Discussiones Mathematicae
Probability and Statistics 20 (2000) 51–62

NUMERICAL METHODS FOR LINEAR MINIMAX ESTIMATION

Norbert Gaffke and Berthold Heiligers

Fakultät für Mathematik, Institut für Mathematische Stochastik
Universität Magdeburg, D–39106 Magdeburg, Germany

Abstract

We discuss two numerical approaches to linear minimax estimation in linear models under ellipsoidal parameter restrictions. The first attacks the problem directly, by minimizing the maximum risk among the estimators. The second method is based on the duality between minimax and Bayes estimation, and aims at finding a least favorable prior distribution.

Keywords and phrases: Bayes estimation, duality, linear model, L-optimality, mean squared error, minimax estimation, non-smooth optimization, parameter restrictions, p-mean, quasi Newton method.

1991 Mathematics Subject Classification: 62F10, 62J05.

1 Introduction

In a linear statistical model with restricted parameter set, alternatives to the Least Squares Estimator are of interest, among those linear minimax estimators have got considerable attention in literature. Nevertheless, explicit solutions to minimax estimation are available only under special circumstances, and thus for an application numerical methods become important.

In this paper we discuss two approaches for computing linear minimax estimators numerically. The first of which attacks the problem directly, minimizing among the estimators the maximum risk, (or an approximation thereto). The second method utilizes the duality between minimax and Bayes estimation, and provides computation of a least favorable prior distribution, finally leading to the desired estimator.

In order to simplify the derivations, we consider a nonsingular linear model with the unitball as parameter set,

$$Y \sim (Xa, V), \qquad a \in \mathcal{A} = \left\{a \in \mathbb{R}^k : \| a \| \leq 1\right\}, \tag{1.1}$$

($\| a \|$ denotes the Euclidian norm of a vector a). That is, Y is the n-dimensional random vector of observations with expectation $E_a[Y] = Xa$ and covariance matrix V; both, the $(n \times k)$-design matrix X and the $(n \times n)$-covariance matrix V are of full rank k and n, respectively, and are supposed to be known. Let $K \in \mathbb{R}^{s \times k}$, $rank(K) = s$, describe the linear parameter function of interest, and denote by $MSE_K[TY, a] = E_a[\| TY - Ka \|^2]$ the mean squared error function of the linear estimator TY for estimating Ka. T^*Y is a *linear minimax estimator* of Ka iff

$$\max_{a \in \mathcal{A}} MSE_K(T^*Y, a) \leq \max_{a \in \mathcal{A}} MSE_K(TY, a) \quad \text{for all} \quad T \in \mathbb{R}^{s \times n}. \tag{1.2}$$

Some comments on the underlying model (1.1) might be appropriate.

The assumptions of full rank matrices X, V, and K are not essential in the following, but avoid a couple of technical difficulties arising otherwise from nonuniqueness of the estimators–actually, nonsingularity of the model ensures unique existence of the linear minimax estimator, cf. e.g. Gaffke & Heiligers (1989). Nevertheless, minimax problems in singular models are always related to those in nonsingular setups, see Heiligers (1985).

Of course, other parameter sets $\mathcal{A}$ than the unit ball may be of interest. However, for linear minimax estimation there is no loss of generality in assuming $\mathcal{A}$ to be compact, convex and symmetric w.r.t. reflection at zero, cf. Heiligers (1993). We note that by the symmetry of $\mathcal{A}$ around zero, for minimax estimation *inhomogeneous* linear estimators are dispensable, cf. Kuks (1972), Remark 1. Taking $\mathcal{A}$ to be an ellipsoid in $\mathbb{R}^k$ centered at zero seems to be the most popular choice in the literature, probably because closed form expressions for the maximum risks of the estimator are available (via eigenvalues). By a linear parameter transformation we can standardize this case to the unit ball as parameter set $\mathcal{A}$, cf. Gaffke & Mathar, 1990, p. 619.

