,N}$ iff $\|\mathrm{bias}\ \beta^*(\eta, h)\|/\|\beta\| \leqq d$ for any nonzero $\beta \in \mathbb{R}^m$.

Proposition 1. $\mathrm{INF}\,(d, N) = \inf\{\Psi_d(M)\colon M \in \mathfrak{M}_N^+\}$.

Proof. For $M \in \mathfrak{M}_N^+$ and $h \in H$ the following series of equivalences hold:

$$b(M, h, \beta) \leqq d^2\beta^{\mathrm{T}}\beta \quad \text{for all} \quad \beta \in \mathbb{R}^m$$

$$\Leftrightarrow \quad h^2 \sup_\beta \frac{\beta^{\mathrm{T}}(M + hI)^{-2} + \beta}{\beta^{\mathrm{T}}\beta} \leqq d^2 \quad (\text{the sup is over } \beta \in \mathbb{R}^m)$$

$$\Leftrightarrow \quad h^2[h + \lambda(M)]^{-2} \leqq d^2$$

$$\Leftrightarrow \quad h \leqq \frac{d}{1-d}\,\lambda(M)\,.$$

Hence $\Gamma_{d,N} = \{(M, h): M \in \mathfrak{M}_N^+, h \leqq (d/1 - d)\,\lambda(M)\}$. Since the function $t(M, h)$ of the variable "h" is strictly decreasing at every fixed nonsingular information matrix M ([3], [6]), it holds:

$$\mathrm{INF}\,(d, N) = \inf\left\{t\left[M, \frac{\mathrm{d}}{1-d}\,\lambda(M)\right] : M \in \mathfrak{M}_N^+\right\}$$
$$= \inf\{\Psi_d(M): M \in \mathfrak{M}_N^+\}\,. \qquad \square$$

Proposition 2. Let $d_1, d_2, d_3, \ldots$ be the sequence of numbers from $\langle 0, 1)$ tending to some $d \in \langle 0, 1)$ and let $M_1, M_2, M_3, \ldots$ be the sequence of matrices from $\mathfrak{M}_N^+$ tending to the matrix M. Then

(a) $M \in \mathfrak{M}_N$;

(b) $\lim\limits_{n\to\infty} \Psi_{d_n}(M_n) = \Psi_d(M)$.

Proof. Evidently $M \in \mathfrak{M}_N$, since $\mathfrak{M}_N$ is a compact set ([2], [6]). Assertion (b) will be proved in two steps. In both the implication ([2]) $M_n \to M \Rightarrow \lambda(M_n) \to \lambda(M)$ will be used.

(1) If M is nonsingular, then the convergence $\Psi_{d_n}(M_n)$ to $\Psi_d(M)$ follows from the continuity of the matrix addition, multiplication and inversion, and from the continuity of the function "Tr".

(2) If M is singular, then $\lambda^{-1}(M_n) \to +\infty$. Since

$$\Psi_{d_n}(M_n) > (1 - d_n)^2\,\lambda^{-1}(M_n)\,,$$

it holds:

$$\lim_{n\to\infty} \Psi_{d_n}(M_n) = +\infty = \Psi_d(M)\,. \qquad \square$$

Proposition 3. There is a matrix $M_d^* \in \mathfrak{M}_N^+$ such that $\Psi_d(M_d^*) = \inf\{\Psi_d(M): M \in \mathfrak{M}_N^+\}$.

Proof. Evidently $\mathfrak{M}_N^+$ is a nonempty set. From this and from the property of the infimum it follows that

$$\inf\{\Psi_d(M): M \in \mathfrak{M}_N^+\} = \lim_{n\to\infty} \Psi_d(M_n) < +\infty \tag{12}$$

for a sequence $M_1, M_2, M_3, \ldots$ of matrices from $\mathfrak{M}_N^+$.

Since $\mathfrak{M}_N$ is compact there exists a strictly increasing sequence $t(1), t(2), t(3), \ldots$ of natural numbers such that the sequence $M_{t(1)}, M_{t(2)}, M_{t(3)}, \ldots$ tends to a matrix $M_d^* \in \mathfrak{M}_N$.

Let the matrix M_d^* be singular. Then according to Proposition 2 $\Psi_d[M_{t(n)}] \to +\infty$; but this is in contradiction to the relation (12). Thus $M_d^* \in \mathfrak{M}_N^+$ and from Proposition 2 it follows that

$$\lim_{n\to\infty} \Psi_d[M_{t(n)}] = \Psi_d(M_d^*)\,. \qquad \square$$

Remark 2. Since Proposition 2 remains true when substituting $\mathfrak{M}_N \mapsto \mathfrak{M}$, Proposition 3 is valid after this substitution as well.

