A MINIMAX METHOD WITH APPLICATION TO THE INITIAL VECTOR CODING PROBLEM (*)

José Mario Martínez (**)  
Lúcio Tunes dos Santos (**)  
Sandru Augusta Santos (**)

Abstract. We consider the problem

\[ \text{Minimize } \max \{ f_1(x), \ldots, f_m(x) \} \quad x \in \Omega \]

where \( f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \) are (generally nonlinear) differentiable functions, \( \Omega \subset \mathbb{R}^n \) and \( n, m \) can be large. We introduce a new algorithm for solving this problem that can be implemented in rather modest computer environments. The new method is based on a fast one-dimensional newtonian procedure applied to the objective value of an auxiliary function. We report numerical experiments, which suggest that the new algorithm, combined with a powerful strategy for minimization on spheres, can be an effective tool for solving initial vector coding problems.

Key words: minimax methods, initial vector coding problem.

C.R. Categories: G.2, J.2

Campinas, November 9, 1993.

(*) Work supported by FAPESP (Grants 90-3724-6 and 91-2441-3), FINEP, CNPq and FAEP-UNICAMP  
(**) Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081 Campinas SP, Brazil
1. Introduction. In this paper we consider the “minimax problem”:

\[
\begin{align*}
\text{Minimize} & \quad f(x) = \max \{ f_1(x), \ldots, f_m(x) \} \\
\text{s.t.} & \quad x \in \Omega \subset \mathbb{R}^n
\end{align*}
\]

(1.1)

where \( f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \) have continuous partial derivatives on an open set containing \( \Omega \). Problem (1.1) is equivalent to

\[
\begin{align*}
\text{Minimize} & \quad \theta \\
\text{s.t.} & \quad f_i(x) \leq \theta, \quad i = 1, \ldots, m \\
\end{align*}
\]

(1.2)

which is a differentiable Nonlinear Programming problem. Problem (1.2) can be solved using Penalization, Barrier, Augmented Lagrangeans, GRG or SQP methods. See [2, 3, 9, 11, 14], etc.

Current implementations of GRG and SQP methods require manipulation of (possibly sparse) matrices and, thus, are not suitable for very large or huge problems. On the other hand, Penalization, Barrier and Augmented Lagrangean methods use penalty parameters which we want to avoid because of their usual lack of “physical meaning”. Moreover, instability is usually associated to large penalty parameters and efficient estimation of Lagrange multipliers can be costly in large-scale problems.

A different approach is to consider (1.1) directly as a nondifferentiable minimization problem, using generalized gradients of \( f \) to generate algorithms. This is the idea behind subgradient techniques and “bundle” methods in nonsmooth optimization. See the excellent survey of Zowe [16]. Finally, regularization algorithms that solve (1.1) using a sequence of differentiable problems have also been developed. See [6, 7, 8].

At the present stage of development, none of the above mentioned approaches deal efficiently with a feature that can be common in practical problems. In fact, the feasible set \( \Omega \), in general different from \( \mathbb{R}^n \), is sometimes simple enough so that efficient algorithms for minimizing differentiable functions on this set are available. Although our approach is general, in this paper we will give special attention to the case where \( \Omega \) is a sphere, for which the initial vector coding problem is an example.

Briefly speaking, in this paper we propose the following strategy for solving (1.1). We consider the problem \( AP(\theta) \) given by

\[
\begin{align*}
\text{Minimize} & \quad \varphi(x, \theta) = \sum_{i=1}^m |f_i(x) - \theta|^2 \\
\end{align*}
\]

(1.3)

where \( z_+ = \max\{0, z\} \).
The objective function of (1.3) has continuous partial derivatives, though its Hessian is generally discontinuous. Let us call \( x(\theta) \) the solution of (1.3). If \( \varphi(x(\theta), \theta) = 0 \) we clearly have that \( f(x(\theta)) \leq \theta \). Conversely, if \( \varphi(x(\theta), \theta) > 0 \) we have that \( f(x) > \theta \) for all \( x \in \Omega \). We deduce that the solution of (1.1) is \( x_* \equiv x(\theta_*) \) where

\[
\theta_* = \sup\{\theta \mid \varphi(x(\theta), \theta) > 0\}.
\]

Given \( \theta_1 \) and \( \theta_2 \) such that \( \varphi(x(\theta_2), \theta_2) = 0 \) and \( \varphi(x(\theta_1), \theta_1) > 0 \), we have that \( \theta_* \in (\theta_1, \theta_2] \).

