Institute of Computer Science
Academy of Sciences of the Czech Republic

Primal interior-point method for large sparse minimax optimization

L.Lukšan, C. Matonoha, J. Vlček

Technical report No. 941

October 2005

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 2 688 42 44, fax: +420 2 858 57 89,
e-mail:e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Primal interior-point method for large sparse minimax optimization

L.Lukšan, C. Matonoha, J. Vlček [1]

Technical report No. 941

October 2005

Abstract:

In this paper, we propose an interior-point method for large sparse minimax optimization. After a short introduction, where various barrier terms are discussed, the complete algorithm is introduced and some implementation details are given. We prove that this algorithm is globally convergent under standard mild assumptions. Thus nonconvex problems can be solved successfully. The results of computational experiments given in this paper confirm efficiency and robustness of the proposed method.

Keywords:
Unconstrained optimization, large-scale optimization, nonsmooth optimization, minimax optimization, interior-point methods, modified Newton methods, computational experiments.

[1]This work was supported by the Grant Agency of the Czech Academy of Sciences, project code IAA1030405, and the institutional research plan No. AV0Z10300504. L.Lukšan is also from the Technical University of Liberec, Hálkova 6, 461 17 Liberec.

1 Introduction

Consider the minimax problem: Minimize function

$$F(x) = \max_{1 \le i \le m} f_i(x), \tag{1}$$

where $f_i : R^n \to R$, $1 \le i \le m$ (m is usually large), are smooth functions depending on a small number of variables (n_i, say) and satisfying Assumption 1 and either Assumption 2 or Assumption 3.

Assumption 1. Functions $f_i(x)$, $1 \le i \le m$, are bounded from below on R^n, i.e., there is $\underline{F} \in R$ such that $f_i(x) \ge \underline{F}$, $1 \le i \le m$, for all $x \in R^n$.

Assumption 2. Functions $f_i(x)$, $1 \le i \le m$, are twice continuously differentiable on the convex hull of level set $\mathcal{L}(\overline{F}) = \{x \in R^n : F(x) \le \overline{F}\}$ for a sufficiently large upper bound $\overline{F}$ and they have bounded the first and second-order derivatives on conv$\mathcal{L}(\overline{F})$, i.e., constants $\overline{g}$ and $\overline{G}$ exist such that $\|\nabla f_i(x)\| \le \overline{g}$ and $\|\nabla^2 f_i(x)\| \le \overline{G}$ for all $1 \le i \le m$ and $x \in \text{conv}\mathcal{L}(\overline{F})$.

Assumption 3 Functions $f_i(x)$, $1 \le i \le m$, are twice continuously differentiable on a sufficiently large convex compact set $\mathcal{D}$.

Since continuous functions attain their maxima on a compact set, Assumption 3 guarantees that constants $\overline{F}$, $\overline{g}$ and $\overline{G}$ exist such that $f_i(x) \le \overline{F}$, $\|g_i(x)\| \le \overline{g}$ and $\|G_i(x)\| \le \overline{G}$ for all $x \in \mathcal{D}$. The choice of $\overline{F}$ and $\mathcal{D}$ will be discussed later (see Assumption 4). Note that set conv$\mathcal{L}(\overline{F})$ used in Assumption 2 need not be compact.

Minimization of F is equivalent to the sparse nonlinear programming problem with $n+1$ variables $x \in R^n$, $z \in R$:

$$\text{minimize} \quad z \quad \text{subject to} \quad f_i(x) \le z, \quad 1 \le i \le m. \tag{2}$$

The necessary first-order (KKT) conditions for a solution of (2) have the form

$$\sum_{i=1}^m u_i \nabla f_i(x) = 0, \quad \sum_{i=1}^m u_i = 1, \quad u_i \ge 0, \quad z - f_i(x) \ge 0, \quad u_i(z - f_i(x)) = 0, \tag{3}$$

where u_i, $1 \le i \le m$, are Lagrange multipliers. Problem (2) can be solved by an arbitrary nonlinear programming method utilizing sparsity (sequential linear programming [4], [8]; sequential quadratic programming [6], [7]; interior-point [14], [18]; nonsmooth equation [5], [15]). In this paper, we introduce a feasible primal interior-point method that utilizes a special structure of the minimax problem (1). The constrained problem (2) is replaced by a sequence of unconstrained problems

$$\text{minimize} \quad B_\mu(x, z) = z + \mu \sum_{i=1}^m \varphi(z - f_i(x)), \tag{4}$$

where $\varphi : (0, \infty) \to R$ is a barrier term, $z > F(x)$ and $\mu > 0$ (we assume that $\mu \to 0$ monotonically). In connection with barrier terms, we will consider the following conditions.

Condition 1. $\varphi(t)$ is a decreasing strictly convex function with a negative third-order derivative such that $\lim_{t\to 0}\varphi(t) = \infty$ and $\lim_{t\to\infty}\varphi'(t) = 0$.

Condition 2. $\varphi(t)$ is a positive function such that $\lim_{t\to\infty}\varphi(t) = 0$.

Condition 1 is essential, we assume its validity for every barrier term. Condition 2 is useful for investigation of the global convergence, since it assures that function $F(x)$ is bounded from above (Lemma 9).

The most known and frequently used logarithmic barrier term

$$\varphi(t) = \log t^{-1} = -\log t, \tag{5}$$

satisfies Condition 1, but it does not satisfy Condition 2, since it is non-positive for $t \ge 1$. Therefore, additional barrier terms have been studied. In [1], a truncated logarithmic barrier term is considered such that $\varphi(t)$ is given by (5) for $t \le \tau$ and $\varphi(t) = a/t^2 + b/t + c$ for $t > \tau$, where a, b, c are chosen in such a way that $\varphi(t)$ is twice continuously differentiable in $(0,\infty)$, which implies that $a = -\tau^2/2$, $b = 2\tau$, $c = -\log\tau - 3/2$. Choosing $\tau = \overline{\tau}$, where $\overline{\tau} = \exp(-3/2)$, one has $\lim_{t\to\infty}\varphi(t) = c = 0$ and Condition 2 is satisfied. In this paper we will investigate four particular barrier terms, which are introduced in Table 1 together with their derivatives.

	$\varphi(t)$	$\varphi'(t)$	$\varphi''(t)$	
B1	$-\log t$	$-1/t$	$1/t^2$	
B2	$-\log t$	$-1/t$	$1/t^2$	$t \le \overline{\tau}$
	$2\overline{\tau}/t - \overline{\tau}^2/(2t^2)$	$-2\overline{\tau}/t^2 + \overline{\tau}^2/t^3$	$4\overline{\tau}/t^3 - 3\overline{\tau}^2/t^4$	$t > \overline{\tau}$
B3	$\log(t^{-1}+1)$	$-1/(t(t+1))$	$(2t+1)/(t(t+1))^2$	
B4	$1/t$	$-1/t^2$	$2/t^3$	

Table 1: Barrier terms ($\overline{\tau} = \exp(-3/2)$).

All these barrier terms satisfy Condition 1 and B2–B4 satisfy also Condition 2. Recall that Condition 1 implies $\varphi'(t) < 0$, $\varphi''(t) > 0$ and $\varphi'''(t) < 0$ for all $t \in (0,\infty)$. Inequality $\varphi'''(t) < 0$ can be easily proved for all barrier terms given in Table 1. For example, considering B3 we can write

$$\varphi'''(t) = -2\frac{3t^2+3t+1}{t^3(t+1)^3} < 0.$$

Note that our theory refers to all barrier terms satisfying Condition 1 (not only B1–B4).

The primal interior-point method described in this paper is iterative, i.e., it generates a sequence of points $x_k \in R^n$, $k \in N$ (N is the set of integers). For proving the global convergence, we need the following assumption concerning function $F(x)$ and sequence $\{x_k\}_1^\infty$.

Assumption 4 Either Assumption 2 holds and $\{x_k\}_1^\infty \in \mathcal{L}(\overline{F})$ or Assumption 3 holds and $\{x_k\}_1^\infty \in \mathcal{D}$.

The primal interior-point method investigated in this paper is a line-search modification of the Newton method. Approximation of the Hessian matrix is computed by the gradient differences which can be carried out efficiently if the Hessian matrix is sparse (see [2]). Since the Hessian matrix need not be positive definite in the non-convex case, the standard line-search realization cannot be used. There are two basic possibilities, either a trust-region approach or the line-search strategy with suitable restarts, which eliminate this insufficiency. We have implemented and tested both these possibilities and our tests have shown that the second possibility, used in Algorithm 1, is more efficient.

The paper is organized as follows. In Section 2, we introduce the interior-point method for large sparse minimax optimization and describe the corresponding algorithm. Section 3 contains more details concerning this algorithm such as the restart strategy and the barrier parameter update. In Section 4 we study theoretical properties of the interior-point method and prove that this method is globally convergent if Assumption 1 and Assumption 4 hold. Section 5 contains a short description of the smoothing method for large sparse minimax approximation, which has been used for a comparison. Finally, in Section 6 we present results of computational experiments confirming the efficiency of the proposed method.

