A Min-Max Algorithm for Non-linear Regression Models

A. Tishler and I. Zang

Faculty of Management
Tel Aviv University
Tel Aviv 69978
ISRAEL

Transmitted by Robert Kalaba

ABSTRACT

We present a simple method for the nonlinear min-max (or L_∞) estimation problem. The method consists of locally smoothing out the nondifferentiabilities in the original L_∞ problem, resulting in an approximate differentiable one that can be estimated using standard gradient techniques. The accuracy of the approximation is determined by a single parameter, whose choice determines *a priori* the length of the uncertainty interval in the maximal absolute error for the solution of the original L_∞ problem. In addition, we present some numerical examples demonstrating the efficiency of the method.

I. INTRODUCTION

In this paper we develop an algorithm for nonlinear min-max estimation problems. Since this problem has discontinuous first order partial derivatives, it is considered to be difficult to solve, especially when the fitted curve is nonlinear in its parameters.

It is well known that the statistical properties of the estimated parameters depend highly upon the underlying distribution of the error terms in the model. Rice and White [10] presented a study of the effectiveness of different norms in estimation models, and discussed the advantages of the min-max estimator for certain distributions.

Efficient algorithms for the min-max problem exist for the case where the fitted curve is linear in the estimated parameters. In particular, Barrodale and Phillips [1, 2], and Hand and Sposito [6, 7] greatly simplified the use of the min-max estimator in the linear case by applying a modification of the simplex

52 Vanderbilt Ave., New York, NY 10017 0096-3003/83/$03.00

method to the primal formulation of the problem as a linear program. There are several algorithms for min-max estimation which may be applied in case the curve is nonlinear in the estimated parameters (see [8]).

In this paper we suggest a new algorithm for nonlinear min-max estimation. This algorithm, which is similar in spirit to the one suggested in [12] for the nonlinear L_1 estimation, approximates the original problem by a continuously differentiable one. Thus, the approximated problem can be solved using efficient gradient (e.g., quasi-Newton) techniques. The accuracy of this approximation is determined by a single parameter, denoted by β. Our approximation replaces the original problem only in some arbitrarily small neighborhoods of the points of discontinuous differentiability. Moreover, one of our main results, Theorem 3.3, shows that it is possible to choose *a priori* a value for the parameter β in a way that guarantees that the minimal value of the objective function of the approximate problem is within a definite predetermined distance of the optimal value of the objective of the original problem. This implies *a priori* bounds on the maximal absolute value of the error.

The method suggested here is simple to use and to program, since it only requires an unconstrained minimization routine using first derivatives (a quasi-Newton method is especially recommended). This last feature makes our method applicable even to linear min-max problems, in case there are many observations and no special LP code [2] is available. We expect, however, that for linear min-max models our method will be less efficient than a specifically designed LP code.

The outline of this article is as follows: Section 2 presents our approximations and method. Some properties of these approximations and their implications are established and discussed in Section 3. Then, in Section 4, we present three nonlinear numerical examples with several parameters, where the number of observations varies between 30 and 500.

2. THE METHOD

The min-max curve fitting problem can be stated as follows: given N observations on the dependent variable y_t and the independent variables $x_t = (x_{1t}, x_{2t}, \ldots, x_{mt})'$, where $t = 1, \ldots, N$, we want to determine the vector of k parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ which minimizes the function $G(\theta)$ given by

$$G(\theta) = \max_t |u_t| = \max_t |y_t - f(x_t, \theta)|, \tag{2.1}$$

where u_t is a random error, and we assume that $N > k$. The function $f(x, \theta)$, by which we specify the functional dependence between x and y, can be linear or nonlinear in x and/or θ. However, for practical purposes, we assume that f possesses everywhere continuous first order partial derivatives with respect to θ, for every x_t, $t = 1, \ldots, N$. Consequently, the sole difficulty in finding the "best" vector of parameters, θ^*, is caused by $G(\theta)$ not being continuously differentiable, due to the presence of the absolute value and max operators in its specification. This property prohibits the use of gradient or gradient based techniques in the estimation process.

The objective function (2.1) can be expressed as follows (see Bertsekas [3]):

$$G(\theta) = |u_1| + q\big[|u_2| - |u_1| + q\big[\cdots + q\big[|u_{N-1}| - |u_{N-2}| + q\big[|u_N| - |u_{N-1}|\big]\big]\cdots\big]\big], \tag{2.2}$$

where

$$q(r) = \max(0, r). \tag{2.3}$$

In a recent paper Tishler and Zang [11] suggested to smooth out derivative discontinuities which are introduced into piecewise regression models by the presence of max (or min) operators. The basic idea in [11] is to smooth (or approximate) the max operator $[q(r) = \max(0, r)]$ by the once continuously differentiable approximation