A very simple and theoretically justified Newton-type procedure is given in order to estimate iteratively \( \theta_* \).

This paper is organized as follows. In Section 2 we describe the main algorithm and present a convergence result. In Section 3 we define a newtonian process to accelerate the main algorithm. In Sections 4 and 5 we present two case studies using the techniques introduced in Sections 2 and 3. In the first case study the set \( \Omega \) is a box in \( \mathbb{R}^n \). We include this case in the present paper to show that our method, despite working in this case, does not seem to be especially efficient. The second case study is the initial vector coding problem, where we take advantage of the structure of \( \Omega \) (a sphere) by using a second order method for minimizing functions on spheres. The objective function in this case is very expensive, so it is essential that the number of evaluations used for its minimization should be moderate. We show that this is what happens in this problem. Finally, some conclusions are stated in Section 6.

2. Main Algorithm. In this section we give a high-level description of the main algorithm considered in this paper. Assume that \( f_1, \ldots, f_m, f, \varphi \) are defined by (1.1)-(1.3). We also assume that, for all \( \theta \in \mathbb{R} \), there exists a global solution \( x(\theta) \) of \( AP(\theta) \). Let us define \( \psi(\theta) = \varphi(x(\theta), \theta) \) for all \( \theta \in \mathbb{R} \). In the following lemma we prove that \( \psi \) is a non-increasing function.

Lemma 2.1. If \( \theta < \theta' \) then \( \psi(\theta) \geq \psi(\theta') \).

Proof. By the definitions, we have that

\[
\psi(\theta) = \varphi(x(\theta), \theta) = \sum_{i=1}^{m} \left[ f_i(x(\theta)) - \theta \right]_+^2 \geq \sum_{i=1}^{m} \left[ f_i(x(\theta)) - \theta' \right]_+^2 = \varphi(x(\theta), \theta') \geq \psi(\theta'). \quad \Box
\]
Lemma 2.2. If \( \theta < \theta' \) and \( \psi(\theta) = \psi(\theta') \) then \( \psi(\theta) = \psi(\theta') = 0 \).

Proof. Observe that, for any \( i \in \{1, \ldots, m\} \) such that \( |f_i(x(\theta)) - \theta|^2 > 0 \), we have \( f_i(x(\theta)) - \theta > 0 \). Moreover, since \( \theta < \theta' \), it follows that \( f_i(x(\theta)) - \theta > f_i(x(\theta)) - \theta' \). This implies that the equality
\[
\sum_{i=1}^{m} [f_i(x(\theta)) - \theta]^2 \geq \sum_{i=1}^{m} [f_i(x(\theta)) - \theta']^2
\]
holds only if \( f_i(x(\theta)) \leq \theta \) for all \( i = 1, \ldots, m \) and the proof is complete. \( \square \)

Algorithm 2.1.

Step 0. Let \( \theta_0^a < \theta_0^b \) be such that \( \psi(\theta_0^a) > 0 \) and \( \psi(\theta_0^b) = 0 \), \( \sigma \in (0, 1/2] \) (\( \sigma \approx 0.1 \)). Let \( k = 0 \).

Step 1. Choose \( \theta_k \in [\sigma \theta_k^a, (1 - \sigma) \theta_k^b] \). Solve problem \( PA(\theta_k) \).

Step 2. If \( \psi(\theta_k) > 0 \), then set \( \theta_k^a = \theta_k \), otherwise set \( \theta_k^b = \theta_k \). \( k = k + 1 \). Go to Step 1.