2 Description of the method

Differentiating $B_\mu(x,z)$ given by (4), we obtain necessary conditions for minimum in the form

$$-\mu \sum_{i=1}^{m} \varphi'(z - f_i(x))\, \nabla f_i(x) = 0, \quad 1 + \mu \sum_{i=1}^{m} \varphi'(z - f_i(x)) = 0, \tag{6}$$

where $\varphi'(z - f_i(x)) < 0$ for all $1 \le i \le m$. Denoting $g_i(x) = \nabla f_i(x)$, $1 \le i \le m$, $A(x) = [g_1(x), \dots, g_m(x)]$ and

$$f(x) = \begin{bmatrix} f_1(x) \\ \dots \\ f_m(x) \end{bmatrix}, \quad u_\mu(x,z) = \begin{bmatrix} -\mu\, \varphi'(z - f_1(x)) \\ \dots \\ -\mu\, \varphi'(z - f_m(x)) \end{bmatrix}, \quad e = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}, \tag{7}$$

we can write (6) in the form

$$A(x) u_\mu(x,z) = 0, \quad 1 - e^T u_\mu(x,z) = 0. \tag{8}$$

Nonlinear equations (8) can be solved by the Newton method. For this purpose, we need second-order derivatives of $B_\mu(x,z)$. One has

$$\begin{aligned} \frac{\partial A(x) u_\mu(x,z)}{\partial x} &= \sum_{i=1}^{m} u_\mu(x,z)_i G_i(x) + \mu \sum_{i=1}^{m} \varphi''(z - f_i(x))\, g_i(x) g_i^T(x) \\ &= G_\mu(x,z) + A(x) V_\mu(x,z) A^T(x), \\ \frac{\partial A(x) u_\mu(x,z)}{\partial z} &= -\mu \sum_{i=1}^{m} \varphi''(z - f_i(x))\, g_i(x) = -A(x) V_\mu(x,z) e, \end{aligned}$$

$$\begin{aligned}\frac{\partial(1-e^T u_\mu(x,z))}{\partial x} &= -\mu\sum_{i=1}^m \varphi''(z-f_i(x))\, g_i^T(x) = -e^T V_\mu(x,z)A^T(x),\\ \frac{\partial(1-e^T u_\mu(x,z))}{\partial z} &= \mu\sum_{i=1}^m \varphi''(z-f_i(x)) = e^T V_\mu(x,z)e,\end{aligned}$$

where $G_i(x) = \nabla^2 f_i(x)$, $1 \le i \le m$, $G_\mu(x,z) = \sum_{i=1}^m u_\mu(x,z)_i G_i(x)$ and $V_\mu(x,z) = \mu\,\mathrm{diag}(\varphi''(z-f_1(x)),\dots,\varphi''(z-f_m(x)))$. Using these expressions, we obtain a set of linear equations corresponding to one step of the Newton method

$$\begin{bmatrix} G_\mu(x,z)+A(x)V_\mu(x,z)A^T(x) & -A(x)V_\mu(x,z)e \\ -e^T V_\mu(x,z)A^T(x) & e^T V_\mu(x,z)e \end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta z\end{bmatrix} = -\begin{bmatrix} A(x)u_\mu(x,z)\\ 1-e^T u_\mu(x,z)\end{bmatrix} \tag{9}$$

or equivalently

$$\left(\begin{bmatrix} G_\mu(x,z) & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix} A(x)\\ -e^T\end{bmatrix} V_\mu(x,z)\,[\,A^T(x)\ \ -e\,]\right)\begin{bmatrix}\Delta x\\ \Delta z\end{bmatrix} = -\begin{bmatrix} A(x)u_\mu(x,z)\\ 1-e^T u_\mu(x,z)\end{bmatrix}. \tag{10}$$

Note that matrix $V_\mu(x,z)$ is positive definite, since $\varphi''(t) > 0$ for $t \in (0,\infty)$ by Condition 1.

Increments Δx and Δz determined from (9) can be used for obtaining new quantities

$$x^+ = x + \alpha\Delta x, \quad z^+ = z + \alpha\Delta z,$$

where $\alpha > 0$ is a suitable step-size. This is a standard way for solving general nonlinear programming problems. For special nonlinear programming problem (2), the structure of $B_\mu(x,z)$ allows us to obtain minimizer $z_\mu(x) \in R$ of function $B_\mu(x,z)$ for a given $x \in R^n$.

Lemma 1. *Let Condition 1 be satisfied. Then function $B_\mu(x,.) : (F(x),\infty) \to R$ (with x fixed) has the unique stationary point, which is its global minimizer. This stationary point is characterized by equation*

$$e^T u_\mu(x,z) = 1. \tag{11}$$

Solution $z_\mu(x)$ of this equation satisfies inequalities

$$F(x) + \underline{t}_\mu = \underline{z}_\mu(x) \le z_\mu(x) \le \overline{z}_\mu(x) = F(x) + \overline{t}_\mu,$$

where values $0 < \underline{t}_\mu \le \overline{t}_\mu$, independent of x, can be obtained as unique solutions of equations

$$1 + \mu\varphi'(\underline{t}_\mu) = 0, \quad 1 + m\mu\varphi'(\overline{t}_\mu) = 0. \tag{12}$$

Moreover

$$e^T u_\mu(x,\overline{z}_\mu(x)) \le 1 \le e^T u_\mu(x,\underline{z}_\mu(x)). \tag{13}$$

Proof. Function $B_\mu(x,.) : (F(x),\infty) \to R$ is convex in $(F(x),\infty)$, since it is a sum of convex functions. Thus if a stationary point of $B_\mu(x,.)$ exists, it is its unique global minimizer. Since $\varphi'(z_\mu(x) - f_i(x)) < 0$ and $\varphi''(z_\mu(x) - f_i(x)) > 0$ for all $1 \le i \le m$ by Condition 1, we can write

$$\begin{aligned}\varphi'(z_\mu(x) - F(x)) &\ge \sum_{i=1}^m \varphi'(z_\mu(x) - f_i(x)) \ge \sum_{i=1}^m \varphi'(z_\mu(x) - F(x)) \\ &= m\varphi'(z_\mu(x) - F(x)).\end{aligned} \tag{14}$$

Thus if we choose $\underline{z}_\mu(x) = F(x) + \underline{t}_\mu$, $\overline{z}_\mu(x) = F(x) + \overline{t}_\mu$ in such a way that (12) hold, we obtain inequalities

$$1 + \sum_{i=1}^m \mu\varphi'(\underline{z}_\mu(x) - f_i(x)) \le 0 \le 1 + \sum_{i=1}^m \mu\varphi'(\overline{z}_\mu(x) - f_i(x)), \tag{15}$$

which are equivalent to (13). Inequalities (13) imply that solution $z_\mu(x)$ of (11) (the stationary point of $B_\mu(x,.)$) exists. Since function $\varphi'(t)$ is increasing, we obtain $F(x) < \underline{z}_\mu(x) \le z_\mu(x) \le \overline{z}_\mu(x)$. The above considerations are correct, since continuous function $\varphi'(t)$ maps $(0, \infty)$ onto $(-\infty, 0)$, which implies that equations (12) have unique solutions. □

Solutions of equations (12) give the most sophisticated bounds for $t_\mu = z_\mu(x) - F(x)$. Unfortunately, sometimes these equations cannot be solved analytically. In these cases we use weaker bounds satisfying inequalities $1 + \mu\varphi'(\underline{t}_\mu) \le 0$ and $1 + m\mu\varphi'(\overline{t}_\mu) \ge 0$. We apply this way for barrier term B2, where (12) lead to cubic equations if $\mu > \overline{\tau}$.

Corollary 1. *Bounds $\underline{t}_\mu$, $\overline{t}_\mu$ for $t_\mu = z_\mu(x) - F(x)$, corresponding to barrier terms B1–B4, are given in Table 2.*

	$\underline{t}_\mu = \underline{z}_\mu - F(x)$	$\overline{t}_\mu = \overline{z}_\mu - F(x)$
B1	μ	$m\mu$
B2	$\min(\mu, \overline{\tau})$	$m\mu$
B3	$2\mu/(1 + \sqrt{1 + 4\mu})$	$2m\mu/(1 + \sqrt{1 + 4m\mu})$
B4	$\sqrt{\mu}$	$\sqrt{m\mu}$

Table 2: Bounds for $t_\mu = z_\mu(x) - F(x)$.

Proof. (a) Consider first the logarithmic barrier term B1. Then $\varphi'(t) = -1/t$, which together with (12) gives $\underline{t}_\mu = \mu$ and $\overline{t}_\mu = m\mu$.

(b) Consider now barrier term B2. Since $\varphi'(t) = -1/t$ for $t \le \overline{\tau}$ and

$$\varphi'(t) + \frac{1}{t} = \left(\frac{\overline{\tau}^2}{t^3} - \frac{2\overline{\tau}}{t^2} + \frac{1}{t}\right) = \frac{1}{t^3}(t^2 - 2\overline{\tau}t + \overline{\tau}^2) = \frac{1}{t^3}(t - \overline{\tau})^2 \ge 0$$

for $t > \overline{\tau}$, we can conclude that $\varphi'(t)$ of B2 is not less than $\varphi'(t)$ of B1, which implies that $\overline{t}_\mu = m\mu$, the upper bound of B1, is also the upper bound of B2. Since $\varphi'(t)$ of B2 is equal to $\varphi'(t)$ of B1 for $t \le \overline{\tau}$, we can set $\underline{t}_\mu = \mu$ if $\mu \le \overline{\tau}$. At the same time, $\underline{t}_\mu = \overline{\tau}$ is a suitable lower bound of B2 if $\mu > \overline{\tau}$.

(c) Consider now barrier term B3. Then $\varphi'(t) = -1/(t^2+t)$ which together with (12) gives $\mu/(\underline{t}_\mu^2 + \underline{t}_\mu) = 1$. Thus $\underline{t}_\mu$ is a positive solution of the quadratic equation $t^2+t-\mu = 0$ and it can be written in the form

$$\underline{t}_\mu = \frac{\sqrt{1+4\mu}-1}{2} = \frac{2\mu}{1+\sqrt{1+4\mu}}.$$

The upper bound can be obtained by the same way.