$$q_1(\beta, r) = \begin{cases} 0 & \text{if} \quad r \leqslant -\beta, \\ (r+\beta)^2/4\beta & \text{if} \quad -\beta \leqslant r \leqslant \beta, \\ r & \text{if} \quad \beta \leqslant r, \end{cases} \tag{2.4}$$

or by the twice continuously differentiable approximation suggested in Zang [13] and given by

$$q_2(\beta, r) = \begin{cases} 0 & \text{if} \quad r \leqslant -\beta, \\ -\dfrac{r^4}{16\beta^3} + \dfrac{3r^2}{8\beta} + \dfrac{r}{2} + \dfrac{3\beta}{16} & \text{if} \quad -\beta \leqslant r \leqslant \beta, \\ r & \text{if} \quad \beta \leqslant r, \end{cases} \tag{2.5}$$

where β is a positive parameter determining the accuracy of the approximations. Moreover, using the identity $|r| = q(r) + q(-r)$, Tishler and Zang [12] obtained the following once and twice continuously differentiable approximations, respectively, to the absolute value operator $|r|$:

$$A_1(\beta, r) = \begin{cases} -r & \text{if} \quad r \leqslant -\beta, \\ (r^2 + \beta^2)/2\beta & \text{if} \quad -\beta \leqslant r \leqslant \beta, \\ r & \text{if} \quad \beta \leqslant r, \end{cases} \tag{2.6}$$

and

$$A_2(\beta, r) = \begin{cases} -r & \text{if} \quad r \leqslant -\beta, \\ -\dfrac{r^4}{8\beta^3} + \dfrac{3r^2}{4\beta} + \dfrac{3\beta}{8} & \text{if} \quad -\beta \leqslant r \leqslant \beta, \\ r & \text{if} \quad \beta \leqslant r. \end{cases} \tag{2.7}$$

The max and absolute value operators and their approximations (2.4)–(2.7) are shown in Figures 1 and 2 respectively. It can be easily seen that both operators are approximated only in the interval where $-\beta \leqslant r \leqslant \beta$ holds, and which can be made arbitrarily small by reducing β. Furthermore,

$$\lim_{\beta \to 0} q_j(\beta, r) = q(r), \qquad j = 1, 2, \tag{2.8}$$

and

$$\lim_{\beta \to 0} A_j(\beta, r) = |r|, \qquad j = 1, 2. \tag{2.9}$$

For additional properties of these approximations see [11], [12], and [13].[1]

We can now approximate problem (2.2) by replacing the max operators (the q's) with their approximation $q_1(\beta, r)$ [or $q_2(\beta, r)$] and the absolute

[1]Similar approximations to the max operators that have higher order continuous derivatives are found in Zang [13].

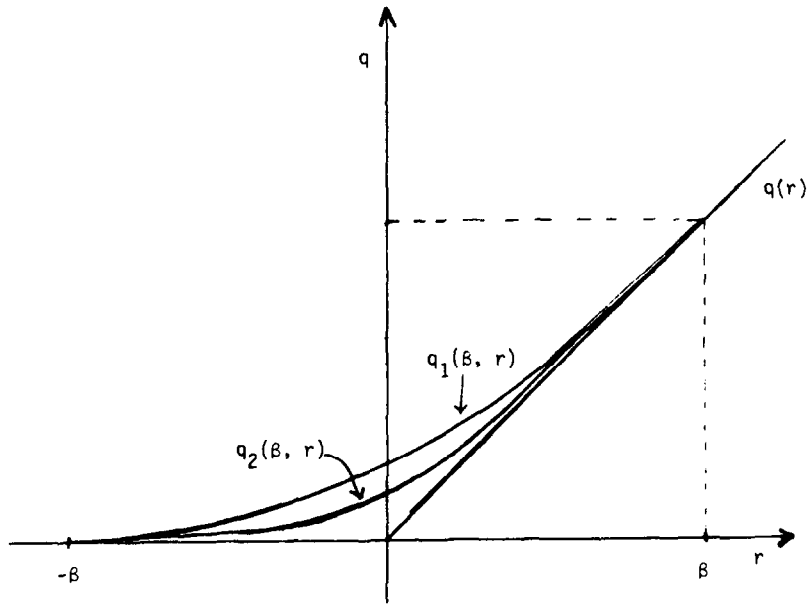

FIG. 1. The function $q(r)$ and its approximations.