We mention in passing that the above setup also covers minimax problems w.r.t. a weighted mean squared error function, (weighted by some nonnegative definite matrix A). By a full rank factorization of A an equivalent minimax problem w.r.t. an unweighted MSE_K is obtained, cf. Heiligers (1993).

The crucial assumption–which is actually needed throughout–is that of a *known* covariance matrix V. There are some attempts in literature to weaken this condition, considering covariances $\sigma^2 V$ with an unknown scalar factor $\sigma^2 > 0$, for example. For the practical point of view, Lauterbach & Stahlecker (1988) report, on the basis of simulation studies, satisfactory results when using unbiased plug-in-estimates of σ^2.

2 Approximate linear minimax estimators

As it is well known (and easily seen), the problem of minimax linear estimation of Ka in model (1.1) writes as

$$(2.1)\quad \text{minimize} \quad \|TV^{1/2}\|^2 + \lambda_{\max}\Big[(K-TX)(K-TX)'\Big] \text{ over } T \in I\!R^{s\times n}\,;$$

here, $\| A \| = \sqrt{trace[AA']}$ is the Euclidian norm of a matrix A, and $\lambda_{\max}[B]$ is the maximum eigenvalue of a square, symmetric matrix B. The variable T in (2.1) is of dimension $s\,n$. An equivalent optimization problem involving a variable of lower dimension $s\,k$, independent of the number n of observations, is given by

$$(2.2)\quad \text{minimize} \quad \Psi(Z) = \|(K-Z)D^{1/2}\|^2 + \lambda_{\max}[ZZ'] \text{ over } Z \in I\!R^{s\times k}\,,$$

where $D = X'V^{-1}X$, for abbreviation. For later reference, we restate here the equivalence of the two problems, (cf. Gaffke & Mathar, 1990, p. 618).

Theorem 1. *The optimal solutions T^* and Z^* to* (2.1) *and* (2.2)*, respectively exist uniquely, and*

$$Z^* \;=\; K - T^*X\,, \qquad T^* \;=\; (K-Z^*)DX'V^{-1}\,. \qquad \blacksquare$$

In general, the objective function Ψ from (2.2) is nondifferentiable. In his diplomathesis, Achtziger (1989) applied a general method from non-smooth convex optimization for solving this problem. His *Fortran* code uses an implementation of the *Bundle trust algorithm* along with special subroutines for computing function values and subgradients of Ψ. In his numerical experiments, which were run on a mainframe (VAX 8600), the algorithm produced accurate solutions.

In order to apply methods from *differentiable* optimization, Stahlecker & Lauterbach (1989) substitute Ψ by a differentiable approximation thereto,

replacing the $\lambda_{\max}[ZZ']$-term by $trace[(ZZ')^p]^{1/p}$ for some large positive integer p. That is, they consider the solution Z_p^* to

(2.2′)

$$\text{minimize} \quad \Psi_p(Z) = \|(K-Z)D^{1/2}\|^2 + trace\Big[(ZZ')^p\Big]^{1/p} \text{ over } Z \in I\!R^{s\times k}$$

as a surrogate to Z^*; the corresponding estimator $T_p^*Y = (K-Z_p^*)DX'V^{-1}Y$ is then viewed as an *approximate minimax estimator* of Ka. The optimal solution to (2.2) exists uniquely, since $\Psi_p(Z) \to \infty$ when $\|Z\| \to \infty$, and Ψ_p is strictly convex.

As a justification of the approximate approach, firstly note that $\lim_{p\to\infty} \Psi_p(Z) = \Psi(Z)$ for all $Z \in I\!R^{s\times k}$. Moreover, $\Psi_p(Z)$ is nonincreasing in p. The following lemma strengthens a result of Lauterbach (1989), Satz 2.16, simultaneously simplifying the proof.