(d) Consider finally barrier term B4. Then $\varphi'(t) = -1/t^2$ which together with (12) gives $\underline{t}_\mu = \sqrt{\mu}$ and $\overline{t}_\mu = \sqrt{m\mu}$. □

Solution $z_\mu(x)$ of nonlinear equation (11) can be obtained by efficient methods proposed in [9], [10], which use localization inequalities (13). Thus we will assume that $z = z_\mu(x)$ with a sufficient precision, which implies that the last elements of the right-hand sides in (9) – (10) are negligible. Assuming $z = z_\mu(x)$, we denote

$$B_\mu(x) = B_\mu(x, z_\mu(x)) = z_\mu(x) + \mu \sum_{i=1}^m \varphi(z_\mu(x) - f_i(x)), \qquad (16)$$

$u_\mu(x) = u_\mu(x, z_\mu(x))$, $V_\mu(x) = V_\mu(x, z_\mu(x))$ and $G_\mu(x) = G_\mu(x, z_\mu(x))$. In this case, barrier function $B_\mu(x)$ depends only on x. In order to obtain minimizer $(x, z) \in R^{n+1}$ of $B_\mu(x, z)$, it suffices to minimize $B_\mu(x)$ over R^n.

Lemma 2. *Consider barrier function (16). Then*

$$\nabla B_\mu(x) = A(x) u_\mu(x), \qquad (17)$$

and

$$\nabla^2 B_\mu(x) = G_\mu(x) + A(x)V_\mu(x)A^T(x) - \frac{A(x)V_\mu(x)ee^TV_\mu(x)A^T(x)}{e^TV_\mu(x)e}. \qquad (18)$$

Solution Δx of the Newton equation

$$\nabla^2 B_\mu(x)\Delta x = -\nabla B_\mu(x) \qquad (19)$$

is equal to the corresponding vector obtained by solving (9) with $z = z_\mu(x)$.

Proof. Differentiating $B_\mu(x)$, we obtain

$$\begin{aligned}
\nabla B_\mu(x) &= \nabla z_\mu(x) + \mu \sum_{i=1}^m \varphi'(z_\mu(x) - f_i(x))\,(\nabla z_\mu(x) - g_i(x)) \\
&= \nabla z_\mu(x)\left(1 + \mu \sum_{i=1}^m \varphi'(z_\mu(x) - f_i(x))\right) - \mu \sum_{i=1}^m \varphi'(z_\mu(x) - f_i(x))\, g_i(x) \\
&= -\mu \sum_{i=1}^m \varphi'(z_\mu(x) - f_i(x))\, g_i(x) = A(x)u_\mu(x),
\end{aligned}$$

since

$$1 - e^T u_\mu(x) = 1 + \mu \sum_{i=1}^{m} \varphi'(z_\mu(x) - f_i(x)) = 0.$$

Differentiating the last equality, one has

$$\mu \sum_{i=1}^{m} \varphi''(z_\mu(x) - f_i(x)) \, (\nabla z_\mu(x) - g_i(x)) = 0,$$

which gives

$$\nabla z_\mu(x) = \frac{A(x) V_\mu(x) e}{e^T V_\mu(x) e}.$$

Thus

$$\begin{aligned}
\nabla^2 B_\mu(x) &= \sum_{i=1}^{m} u_\mu(x)_i \, G_i(x) + \mu \sum_{i=1}^{m} \varphi''(z_\mu(x) - f_i(x)) \, (g_i(x) - \nabla z_\mu(x)) \, g_i^T(x) \\
&= G_\mu(x) + A(x) V_\mu(x) A^T(x) - \frac{A(x) V_\mu(x) e e^T V_\mu(x) A^T(x)}{e^T V_\mu(x) e}.
\end{aligned}$$

Using the second equation of (9) with $e^T u_\mu(x) = 1$, we obtain

$$\Delta z = \frac{e^T V_\mu(x) A^T(x)}{e^T V_\mu(x) e} \Delta x,$$

which after substituting into the first equation gives

$$\left(G_\mu(x) + A(x) V_\mu(x) A^T(x) - \frac{A(x) V_\mu(x) e e^T V_\mu(x) A^T(x)}{e^T V_\mu(x) e} \right) \Delta x = -A(x) u_\mu(x).$$

This is exactly equation (19). □

Note that we use (9) rather than (19) for direction determination, since nonlinear equation (11) is solved with precision $\underline{\delta}$ and, therefore, in general $|1 - e^T u_\mu(x)|$ differs from zero.

Lemma 3. *Let Δx solve (19) (or (9) with $z = z_\mu(x)$). If matrix $G_\mu(x)$ is positive definite, then $(\Delta x)^T \nabla B_\mu(x) < 0$ (direction vector Δx is descent for $B_\mu(x)$).*

Proof. Equation (19) implies

$$(\Delta x)^T \nabla^2 B_\mu(x) \Delta x = -(\Delta x)^T \nabla B_\mu(x).$$

Thus $(\Delta x)^T \nabla B_\mu(x) < 0$ if $\nabla^2 B_\mu(x)$ is positive definite. But

$$\begin{aligned}
v^T \nabla^2 B_\mu(x) v &= v^T G_\mu(x) v + \left(v^T A(x) V_\mu(x) A^T(x) v - \frac{(v^T A(x) V_\mu(x) e)^2}{e^T V_\mu(x) e} \right) \\
&\geq v^T G_\mu(x) v
\end{aligned}$$

for an arbitrary $v \in R^n$ by (18) and by the Schwarz inequality (since $V_\mu(x)$ is positive definite). Thus $(\Delta x)^T \nabla B_\mu(x) < 0$ if $G_\mu(x)$ is positive definite. □

Consider the logarithmic barrier function B1. Then

$$V_\mu(x) = \frac{1}{\mu} U_\mu^2(x),$$

where $U_\mu(x) = \mathrm{diag}(u_\mu(x)_1, \ldots, u_\mu(x)_m)$, which implies that $\|V_\mu(x)\| \to \infty$ as $\mu \to 0$. Thus $\nabla^2 B_\mu(x)$ can be ill-conditioned (see (18)). For this reason, it is necessary to use lower bound $\underline{\mu}$ for μ (more details are given in Section 3). The following lemma gives upper bounds for $\|\nabla^2 B_\mu(x)\|$ if

$$\varphi'(t)\varphi'''(t) \geq \varphi''(t)^2 \quad \forall t > 0. \tag{20}$$

This inequality holds for barrier terms B1–B4 as can be verified by using Table 1.

Lemma 4. *Let Assumption 4 hold and Condition 1 together with (20) be satisfied. If* $\mu \geq \underline{\mu} > 0$, *then*

$$\|\nabla^2 B_\mu(x)\| \leq m(\overline{G} + \overline{g}^2 \|V_\mu(x)\|) \leq m(\overline{G} + \overline{g}^2 \overline{V}),$$

where $\overline{V} = \underline{\mu}\varphi''(\underline{t}_{\underline{\mu}})$.

Proof. Using (18) and Assumption 4, we obtain

$$\begin{aligned} \|\nabla^2 B_\mu(x)\| &\leq \left\| G_\mu(x) + A(x) V_\mu(x) A^T(x) \right\| \\ &\leq \left\| \sum_{i=1}^m u_\mu(x)_i G_i(x) \right\| + \left\| \sum_{i=1}^m V_\mu(x)_i g_i(x) g_i^T(x) \right\| \\ &\leq m\overline{G} + m\overline{g}^2 \|V_\mu(x)\|. \end{aligned}$$

Since $V_\mu(x)$ is diagonal and $f_i(x) \leq F(x)$ for all $1 \leq i \leq m$, one has

$$\|V_\mu(x)\| = \mu\varphi''(z_\mu(x) - F(x)) \leq \mu\varphi''(\underline{t}_\mu).$$

Now we prove that $\mu\varphi''(\underline{t}_\mu)$ is a non-increasing function of μ, which implies that $\mu\varphi''(\underline{t}_\mu) \leq \underline{\mu}\varphi''(\underline{t}_{\underline{\mu}})$. Differentiating (12) by μ, we obtain

$$\varphi'(\underline{t}_\mu) + \mu\varphi''(\underline{t}_\mu)\underline{t}'_\mu = 0 \quad \Rightarrow \quad \underline{t}'_\mu = -\frac{\varphi'(\underline{t}_\mu)}{\mu\varphi''(\underline{t}_\mu)} > 0, \tag{21}$$

where $\underline{t}'_\mu$ is a derivative of $\underline{t}_\mu$ by μ. Thus we can write

$$\frac{\mathrm{d}(\mu\varphi''(\underline{t}_\mu))}{\mathrm{d}\mu} = \varphi''(\underline{t}_\mu) + \mu\varphi'''(\underline{t}_\mu)\underline{t}'_\mu = \varphi''(\underline{t}_\mu) - \varphi'''(\underline{t}_\mu)\frac{\varphi'(\underline{t}_\mu)}{\varphi''(\underline{t}_\mu)} \leq 0$$

by (20) and the fact that $\varphi''(\underline{t}_\mu) > 0$. □

Corollary 2. *Bounds* $\overline{V} = \underline{\mu}\varphi''(\underline{t}_\mu)$, *corresponding to barrier terms B1–B4, are given in Table 3, where we assume that* $\underline{\mu} \le \overline{\tau}$ *for B2 and* $\underline{\mu} \le 1/2$ *for B3 (this is possible, since* $\underline{\mu}$ *is a small number).*

	Bounds	Restrictions
B1	$\overline{V} = \underline{\mu}^{-1}$	
B2	$\overline{V} = \underline{\mu}^{-1}$	$\underline{\mu} \le \overline{\tau}$
B3	$\overline{V} \le 2\underline{\mu}^{-1}$	$\underline{\mu} \le 1/2$
B4	$\overline{V} = 2\underline{\mu}^{-1/2}$	

Table 3.

Proof. We use expressions for $\varphi''(t)$ given in Table 1 and formulas for $\underline{t}_\mu$ given in Table 2.

(a) Consider first the logarithmic barrier term B1. In this case $\mu\varphi''(\underline{t}_\mu) = \mu\varphi''(\mu) = \mu^{-1}$.

(b) Consider now barrier term B2. Assuming $\mu \le \overline{\tau}$, we obtain the same bound as in the previous case.