Let us prove that Algorithm 2.1 approximates a solution of (1.1).

Theorem 2.3. The sequence of real numbers \( \{\theta_k\} \) given by Algorithm 2.1 is well defined and has a limit point \( \theta_* \) such that \( x(\theta_*) \) is a solution of (1.1).

Proof. The sequence is well defined because we assumed that \( PA(\theta) \) has a solution for all \( \theta \in \mathbb{R} \). By construction, \( \theta_k \in I_k \) for all \( k \in \mathbb{N} \), where \( I_k \) is a closed interval, \( I_k \subseteq I_{k+1} \) and diameter \( (I_{k+1}) \leq (1 - 2\sigma) \text{diameter}(I_k) \). Therefore, \( \theta_k \) converges to a limit point \( \theta_* \) such that \( \theta_* \in I_k \) for all \( k \in \mathbb{N} \). But \( I_k = [\theta_k^a, \theta_k^b] \) where \( \psi(\theta_k^a) > 0 \) and \( \psi(\theta_k^b) = 0 \) for all \( k \in \mathbb{N} \). By Lemmas 2.1 and 2.2 and the fact that \( \lim \theta_k^a = \lim \theta_k^b = \theta_* \), this implies that \( \psi(\theta) > 0 \) for all \( \theta < \theta_* \) and \( \psi(\theta) = 0 \) for all \( \theta > \theta_* \).

Assume that \( \theta > \theta_* \). Since \( \psi(\theta) = 0 \) we have that \( f_i(x(\theta)) \leq \theta \) for all \( i = 1, \ldots, m \). However, \( \theta' = (\theta + \theta_*)/2 > \theta_* \), so \( \psi(\theta') = 0 \), \( \theta' < \theta \) and \( f_i(x(\theta')) \leq \theta' \) for all \( i = 1, \ldots, m \). This means that \( \theta \) is not the minimum of problem (1.2).

Assume now that \( \theta < \theta_* \). Since \( \psi(\theta) > 0 \) we have that \( \max\{f_1(x), \ldots, f_m(x)\} > \theta \) for all \( x \in \Omega \). Therefore, the minimum of (1.1) is greater than or equal to \( \theta \). This completes the proof. \( \square \)

3. A Newtonian Acceleration Procedure. The most straightforward implementation of Algorithm 2.1 consists of setting \( \sigma = 1/2 \). So, \( \theta_k = (\theta_k^a + \theta_k^b)/2 \) for all \( k \in \mathbb{N} \) and the method is
a bisection procedure for finding the “least root” of $\psi(\theta) = 0$. This process can be unacceptably slow, so it is needed some type of acceleration that takes into account the differentiable structure of $\varphi$.

Before defining a newtonian process to accelerate Algorithm 2.1, we make the following assumptions.

**Regularity Assumptions.** Suppose that $\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ where $g : \mathbb{R}^n \to \mathbb{R}^q$ has continuous partial derivatives. Assume that there exists a neighborhood $(\theta_0, \theta_1)$ of $\theta$, such that, for all $\theta \in (\theta_0, \theta_1)$, the active constraints at $x(\theta)$ remain the same. In other words, if $A \subset \{1, \ldots, q\}$ is the set of indices of the active constraints, we have that $g_i(x(\theta)) = 0$, $i \in A$, $\theta \in (\theta_0, \theta_1)$ and $g_i(x(\theta)) < 0$, $i \notin A$, $\theta \in (\theta_0, \theta_1)$. To simplify the notation, $\mathcal{I}$ denotes the set of components of $g$ corresponding to $A$. We also assume that

$$
\|x(\theta) - x(\theta')\| \leq L|\theta - \theta'|, \quad (3.1)
$$

for all $\theta_0 < \theta$, $\theta' < \theta_1$.