Lemma 2. *For the optimal solutions Z_p^* and Z^* to (2.2′) and (2.2), respectiveli, we have*

$$\lim_{p\to\infty} Z_p^* = Z^*, \quad \text{and} \quad \lim_{p\to\infty} \Psi_p(Z_p^*) = \Psi(Z^*).$$

Proof. Since $\Psi_p(Z)$ is nonincreasing in $p \geq 1$ for any fixed Z, we obtain $\Psi_p(Z_p^*) \geq \Psi_{p+1}(Z_p^*) \geq \Psi_{p+1}(Z_{p+1}^*)$. Hence, the sequence $\Psi_p(Z_p^*)$, $p \in I\!N$, is nonincreasing. In particular, $\Psi_1(Z_1^*) \geq \Psi_p(Z_p^*) \geq \| (K - Z_p^*)D^{1/2} \|^2$; thus $Z_p^*D^{1/2}$ is in the ball with center $KD^{1/2}$ and radius $\sqrt{\Psi_1(Z_1^*)}$. This shows boundedness of the sequence $Z_p^*D^{1/2}$ and, by the positive definiteness of D, boundedness of the sequence Z_p^* follows (e.g. via $\|ZD^{1/2}\|^2 \geq \lambda_{\min}[D]\,\|Z\|^2$, where $\lambda_{\min}[D]$ denotes the minimum eigenvalue of D^{-1}).

Fix a sequence of positive integers p_ν, $\nu \in I\!N$, with $\lim_{\nu\to\infty} p_\nu = \infty$ and such that the associated subsequence $Z_{p_\nu}^*$, $\nu \in I\!N$, converges to some $(s \times k)$-matrix $\widetilde{Z}$. Since for all Z and p,

$$s^{-1/p}\, trace\Big[(ZZ')^p\Big]^{1/p} \leq \lambda_{\max}\Big[ZZ'\Big] \leq trace\Big[(ZZ')^p\Big]^{1/p},$$

and s^{-1/p_ν} tends to 1, we obtain (also using continuity of Ψ)

$$\begin{aligned}
&\limsup_{\nu\to\infty} \Psi_{p_\nu}(Z_{p_\nu}^*) \\
&\quad = \limsup_{\nu\to\infty} \Big(\| (Z_{p_\nu}^* - K)D^{1/2} \|^2 + s^{-1/p_\nu}\, trace\Big[(Z_{p_\nu}^* Z_{p_\nu}^{*\prime})^{p_\nu}\Big]^{1/p_\nu} \Big) \\
&\quad \leq \lim_{\nu\to\infty} \Psi(Z_{p_\nu}^*) = \Psi(\widetilde{Z}) \leq \liminf_{\nu\to\infty} \Psi_{p_\nu}(Z_{p_\nu}^*),
\end{aligned}$$

hence

$$\lim_{\nu\to\infty} \Psi_{p_\nu}(Z^*_{p_\nu}) = \Psi(\widetilde{Z})\,. \tag{2.3}$$

From (2.3) we get for any Z,

$$\Psi(Z) = \lim_{\nu\to\infty} \Psi_{p_\nu}(Z) \geq \lim_{\nu\to\infty} \Psi_{p_\nu}(Z^*_{p_\nu}) = \Psi(\widetilde{Z})\,,$$

i.e., $\widetilde{Z}$ solves (2.2), and consequently $\widetilde{Z} = Z^*$. We have thus proved that any convergent subsequence of the bounded sequence Z^*_p, $p \in \mathbb{N}$, converges to Z^*, and hence $\lim_{p\to\infty} Z^*_p = Z^*$. Convergence of $\Psi_p(Z^*_p)$ to $\Psi(Z^*)$ now follows from (2.3). ∎

Of course, close approximations of Ψ by Ψ_p will require large values of p. For example, if $K = D = I_2$, then $Z^* = \frac{2}{3}\, I_2$ and $\Psi(Z^*) = \frac{1}{3}$, see Hoffmann (1979), Section 4, but $\Psi_p(Z^*) - \Psi(Z^*) \leq \epsilon\Psi(Z^*)$ for some desired precision $\epsilon > 0$ if and only if $p \geq ln(2)/ln(1 + \frac{3}{2}\epsilon)$. More explicitly, for $\epsilon = 10^{-3}$ this means $p > 462$; even larger values of p are necessary to guarantee $\|Z^* - Z^*_p\| \leq \epsilon$. As a consequence, numerical instabilities may be expected when evaluating function values or gradients of Ψ_p. Moreover, when solving (2.2′), a *numerical* nondifferentiability of Ψ_p might take the regime for large p and make an algorithm for solving *smooth* problems fail.