(c) Consider now barrier term B3. Assuming $\mu \le 1/2$, we can write

$$\mu\varphi''(\underline{t}_\mu) = \mu\frac{2\underline{t}_\mu + 1}{(\underline{t}_\mu^2 + \underline{t}_\mu)^2} = \frac{1}{\mu}\left(1 + \frac{4\mu}{1+\sqrt{1+4\mu}}\right) \le \frac{2}{\mu},$$

since $\underline{t}_\mu^2 + \underline{t}_\mu = \mu$ (see proof of Corollary 1).

(d) Consider finally barrier term B4. In this case $\mu\varphi''(\underline{t}_\mu) = \mu\varphi''(\sqrt{\mu}) = 2\mu/\mu^{3/2} = 2\mu^{-1/2}$.
□

As we can deduce from Table 2 and Table 3, properties of barrier function B4 depend on $\sqrt{\mu}$ instead of μ. For this reason, we have used μ^2 instead of μ in barrier function B4 in our computational experiments.

Now we return to the direction determination. To simplify the notation, we can write equation (9) in the form

$$\begin{bmatrix} H & -a \\ -a^T & \alpha \end{bmatrix} \begin{bmatrix} d \\ \delta \end{bmatrix} = \begin{bmatrix} b \\ \beta \end{bmatrix} \tag{22}$$

where

$$H = G_\mu(x,z) + A(x)V_\mu(x,z)A^T(x), \tag{23}$$

and $a = A(x)V_\mu(x,z)e$, $\alpha = e^TV_\mu(x,z)e$, $b = -A(x)u_\mu(x,z)$, $\beta = -(1 - e^Tu_\mu(x,z))$. Since

$$\begin{bmatrix} H & -a \\ -a^T & \alpha \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} - H^{-1}a(a^TH^{-1}a - \alpha)^{-1}a^TH^{-1} & -H^{-1}a(a^TH^{-1}a - \alpha)^{-1} \\ -(a^TH^{-1}a - \alpha)^{-1}a^TH^{-1} & -(a^TH^{-1}a - \alpha)^{-1} \end{bmatrix},$$

we can write

$$\begin{bmatrix} d \\ \delta \end{bmatrix} = \begin{bmatrix} H & -a \\ -a^T & \alpha \end{bmatrix}^{-1} \begin{bmatrix} b \\ \beta \end{bmatrix} = \begin{bmatrix} H^{-1}(b + a\delta) \\ \delta \end{bmatrix}, \tag{24}$$

where

$$\delta = -(a^T H^{-1} a - \alpha)^{-1} (a^T H^{-1} b + \beta).$$

Matrix H is sparse if $A(x)$ has sparse columns. If H is not positive definite, it is advantageous to change it before computation of the direction vector. Thus we use the sparse Gill-Murray decomposition

$$H + E = LDL^T, \tag{25}$$

where E is a positive semidefinite diagonal matrix that assures positive definiteness of LDL^T. Using the Gill-Murray decomposition, we solve two equations

$$LDL^T c = a, \quad LDL^T v = b \tag{26}$$

and set

$$\delta = -\frac{a^T v + \beta}{a^T c - \alpha}, \quad d = v + \delta c. \tag{27}$$

Now we are in position to describe the basic algorithm, in which the direction vector is modified in such a way that

$$-g^T d \geq \varepsilon_0 \|g\| \|d\|, \quad \underline{c}\|g\| \leq \|d\| \leq \overline{c}\|g\|, \tag{28}$$

where ε_0, $\underline{c}$, $\overline{c}$ are suitable constants.

Algorithm 1.

Data: Termination parameter $\underline{\varepsilon} > 0$, precision for the nonlinear equation solver $\underline{\delta} > 0$, minimum value of the barrier parameter $\underline{\mu} > 0$, rate of the barrier parameter decrease $0 < \lambda < 1$, restart parameters $0 < \underline{c} < \overline{c}$ and $\varepsilon_0 > 0$, line search parameter $\varepsilon_1 > 0$, rate of the step-size decrease $0 < \beta < 1$, step bound $\overline{\Delta} > 0$.

Input: Sparsity pattern of matrix $A(x)$. Initial estimation of vector x.

Step 1: *Initiation.* Choose value $\mu > 0$ (e.g. $\mu = 1$). Determine the sparsity pattern of matrix $H(x)$ from the sparsity pattern of matrix $A(x)$. Carry out symbolic decomposition of $H(x)$. Compute values $f_i(x)$, $1 \leq i \leq m$, and $F(x) = \max_{1 \leq i \leq m} f_i(x)$. Set $k := 0$ (iteration count) and $r := 0$ (restart indicator).

Step 2: *Termination.* Solve nonlinear equation (11) with precision $\underline{\delta}$ to obtain value $z_\mu(x)$ and vector $u_\mu(x) = u_\mu(x, z_\mu(x))$. Compute matrix $A := A(x)$ and vector $g := g_\mu(x) = A(x) u_\mu(x)$. If $\mu \leq \underline{\mu}$ and $\|g\| \leq \underline{\varepsilon}$, then terminate the computation. Otherwise set $k := k + 1$.

Step 3: *Approximation of the Hessian matrix.* Compute approximation G of Hessian matrix $G_\mu(x)$ by using differences $A(x + \delta v) u_\mu(x) - g_\mu(x)$ for a suitable set of vectors v (see [2]).

Step 4: *Direction determination.* Determine matrix H by (23). Determine vector d from (26)-(27) by using the Gill-Murray decomposition (25) of matrix H.

Step 5: *Restart.* If $r = 0$ and (28) does not hold, choose positive definite diagonal matrix D by formula (31) introduced in Section 3, set $G := D$, $r := 1$ and go to Step 4. If $r = 1$ and (28) does not hold, set $d := -g$ (steepest descent direction) and $r := 0$.

Step 6: *Step-length selection.* Define maximum step-length $\overline{\alpha} = \min(1, \overline{\Delta}/\|d\|)$. Find the minimum integer $l \geq 0$ such that $B_\mu(x + \beta^l \overline{\alpha} d) \leq B_\mu(x) + \varepsilon_1 \beta^l \overline{\alpha} g^T d$ (note that nonlinear equation (11) has to be solved at all points $x + \beta^j \overline{\alpha} d$, $0 \leq j \leq l$). Set $x := x + \beta^l \overline{\alpha} d$. Compute values $f_i(x)$, $1 \leq i \leq m$, and $F(x) = \max_{1 \leq i \leq m} f_i(x)$.

Step 7: *Barrier parameter update.* Determine a new value of the barrier parameter $\mu \geq \underline{\mu}$ (not greater than the current one) by the procedure described in Section 3. Go to Step 2.

The above algorithm requires several notes. The restart strategy in Step 5 implies that the direction vector d is uniformly descent and gradient-related (see (28)). Since function $B_\mu(x)$ is smooth, the line search utilized in Step 6 always finds a step size satisfying the Armijo condition $B_\mu(x+\alpha d) \leq B_\mu(x)+\varepsilon_1 \alpha g^T d$. The use of the maximum step-length $\overline{\Delta}$ has no theoretical significance but is very useful for practical computations. First, the problem functions can sometimes be evaluated only in a relatively small region (if they contain exponentials) so that the maximum step-length is necessary. Secondly, the problem can be very ill-conditioned far from the solution point, thus large steps are unsuitable. Finally, if the problem has more local solutions, a suitably chosen maximum step-length can cause a better local solution to be reached. Therefore, maximum step-length $\overline{\Delta}$ is a parameter, which is most frequently tuned.

The important part of Algorithm 1 is the update of barrier parameter μ. There are several influences that should be taken into account, which make the updating procedure rather complicated. More details are given in Section 3.

Note finally, that the proposed interior-point method is very similar algorithmically (but not theoretically) to the smoothing method described in [17] and [19]. Thus Algorithm 1 can be easily adapted to one implementing the smoothing method (see Section 5). These methods are compared in Section 6.

3 Implementation details

In Section 2, we have proved (Lemma 3) that direction vector d obtained by solving (22) is descent for $\nabla B_\mu(x)$ if matrix $G_\mu(x)$ is positive definite. Unfortunately, positive definiteness of this matrix is not assured. A similar problem arises in connection with the Newton method for smooth unconstrained optimization, where trust-region methods were developed for this reason. Since we use the line-search approach, we solve this problem by suitable restarts. In this case, matrix $G_\mu(x)$ is replaced by a positive definite diagonal matrix $D = \mathrm{diag}(D_{ii})$ if (28) (with $g = A(x)u_\mu(x)$) does not hold . Thus the Hessian matrix $\nabla^2 B_\mu(x)$ is replaced by the matrix

$$B = D + A(x)V_\mu(x)A^T(x) - \frac{A(x)V_\mu(x)ee^T V_\mu(x)A^T(x)}{e^T V_\mu(x)e} \tag{29}$$

(see (18)). Let $0 < \underline{\gamma} \le D_{ii} \le \overline{\gamma}$ for all $1 \le i \le n$. Then the minimum eigenvalue of B is not less than $\underline{\gamma}$ (see proof of Lemma 3) and, using the same way as in the proof of Lemma 4, we can write

$$\|B\| \le \left\| D + A(x)V_\mu(x)A^T(x) \right\| \le \overline{\gamma} + m\overline{g}^2 \|V_\mu(x)\| \le \overline{\gamma} + m\overline{g}^2 \overline{V}, \tag{30}$$

where bounds $\overline{V}$ for individual barrier terms can be found in Table 3 (procedure used in Step 7 of Algorithm 1 assures that $\underline{\mu} \le \mu \le \mu_1$). Using these inequalities, we can write

$$\kappa(B) \le \frac{1}{\underline{\gamma}}(\overline{\gamma} + m\overline{g}^2 \overline{V}).$$

If d solves equation $Bd + g = 0$, then (28) hold with $\varepsilon_0 \le 1/\kappa(B)$, $\underline{c} \le \underline{\gamma}$ and $\overline{c} \ge \underline{\gamma}\kappa(B)$ (see [3]). If these inequalities are not satisfied, the case when (28) does not hold can appear. In this case we simply set $d = -g$ (this situation appears rarely when ε_0, $\underline{c}$ and $1/\overline{c}$ are sufficiently small).