**Theorem 3.1.** Under the above assumptions, we have, for all $\theta_0 < \theta < \theta_1$, that

$$
\psi'(\theta) = \frac{\partial \varphi}{\partial \theta}(x(\theta), \theta) = -2 \sum_{i=1}^{m} [f_i(x(\theta)) - \theta]_+. \quad (3.2)
$$

**Proof.** By definition,

$$
\psi'(\theta) = \lim_{h \to 0} \frac{\psi(\theta + h) - \psi(\theta)}{h} = \lim_{h \to 0} \frac{\varphi(x(\theta + h), \theta + h) - \varphi(x(\theta), \theta)}{h} = \lim_{h \to 0} \frac{\varphi(x(\theta + h), \theta + h) - \varphi(x(\theta), \theta)}{h} + \lim_{h \to 0} \frac{\varphi(x(\theta + h), \theta) - \varphi(x(\theta), \theta)}{h}.
$$

Since $\varphi$ is $C^1$ in both arguments, there exist $\xi$, $\eta \in [0, 1]$ such that

$$
\varphi(x(\theta + h), \theta + h) - \varphi(x(\theta + h), \theta) = \frac{\partial \varphi}{\partial \theta}(x(\theta + h), \theta + \xi h) h
$$

and

$$
\varphi(x(\theta + h), \theta) - \varphi(x(\theta), \theta) = \nabla_x \varphi(x(\theta) + \eta(x(\theta + h) - x(\theta)), \theta) v_h(\theta),
$$

where

$$
v_h(\theta) = x(\theta + h) - x(\theta).
$$

5
Therefore, we have
\[ \psi'(\theta) = \frac{\partial \varphi}{\partial \theta}(x(\theta), \theta) + \lim_{h \to 0} \nabla_x \varphi(x(\theta), \theta) \frac{\nu_h(\theta)}{h}. \quad (3.3) \]

Now, by the first-order necessary condition for inequality constrained minimization (see [11]) and the regularity assumptions, there exists \( \lambda(\theta) \in \mathbb{R}^q \) such that
\[ \nabla_x \varphi(x(\theta), \theta) = \nabla g(x(\theta)) \lambda(\theta). \quad (3.4) \]

Substituting (3.4) in (3.3) we have
\[ \psi'(\theta) = \frac{\partial \varphi}{\partial \theta}(x(\theta), \theta) + \lambda(\theta)^T \lim_{h \to 0} \nabla g(x(\theta)) \frac{\nu_h(\theta)}{h}. \quad (3.5) \]

Let \( h \) be such that \( \theta_0 < \theta + h < \theta_1 \). Then, using the first-order Taylor approximation in \( g \),
\[ g(x(\theta + h)) = g(x(\theta)) + \nabla g(x(\theta))^T \nu_h(\theta) + o(\|\nu_h(\theta)\|) \]
and using the fact that \( g(x(t)) = 0 \), \( \theta_0 < t < \theta_1 \), we have
\[ \frac{\nabla g(x(\theta))^T \nu_h(\theta)}{\|\nu_h(\theta)\|} + o(\|\nu_h(\theta)\|) = 0. \quad (3.6) \]

Therefore, by (3.6) and the regularity assumptions,
\[ \lim_{h \to 0} \left| \frac{\nabla g(x(\theta))^T \nu_h(\theta)}{h} \right| = \lim_{h \to 0} \left| \frac{\nabla g(x(\theta))^T \frac{\nu_h(\theta)}{h}}{\|\nu_h(\theta)\|} \right| \leq \lim_{h \to 0} \left| \frac{o(\|\nu_h(\theta)\|)}{\|\nu_h(\theta)\|} \right| L = 0. \]

So,
\[ \psi'(\theta) = \frac{\partial \varphi}{\partial \theta}(x(\theta), \theta). \quad \square \]

By (3.2) we see that \( \psi'(\theta_*) = \psi(\theta_*) = 0 \). This property suggests that a suitable iterative process for finding \( \theta_* \), given \( \theta_k < \theta_* \), should be Newton’s method with a relaxation factor of 2:
\[ \theta_{k+1} = \theta_k - 2 \frac{\psi(\theta_k)}{\psi'(\theta_k)}. \quad (3.7) \]