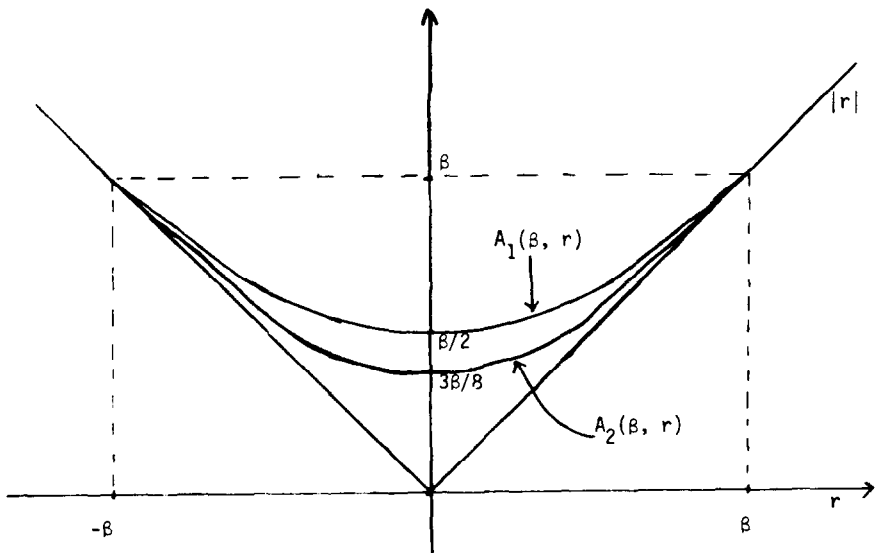

FIG. 2. The absolute value function and its approximations.

value operators by $A_1(\beta, r)$ [or $A_2(\beta, r)$]. That is,

$$G_j(\beta,\theta) = A_j(\beta,u_1) + q_j\big[\beta, A_j(\beta,u_2) - A_j(\beta,u_1) + q_j\big[\beta,\cdots + q_j\big[\beta, A_j(\beta,u_{N-1}) - A_j(\beta,u_{N-2}) + q_j\big[\beta, A_j(\beta,u_N) - A_j(\beta,u_{N-1})\big]\big]\cdots\big]\big], \quad j=1,2, \tag{2.10}$$

where u_t, $t = 1,\ldots,N$, are defined by (2.1).

The function $G_j(\beta,\theta)$ can be made arbitrarily close to $G(\theta)$ by an appropriate reduction of β, since by (2.8) and (2.9)

$$\lim_{\beta\to 0} G_j(\beta,\theta) = G(\theta), \qquad j=1,2, \tag{2.11}$$

must hold. Moreover, $G_1(\beta,\theta)$ is once continuously differentiable, which makes possible the use of some efficient gradient techniques, such as quasi-Newton methods [9], for its minimization [this is impossible for problem (2.1)]. In case $f(x_t,\theta)$ is twice continuously differentiable with respect to θ for every x_t, $t = 1,\ldots,N$, then so will be $G_2(\beta,\theta)$. Consequently, it is possible to find the minimum of this function using second order techniques such as Newton's method or that suggested by Goldfeld, Quandt, and Trotter [5].

Also note that our approximations replace the original problem only in some arbitrarily small (via β) neighborhoods of the points where $G(\theta)$ has discontinuous derivatives. Everywhere else $G_j(\beta,\theta) = G(\theta)$, and consequently the original problem remains unchanged.

It is now possible to find the minimum of $G(\theta)$ via minimizing $G_j(\beta,\theta)$ ($j=1$ or 2) for some small enough value of β. One can even do better, by using the following simple algorithmic scheme:

Stage 1. Assume θ^1 and β^1 are given. Solve

$$\min_{\theta} G_j(\beta^1,\theta), \qquad j=1 \text{ or } 2, \tag{2.12}$$

starting from θ^1. Let θ^2 be the solution point to problem (2.12). Take $\beta^2 < \beta^1$. Go to stage 2.

Stage l ($l = 2,3,\ldots$). Given θ^l and β^l, solve

$$\min_{\theta} G_j(\beta^l, \theta) \qquad j = 1 \text{ or } 2, \tag{2.13}$$

starting from θ^l. Let θ^{l+1} be the solution point to the problem (2.13). In case

$$\|\theta^{l+1} - \theta^l\| < \varepsilon, \tag{2.14}$$

where ε is a small predetermined tolerance, then declare θ^{l+1} as an optimal solution and stop. In any other case, take $\beta^{l+1} < \beta^l$ and go to stage $l+1$.

In the next section (Theorem 3.4), we will establish convergence of the above iterative process. However, for practical purposes it will generally suffice to use only one iteration, choosing a value of β which is substantially smaller than the expected maximal absolute error (MXAE). This recommendation will later be reinforced by Theorem 3.3. In our experiments, which are described in Section 4, we used the value of 0.1 for β, which gave satisfactory results. The user of the method should be cautioned, however, not to choose too small a value for β; for such a value $G_j(\beta, \theta)$ becomes too close to $G(\theta)$, which has discontinuous first order partial derivatives. This ill conditioning may cause some numerical difficulties in the solution process.