The choice of matrix D in restarts strongly affects efficiency of Algorithm 1 for problems with indefinite Hessian matrices. We have tested various possibilities including the simple choice $D = I$, which proved to be unsuitable. The best results were obtained by the heuristic procedure proposed in [13] for equality constrained optimization and used in [14] in connection with interior-point methods for nonlinear programming. This procedure uses formulas

$$\begin{aligned} D_{ii} &= \underline{\gamma}, \quad \text{if} \quad \frac{\|g\|}{10}|H_{ii}| < \underline{\gamma}, \\ D_{ii} &= \frac{\|g\|}{10}|H_{ii}|, \quad \text{if} \quad \underline{\gamma} \le \frac{\|g\|}{10}|H_{ii}| \le \overline{\gamma}, \\ D_{ii} &= \overline{\gamma}, \quad \text{if} \quad \overline{\gamma} < \frac{\|g\|}{10}|H_{ii}|, \end{aligned} \tag{31}$$

where $\underline{\gamma} = 0.005$, $\overline{\gamma} = 500.0$ and H is given by (23).

Lemma 5. *Direction vectors d_k, $k \in N$, generated by Algorithm 1 are uniformly descent and gradient-related ((28) hold for all $k \in N$). If Assumption 4 holds, then the Armijo line search (Step 6 of Algorithm 1) assures that a constant c exists such that*

$$B_{\mu_k}(x_{k+1}) - B_{\mu_k}(x_k) \le -c\|g_{\mu_k}(x_k)\|^2 \quad \forall k \in N. \tag{32}$$

Proof. Inequalities (28) are obvious (they are assured by the restart strategy) and inequality (32) is proved, e.g., in [3] (note that $\nabla B_{\mu_k}(x_k) = g_{\mu_k}(x_k)$ by Lemma 2). □

A very important part of Algorithm 1 is the update of the barrier parameter μ. There are two requirements, which play opposite roles. First $\mu \to 0$ should hold, since this is the main property of every interior-point method. On the other hand, round-off errors can cause that $z_\mu(x) = F(x)$ when μ is too small (since $F(x) + \underline{t}_\mu \le z_\mu(x) \le F(x) + \overline{t}_\mu$ and $\overline{t}_\mu \to 0$ as $\mu \to 0$ by Lemma 1), which leads to a breakdown (division by $z_\mu(x) - F(x) = 0$). Thus a lower bound $\underline{\mu}$ for the barrier parameter has to be used (we recommend value $\underline{\mu} = 10^{-10}$ in double precision arithmetic).

Algorithm 1 is also sensitive to the way in which the barrier parameter decreases. Considering logarithmic barrier function B1 and denoting by $s_\mu(x) = z_\mu(x)e - f(x)$ vector of slack variables, we can see from (7) that $u_\mu(x)_i s_\mu(x)_i = \mu \ \forall 1 \le i \le m$. In this case, interior-point methods assume that μ decreases linearly (see [18]). We have tested various possibilities for the barrier parameter update including simple geometric sequences, which were proved to be unsuitable. Better results were obtained by the following procedure, which consists of two phases:

Phase 1: If $\|g_k\| \ge \underline{g}$, we set $\mu_{k+1} = \mu_k$, i.e., barrier parameter is not changed.

Phase 2: If $\|g_k\| < \underline{g}$, we set

$$\mu_{k+1} = \max\left(\tilde{\mu}_{k+1},\ \underline{\mu},\ 10\,\varepsilon_M |F(x_{k+1})|\right), \qquad (33)$$

where $F(x_{k+1}) = \max_{1 \le i \le m} f_i(x_{k+1})$, ε_M is the machine precision and

$$\tilde{\mu}_{k+1} = \min\left[\max(\lambda\mu_k,\ \mu_k/(\gamma\mu_k + 1)),\ \max(\|g_{\mu_k}(x_k)\|^2,\ 10^{-2k})\right], \qquad (34)$$

where $g_{\mu_k}(x_k) = A(x_k)u_{\mu_k}(x_k)$. Values $\underline{\mu} = 10^{-10}$, $\lambda = 0.85$ and $\gamma = 100$ were chosen as defaults.

Usually $\underline{g} = \gamma_M$ (the maximum machine number), but sometimes a lower value (e.g., $\underline{g} = 1$) is more suitable. The reason for using formula (33) was mentioned above. Formula (34) requires several notes. The first argument of the minimum controls the rate of the barrier parameter decrease, which is linear (geometric sequence) for small k (term $\lambda\mu_k$) and sublinear (harmonic sequence) for large k (term $\mu_k/(\gamma\mu_k + 1)$). Thus the second argument, which assures that μ is small in the neighborhood of the solution, plays an essential role for large k. Term 10^{-2k} assures that $\mu = \underline{\mu}$ does not hold for small k. This situation can arise, when $\|g_{\mu_k}(x_k)\|$ is small, even if x_k is far from the solution.

4 Global convergence

In the subsequent considerations, we will assume that $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$ and all computations are exact ($\varepsilon_M = 0$ in (33)). We will investigate infinite sequence $\{x_k\}_1^\infty$ generated by Algorithm 1.

Lemma 6 *Let Assumption 1, Assumption 4 hold and Condition 1 be satisfied. Then values* $\{\mu_k\}_1^\infty$*, generated by Algorithm 1, form a non-increasing sequence such that* $\mu_k \to 0$.

Proof. We prove that the number of consecutive steps in Phase 1 is finite. Thus the number of steps of Phase 2 is infinite. Since μ_k decreases either by geometric or by harmonic sequence in Phase 2, one has $\mu_k \to 0$.

(a) First we prove that $B_\mu(x)$ is bounded from below if μ is fixed. This assertion holds trivially if Condition 2 is satisfied. If this is not the case, then Lemma 1 imply that constants $0 < \underline{t}_\mu \le \overline{t}_\mu$ exist such that $F(x) + \underline{t}_\mu \le z_\mu(x) \le F(x) + \overline{t}_\mu$. Thus $z_\mu(x) - f_i(x) \le F(x) + \overline{t}_\mu - \underline{F}$ for all $1 \le i \le m$ by Assumption 1 and we can write

$$B_\mu(x) \ge F(x) + \underline{t}_\mu + \mu \sum_{i=1}^{m} \varphi(F(x) + \overline{t}_\mu - \underline{F}) \ge \psi(F(x)),$$

where $\psi(F) = F + m\mu\varphi(F + \overline{t}_\mu - \underline{F})$. Since

$$\psi'(F) = 1 + m\mu\varphi'(F + \overline{t}_\mu - \underline{F}) \ge 1 + \mu \sum_{i=1}^{m} \varphi'(z_\mu(x) - f_i(x)) = 0$$

by (6) and monotonicity of $\varphi'(t)$, value $\psi(F(x))$ is not less than $\psi(\underline{F})$. Thus

$$B_\mu(x) \ge \underline{F} + m\mu\varphi(\overline{t}_\mu). \tag{35}$$

(b) In Phase 1, value μ is fixed. Since function $B_\mu(x)$ is continuous, bounded from below by (a) and since (32) (with $\mu_k = \mu$) holds, it can be proved (see [3]) that $\|g_\mu(x_k)\| \to 0$ if Phase 1 contains an infinite number of consecutive steps (proof of this assertion is identical with part (a) of the proof of Theorem 1 assuming that $L = 0$). Thus a step (with index l) belonging to Phase 1 exists such that $\|g_\mu(x_l)\| < \underline{g}$, which is a contradiction with the definition of this phase. □

Now we will prove that

$$B_{\mu_{k+1}}(x_{k+1}) \le B_{\mu_k}(x_{k+1}) - L(\mu_{k+1} - \mu_k) \quad L \in R, \tag{36}$$

for some $L \in R$, where $L \le 0$ if Condition 2 is satisfied. For this purpose, we consider that $z_\mu(x)$ and $B_\mu(x)$ are functions of μ and we write $z(x, \mu) = z_\mu(x)$ and $B(x, \mu) = B_\mu(x)$.

Lemma 7. *Let $z(x, \mu)$ be the solution of equation (11) (for fixed x and variable μ), i.e., $1 - e^T u(x, z(x, \mu)) = 0$ and let $B(x, \mu) = B_\mu(x, z(x, \mu))$. Then*

$$\frac{\partial z(x, \mu)}{\partial \mu} > 0, \qquad \frac{\partial B(x, \mu)}{\partial \mu} = \sum_{i=1}^{m} \varphi(z(x, \mu) - f_i(x)).$$

Proof. Differentiating equation (11), which has the form

$$1 + \mu \sum_{i=1}^{m} \varphi'(z(x, \mu) - f_i(x)) = 0,$$

we obtain

$$\sum_{i=1}^{m} \varphi'(z(x,\mu) - f_i(x)) + \mu \sum_{i=1}^{m} \varphi''(z(x,\mu) - f_i(x)) \frac{\partial z(x,\mu)}{\partial \mu} = 0,$$

which gives

$$\frac{\partial z(x,\mu)}{\partial \mu} = \frac{1}{\mu^2 \sum_{i=1}^{m} \varphi''(z(x,\mu) - f_i(x))} > 0.$$

Differentiating function

$$B(x,\mu) = z(x,\mu) + \mu \sum_{i=1}^{m} \varphi(z(x,\mu) - f_i(x)),$$

one has

$$\begin{aligned}
\frac{\partial B(x,\mu)}{\partial \mu} &= \frac{\partial z(x,\mu)}{\partial \mu} + \sum_{i=1}^{m} \varphi(z(x,\mu) - f_i(x)) + \mu \sum_{i=1}^{m} \varphi'(z(x,\mu) - f_i(x)) \frac{\partial z(x,\mu)}{\partial \mu} \\
&= \frac{\partial z(x,\mu)}{\partial \mu} \left(1 + \mu \sum_{i=1}^{m} \varphi'(z(x,\mu) - f_i(x))\right) + \sum_{i=1}^{m} \varphi(z(x,\mu) - f_i(x)) \\
&= \sum_{i=1}^{m} \varphi(z(x,\mu) - f_i(x)).
\end{aligned}$$