The relaxation factor 2 guarantees quadratic convergence to the root \( \theta_* \). In fact, if the root has multiplicity \( w \), the Newton’s method with a relaxation factor of \( w \) has quadratic convergence to the root. See [15].
The application of (3.7) as a complement of Algorithm 2.1 is straightforward. When we choose \( \theta_{k+1} \in [\sigma \theta_k^a, (1 - \sigma) \theta_k^b] \) we try first (3.7). If this \( \theta_{k+1} \) does not belong to \( I_{k+1} \) we replace it by \( (\theta_k^a + \theta_k^b)/2 \).

4. Minimax Discrimination. In this section we present a case study using the techniques of Sections 2 and 3. Given two finite sets of \( \mathbb{R}^n \), \( U \) and \( L \), we wish to determine a hyperplane with nonnegative coefficients of the form \( a_1y_1 + \cdots + a_ny_n = \beta \) \((\beta \text{ is given})\) that separates both sets in the sense that \( a^T y \geq \beta \) for all \( y \in U \) and \( a^T y \leq \beta \) for all \( y \in L \). In general, this is impossible, so we wish to determine a hyperplane \( H \) and \( \epsilon > 0 \) such that

\[
a^T y \geq \beta - \epsilon
\]

for all \( y \in U \), and

\[
a^T y \leq \beta + \epsilon
\]

for all \( y \in L \). Among the pairs \((a, \epsilon)\) which satisfy (4.1) and (4.2), we wish to choose the pair that minimizes \( \epsilon \). This problem can be written in the form (1.1) if, without loss of generality, we define

\[
U = \{y^1, \ldots, y^\theta\}, \quad L = \{y^{q+1}, \ldots, y^m\},
\]

\[
f_i(a) = \beta - a^T y_i, \quad i = 1, \ldots, q,
\]

\[
f_i(a) = a^T y_i - \beta, \quad i = q+1, \ldots, m.
\]

In this case, the problem \( AP(\theta) \) is:

\[
\text{Minimize } \sum_{i=1}^q [\beta - a^T y_i - \theta]^2 + \sum_{i=q+1}^m [a^T y_i - \beta - \theta]^2.
\]

For solving (4.3) we use a trust-region method (FMS) introduced by Friedlander, Martínez and Santos [4, 5]. The FMS method generates a sequence of approximations \( a_k \in \mathbb{R}^n \) such that every accumulation point is stationary for (4.3). Since (4.3) is a convex problem, this means that we are able to obtain the solution of it with an arbitrary precision.

Let us describe our test problems. We generated the vectors \( y_1, \ldots, y_m \) with each coordinate randomly chosen in \([0,1]\). We set \( \beta = 100, \quad m = 80, \quad q = 40 \) and we ran our algorithm for different values of \( n \). Initially, we set \( \theta_0 = 0 \). In all cases the acceleration procedure worked very satisfactorily. This allowed us to use the stopping criteria \( |\psi'(\theta_k)| \leq 10^{-8}, \quad 0 < \psi(\theta_k) \leq 10^{-3} \). The results are shown in Table 1, where FMS means the number of iterations performed by this method. These tests were run at a PC-AT 286 computer. Looking at the numerical results, we
see that, though the number of auxiliary problems is generally small, the number of iterations used by the algorithm FMS for solving some problems could be very large. In fact, the FMS method was designed for solving problems with many variables but it is not a second order method. This means that neither convergence to second order stationary points nor quadratic convergence is expected. We will see in the following section that a different situation occurs when the initial vector coding problem is considered.