As a possible way out, Lauterbach (1989) proposes not to solve (2.2′) with *one* prescribed p, but instead to go through a series of problems (2.2′), with increasing p, taking iteratively the obtained numerical solution for one p as a starting point for the next p. More formally, the method looks as follows.

Approximate minimax algorithm

Step 1.(*Initialization*) Choose a starting point $Z \in \mathbb{R}^{s\times k}$ and a $p \in \mathbb{N}$, and go to Step 2.

Step 2. (*Termination*) If Z is acceptable as an optimal solution to (2.2), terminate (with the nearly optimal solution $\widetilde{Z}^* = Z$ to (2.2)); otherwise go to Step 3.

Step 3.(Ψ_p-*optimization*) Compute a numerical solution Z_p to (2.2′) (with the current p).

Step 4. (*Update* p) Set $Z = Z_p$, choose an integer $\tilde{p} > p$, set $p = \tilde{p}$, and go to Step 2.

For a realization of this algorithm, gradients of Ψ_p will be needed, which for a positive *integer* p are given by

$$\nabla\Psi_p[Z] = 2\,(Z-K)D + 2\,c^{(1-p)/p}\,(ZZ')^{p-1}Z\,, \quad (c = trace[ZZ'])\,.$$

Moreover, the Hessian of Ψ_p at $Z \neq 0$, as a bilinear real functional on $I\!R^{s\times k} \times I\!R^{s\times k}$, is obtained as

$$\begin{aligned}\nabla^2\Psi_p[Z](A,B) &= 2\,trace[ADB'] \\ &+ c^{(1-2p)/p}\Big(4(1-p)\,trace[M_{p-1}A']\,trace[M_{p-1}B'] \\ &+ c\sum_{i=0}^{p-2} trace[M_iA'M_{p-2-i}ZB' + M_iB'M_{p-2-i}ZA'] \\ &+ c\sum_{i=0}^{p-1} trace[N_iAL_{p-1-i}B' + N_iBL_{p-1-i}B']\Big)\,, \quad A,B \in I\!R^{s\times k}\,;\end{aligned}$$

here we abbreviate, for $0 \le i \le p-1$, $M_i = (ZZ')^iZ$, $N_i = (ZZ')^i$, and $L_i = (Z'Z)^i$.

For the Ψ_p-optimization in Step 3 of the algorithm, Stahlecker & Lauterbach (1989) propose the steepest descent method, combined with an exact line search. Convergence to the optimum of (2.2′) is well-known (though at a bad final rate) *for fixed* p. However, convergence to the optimum of (2.2) of the above overall algorithm seems to be difficult to prove. This would require appropriate termination rules in Step 3, other than the heuristical ones employed by Stahlecker & Lauterbach (1989), p. 2772. Although a couple of bounds for the deviation from the optimum is derived in that paper, it is proved only that for a given $\epsilon > 0$ there exists a p (sufficiently large) such that a convergent fixed-p algorithm can approximate the optimum of (2.2) up to ϵ, cf. Stahlecker & Lauterbach (1989), Corollary 2. This is also an immediate consequence from Lemma 2. So we are left to look at the *practical* behavior of their algorithm, and this turns out to be bad even in fairly simple examples, (see Section 4).