□

Lemma 8. *Let Assumption 1, Assumption 4 hold and Condition 1 be satisfied. Then (36) holds with some $L \in R$.*

Proof. Using Lemma 7, the mean value theorem, Lemma 1 and (21), we can write

$$\begin{aligned}
B(x_{k+1}, \mu_{k+1}) - B(x_{k+1}, \mu_k) &= \sum_{i=1}^{m} \varphi(z(x_{k+1}, \tilde{\mu}_k) - f_i(x_{k+1}))(\mu_{k+1} - \mu_k) \\
&\le \sum_{i=1}^{m} \varphi(F(x_{k+1}) + \bar{t}_{\tilde{\mu}_k} - f_i(x_{k+1}))(\mu_{k+1} - \mu_k) \\
&\le m\varphi(\overline{F} - \underline{F} + \bar{t}_{\mu_1})(\mu_{k+1} - \mu_k) \\
&\triangleq -L(\mu_{k+1} - \mu_k),
\end{aligned}$$

where $0 < \mu_{k+1} \le \tilde{\mu}_k \le \mu_k \le \mu_1$. Constant L is non-positive (thus the inequality holds also with $L = 0$) if Condition 2 is satisfied. □

Using (36) with $L = 0$, we can prove the following lemma.

Lemma 9. *Let Assumption 1, Assumption 4 hold and Condition 1, Condition 2 be satisfied. Then if $\overline{F} \ge B_{\mu_1}(x_1)$, one has $x_k \in \mathcal{L}(\overline{F})$ for all $k \in N$.*

Proof. Condition 2 implies that $F(x_k) < z_{\mu_k}(x_k) \le B_{\mu_k}(x_k)$. Moreover Condition 2, inequality (36) with $L = 0$ and inequality (32) imply that that the sequence $\{B_{\mu_k}(x_k)\}_1^\infty$ is non-increasing (since sequence $\{\mu_k\}_1^\infty$ is non-increasing). This fact implies that

$$F(x_k) < z_{\mu_k}(x_k) \le B_{\mu_k}(x_k) \le B_{\mu_1}(x_1)$$

and if $\overline{F} \ge B_{\mu_1}(x_1)$, one has $x_k \in \mathcal{L}(\overline{F})$ for all $k \in N$. □

If Condition 2 is not satisfied, the existence of upper bound $\overline{F}$ is not assured. In this case, we have to assume that sequence $\{x_k\}_1^\infty$ is bounded and replace Assumption 2 by Assumption 3. Now we are in a position to prove the main theorem.

Theorem 1. *Let Assumption 1, Assumption 4 hold and Condition 1 be satisfied. Consider sequence $\{x_k\}_1^\infty$, generated by Algorithm 1 such that $x_k \in \mathcal{L}(\overline{F})$. Then*

$$\lim_{k\to\infty} \sum_{i=1}^m u_{\mu_k}(x_k)_i g_i(x_k) = 0, \quad \sum_{i=1}^m u_{\mu_k}(x_k)_i = 1$$

and

$$u_{\mu_k}(x_k)_i \ge 0, \quad z_{\mu_k}(x_k) - f_i(x_k) \ge 0, \quad \lim_{k\to\infty} u_{\mu_k}(x_k)_i (z_{\mu_k}(x_k) - f_i(x_k)) = 0$$

for $1 \le i \le m$.

Proof. Equality $1 - e^T u_{\mu_k}(x_k) = 0$ holds since computations are exact. Inequalities $u_{\mu_k}(x_k)_i \ge 0$, $z_{\mu_k}(x_k) - f_i(x_k) \ge 0$ follow from (7) (since $\varphi'(t)$ is negative for $t > 0$) and from Lemma 1.

(a) Since (32) and (36) hold, we can write

$$\begin{aligned} B_{\mu_{k+1}}(x_{k+1}) - B_{\mu_k}(x_k) &= (B_{\mu_{k+1}}(x_{k+1}) - B_{\mu_k}(x_{k+1})) + (B_{\mu_k}(x_{k+1}) - B_{\mu_k}(x_k)) \\ &\le -L(\mu_{k+1} - \mu_k) - c\|g_{\mu_k}(x_k)\|^2, \end{aligned}$$

which implies

$$\begin{aligned} \underline{B} &\le \lim_{k\to\infty} B_{\mu_k}(x_k) \le B_{\mu_1}(x_1) - L\sum_{k=1}^\infty (\mu_{k+1} - \mu_k) - c\sum_{k=1}^\infty \|g_{\mu_k}(x_k)\|^2 \\ &= B_{\mu_1}(x_1) + L\mu_1 - c\sum_{k=1}^\infty \|g_{\mu_k}(x_k)\|^2, \end{aligned}$$

where $\underline{B} = \underline{F} + \min_{\underline{\mu} \le \mu \le \mu_1} m\mu\varphi(\overline{t}_\mu)$ (see (35)). Thus one has

$$\sum_{k=1}^\infty \|g_{\mu_k}(x_k)\|^2 \le \frac{1}{c}(B_{\mu_1}(x_1) - \underline{B} + L\mu_1) < \infty,$$

which implies $g_{\mu_k}(x_k) = \sum_{i=1}^m u_{\mu_k}(x_k)_i g_i(x_k) \to 0$.

(b) Let index $1 \le i \le m$ be chosen arbitrarily. Since $u_{\mu_k}(x_k)_i \ge 0$, $z_{\mu_k}(x_k) - f_i(x_k) \ge 0$, one has

$$\limsup_{k\to\infty} u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \ge \liminf_{k\to\infty} u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \ge 0.$$

It suffices to prove that these inequalities are satisfied as equalities. Assume on the contrary that there is an infinite subset $N_1 \subset N$ such that $u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \ge \varepsilon \ \forall k \in N_1$, where $\varepsilon > 0$. Since $\underline{F} \le f_i(x_k) \le F(x_k) \le \overline{F} \ \forall k \in N_1$, there exists an infinite subset $N_2 \subset N_1$ such that sequence $F(x_k) - f_i(x_k)$, $k \in N_2$, converges.

(c) Assume first that $F(x_k) - f_i(x_k) \stackrel{N_2}{\to} \delta > 0$. Since

$$z_{\mu_k}(x_k) - f_i(x_k) \ge F(x_k) - f_i(x_k) \ge \delta/2$$

for sufficiently large $k \in N_2$, one has

$$u_{\mu_k}(x_k)_i = -\mu_k \varphi'(z_{\mu_k}(x_k) - f_i(x_k)) \le -\mu_k \varphi'(\delta/2) \stackrel{N_2}{\to} 0,$$

since $\mu_k \to 0$ by Lemma 6. Since sequence $z_{\mu_k}(x_k) - f_i(x_k)$, $k \in N_2$, is bounded (see proof of Lemma 8), we obtain $u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \stackrel{N_2}{\to} 0$, which is a contradiction.

(d) Assume now that $F(x_k) - f_i(x_k) \stackrel{N_2}{\to} 0$. Since $z_{\mu_k}(x_k) - F(x_k) \to 0$ as $\mu_k \to 0$ by Lemma 1, we can write

$$z_{\mu_k}(x_k) - f_i(x_k) = (z_{\mu_k}(x_k) - F(x_k)) + (F(x_k) - f_i(x_k)) \stackrel{N_2}{\to} 0.$$

At the same time, (6) and (7) imply that sequence $u_{\mu_k}(x_k)_i$, $k \in N_2$, is bounded. Thus $u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \stackrel{N_2}{\to} 0$, which is a contradiction. □

Corollary 3. *Let assumptions of Theorem 1 hold. Then every cluster point $x \in R^n$ of sequence $\{x_k\}_1^\infty$ satisfies KKT conditions (3), where $u \in R^m$ is a cluster point of sequence $\{u_{\mu_k}(x_k)\}_1^\infty$.*

Assuming that values $\underline{\delta}$, $\underline{\varepsilon}$, $\underline{\mu}$ are nonzero and logarithmic barrier term B1 is used, we can prove the following theorem informing us about the precision obtained, when Algorithm 1 terminates.

Theorem 2. *Consider sequence $\{x_k\}_1^\infty$ generated by Algorithm 1 with logarithmic barrier term B1. Let Assumption 1 and Assumption 4 hold. Then, choosing $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, $\underline{\mu} > 0$ arbitrarily, there is an index $\overline{k} \ge 1$ such that*

$$\|g_{\mu_k}(x_k)\| \le \underline{\varepsilon}, \quad |1 - e^T u_{\mu_k}(x_k)| \le \underline{\delta},$$

and

$$u_{\mu_k}(x_k)_i \ge 0, \quad z_{\mu_k}(x_k) - f_i(x_k) \ge 0, \quad u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) \le \underline{\mu}, \quad 1 \le i \le m,$$

for all $k \ge \overline{k}$.

Proof. Equality $|1 - e^T u_{\mu_k}(x_k)| \le \underline{\delta}$ follows immediately from the fact that equation $1-e^T u_{\mu_k}(x_k)$ is solved with precision $\underline{\delta}$. Inequalities $u_{\mu_k}(x_k)_i \ge 0$, $z_{\mu_k}(x_k)-f_i(x_k) \ge 0$ follow from (7) and Lemma 1 as in the proof of Theorem 1. Since $g_{\mu_k}(x_k) \to 0$ by Theorem 1, there is an index $\overline{k} \ge 1$ such that $\|g_{\mu_k}(x_k)\| \le \underline{\varepsilon}$ for all $k \ge \overline{k}$. Using (7) and the fact that $\varphi'(t) = -1/t$ for B1, we obtain

$$u_{\mu_k}(x_k)_i(z_{\mu_k}(x_k) - f_i(x_k)) = \frac{\mu_k}{z_{\mu_k}(x_k) - f_i(x_k)}(z_{\mu_k}(x_k) - f_i(x_k)) = \mu_k \le \underline{\mu}$$

□

Theorem 2 also holds for B2 and B3, since $\varphi'(t) \ge \varphi'_{B1}(t)$ for these barrier terms (see proof of Corollary 1). For B4 the upper bound is proportional to $\sqrt{\underline{\mu}}$, which again indicates that we should use μ^2 instead of μ in (16) in this case.