\[
\begin{array}{|c|c|c|c|c|}
\hline
n & k & \theta_k & \psi(\theta_k) & \psi'(\theta_k) & \text{FMS} \\
\hline
5 & 0 & 0 & 0.18E5 & -0.14E4 & 33 \\
 & 1 & 26.53 & 0.12E4 & -0.19E3 & 74 \\
 & 2 & 39.40 & 0.18E2 & -0.19E2 & 65 \\
 & 3 & 41.30 & 0.47E-1 & -0.84E0 & 42 \\
 & 4 & 41.42 & 0.35E-7 & -0.61E-3 & 22 \\
 & 5 & 41.42 & 0.13E-8 & -0.72E-4 & 0 \\
\hline
10 & 0 & 0 & 0.88E4 & -0.95E3 & 40 \\
 & 1 & 18.48 & 0.42E3 & -0.14E3 & 69 \\
 & 2 & 24.61 & 0.28E1 & -0.96E1 & 69 \\
 & 3 & 25.20 & 0.55E-5 & -0.12E-1 & 107 \\
 & 4 & 25.20 & 0.19E-7 & -0.38E-3 & 2 \\
 & 5 & 25.20 & 0.00E0 & -0.00E0 & 1 \\
\hline
20 & 0 & 0 & 0.25E4 & -0.54E3 & 91 \\
 & 1 & 9.31 & 0.70E2 & -0.63E2 & 55 \\
 & 2 & 11.51 & 0.20E0 & -0.31E1 & 185 \\
 & 3 & 11.64 & 0.20E-3 & -0.81E-1 & 23 \\
 & 4 & 11.64 & 0.38E-9 & -0.39E-4 & 8 \\
\hline
30 & 0 & 0 & 0.19E4 & -0.48E3 & 78 \\
 & 1 & 8.08 & 0.27E2 & -0.43E2 & 911 \\
 & 2 & 9.32 & 0.44E-1 & -0.10E0 & 158 \\
 & 3 & 9.37 & 0.38E-4 & -0.34E-1 & 16 \\
 & 4 & 9.38 & 0.00E0 & -0.00E0 & 2 \\
\hline
40 & 0 & 0 & 0.34E2 & -0.56E2 & 486 \\
 & 1 & 1.20 & 0.59E-2 & -0.73E0 & 190 \\
 & 2 & 1.21 & 0.13E-3 & -0.68E-1 & 4 \\
 & 3 & 1.22 & 0.46E-6 & -0.14E-2 & 2 \\
 & 4 & 1.22 & 0.00E0 & -0.89E-15 & 0 \\
\hline
50 & 0 & 0 & 0.34E2 & -0.56E2 & 486 \\
 & 1 & 1.20 & 0.59E2 & -0.73E0 & 190 \\
 & 2 & 1.21 & 0.13E-3 & -0.68E-1 & 4 \\
 & 3 & 1.22 & 0.46E-6 & -0.14E-2 & 2 \\
 & 4 & 1.22 & 0.00E0 & -0.00E0 & 0 \\
\hline
75 & 0 & 0 & 0.37E2 & -0.54E0 & 86 \\
 & 1 & 0.01 & 0.17E-3 & -0.59E-1 & 4 \\
 & 2 & 0.02 & 0.00E0 & -0.00E0 & 1 \\
\hline
100 & 0 & 0 & 0.11E-2 & -0.23E0 & 47 \\
 & 1 & 0.97E-2 & 0.00E0 & -0.00E0 & 1 \\
\hline
\end{array}
\]

Table 1. Experiments with minimax discrimination.

5. Initial Vector Coding Problems. The Initial Vector Coding Problem in Coding Theory (see [10]) can be described as follows. Let \( G = \{G_1, \ldots, G_m\} \) be a group of permutation matrices of order \( n \) where we have excluded the Identity and we also eliminated one of the matrices of each
pair \( \{ G, G^{-1} \} \). We wish to find \( x \in \mathbb{R}^n \), \( \| x \| = 1 \) such that \( \max \{ x^T G_i x, G_i \in \mathcal{G} \} \) is minimum (\( \| \cdot \| \) denotes the Euclidean norm). Therefore, this problem is of type (1.1), with \( f_i(x) = x^T G_i x \) and \( \Omega = \{ x \in \mathbb{R}^n \mid \| x \| = 1 \} \). In our experiments, \( \mathcal{G} \) was generated using the group of all permutation matrices (symmetric group). Therefore, the number of elements \( m \) grows rapidly with \( n \). When these problems are solved using standard Nonlinear Programming techniques, they produce nonlinear optimization problems with a very large number of constraints and/or variables (see [10]). With the choice we have made for \( \mathcal{G} \), an analytic solution of the problem is known (see [1]), so we can test if the numerical solutions obtained by our method are reliable. In this case, the problem \( AP(\theta) \) takes the form