3. SOME PROPERTIES OF THE ALGORITHM

In this section we establish some properties of the approximate objective function $G_j(\beta, \theta)$. First we show that it retains some convexity properties in case these properties are possessed by $G(\theta)$:

THEOREM 3.1. *Let* $|y_t - f(x_t, \theta)|$ *be convex functions of* θ *for* $t = 1,\ldots,N$ *(and consequently* $G(\theta)$ *be convex). Then for every* $\beta > 0$, $G_j(\beta, \theta)$ *are convex functions of* θ.

PROOF. Theorem 3.1 in [12] implies the convexity of $A_j(\beta, u_t(\theta))$ with respect to θ. Then Theorem 3.8 in [13] implies the convexity of $G_j(\beta, \theta)$ with respect to θ.

Note that in case G_j is nonconvex, then several nonglobal local minima may exist. In this case the solution obtained depends on the starting point θ^1, and it is recommended to solve the problem repeatedly using different values

for θ^1. Next we establish bounds on the difference between the optimal values of the approximate and the original problem. First, we need the following lemma:

LEMMA 3.2. *The inequalities*

$$0 \leqslant G_1(\beta, \theta) - \max_t A_1(\beta, u_t(\theta)) \leqslant \min\left\{ \frac{(N-1)\beta}{4}, \beta \right\} \tag{3.1}$$

and

$$0 \leqslant G_2(\beta, \theta) - \max_t A_2(\beta, u_t(\theta)) \leqslant \min\left\{ \frac{3(N-1)\beta}{16}, \beta \right\} \tag{3.2}$$

hold, where $A_j(\beta, u_t(\theta))$, $j=1,2$, *are defined by* (2.6) *and* (2.7) *respectively, and* $u_t(\theta) = y_t - f(x_t, \theta)$.

PROOF. Follows from Proposition 3.3 in Zang [13].

Using the above lemma we can prove

THEOREM 3.3. *Let* θ^* *be a global minimum point of* $G(\theta)$, *and let* θ_j *be global minimum points of* $G_j(\beta, \theta)$, $j=1,2$, *respectively. Then*

$$0 \leqslant G_1(\beta, \theta_1) - G(\theta^*) \leqslant \min\left\{ \frac{(N+1)\beta}{4}, \frac{3\beta}{2} \right\} \tag{3.3}$$

and

$$0 \leqslant G_2(\beta, \theta_2) - G(\theta^*) \leqslant \min\left\{ \frac{3(N+1)\beta}{16}, \frac{11\beta}{8} \right\}. \tag{3.4}$$

PROOF. We only prove (3.3). The proof of (3.4) is similar. First we note that the inequality

$$0 \leqslant A_1(\beta, u_t(\theta)) - |u_t(\theta)| \leqslant \beta/2 \tag{3.5}$$

must hold for every t and all θ. This follows from the inequality

$$0 \leqslant A_1(\beta, r) - |r| \leqslant \beta/2, \tag{3.6}$$

which can be easily shown to hold for every r, since at $r = 0$ the difference between $A_1(\beta, r)$ and the absolute value operator is maximal. Next we note that (3.5) and Theorem 1 in Geoffrion [4] will imply

$$0 \leqslant \max_t A_1(\beta, u_t(\theta)) - \max_t |u_t(\theta)| \leqslant \frac{\beta}{2} \tag{3.7}$$

for every θ. Adding now inequalities (3.1), (3.7) and substituting $G(\theta) = \max_t |u_t(\theta)|$, we obtain that

$$0 \leqslant G_1(\beta, \theta) - G(\theta) \leqslant \min\left\{ \frac{(N-1)\beta}{4}, \beta \right\} + \frac{\beta}{2} \tag{3.8}$$

must hold for every θ. Inequality (3.3) follows (3.8) by reapplying Theorem 1 in [4].

We consider Theorem 3.3 as our main result. It shows that it is possible to determine *a priori* the length of the uncertainty interval in $G(\theta^*)$ by the choice of the value for β. For example, if $N \geqslant 5$ and $j = 1$, then by (3.3)

$$G_1(\beta, \theta_1) - 3\beta/2 \leqslant G(\theta^*) \leqslant G_1(\beta, \theta_1) \tag{3.9}$$

must hold, and consequently the length of the uncertainty interval in the maximal absolute error (MXAE), given by $G(\theta^*)$, is $3\beta/2$. In the same manner, an uncertainty interval of $11\beta/8$ for the MXAE is obtained whenever $N \geqslant 8$ and $j = 2$ are used.[2] This property supports our previous suggestion that in many cases one iteration of the algorithmic scheme may be sufficient for determining satisfactory parameters, provided β was properly chosen. Still, it may happen that the objective function $G(\theta)$ is flat in a neighborhood of θ^* (due to multicollinearity). In such cases, a value of β, giving reasonable *a priori* bounds on the length of the uncertainty interval of the MXAE, may not produce a vector of parameters θ_j which is close enough to θ^*. This is, however, a situation where other methods may produce unsatisfactory results

[2] For $j = 1$ and $N < 5$ or $j = 2$ and $N < 8$ the length of this uncertainty interval becomes even smaller.

as well, and in our case a further decrease of β, according to the algorithmic scheme given in Section 2, may be necessary.