B. Further Basic Propositions

In this section we shall prove two basic propositions connected with minimization of the function $\Psi_d(M)$.

Proposition 4. The function $g_N: d \in \langle 0, 1) \mapsto \min_{M \in \mathfrak{M}_N} \Psi_d(M)$ has the following properties:

(a) is strictly decreasing on the interval $\langle 0, 1)$ from the value $g_N(0) = \min \{\operatorname{Tr} M^{-1}: M \in \mathfrak{M}_N^+\}$ to the limit value zero;

(b) is continuous.

Proof. It is easy to verify that $\Psi_{d_1}(M) > \Psi_{d_2}(M)$ for every $0 \leqq d_1 < d_2 < 1$ and $M \in \mathfrak{M}_N^+$. Consequently, using the notation

$$M_d^* = \arg \min_{M \in \mathfrak{M}_N^+} \Psi_d(M) \tag{13}$$

we have:

$$g_N(d_1) = \Psi_{d_1}(M_{d_1}^*) > \Psi_{d_2}(M_{d_1}^*) \geqq g_N(d_2) .$$

Let now $M^0 \in \mathfrak{M}_N^+$ be an arbitrary matrix and let $d_1, d_2, d_3, \ldots$ be an arbitrary sequence of numbers from $\langle 0, 1)$ tending to 1. Since $0 < g_N(d_n) \leqq \Psi_{d_n}(M^0)$ for all positive integers "n" and $\lim_{n \to \infty} \Psi_{d_n}(M^0) = 0$, we have also $\lim_{n \to \infty} g_N(d_n) = 0$.

Thus the first part of the proposition is proved.

For the proof of the second part, let $d \in (0, 1)$. If $d_1, d_2, d_3, \ldots$ is a sequence of numbers from $\langle 0, 1)$ strictly increasing to "d", then, according to the first now proved part of this proposition, the inequality $g_N(d) < g_N(d_n)$ holds for all natural numbers "n".

Let $\delta > 0$ be fixed, but arbitrary. According to Proposition 2, there is $n(\delta)$ such that for all positive integers "n", $n \geqq n(\delta)$, there holds: $\Psi_d(M_d^*) + \delta > \Psi_{d_n}(M_d^*)$. Thus

$$0 < g_N(d_n) - g_N(d) \leqq \Psi_{d_n}(M_d^*) - \Psi_d(M_d^*) < \delta$$

for all $n \geqq n(\delta)$. This is equivalent to the continuity of g_N from the left.

The continuity from the right will be proved as follows:

Let $d \in \langle 0, 1)$ and let $d_1, d_2, d_3, \ldots$ be a strictly decreasing sequence of numbers from $(0, 1)$ tending to "d". Since

$$g_N(d_n) < g_N(d_{n+1}) < g_N(d)$$

for all "n", we obtain:

$$\lim_{n \to \infty} g_N(d_n) \leqq g_N(d) . \tag{14}$$

Set $M_n^* \equiv M_{d_n}^*$. Since $M_n^* \in \mathfrak{M}_N^+$, there exist a strictly increasing sequence $q(1), q(2), q(3), \ldots$ of natural numbers and a matrix M^* from $\mathfrak{M}_N$ such that $M_{q(n)}^* \to M^*$. Suppose that M^* is singular. Then, according to Proposition 2, $\lim_{n \to \infty} g_N[d_{q(n)}] = +\infty$. But this is a contradiction since

$$\lim_{n \to \infty} g_N[d_{q(n)}] = \lim_{n \to \infty} g_N(d_n)$$

and the last limit is, according to (14), bounded from above. Thus $M^* \in \mathfrak{M}_N^+$ and

$$\lim_{n\to\infty} g_N(d_n) = \lim_{n\to\infty} g_N[d_{q(n)}] = \lim_{n\to\infty} \Psi_{d_{q(n)}}[M^*_{q(n)}] = \Psi_d(M^*) \geqq g_N(d),$$

which together with (14) give the required continuity from the right. □

Remark 3. For the function $g: d \in \langle 0, 1) \mapsto \min\{\Psi_d(M): M \in \mathfrak{M}\}$ we can formulate a proposition analogous to Proposition 4.