\[
\text{Minimize} \quad \sum_{i=1}^{m} [x^T G_i x - \theta]^2_+.
\]

(5.1)

For obtaining \( \theta_0 \) at the beginning of Algorithm 2.1, we solved the following problem

\[
\text{Minimize} \quad \left( \sum_{i=1}^{m} [x^T G_i x - 2]^\nu(n) \right)^{1/\nu(n)}
\]

(5.2)

where \( \nu(n) \) is a relatively large integer.

Problems (5.1) and (5.2) were solved using a trust region method for minimization on arbitrary domains due to Martínez and Santos [13]. At each iteration of this method, a trust region problem is defined that consists on the minimization of the quadratic second order approximation of the function on the intersection of the \( n \)-sphere and the trust Euclidean ball. This trust-region subproblem is solved using a characterization of local minimizers of quadratics on spheres given recently by Martínez [12]. The use of this trust-region subproblem guarantees that the algorithm actively tries to find a region where second order conditions of minimizers are satisfied, escaping from regions of “negative curvature”. In addition, locally, the algorithm exhibits quadratic convergence.

The initial interval \( [\theta_0^L, \theta_0^U] \) was obtained as follows. From the minimizer of problem (5.2) we calculated \( \theta_0 = \max_{1 \leq i \leq p} x^T G_i x \). Given \( \varepsilon > 0 \), we found \( [\theta_0^L, \theta_0^U] \subset [-1, \theta_0] \) such that \( \psi(\theta_0^L) > \varepsilon \), \( \psi(\theta_0^U) \leq \varepsilon \) and \( |\theta_0^L - \theta_0^U| = 0.05|\theta_0 + 1| \).

The initial approximation for solving (5.2) was randomly chosen on the unitary sphere. For each problem (5.1) solved, the initial guess was the solution found at the previous problem.

The numerical results are presented in Table 2, where MS means the number of iterations performed by the trust region algorithm mentioned above and Time indicates the total time spent to solve (5.1), including the resolution of (5.2) to initialize Algorithm 2.1. These
experiments were run at a SUN Sparc Station 2, under the Fortran 77 compiler.