In his diplomathesis, Dietz (1993) could significantly improve the algorithm. He employed in Step 3 the Polak-Ribière conjugate gradient or the BFGS methods combined with the (inexact) line search procedure from Press et al. (1988), pp. 384–386. A numerically stable procedure for computing function values and gradients of Ψ_p for large p is included in his *C*-implementation. He found that the sequence of Fibonacci numbers,

$$p_1 = 1\,, \quad p_2 = 2\,, \quad p_i = p_{i-1} + p_{i-2}\,,\ (i \ge 3)\,,$$

provides a good update rule for the p values in Step 4 (combined with the priori approximation bound $s^{1/p} - 1 \leq \epsilon$ as stopping criterion). This is motivated by the idea that for small p problem (2.2′) is numerically stable, and thus precise numerical solutions can be expected, while for larger p the Ψ_p-functions nearly coincide, and thus the (numerical) solution to (2.2′) for large p might be good starting point when solving the problem with $\tilde{p} \gg p$.

3 Bayes estimators

Linear minimax estimation is paired with the dual problem of finding a least favorable prior distribution for linear Bayes estimation in model (1.1). This duality is established by observing that the solution Z^* to (2.2) is a minimax solution to the convex-concave (actually, convex-linear) kernel

$$trace\Big[(K - Z)D(K - Z)' \,+\, ZMZ'\Big]\,, \quad Z \in I\!R^{s\times k},\ M \in \mathcal{M}\,,$$

where $\mathcal{M}$ consists of the nonnegative definite $(k \times k)$-matrices M with $trace[M] \leq 1$. Applying a minimax theorem to that kernel, gives

$$\min_{Z\in I\!R^{s\times k}} \Psi(Z) \;=\; \max_{M\in\mathcal{M}} trace[KDK' \,-\, KD(D+M)^{-1}DK']\,;$$

note that $trace[KDK' - KD(D+M)^{-1}DK']$ is the Bayes risk of the Bayes estimator $T_M Y = (K - Z_M)DX'V^{-1}Y$, $Z_M = KD(D+M)^{-1}$, of Ka w.r.t. a prior distribution μ over $\mathcal{A}$ with moment matrix $\int_{\mathcal{A}} aa' \mathrm{d}\mu = M$. Hence, the minimax risk can be obtained by finding a least favorable moment matrix, i.e., by solving the convex optimization problem

$$\text{(3.1)} \qquad \text{minimize} \quad \Phi(M) \;=\; trace\Big[C(D+M)^{-1}\Big] \;\text{ over }\; M \in \mathcal{M}\,,$$

where $C = DK'KD$, for abbreviation. The strong duality of (2.2) and (3.1) has been proved by Pilz (1986), Gaffke & Heiligers (1989), and Gaffke & Mathar (1990), which we summarize as follows:

Theorem 3. *The optimal solution M^* of* (3.1) *exists uniquely, and M^* gives the optimal solution Z^* of* (2.2) *by*

$$Z^* \;=\; KD(D+M^*)^{-1}\,. \qquad \blacksquare$$

A favorable effect of the Bayesian approach to minimax estimation is that the objective function Φ is twice continuously differentiable, though (3.1) is a restricted optimization problem. For the gradient $G = \nabla\Phi[M]$ and the Hessian $H = \nabla^2\Phi[M]$ of Φ at M (as a bilinear real functional on $Sym(k) \times Sym(k)$) we have,

$$\begin{aligned} G &= -(D+M)^{-1}C(D+M)^{-1}\,, \quad \text{and} \\ H(A,B) &= -trace[(D+M)^{-1}(AGB+BGA)]\,, \quad A,B \in Sym(k)\,. \end{aligned}$$

As pointed out by Pilz (1986), (3.1) is closely related to Bayes-L-optimality in a particular experimental design setting, and thus algorithms developed there can be utilized in the present context. We briefly describe here the algorithm from Gaffke & Heiligers (1996) for solving (3.1).

Least favorable prior algorithm

Step 1. (*Initialization*) Choose a starting point $M_0 \in \mathcal{M}$, set $n = 0$, and go to Step 2.