Proposition 5. Let $d \in \langle 0, 1)$ and let $d_1, d_2, d_3, \ldots$ be a sequence of numbers from $\langle 0, 1)$ converging to "d". Let further

$$M^*_{d_1}, M^*_{d_2}, M^*_{d_3}, \ldots \tag{15}$$

be the sequence of matrices given by the relation (13). There holds: if M^* is a cluster point of the sequence (15), then $M^* \in \mathfrak{M}_N^+$ and $\Psi_d(M^*) = g_N(d)$.

Proof. Let M^* be a cluster point of the sequence (15) (at least one such point exists). Let $M^*_{d_{q(1)}}, M^*_{d_{q(2)}}, M^*_{d_{q(3)}}, \ldots$ be a subsequence of the sequence (15), converging to M^*. In the following series of equalities, the first is a consequence of the continuity of g_N (Proposition 4), the second follows from the definition of g_N, the third holds trivially and the fourth is evident from Proposition 2:

$$g_N(d) = \lim_{n\to\infty} g_N(d_n) = \lim_{n\to\infty} \Psi_{d_n}(M^*_{d_n}) = \lim_{n\to\infty} \Psi_{d_{q(n)}}[M^*_{d_{q(n)}}] = \Psi_d(M^*). \quad \square$$

Remark 4. For $d \in \langle 0, 1)$ denote the argument of the minimum of the function Ψ_d on $\mathfrak{M}$ by the symbol $M^*(d)$. Let $d_1, d_2, d_3, \ldots$ be a sequence of numbers from $\langle 0, 1)$ tending to "d". Evidently, on $\mathfrak{M}$ the following analogy of Proposition 5 is valid:

Every cluster point M^* of the sequence $M^*(d_1), M^*(d_2), M^*(d_3), \ldots$ belongs to the set $\mathfrak{M}^+$ and $g(d) = \Psi_d(M^*)$.

Since the function $\operatorname{Tr} M^{-1}$ is strictly convex on $\mathfrak{M}^+$ (cf. [2], Proposition IV.3.), in the special case of $d = 0$ the matrix M^* is determined unambiguously (compare with Proposition IV.32. in [2]).

C. Comparison with the A-optimality Criterion

Introduce the following notations:

$$\varphi_0(\beta) \equiv \beta^T\beta/\operatorname{Tr}(M_0^*)^{-1};$$

$$b^*(d, \beta) \equiv b\left[M_d^*, \frac{d}{1-d}\lambda(M_d^*), \beta\right];$$

$$w^*(d, \beta) \equiv w\left[M_d^*, \frac{d}{1-d}\lambda(M_d^*), \beta\right].$$

(We recall that $d \in \langle 0, 1)$ and M_d^* is defined by the relation (13); hence it is an A_d-optimum matrix.)

Proposition 6. For all $0 < d \leqq 2[1 + m\,\varphi_0(\beta)]^{-1}$ we have:

$$w^*(d, \beta) < \mathrm{Tr}\,(M_0^*)^{-1}\,.$$

Proof. Since $b^*(d, \beta) \leqq d^2\beta^{\mathrm{T}}\beta$ for all $\beta \in \mathbb{R}^m$, we have:

$$w^*(d, \beta) = \Psi_d(M_d^*) + b^*(d, \beta) \leqq \Psi_d(M_0^*) + d^2\beta^{\mathrm{T}}\beta\,. \tag{16}$$

If now $v_1 \geqq \ldots \geqq v_m$ are eigenvalues of the matrix M_0^* and "d" is chosen as in the assumption of this proposition, then

$$\Psi_d(M_0^*) + d^2\beta^{\mathrm{T}}\beta = \sum_{k=1}^{m} v_k \left(v_k + \frac{d}{1-d}\,v_m\right)^{-2} + d^2\,\varphi_0(\beta) \sum_{k=1}^{m} v_k^{-1} <$$

$$< \left[\sum_{k=1}^{m} v_k^{-1} + (1-d)^2\, v_m^{-1}\right] + md^2\,\varphi_0(\beta)\, v_m^{-1} =$$

$$= \sum_{k=1}^{m-1} v_k^{-1} + \left[-1 + (1-d)^2 + md^2\,\varphi_0(\beta)\right] v_m^{-1} < \mathrm{Tr}\,(M_0^*)^{-1}\,,$$

because $-1 + (1-d)^2 + md^2\,\varphi_0(\beta) < 0$. □

Then the ridge estimate $\beta^*(\eta_d^*, h_d^*)$ with $h_d^* \equiv (d/1-d)\,\lambda[M(\eta_d^*)]$ and $0 < d \leqq 2[1 + m\,\varphi_0(\beta)]^{-1}$ is, according to the loss function "w", better than an arbitrary least-squares estimate from the class $\{\beta^*(\eta, 0): \eta \in \mathfrak{P}_N^+\}$.