Let us now establish the convergence of the algorithm scheme, suggested at the end of Section 2.

THEOREM 3.4. *Let $\{\beta^l\} \to 0$ be a sequence of monotonically decreasing positive numbers, and assume that θ_j^l is a solution to*

$$\min G_j(\beta^l, \theta), \qquad j = 1,2. \tag{3.10}$$

Also let θ', θ'' be any accumulation points of the sequences $\{\theta_1^l\}$ and $\{\theta_2^l\}$ respectively. Then

$$G(\theta') = G(\theta'') = G(\theta^*). \tag{3.11}$$

PROOF. The proof follows directly from (3.3), (3.4), and (2.11).

We now show that a local minimum θ^* of $G(\theta)$ satisfying $G(\theta^*) = |y_t - f(x_t, \theta^*)|$ for only one value of $1 \leqslant t \leqslant N$ is also a local minimum of $G_j(\beta, \theta)$, $j = 1,2$, for a sufficiently small value of β. Thus, using the algorithmic scheme given in Section 2 with $\{\beta^l\} \to 0$, it is possible to locate such points precisely (as far as the unconstrained subproblem allows) in a finite number of steps.

THEOREM 3.5. *Let θ^* be a local minimum of $G(\theta)$, where $G(\theta^*) = |y_\tau - f(x_\tau, \theta^*)|$ for only one value of $1 \leqslant \tau \leqslant N$. Then there exists a positive real number $\bar{\beta}$ such that θ^* is a local minimum of $G_j(\beta, \theta)$ for every β satisfying $0 < \beta < \bar{\beta}$.*

PROOF. We may suppose that $|u_\tau(\theta^*)| = |y_\tau - f(x_\tau, \theta^*)| > 0$, since otherwise we must have $y_t - f(x_t, \theta^*) = 0$ for all $t = 1, \ldots, N$. Then, in view of (2.9) and by continuity, for sufficiently small β, say $0 < \beta < \tilde{\beta}$, we have $A_j(\beta, u_\tau(\theta^*)) = |u_\tau(\theta^*)|$ and

$$A_j(\beta, u_\tau(\theta^*)) = \max_t A_j(\beta, u_t(\theta^*)).$$

The remainder of the proof is similar to that of Proposition 3.6 in [13], with the modification that the $\bar{\beta}$ must satisfy $\bar{\beta} \leqslant \tilde{\beta}$.

4. NUMERICAL EXAMPLES

To test the algorithm of Section 2, we solved the three numerical examples which were presented in [12]. For all three, the x_{it} and u_t were arbitrarily chosen (see details below). Then, using a given set of predetermined parameters $\hat{\theta}$ and a randomly generated error, we computed "observed " y_t's. For each example, we performed the estimation for three sample sizes ($N = 30, 100, 500$) and two distributions of the u_t's (normal and uniform distributions), denoted by $N(0, \sigma^2)$ and $U(-a, a)$ respectively.

EXAMPLE 1.

$$y_t = \frac{\sum_{i=1}^{4} \theta_i x_{it}}{1 + \sum_{i=5}^{6} \theta_i x_{it}} + u_t, \tag{4.1}$$

where $\hat{\theta} = (1.5, 1.0, 2.7, 0.75, -0.35, 0.7)'$, and x_i were uniformly distributed in the ranges (65,75), (10,20), (0,30), (−10,90), (0,2), and (0,4) respectively. The distributions of the u_t's were $N(0, 12^2)$, and $U(-15, 15)$.

EXAMPLE 2.

$$y_t = \theta_1 x_{1t}^2 + \theta_2 x_{2t}^2 + \frac{\theta_3}{\theta_1 + \theta_2} x_{1t} x_{2t} + u_t, \tag{4.2}$$

where $\hat{\theta} = (1.5, 3.0, 2.25)'$, and the x_i were uniformly distributed in the ranges (3,9) and (3,5) respectively. The distributions of the u_t's were $N(0, 8^2)$ and $U(-12, 12)$.

EXAMPLE 3 (C.E.S. production function).

$$y_t = \theta_1\left[\theta_2 x_{1t}^{-\theta_3} + (1-\theta_2)x_{2t}^{-\theta_3}\right]^{-\theta_4/\theta_3} + u_t, \tag{4.3}$$

where $\hat{\theta} = (1.5, 0.6, 0.5, 0.8)'$, and

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 550 \\ 450 \end{pmatrix}, \begin{pmatrix} 10000 & 5500 \\ 5500 & 8000 \end{pmatrix}\right] \tag{4.4}$$

The distributions of the u_t were $N(0, 3^2)$ and $U(-8, 8)$.