5 Smoothing method for large sparse minimax optimization

In this section, we briefly describe a smoothing method for large sparse minimax optimization, which is algorithmically very similar to the proposed interior-point method and which will be used for a comparison. This smoothing method investigated in [17] and [19] (and in other papers quoted therein), uses smoothing function

$$S_\mu(x) = \mu \log \sum_{i=1}^{m} \exp\left(\frac{f_i(x)}{\mu}\right) = F(x) + \mu \log \sum_{i=1}^{m} \exp\left(\frac{f_i(x) - F(x)}{\mu}\right), \tag{37}$$

where $F(x)$ is given by (1) and $\mu > 0$ (we assume that $\mu \to 0$ monotonically). The following result is proved in [17].

Lemma 10. *Consider smoothing function (37). Then*

$$\nabla S_\mu(x) = A(x)\tilde{U}_\mu(x)e, \tag{38}$$

and

$$\nabla^2 S_\mu(x) = \tilde{G}_\mu(x) + \frac{1}{\mu}A(x)\tilde{U}_\mu(x)A^T(x) - \frac{1}{\mu}A(x)\tilde{U}_\mu(x)ee^T\tilde{U}_\mu(x)A^T(x), \tag{39}$$

where $\tilde{G}_\mu(x) = \sum_{i=1}^m \tilde{u}_\mu(x)_i G_i(x)$, $\tilde{U}_\mu(x) = \operatorname{diag}(\tilde{u}_\mu(x)_1, \dots, \tilde{u}_\mu(x)_m)$ *and*

$$\tilde{u}_\mu(x)_i = \frac{\exp(f_i(x)/\mu)}{\sum_{j=1}^m \exp(f_j(x)/\mu)} = \frac{\exp((f_i(x) - F(x))/\mu)}{\sum_{j=1}^m \exp((f_j(x) - F(x))/\mu)} \tag{40}$$

for $1 \le i \le m$, *which implies* $e^T\tilde{u}_\mu(x) = 1$.

Note that (39) together with the Schwarz inequality implies

$$v^T\nabla^2 S_\mu(x)v = v^T\tilde{G}_\mu(x)v + \frac{1}{\mu}\left(v^TA(x)\tilde{U}_\mu(x)A^T(x)v - \frac{(v^TA(x)\tilde{U}_\mu(x)e)^2}{e^T\tilde{U}_\mu(x)e}\right) \ge v^T\tilde{G}_\mu(x)v.$$

Thus $\nabla^2 S_\mu(x)$ is positive definite if $\tilde{G}_\mu(x)$ is positive definite.

Using Lemma 10, we can write one step of the Newton method in the form $x^+ = x + \alpha d$, where $\nabla^2 S_\mu(x) d = -\nabla S_\mu(x)$ or

$$\left(\tilde{H} - \frac{1}{\mu} \tilde{g} \tilde{g}^T \right) d = -\tilde{g}, \tag{41}$$

where

$$\tilde{H} = \tilde{G}_\mu(x) + \frac{1}{\mu} A(x) \tilde{U}_\mu(x) A^T(x) \tag{42}$$

and $\tilde{g} = A(x)\tilde{U}_\mu(x)e$. It is evident that matrix $\tilde{H}$ has the same sparsity pattern as H in (23). Since

$$\left(\tilde{H} - \frac{1}{\mu} \tilde{g} g^T \right)^{-1} = \tilde{H}^{-1} + \frac{\tilde{H}^{-1} \tilde{g} \tilde{g}^T \tilde{H}^{-1}}{\mu - \tilde{g}^T \tilde{H}^{-1} \tilde{g}},$$

the solution of (41) can be written in the form

$$d = \frac{\mu}{\tilde{g}^T \tilde{H}^{-1} \tilde{g} - \mu} \tilde{H}^{-1} \tilde{g}. \tag{43}$$

If $\tilde{H}$ is not positive definite, it is advantageous to change it before computation of the direction vector. Thus we use the sparse Gill-Murray decomposition $\tilde{H} + \tilde{E} = \tilde{L}\tilde{D}\tilde{L}^T$, solve equation

$$\tilde{L}\tilde{D}\tilde{L}^T v = \tilde{g} \tag{44}$$

and set

$$d = \frac{\mu}{\tilde{g}^T v - \mu} v. \tag{45}$$

More details concerning the smoothing method can be found in [17] and [19], where the proof of its global convergence is introduced.

The above considerations and formulas form a basis for the algorithm, which is very similar to Algorithm 1. This algorithm differs from Algorithm 1 in Step 2, where nonlinear equation is not solved (since vector $\tilde{u}_\mu(x)$ is computed directly from (40)), in Step 4, where (26)-(27) are replaced by (44)-(45), and in Step 6, where $B_\mu(x)$ is replaced by $S_\mu(x)$. Note that μ in (37) has a different meaning from μ in (16). Thus procedures for updating these parameters need not be identical. Nevertheless, the procedure described in Section 3 was successful in connection with the smoothing method (we have also tested procedures proposed in [17] and [19], but they were less efficient). Note finally, that the smoothing method described in this section has also insufficiencies concerning finite precision computations. If μ is small, than many evaluations of exponentials lead to underflows. This effect decreases the precision of computed gradients, which brings a problem with the termination of iterative process. For this reason, a lower bound $\underline{\mu}$ has to be used, which is usually greater than corresponding bound in the interior point method (we recommend value $\underline{\mu} = 10^{-6}$ for the smoothing method).

6 Computational experiments

The primal interior-point method was tested by using the collection of relatively difficult problems with optional dimension chosen from [12], which can be downloaded (together with the above report) from `www.cs.cas.cz/~luksan/test.html` as Test 14 and Test 15. In [12], functions $f_i(x)$, $1 \le i \le m$, are given, which serve for defining objective functions

$$F(x) = \max_{1 \le i \le m} f_i(x) \tag{46}$$

and

$$F(x) = \max_{1 \le i \le m} |f_i(x)| = \max_{1 \le i \le m} \left[\max(f_i(x), -f_i(x)) \right]. \tag{47}$$

Function (46) could not be used in connection with Test 15, since Assumption 1 is not satisfied (sometimes $F(x) \to -\infty$) in this case. We have used parameters $\underline{\varepsilon} = 10^{-6}$, $\underline{\delta} = 10^{-6}$, $\underline{\mu} = 10^{-10}$ for the interior-point method and $\underline{\mu} = 10^{-6}$ for the smoothing method, $\lambda = 0.85$ for the interior-point method and $\lambda = 0.95$ for the smoothing method, $\gamma = 100$, $\underline{c} = 10^{-10}$, $\overline{c} = 10^{10}$, $\varepsilon_0 = 10^{-8}$, $\varepsilon_1 = 10^{-4}$, $\beta = 0.5$, $\overline{\Delta} = 1000$ in Algorithm 1 as defaults (values $\underline{\mu}$ and $\overline{\Delta}$ were sometimes tuned).

The first set of the tests concerns comparison of the primal interior point method with various barrier terms and the smoothing method proposed in [17] and [19]. Medium-size test problems with 200 variables were used. The results of computational experiments are reported in three tables, where only summary results (over all 22 test problems) are given. Here M is the merit function used: B1-B4 (see Table 1) or S (see (37)); NIT is the total number of iterations, NFV is the total number of function evaluations, NFG is the total number of gradient evaluations, NR is the total number of restarts, NF is the number of problems, for which the global minimizer was not found (either a worse local minimum was obtained or the method failed even if its parameters were tuned), NT is the number of problems for which parameters were tuned and TIME is the total computational time in seconds.

M	NIT	NFV	NFG	NR	NF	NT	TIME
B1	1347	2676	9110	88	-	5	1.56
B2	2001	6098	13027	624	-	10	2.33
B3	1930	6578	11875	602	-	7	1.95
B4	5495	9947	32221	623	1	8	4.08
S	8520	15337	49551	811	2	7	12.64

Table 4: Test 14: Function (46) with 200 variables

M	NIT	NFV	NFG	NR	NF	NT	TIME
B1	1833	5137	15779	357	-	7	3.39
B2	4282	17088	31225	2929	-	5	6.68
B3	3029	13456	21892	1342	-	7	4.47
B4	5679	31949	51350	3256	-	9	11.02
S	10634	26101	60246	1416	2	9	26.09

Table 5: Test 14: Function (47) with 200 variables

M	NIT	NFV	NFG	NR	NF	NT	TIME
B1	8247	10260	43555	2412	-	14	10.13
B2	7757	12701	38339	5461	2	10	9.52
B3	9495	11329	45911	6464	1	16	13.66
B4	13419	20103	72326	10751	1	14	12.61
S	13160	24334	79099	6682	5	9	47.84

Table 6: Test 15: Function (47) with 200 variables

The results introduced in these tables indicate that the logarithmic barrier function B1 is the best choice for implementation of the primal interior-point method and that primal interior-point methods are more efficient than the smoothing method. The primal interior-point method with logarithmic barrier term B1 is less sensitive to the choice of parameters and requires lower number of restarts, lower number of iterations and shorter computational time in comparison with the other methods. This method also finds the global minimum (if the minimax problems has several local solutions) more frequently than the other methods (see column NF in the above tables).