If $m\,\varphi_0(\beta) \leqq 1$, i.e. if $m\beta^{\mathrm{T}}\beta \leqq \mathrm{Tr}\,(M_0^*)^{-1}$ (which is valid when $\lambda(M_0^*)$ is sufficiently small), then it follows from Proposition 6 that $w^*(d, \beta) < \mathrm{Tr}\,(M_0^*)^{-1}$ for all $0 < d < 1$.

In the following proposition we will attempt the "percentual" reply. Hence we introduce these new notations:

$$c \equiv d(1-d)^{-1}\,; \quad T \equiv \langle 1, +\infty)\,;$$

$$\varrho_{m,d}(t) \equiv \frac{\dfrac{t}{(t+c)^2} + \dfrac{1}{(m-1)(1+c)^2}}{\dfrac{1}{t} + \dfrac{1}{m-1}}\,, \quad t \in T\,;$$

$$a_m(d) \equiv \sup_{t \in T} \varrho_{m,d}(t)\,.$$

Proposition 7. The following inequality

$$\frac{w^*(d, \beta)}{\mathrm{Tr}\,(M_0^*)^{-1}} \leqq a_m(d) + d^2\,\varphi_0(\beta)$$

is true for every $d \in \langle 0, 1)$.

Proof. Using (16), we have

$$\frac{w^*(d, \beta)}{\mathrm{Tr}\,(M_0^*)^{-1}} \leqq \frac{\Psi_d(M_0^*)}{\mathrm{Tr}\,(M_0^*)^{-1}} + d^2\,\varphi_0(\beta)\,. \tag{17}$$

Let now $M \in \mathfrak{M}_N^+$ be an arbitrary matrix and let $v_1 \geqq \ldots \geqq v_m$ be its eigenvalues.

Put $t_k \equiv v_k/v_m$, $k = 1, \ldots, m - 1$. We have

$$(18) \qquad \frac{\Psi_d(M)}{\operatorname{Tr} M^{-1}} = \frac{\dfrac{t_1}{(t_1 + c)^2} + \ldots + \dfrac{t_{m-1}}{(t_{m-1} + c)^2} + \dfrac{1}{(1 + c)^2}}{\dfrac{1}{t_1} + \ldots + \dfrac{1}{t_{m-1}} + 1}.$$

Therefore

$$(19) \qquad \sup_{M \in \mathfrak{M}_N^+} \frac{\Psi_d(M)}{\operatorname{Tr} M^{-1}} \leqq A_m(d) \equiv \sup_{\substack{t_k \in T \\ 1 \leqq k \leqq m-1}} \varrho_d(t_1, \ldots, t_{m-1}),$$

where $\varrho_d(t_1, \ldots, t_{m-1})$ is the function of variables $t_1, \ldots, t_{m-1}$ $(t_1 \in T, \ldots, t_{m-1} \in T)$ defined by the right-hand side of (18). Evidently

$$\varrho_d(t_1, \ldots, t_{m-1}) = \frac{\sum_{k=1}^{m-1} \{t_k(t_k + c)^{-2} + [(m - 1)(1 + c)^2]^{-1}\}}{\sum_{k=1}^{m-1} [t_k^{-1} + (m - 1)^{-1}]}.$$

Since

$$\frac{t_k(t_k + c)^{-2} + [(m - 1)(1 + c)^2]^{-1}}{t_k^{-1} + (m - 1)^{-1}} \leqq a_m(d) \leqq 1$$

for all $t_k \in T$ and every "k", $1 \leqq k \leqq m - 1$, is valid, then

$$t_k(t_k + c)^{-2} + [(m - 1)(1 + c)^2]^{-1} \leqq a_m(d)\,[t_k^{-1} + (m - 1)^{-1}]$$

for all $t_k \in T$ and every "k", $1 \leqq k \leqq m - 1$, is valid, too. Hence

$$\sum_{k=1}^{m-1} \{t_k(t_k + c)^{-2} + [(m - 1)(1 + c)^2]^{-1}\} \leqq a_m(d) \sum_{k=1}^{m-1} [t_k^{-1} + (m - 1)^{-1}].$$