Step 2. (*Termination*) If M_n is acceptable as an optimal solution to (3.1), terminate (with the approximately optimal solution $\widetilde{Z}^* = KD(D+M_n)^{-1}$ to (2.2)); otherwise go to Step 3.

Step 3. (*Search Direction*) Choose $\overline{M}_n \in \mathcal{M}$ with $trace[G_n\,(\overline{M}_n - M_n)] < 0$, where G_n is the gradient of Φ at M_n; go to Step 4.

Step 4. (*Line Search*) Choose $0 < \alpha_n < 1$ such that $\Phi((1-\alpha_n)M_n + \alpha_n\overline{M}_n) < \Phi(M_n)$. Take $M_{n+1} = (1-\alpha_n)M_n + \alpha_n\overline{M}_n$, replace n by $n+1$, and go to Step 2.

The absolute steepest descent value ε_n of Φ at $M_n \in \mathcal{M}$,

$$\varepsilon_n \;=\; -\min_{N\in\mathcal{M}} trace\Big[G_n\,(N-M_n)\Big] \;=\; -\lambda_{\min}[G_n] + trace[G_n\,M_n]\,, \tag{3.2}$$

can be utilized to base stopping criteria in Step 2 on the (relative) efficiencies of M_n for (3.1) *and* of the associated $Z_n = KD(D+M_n)^{-1}$ for (2.2), defined by

$$Eff_{\Phi}(M_n) = \Phi(M^*)/\Phi(M_n)\,, \quad Eff_{\Psi}(Z_n) = \Psi(Z^*)/\Psi(Z_n)\,.$$

Of course, the optimum values $\Phi(M^*)$ and $\Psi(Z^*)$ will be not known. However, the following lower efficiency bounds are available, which become asymptotically sharp during the iterations.

Lemma 4. *Let ε be the absolute steepest descent value of Φ at $M \in \mathcal{M}$, and $Z = KD(D+M)^{-1}$. Then*

$$\mathit{Eff}_{\Phi}(M) \geq \frac{\Phi(M)}{\Phi(M)+\varepsilon}, \quad \text{and} \quad \mathit{Eff}_{\Psi}(Z) \geq \frac{\Phi(0)-\Phi(M)}{\Phi(0)-\Phi(M)+\varepsilon}.$$

Proof. Since Φ is a positive convex function and $1/\Phi$ is concave, the bound for Eff_{Φ} follows from Gaffke & Mathar (1992). Let $G = \nabla\Phi[M]$. From $\lambda_{\max}[ZZ'] = -\lambda_{\min}[G]$ and $\mathit{trace}[(D+M)G] = -\Phi(M)$ one gets

$$\begin{aligned}\Psi(Z) &= \mathit{trace}[(K-Z)D(K-Z)'] + \lambda_{\max}[ZZ'] \\ &= \Phi(0) - 2\Phi(M) - \mathit{trace}[DG] - \lambda_{\min}[G] \\ &= \Phi(0) - \Phi(M) + \mathit{trace}[MG] - \lambda_{\min}[G] = \Phi(0) - \Phi(M) + \varepsilon,\end{aligned}$$

and $\Psi(Z^*) = \min_Z \Psi(Z) = \max_N(\Phi(0) - \Phi(N)) \geq \Phi(0) - \Phi(M)$ yields the asserted minimax efficiency bound. ∎

Since both, the gradient G_n and the Hessian H_n of Φ at M are easily accessible, it would be most appealing to realize the algorithm as a second order (Newton) method, i.e., to take as search direction $\overline{M}_n$ a minimizer over $\mathcal{M}$ of the quadratic approximation to Φ at M_n,

$$\Phi(M_n) + \mathit{trace}\Big[G_n\,(M-M_n)\Big] + \frac{1}{2}\mathit{trace}\Big[H_n\Big(M-M_n, M-M_n\Big)\Big],$$