For each function, sample size, and distribution, we estimated the parameters three times, using the methods of least absolute errors which was developed in [12] (denoted L_1), the method of least squares (denoted L_2), and the min-max algorithm developed in this paper (denoted L_∞) for $j = 1$ (the results for $j = 2$ were almost identical to those for $j = 1$, and hence they are omitted). For both the L_1 and L_∞ methods a fixed value of $\beta = 0.1$ was chosen. In view of the discussion that followed Theorem 3.3, we have that the MXAE corresponding to the original problem (2.1) must be within $3\beta/2 = 0.15$ below the reported value of MXAE for the L_∞ estimator in Tables 1–7. The starting points θ^1 were $(1.6, 1.1, 2.6, 0.7, -0.33, 0.72)$, $(1.6, 2.6, 3.6)$, and $(1.4, 0.5, 0.8, 0.9)$ for Examples 1, 2, and 3 respectively.[3] The minimization was carried out by subroutine VA13AD of the Harwell subroutine library, which is a quasi-Newton method [9]. The computations were carried out on an IBM 370/168 computer, located at the University of Southern California.

The results of the experiments are contained in Tables 1–6. Beside giving the values of the true and estimated parameters, we also report the RMSE, the mean absolute error (MAE), the maximal absolute error[4] (MXAE), the number of function calls in the optimization process (NF), and the CPU time used for the optimization. The two distributions considered are denoted by N (normal) and U (uniform) respectively.

To demonstrate the performance of the algorithmic scheme of Section 2 and the discussion concerning the choice of β, we show in Table 7 the results of an experiment carried out with the above scheme, using Example 1, $N = 100$, $j = 1$, and errors generated according to the uniform distribution. In

[3] Somewhat more distant θ^1 may affect the method (as well as all other methods) to converge to local minima other than the global one. This may happen to Example 1 and 3 which are nonconvex minimization problems. We preferred to avoid this phenomenon.

[4] That is, $G_1(0.1, \theta_1)$, where θ_1 is the vector of optimal parameters obtained (applicable, of course, to the L_∞ estimator only).

TABLE 1

ESTIMATION RESULTS: EXAMPLE 1 (NORMAL DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	1.2042	1.3630	1.2693	1.3624	1.3793	1.3516	1.4833	1.4975	1.4941
θ_2	1.0	1.8626	1.9280	4.7819	1.2951	1.2941	1.5700	1.0598	1.0271	0.0815
θ_3	2.7	2.9468	2.5114	1.8551	2.8384	2.6804	2.6114	2.6691	2.7273	2.9406
θ_4	0.75	0.7275	0.6485	0.5382	0.7263	0.7488	0.6217	0.7708	0.7655	0.7980
θ_5	− 0.35	− 0.3620	− 0.3585	− 0.2970	− 0.3624	− 0.3624	− 0.3392	− 0.3518	− 0.3505	− 0.3697
θ_6	0.7	0.7031	0.7310	0.7859	0.7029	0.7107	0.6769	0.7108	0.7180	0.6824
RMSE		9.56	9.06	11.32	10.09	10.00	11.52	11.15	11.14	11.95
MAE		6.67	6.83	9.32	7.38	7.48	9.28	8.31	8.33	9.05
MXAE				17.62			21.35			31.50
CPU		1.03	0.75	1.35	1.81	1.13	2.35	5.79	3.61	12.49
FC		60	22	85	46	20	72	40	21	96

TABLE 2

ESTIMATION RESULTS: EXAMPLE 1 (UNIFORM DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	1.8605	1.7524	1.6800	1.4859	1.4153	1.4187	1.5749	1.5209	1.4917
θ_2	1.0	−1.0151	−0.4306	−0.5619	2.2029	1.8831	1.1269	0.9612	1.0370	1.0183
θ_3	2.7	3.1586	3.2800	2.9802	2.4503	2.6911	2.7546	2.6741	2.6962	2.7143
θ_4	0.75	0.9355	0.8640	0.6781	0.6493	0.6982	0.7446	0.7664	0.7555	0.7495
θ_5	−0.35	−0.3409	−0.3407	−0.3727	−0.3217	−0.3407	−0.3569	−0.3452	−0.3477	−0.3503
θ_6	0.7	0.7432	0.7352	0.6709	0.7215	0.7254	0.6957	0.7190	0.7123	0.7004
RMSE		7.64	7.46	8.85	9.12	8.77	9.20	8.97	8.91	8.94
MAE		5.60	5.97	7.93	7.25	7.52	8.12	7.70	7.74	7.79
MXAE				13.16			14.36			14.94
CPU		0.95	0.70	1.12	1.65	1.11	2.47	5.18	3.96	8.85
FC		50	20	73	40	19	67	34	24	61