Thus $\varrho_d(t_1, \ldots, t_{m-1}) \leqq a_m(d)$ for all $t_1 \in T, \ldots, t_{m-1} \in T$, and consequently

$$(20) \qquad A_m(d) \leqq a_m(d).$$

On the other hand, for arbitrary $t \in T$ we have: $\varrho_d(t, \ldots, t) = \varrho_{m,d}(t)$. From this it follows that

$$(21) \qquad A_m(d) \geqq \sup_{t \in T} \varrho_d(t, \ldots, t) = a_m(d).$$

Comparing (20) and (21) we obtain the equality $A_m(d) = a_m(d)$, wherefrom, by means of (17) and (19), we shall easily conclude the proof. □

Remark 5. In the interval $\langle 0,1)$, the function $a_m(d)$ is strictly decreasing from the value 1 to the limit value 0. The numbers $a_m(d)$ can be calculated by minimizing the function $\varrho_{m,d}(t)$ on T([6]).

D. Computation of A_d-optimum Asymptotical Design

According to the currently accepted terminology, we shall call an A_d-optimum asymptotical design (in the following only A_d-OAD) the asymptotical design which leads to the matrix $M^*(d)$ minimizing the function Ψ_d on $\mathfrak{M}$ (see Remark 4). The problems of the computation of A_d-OAD will be discussed here only briefly. Further details, especially the proofs of the following statements, can be found in [6].

Since for $a > 0$ and $A \in R^{0,+}_{m \times m}$ the equality $\Psi_d(aA) = a^{-1} \Psi_d(A)$ holds, and Ψ_d attains on $\mathfrak{M}$ minimum (Remark 2), the transition from the exact to the asymptotical designs is justified (for the same reasons as in the special case of $d = 0$).

In opposite to the function Ψ_0, the function Ψ_d for $d \neq 0$ is rather complicated. Therefore, a direct computation of the A_d-OAD is possible only in very simple regression models of the experiment.

Example. Consider the regression model (9) in the form:

$$\mathsf{E}\, y(x) = \begin{cases} ax & \text{if} \quad -2 \leqq x \leqq 0 . \\ bx & \text{if} \quad 0 \leqq x \leqq 1 . \end{cases}$$

Evidently

$$\mathsf{E}\, y(x) = [f_1(x), f_2(x)] \begin{pmatrix} a \\ b \end{pmatrix},$$

where

$$f_1(x) = \begin{cases} x & \text{if} \quad -2 \leqq x \leqq 0 \\ 0 & \text{if} \quad 0 \leqq x \leqq 1 \end{cases}; \qquad f_2(x) \equiv \begin{cases} 0 & \text{if} \quad -2 \leqq x \leqq 0 \\ x & \text{if} \quad 0 \leqq x \leqq 1 \end{cases}.$$

For the sake of simplicity of computations perform the designing on the finite set $X^0 \equiv \{-2, 1\}$.

Every asymptotical design ξ on X^0 leads for any "p", $0 \leqq p \leqq 1$, to the matrix $M(p) \equiv \begin{pmatrix} 4p & 0 \\ 0 & 1-p \end{pmatrix}$. We have:

$$\Psi_d[M(p)] = \begin{cases} \dfrac{1-p}{(1-p+4cp)^2} + \dfrac{1}{(1+c)^2} \dfrac{1}{4p} & \text{if} \quad 0 < p < \frac{1}{5} \\ \dfrac{4p}{[4p + c(1-p)]^2} + \dfrac{1}{(1+c)^2} \dfrac{1}{1-p} & \text{if} \quad \frac{1}{5} \leqq p < 1 . \\ +\infty & \text{if} \quad p(1-p) = 0 \end{cases}$$

Denote $p^*(d) = \arg\min\limits_{0<p<1} \Psi_d[M(p)]$. After some elementary computations it is possible to verify that $p^*(d) \in \langle 5^{-1}, 1)$ for all $d \in \langle 0, 1)$. It is also possible to verify that on the interval $\langle 5^{-1}, 1)$ it holds:

(1) For $c \leqq \frac{4}{11}$ there is $\partial^2 \Psi_d[M(p)]/\partial p^2 > 0$, which denotes the strictly convexity of Ψ_d and the unambiguous determination of $p^*(d)$.