equivalently, to choose $\overline{M}_n$ as solution to the problem

$$\text{(3.3)}\quad \text{minimize} \quad \mathit{trace}\Big[A_n M\Big] - \mathit{trace}\Big[(D+M_n)^{-1}MG_nM\Big] \text{ over } M \in \mathcal{M},$$

where $A_n = G_n + 2\,(D+M_n)^{-1}M_nG_n$. However, (3.3) is too difficult to be solved with sufficient accuracy at each iteration. What we can do is to substitute the feasibility set $\mathcal{M}$ by finitely many sideconditions. Actually, Gaffke & Heiligers (1996) establish convergence of the algorithm to the optimum when $\mathcal{M}$ in (3.3) is replaced by certain *polyhydral* subsets $\mathcal{M}_n \subset \mathcal{M}$ (along with the sectioning/bracketing line search from Fetcher (1987), Section 2.6, in Step 4): The only conditions to be imposed on $\mathcal{M}_n$ are that it contains the current point M_n and a feasible point at which the absolute steepest descent value ϵ_n from (3.2) is attained.

4 Examples

We tested the above algorithms in a series of examples. For computing the search direction $\overline{M}_n$ in the least favorable prior algorithm we replaced $\mathcal{M}$ in (3.3) by

$$\mathcal{M}_n \;=\; conv\Big(\Big\{x_1x_1', ..., x_sx_s', ..., y_1y_1', ..., y_ky_k'\Big\}\Big)\,,$$

where the $x_1, ..., x_s$ and $y_1, ..., y_k$ are orthonormal eigenvectors of G_n and M_n, respectively, (these also depend on n, but this dependence is dropped in our notation). The method from Higgins & Polak (1990) was utilized for solving (3.3) (with $\mathcal{M}_n$ instead of $\mathcal{M}$).

One of the findings from our numerical experiments is that the least favorable prior algorithm and Dietz's version of the approximate minimax algorithm work roughly equally good for problem sizes up to $k = 40$, as they find very accurate solutions in reasonable computation times. In most test examples the least favorable prior algorithm found estimators with guaranteed minimax efficiency ≥ 0.99995 in not more than 100 iterations, while Dietz's algorithm needed roughly twice as much iterations to achieve the same efficiency. Note, however, that the approximate minimax approach does not directly provide minimax efficiency bounds. On the other hand, Stahlecker & Lauterbach's version of the approximate minimax algorithm appears to run more quickly into numerical problems, in particular when nonsingularity comes in. We suppose that this problem is basically caused by inadequate stopping criteria.

For illustration, we report here numerical results obtained by the three algorithms for some small setups taken from Stahlecker & Lauterbach (1989). In all these examples we have $k = 5$, and $V = I_n$.

The design matrix X is chosen such that, for some prefixed $0 \leq \rho < 1$,

$$X'X \;=\; \Big(\,\rho^{|i-j|}\,\Big)_{1\leq i,j\leq 5}\,;$$

actually we took $\rho = 0.0, 0.1, 0.5, 0.75, 0.99$. Note that increasing ρ reflects increasing collinearity in the model, and $\rho = 0.0$ corresponds to an orthogonal design. As parameter functions Ka we tested

$$K_1 \;=\; I_5\,,$$
$$K_2 \;=\; diag_{1\leq i\leq 5}\,(2^{(i-1)/2})\,,$$

$$K_3 \;=\; 0.1\, I_5 \;-\; \frac{0.1+\sqrt{4.96}}{5}\,\mathbf{1}_{5\times 5}\,, \quad \text{thus} \quad K_3' K_3 \;=\; 0.01\, I_5 + 0.99\,\mathbf{1}_{5\times 5}\,,$$

$$K_4 \;=\; \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix};$$

these are equivalent to the weight matrices $A_1, ..., A_4$ from Stahlecker & Lauterbach (1989), p. 2773.

Table 1 gives the minimax risks obtained in the different setups. The first rows correspond to the least favorable prior algorithm, the seconds are from Dietz's version of the approximate minimax approach. The thirds rows are taken from Stahlecker & Lauterbach (1989), for reference. Similar results are obtained in other setups.