The second set of the tests concerns a comparison of the primal interior-point method with logarithmic barrier function B1 with the smoothing method proposed in [17] and [19]. Large-scale test problems with 1000 variables were used. The results of computational experiments are given in three tables, where P is the problem number, NIT is the number of iterations, NFV is the number of function evaluations, NFG is the number of gradient evaluations and F is the function value reached. The last row of every table contains summary results including the total computational time.

	Interior-point method				Smoothing method			
P	NIT	NFV	NFG	F	NIT	NFV	NFG	F
1	76	134	308	0.9898964793	737	1556	2952	0.9898966944
2	80	282	567	0.1276877391E-24	119	187	840	0.5138610463E-08
3	22	23	161	0.4568001350E-07	75	104	532	0.1053071878E-10
4	40	53	287	0.5427614641	192	562	1351	0.5427618250
5	9	10	60	0.2352121219E-09	55	67	336	0.1480219031E-07
6	107	116	1512	0.4246457820E-07	77	183	1092	0.1013842348E-06
7	33	62	306	0.2601625837	164	685	1485	0.2601629017
8	8	18	162	1556.501961	5	93	90	1556.501961
9	427	1125	7704	0.7152533801	1474	1489	26550	0.7159831998
10	322	1804	5814	-0.4959471915E-01	251	621	4536	-0.3400601681E-01
11	13	19	84	0.5388293945E-01	306	311	1842	0.5388293945E-01
12	208	231	836	0.9964922393	2388	6415	9556	0.9964925766
13	7	8	32	0.9060203997E-36	6	7	28	0.6007356980E-13
14	5	6	36	0.1014203886E-16	4	5	30	0.3701070007E-16
15	14	20	60	0.3996002833E-02	112	234	452	0.4041319466E-02
16	163	558	984	-0.6408315668E-03	2017	8788	12108	-0.6376649711E-03
17	25	29	156	0.1971314267E-08	10	11	66	0.5855792366E-06
18	32	88	198	0.1360346290E-07	381	1417	2292	0.5909422344E-06
19	6	16	42	42.50757784	3	45	18	42.50757784
20	9	25	60	-0.9990009990E-03	77	88	468	-0.9990009476E-03
21	12	18	78	0.5993990042E-02	204	851	1230	0.5993990042E-02
22	110	784	666	0.1156346235E-02	156	564	942	0.1156346235E-02
Σ	1728	5429	20113	TIME=23.97	8813	24283	68796	TIME=92.32

Table 7: Test 14: Function (46) with 1000 variables

	Interior-point method				Smoothing method			
P	NIT	NFV	NFG	F	NIT	NFV	NFG	F
1	54	67	220	0.9898964855	576	960	2308	0.9898966944
2	111	273	784	0.1917953685E-07	110	174	777	0.8727104889E-07
3	43	44	308	0.4022970814E-07	67	99	476	0.4165892164E-07
4	46	59	329	0.5427614641	184	532	1295	0.5427618250
5	119	125	720	0.7962952760E-07	312	1142	1878	0.3773799412E-06
6	149	1309	2100	0.1051683837E-06	199	642	2800	0.3130531293E-06
7	36	90	333	0.2601625837	180	722	1629	0.2601629017
8	7	8	144	1556.501961	5	93	90	1556.501961
9	705	1008	12708	0.7124668996	1569	1573	28260	0.7159831562
10	310	899	5580	0.4809503605E-02	324	863	5850	0.3045714174E-02
11	27	28	168	0.5388293945E-01	307	309	1848	0.5388293945E-01
12	243	278	976	0.9964922393	1598	4877	6396	0.9964925766
13	18	19	76	0.2497152305E-07	7	8	32	0.2185240307E-11
14	5	31	36	0.3936489854E-11	1	22	6	0.3936493419E-11
15	60	207	244	0.3996002833E-02	542	1178	2172	0.4000515611E-02
16	101	235	612	0.8665662046E-14	350	669	2106	0.2879207280E-10
17	84	519	510	0.1583674449E-06	78	228	474	0.2214113588E-06
18	234	1780	1410	0.4211008551E-07	221	757	1332	0.3548586285E-06
19	6	16	42	42.50757784	3	45	18	42.50757784
20	298	2245	1794	0.9124286790E-04	1094	2722	6570	0.1019502822E-03
21	10	12	66	0.5993990042E-02	25	52	156	0.5993990042E-02
22	77	449	468	0.1156346235E-02	415	1493	2496	0.1156346235E-02
Σ	2743	9701	29628	TIME=35.15	8167	19160	68969	TIME=113.83

Table 8: Test 14: Function (47) with 1000 variables

The results introduced in these tables confirm conclusions following from the previous tables. The primal interior-point method seems to be more efficient than the smoothing method in all indicators. Especially, the computational time is much shorter and also the number of global minima attained is greater in the case of the primal interior-point method. We believe that the efficiency of the primal interior-point method could be improved by using a better procedure for the barrier parameter update, another strategy for restarts or a suitable trust region realization.

	Interior-point method				Smoothing method			
P	NIT	NFV	NFG	F	NIT	NFV	NFG	F
1	3542	4253	14172	0.4440892099E-14	204	653	820	0.9000002516
2	107	121	540	0.1931788063E-12	109	216	550	1.741886510
3	19	20	120	0.1116768726E-07	39	75	240	0.1001835213E-06
4	46	95	188	0.4450717246	131	384	528	0.4450720864
5	8	12	54	0.2581101999E-11	17	21	108	0.1241754477E-09
6	13	15	196	0.3330669074E-15	30	37	434	0.1348121614E-09
7	39	80	160	8.733141278	103	348	416	8.733141485
8	106	163	321	0.9994462557E-06	101	224	306	0.1220896225E-05
9	254	744	1785	1.867650789	1427	4669	9996	1.870311605
10	95	260	576	0.6790301632	2363	2401	14184	0.6790304227
11	5006	6498	20028	0.2169375790E-11	279	896	1120	1.794696397
12	118	164	714	9.235413645	264	579	1590	9.235413761
13	129	473	780	18.79734633	323	1116	1944	21.96829628
14	49	127	350	0.4599900248	38	75	273	0.4599902005
15	54	91	385	0.2306075954	258	287	1813	0.2306077911
16	70	229	497	0.6666672280E-01	64	129	455	0.6666692037E-01
17	1049	2035	9450	0.1952639701E-02	46	72	423	0.6964914756E-15
18	2553	2908	15324	0.3108624469E-14	5469	9376	32820	0.1498356994E-11
19	35	58	396	0.5496714195E-10	100	292	1111	0.3213034283E-09
20	283	284	1704	0.9999665936	720	1929	4326	0.9999703617
21	32	82	198	0.9000000694	165	450	996	0.9000002597
22	82	290	492	2.354873828	590	634	3546	2.407431606
Σ	13689	19002	68430	TIME=65.84	12840	24863	77999	TIME=228.78

Table 9: Test 15: Function (47) with 1000 variables

References

[1] A.Ben-Tal, G.Roth: A truncated log barrier algorithm for large-scale convex programming and minimax problems: Implementation and computational results. Optimization Methods and Software 6 (1996) 283-312.

[2] T.F.Coleman, J.J.Moré: Estimation of sparse Hessian matrices and graph coloring problems. Mathematical Programming 28 (1984) 243-270.

[3] R.Fletcher: Practical Methods of Optimization (second edition). Wiley, New York, 1987.

[4] R.Fletcher, E.Sainz de la Maza: Nonlinear programming and nonsmooth optimization by successive linear programming. Mathematical Programming (43) (1989) 235-256.

[5] Y.Gao, X.Li: Nonsmooth equations approach to a constrained minimax problem. Applications of Mathematics 50 5-130.

[6] P.E.Gill, W.Murray, M.A.Saunders: SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Review 47 (2005) 99-131.

[7] S.P.Han: Variable metric methods for minimizing a class of nondifferentiable functions. Mathematical programming 20 (1981) 1-13.

[8] K.Jónasson, K.Madsen: Corrected sequential linear programming for sparse minimax optimization. BIT 34 (1994) 372-387.

[9] D.Le: Three new rapidly convergent algorithms for finding a zero of a function. SIAM J. on Scietific and Statistical Computaions 6 (1985) 193-208.

[10] D.Le: An efficient derivative-free method for solving nonlinear equations. ACM Transactions on Mathematical Software 11 (1985) 250-262.

[11] L.Lukšan: An Implementation of Recursive Quadratic Programming Variable Metric Methods for Linearly Constrained Nonlinear Minimax Approximation, Kybernetika 21 (1985) 22-40.

[12] L.Lukšan, J.Vlček: Sparse and partially separable test problems for unconstrained and equality constrained optimization, Report V-767, Prague, ICS AS CR, 1998.

[13] L.Lukšan, J.Vlček: Indefinitely Preconditioned Inexact Newton Method for Large Sparse Equality Constrained Nonlinear Programming Problems. Numerical Linear Algebra with Applications 5 (1998) 219-247.

[14] L.Lukšan, C.Matonoha, J.Vlček: Interior point method for nonlinear nonconvex optimization. Numerical Linear Algebra with Applications 11 (2004) 431-453.

[15] L.Lukšan, C.Matonoha, J.Vlček: Nonsmooth equation method for nonlinear nonconvex optimization. In: Conjugate Gradient Algorithms and Finite Element Methods (M.Křížek, P.Neittaanmäki, R.Glovinski, S.Korotov eds.). Springer Verlag, Berlin 2004.

[16] J.Vlček, L.Lukšan: Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization. Journal of Optimization Theory and Applications 111 (2001) 407-430.

[17] E.Polak, J.O.Royset, R.S.Womersley: Algorithm with adaptive smoothing for finite minimax problems. Journal of Optimization Theory and Applications 119 (2003) 459-484.

[18] J.Vanderbei, D.F.Shanno: An interior point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications, 13, 231-252, 1999.

[19] S.Xu: Smoothing methods for minimax problems. Computational Optimization and Applications 20 (2001) 267-279.