(2) For $c = \frac{1}{2}$ the function Ψ_d is strictly convex at the point $p = \frac{2}{7}$ and strictly concave in the right-hand neighbourhood of the point $p = \frac{1}{5}$; consequently, its

course changes. But $\partial \Psi_d[M(p)]/\partial p$ is always positive. It follows from this that in this case $p^*(d) = \frac{1}{5}$.

In the following table we introduce for the same choice of "biased" values of "c" the adequate approximate values of $p^*(d)$ and $\Psi_d[M(p^*(d))]$.

c	$p^*(d)$	$\Psi_d[M(p^*(d))]$
0·00	0·3333	2·250 000 011
0·02	0·3321	2·176 961 919
0·05	0·3301	2·074 314 135
0·10	0·3258	1·919 526 063
0·16	0·3191	1·756 518 467
0·22	0·3101	1·613 843 345
0·27	0·3002	1·507 696 314
0·50	0·2000	1·111 111 111

The choice of an iterative procedure suitable for the (approximate) determination of the A_d-OAD is not simple. In fact, the function Ψ_d on $\mathfrak{M}$ is in general not convex and neither must the function, in addition, have the gradient everywhere on $\mathfrak{M}^+$. We partly reduce these difficulties by means of the function $\overline{\Psi}_{d,q}$ defined on $R^0_{m \times m}$ by

$$\overline{\Psi}_{d,q}(A) \equiv \begin{cases} \operatorname{Tr}\left[A + 2c\,\tau_q(A)\,I\right]^{-1} & \text{if} \quad A \in R^{0,+}_{m \times m} \\ +\infty & \text{if} \quad A \notin R^{0,+}_{m \times m} \end{cases},$$

where "q" is a positive integer and $\tau_q(A) \equiv \left[\operatorname{Tr} A^{-q}\right]^{-1/q}$.

This function $\overline{\Psi}_{d,q}$ is strictly convex and differentiable on $R^{0,+}_{m \times m}$.

Let $M \in \mathfrak{M}^+$. Set $\overline{\Psi}_d(M) \equiv \operatorname{Tr}\left[M + 2c\,\lambda(M)\,I\right]^{-1}$. It holds:

$$(22) \qquad 0 \leqq \overline{\Psi}_d(M) - \Psi_d(M) \leqq c^2\,\Psi_d(M)\,.$$

In addition we have:

$$(23) \qquad \lim_{q \to \infty} \tau_q(M) = \lambda(M)\,.$$

From (22) and (23) it follows that the function $\overline{\Psi}_{d,q}$ for a sufficiently large "q" and for a sufficiently small "d" (e.g. for $d \leqq \frac{1}{6}$, which corresponds to $c^2 \leqq 0{\cdot}04$) is a good approximation to the function Ψ_d on $\mathfrak{M}^+$. Therefore, the problem of the minimization of the function Ψ_d on $\mathfrak{M}^+$ can be replaced by the problem of the minimization of the function $\overline{\Psi}_{d,q}$ on $\mathfrak{M}^+$. For this new (approximate) problem, the optimum matrix (design) can be computed by means of Atwood's iterative procedure ([2]).

(Received September 24, 1987.)

REFERENCES

[1] A. E. Hoerl and R. W. Kennard: Ridge regression: Biased estimation for nonorthogonal problems; Application to nonorthogonal problems. Technometrics *12* (1970), 55–67; 69–82.

[2] A. Pázman: Foundations of Optimum Experimental Design. D. Reidel Publishing Company, Dordrecht — Boston — Lancaster — Tokyo 1986.

[3] L. Kubáček: Základy teórie odhadu (Foundation of Estimation Theory). Veda, Bratislava 1983.

[4] G. A. F. Seber: Linear Regression Analysis. J. Wiley and Sons, New York—London—Sydney—Toronto 1977.
[5] E. Z. Demienko: Linejnaja i nelinejnaja regressii. Finansy i statistika, Moskva 1981.
[6] E. Horváth: Navrhovanie optimálneho regresného experimentu pre hrebeňové odhady (On Optimum Experimental Design for Ridge Estimates). Ph. D. Thesis, Mathematical Institute, Slovak Academy of Sciences, Bratislava 1987.

RNDr. Emil Horváth, Ústav výpočtovej techniky VŠ pri UK (Institute of Computer Science — Comenius University), Mlynská dolina, 842 43 Bratislava. Czechoslovakia.