TABLE 3

ESTIMATION RESULTS: EXAMPLE 2 (NORMAL DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	1.2524	1.2389	0.4696	1.4747	1.4484	0.8239	1.5722	1.5635	1.3619
θ_2	3.0	2.3565	2.1046	0.3483	2.6397	2.6416	1.3650	3.0906	3.0539	3.1430
θ_3	2.25	4.7391	5.0649	3.2301	3.4341	3.6163	6.0383	1.5592	1.8062	2.7345
RMSE		6.24	6.14	6.82	7.72	7.72	8.25	7.78	7.77	8.39
MAE		4.72	4.86	5.35	5.72	5.74	6.44	5.90	5.91	6.47
MXAE				12.37			22.17			24.46
CPU		0.68	0.73	3.03	1.04	0.83	4.17	3.28	2.62	5.80
FC		28	26	341	26	15	161	29	23	49

TABLE 4

ESTIMATION RESULTS: EXAMPLE 2 (UNIFORM DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	1.7387	1.8053	2.0058	1.4228	1.5335	1.6576	1.5198	1.4790	1.5213
θ_2	3.0	4.0453	4.0468	4.4332	3.2528	3.2754	3.2709	3.0858	2.9793	3.0356
θ_3	2.25	− 3.4082	− 4.2442	− 8.5931	2.1419	1.1948	2.2606	1.8092	2.3986	2.0152
RMSE		6.28	6.20	6.29	6.36	6.24	6.42	6.63	6.62	6.63
MAE		5.28	5.39	5.47	5.21	5.28	5.44	5.66	5.66	5.67
MXAE				10.00			11.91			11.99
CPU		0.82	0.71	2.53	1.05	0.93	2.23	3.07	2.50	3.35
FC		44	30	263	26	22	77	26	20	28

TABLE 5

ESTIMATION RESULTS: EXAMPLE 3 (NORMAL DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	1.4803	1.3964	1.1603	1.4123	1.4268	1.3358	1.5251	1.5415	1.6671
θ_2	0.6	0.7026	0.6451	0.6156	0.5885	0.5820	0.5937	0.5927	0.5907	0.5620
θ_3	0.5	2.6638	1.3757	1.3796	0.7813	0.5408	0.9497	0.3633	0.3343	−0.0455
θ_4	0.8	0.8008	0.8105	0.8406	0.8101	0.8083	0.8191	0.7974	0.7956	0.7828
RMSE		2.38	2.24	2.61	2.74	2.72	2.79	2.96	2.96	3.12
MAE		1.78	1.87	2.24	2.17	2.18	2.20	2.31	2.32	2.48
MXAE				4.22			5.93			8.43
CPU		2.49	1.68	2.74	4.04	3.57	4.64	17.92	18.35	32.2
FC		72	44	88	39	35	49	40	42	84

TABLE 6

ESTIMATION RESULTS: EXAMPLE 3 (UNIFORM DISTRIBUTION)

	True Values	$N = 30$			$N = 100$			$N = 500$		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
θ_1	1.5	3.4542	1.9434	1.6566	1.6483	1.5830	1.5094	1.3009	1.3863	1.4790
θ_2	0.6	0.3243	0.5836	0.6375	0.6055	0.5954	0.6113	0.5557	0.5874	0.6045
θ_3	0.5	−0.5329	0.5894	0.6189	0.5696	0.5119	0.6123	0.2607	0.4286	0.5285
θ_4	0.8	0.6835	0.7620	0.7820	0.7853	0.7922	0.7990	0.8247	0.8131	0.8020
RMSE		5.06	4.47	4.66	4.62	4.62	4.64	4.76	4.74	4.75
MAE		3.69	3.89	4.02	4.03	4.04	4.09	4.14	4.15	4.18
MXAE				7.04			7.73			7.93
CPU		3.89	2.09	2.27	4.59	4.13	3.73	28.06	20.04	14.96
FC		121	57	70	46	41	42	64	46	39

TABLE 7

RESULTS OBTAINED FOR DIFFERENT VALUES OF β IN EXAMPLE 1

β	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	NABS	NMAX	RMSE	MAE	MXAE	FC
			True Values									
	1.5	1.0	2.7	0.75	−0.35	0.7						
			Estimated Values									
20.0	1.4624	1.4310	2.7371	0.7315	−0.3484	0.7330	100	65	8.89	7.80	32.15	23
10.0	1.4684	1.2204	2.6795	0.7452	−0.3526	0.7198	60	31	9.03	8.02	21.82	24
5.0	1.4471	1.1431	2.7278	0.7508	−0.3546	0.7094	32	19	9.10	8.08	17.78	24
1.0	1.4253	1.1424	2.7554	0.7474	−0.3557	0.6990	4	10	9.14	8.07	14.87	32
0.1	1.4187	1.1269	2.7546	0.7446	−0.3569	0.6957	0	7	9.20	8.12	14.36	34
0.01	1.4179	1.1209	2.7553	0.7442	−0.3571	0.6952	0	6	9.21	8.12	14.33	59
0.001	1.4178	1.1201	2.7554	0.7442	−0.3571	0.6951	0	6	9.21	8.13	14.32	52

Table 7, NABS and NMAX denote the numbers of observations which satisfy $-\beta \leqslant r \leqslant \beta$ in (2.6) and (2.4) respectively at the optimal solution.