<table>
<thead>
<tr>
<th>n</th>
<th>ν(n)</th>
<th>p</th>
<th>k</th>
<th>( \theta_k )</th>
<th>( \psi(\theta_k) )</th>
<th>( \psi'(\theta_k) )</th>
<th>MS</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>10</td>
<td>16</td>
<td>0</td>
<td>0.7373</td>
<td>0.58E-2</td>
<td>-0.13E0</td>
<td>4</td>
<td>0.5s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9001</td>
<td>0.00E0</td>
<td>-0.00E0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>72</td>
<td>0</td>
<td>0.8618</td>
<td>0.28E-2</td>
<td>-0.15E0</td>
<td>5</td>
<td>4s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9001</td>
<td>0.45E-11</td>
<td>-0.42E-1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9001</td>
<td>0.15E-27</td>
<td>-0.18E-13</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>397</td>
<td>0</td>
<td>0.8742</td>
<td>0.11E-1</td>
<td>-0.34E0</td>
<td>7</td>
<td>14s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9087</td>
<td>0.41E-4</td>
<td>-0.20E-1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9429</td>
<td>0.62E-10</td>
<td>-0.16E-4</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.9429</td>
<td>0.47E-14</td>
<td>-0.14E-6</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.9429</td>
<td>0.62E-32</td>
<td>-0.11E-15</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>2335</td>
<td>0</td>
<td>0.8796</td>
<td>0.29E-1</td>
<td>-0.95E0</td>
<td>7</td>
<td>1.40s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9414</td>
<td>0.14E-2</td>
<td>-0.13E0</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9645</td>
<td>0.61E-10</td>
<td>-0.22E-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.9645</td>
<td>0.39E-12</td>
<td>-0.13E-5</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.9645</td>
<td>0.62E-32</td>
<td>-0.11E-15</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>2051</td>
<td>0</td>
<td>0.8866</td>
<td>0.66E-1</td>
<td>-0.23E+1</td>
<td>9</td>
<td>5s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9441</td>
<td>0.36E-2</td>
<td>-0.28E0</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9702</td>
<td>0.12E-3</td>
<td>-0.38E-1</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.9702</td>
<td>0.80E-13</td>
<td>-0.88E-6</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.9702</td>
<td>0.58E-15</td>
<td>-0.54E-7</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.9702</td>
<td>0.30E-16</td>
<td>-0.77E-8</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>18249</td>
<td>0</td>
<td>0.8896</td>
<td>0.17E0</td>
<td>-0.67E+1</td>
<td>4</td>
<td>25s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9416</td>
<td>0.12E-1</td>
<td>-0.85E0</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9694</td>
<td>0.70E-3</td>
<td>-0.10E0</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.9830</td>
<td>0.41E-6</td>
<td>-0.25E-2</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.9833</td>
<td>0.45E-11</td>
<td>-0.72E-5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.9833</td>
<td>0.31E-13</td>
<td>-0.40E-6</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0.9833</td>
<td>0.30E-15</td>
<td>-0.35E-7</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>0.9833</td>
<td>0.00E0</td>
<td>-0.00E0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>181947</td>
<td>0</td>
<td>0.8931</td>
<td>0.47E0</td>
<td>-0.21E+2</td>
<td>7</td>
<td>5h</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.9388</td>
<td>0.38E-1</td>
<td>-0.28E+1</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.9664</td>
<td>0.27E-2</td>
<td>-0.36E-0</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.9813</td>
<td>0.17E-3</td>
<td>-0.52E-1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.9879</td>
<td>0.24E-10</td>
<td>-0.18E-4</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.9879</td>
<td>0.22E-12</td>
<td>-0.11E-5</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0.9879</td>
<td>0.35E-14</td>
<td>-0.84E-7</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>0.9879</td>
<td>0.00E0</td>
<td>-0.00E0</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Experiments with Initial Vector Coding Problems.

6. Conclusions. In this paper we presented a new method for solving nonlinear minimax problems. The idea is to solve these problems using a sequence of differentiable problems, for which we can use any available algorithm. A single parameter must be updated at each step of the process. We developed a simple newtonian rule for this updating. Our numerical experiments show that, in general, no more than five differentiable problems must be solved for each minimax problem.

In principle, the new method allows us to solve some problems which could be very large (either in number of constraints, number of slack variables, or both) if attacked using
conventional Nonlinear Programming approaches.

We do not claim that the new method is more efficient than other existing minimax methods for most problems. Instead, we would like to stress that, when one has a powerful method for minimizing a differentiable function subject to the feasible set \( \Omega \), and the function is very expensive, our approach seems to work well, exploiting the characteristics of \( \Omega \) and saving functional evaluations. These features are present in the initial vector coding problem. The regularization approach of Gígola and Gómez is related to ours, in the sense that a sequence of differentiable problems is considered, but we do not know yet how to exploit the shape of \( \Omega \) in their case.

An open problem that needs to be considered both from the theoretical and the practical point of view concerns the stopping criterion for the differentiable subproblems. In fact, it has to be kept in mind the danger of oversolving them. In our tests the subproblems were solved through an iterative procedure within a prescribed accuracy. The degree of such an accuracy, however, certainly affects the performance of the algorithm.

References