In summary, for computing linear minimax estimators we propose second order methods, and we particularly recommend the Bayesian approach to the problem, since this does not only provide a quick and stable method for computation, but also a well founded stopping criterion.

ρ	K_1	K_2	K_3	K_4
0.0	0.8333	8.4575	2.4796	2.1143
	0.8333	8.4582	2.4797	2.1153
	0.8333	8.4575	2.4800	2.1144
0.1	0.8341	8.4542	2.2717	2.0109
	0.8342	8.4548	2.2715	2.0116
	0.8356	8.4875	2.2841	2.0188
0.5	0.8518	8.3717	1.2672	1.6287
	0.8520	8.3724	1.2671	1.6297
	0.8846	9.2393	1.5543	1.8035
0.75	0.8699	8.2616	0.5858	1.4460
	0.8701	8.2622	0.5857	1.4465
	0.9386	10.4689	1.1545	1.8247
0.99	0.8881	8.0992	0.0274	1.3242
	0.8883	8.0998	0.0274	1.3243
	0.9976	13.6940	0.8378	2.0580

REFERENCES

[1] W. Achtziger, *Nichtglatte Optimierung: Ein spezielles Problem bei der Durchführung von Versuchen*, Diplomathesis, Universität Bayreuth, Germany 1989 (in German).

[2] A. Dietz, *Implementierung eines Algorithmus zur Berechnung linearer Minimaxschätzer*, Diplomathesis, Universität Augsburg, Germany 1993 (in German).

[3] R. Fletcher, *Practical Methods of Optimization*, (2nd ed.) Wiley, New York 1987.

[4] N. Gaffke and B. Heiligers, *Bayes, admissible and minimax linear estimators in linear models with restricted parameter space*, Statistics **20** (1989), 487–508.

[5] N. Gaffke and B. Heiligers, *Second order methods for solving extremum problems from optimal linear regression design*, Optimization **36** (1996), 41–57.

[6] N. Gaffke and R. Mathar, *Linear minimax estimation and related Bayes L-optimal design*, Methods of Operations Research **60** (1990), 617–628.

[7] N. Gaffke and R. Mathar, *On a class of algorithms from experimental design theory*, Optimization **24** (1992), 91–126.

[8] B. Heiligers, *Affin-lineare Minimaxschätzer im linearen statistischen Modell bei eingeschränktem Parameterbereich*, Diplomathesis, RWTH Aachen, Germany 1985 (in German).

[9] B. Heiligers, *Linear Bayes and minimax estimators in linear models with partially restricted parameter space*, J. Statist. Plann. Inference **36** (1993), 175–183.

[10] J.E. Higgins and E. Pollack, *Minimizing pseudoconvex functions on convex compact sets*, J. Optim. Theory and Appl. **65** (1990), 1–27.

[11] K. Hoffmann, Characterization of minimax linear estimators in linear regression, Math. Operationsforsch. u. Statist., Ser. Statist. **8** (1979), 425–438.

[12] J. Kuks, *Minimax estimation of regression coefficients*, Izv. Akad. Nauk Eston SSR **21** (1972), 73–78 (in Russian).

[13] J. Lauterbach, *Zur Berechnung approximativer Minimax-Schätzer im linearen Regressionsmodell*, Dissertationthesis, Universität Hannover, Germany 1989 (in German).

[14] J. Lauterbach and P. Stahlecker, *Approximate minimax estimation in linear regression: A simulation study*, Comm. Statist., Simulation **17** (1988), 209–227.

[15] J. Pilz, *Minimax linear regression estimation with symmetric parameter restrictions*, J. Statist. Plann. Inference **13** (1986), 297–318.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press, Cambridge 1988.

[17] P. Stahlecker and J. Lauterbach, *Approximate linear minimax estimation in regression analysis with ellipsoidal constraints*, Comm. Statist., Theory Meth. **18** (1989), 2755–2784.

Received 25 November 1998