The experiments reported in Tables 1–6 were designed mainly to evaluate the numerical properties of the new min-max method developed in this paper. The results reveal that, as expected, the min-max estimator is superior to the least absolute error (L_1) and the least squares (L_2) estimators for the uniform distribution. For the normal distribution, the opposite is true. Finally, the L_2 estimator is somewhat better than the L_1 for the uniform distribution. All these results are in agreement with the statistical properties of the above distributions (see Rice and White [10]). Another conclusion that can be inferred from the tables and additional experiments which are not reported here is that for large samples the three estimators tend to be similar, if the "true" error distribution is either normal or uniform. However, the convergence of these estimators to the "true" ones is not monotone. The execution times obtained for the min-max estimator are quite reasonable. For the normal distribution the L_2 method is in general faster than the L_1 method, which is, in turn, faster than the min-max method. However, for the uniform distribution, the min-max may be faster than the L_2 and L_1 methods. This is because our algorithm does not use all the observations in the computation of the derivatives in each iteration. Rather, we exploit the sequential form of (2.10) and the fact that if the absolute value of $u_t(\theta)$ is greater than β, then the partial derivatives of $A_j(\beta, u_t(\theta))$ with respect to θ are zero.

Finally, we would like to note that in empirical application it is not clear when the min-max is superior (or inferior) to the L_1 or L_2 estimators, since its statistical properties are not yet known. Our experiments show that in small samples, the min-max estimator is quite different from the L_2 and L_1 estimators, and for uniformly distributed errors it is superior to them. Since the numerical computation of the min-max estimator does not seem to be difficult, it will be useful to explore its statistical properties further in the future.

REFERENCES

1 I. Barrodale and C. Phillips, An improved algorithm for discrete Chebyshev linear approximation, in *Proceedings of the Fourth Manitoba Conference on Numerical Mathematics*, Vol. 4, pp. 177–190.

2 I. Barrodale and C. Phillips, Algorithm 495: Solution of an overdetermined system of linear equations in the Chebyshev norm, *ACM Trans. Math. Software* 1(3):264–79 (1975).

3 D. P. Bertsekas, Nondifferentiable optimization via approximation, *Math. Programming Stud.* 3:1–25 (1975).

4 A. M. Geoffrion, Objective function approximations in mathematical programming, *Mathematical Programming* 13:23–37 (1977).
5 S. M. Goldfeld, R. E. Quandt, and H. F. Trotter, Maximization by quadratic hill-climbing, *Econometrica* 34:541–551 (1966).
6 M. L. Hand and V. A. Sposito, An efficient linear programming algorithm for least maximum residual estimation/approximation, mimeo, Iowa State Univ.
7 M. L. Hand and V. A. Sposito, Using the least squares estimator in Chebyshev estimation, *Comm. Statist. — Simulation Comput.* B9(1):43–49 (1980).
8 W. Murray and M. L. Overton, A projected Lagrangian algorithm for nonlinear minimax optimization, *SIAM J. Sci. Statist. Comput.* 1:345–370 (1980).
9 M. J. D. Powell, A view of unconstrained optimization, in *Optimization in Action* (L. C. W. Dixon Ed.), Academic, 1976, pp. 117–152.
10 J. R. Rice and J. S. White, Norms for smoothing and estimation, *SIAM Rev.* 6:243–256 (1964).
11 A. Tishler and I. Zang, A new maximum likelihood algorithm for piecewise regression, *J. Amer. Statis. Assoc.* 76:980–987 (1981).
12 A. Tishler and I. Zang, An absolute deviations curve fitting algorithm for non-linear models, in *Optimization in Statistics*, TIMS Studies in the Management Sciences, Vol. 19 (S. H. Zanakis and J. S. Rustagy Eds.), North Holland, 1982, pp. 81–103.
13 I. Zang, A smoothing-out technique for min-max optimization, *Math. Programming* 19:61–77 (1980).

Note added in proof: This research was carried out while A. Tishler was visiting at the Economics department, at the University of Southern California, and I. Zang was visiting at the Faculty of Commerce and Business Administration, the University of British Columbia. Research was partially supported by the Israel Institute of Business Research at Tel-Aviv University.