KYBERNETIKA — VOLUME 46 (2010), NUMBER 4, PAGES 697–721

PRIMAL INTERIOR POINT METHOD FOR MINIMIZATION OF GENERALIZED MINIMAX FUNCTIONS

Ladislav Lukšan, Ctirad Matonoha and Jan Vlček

In this paper, we propose a primal interior-point method for large sparse generalized minimax optimization. After a short introduction, where the problem is stated, we introduce the basic equations of the Newton method applied to the KKT conditions and propose a primal interior-point method. Next we describe the basic algorithm and give more details concerning its implementation covering numerical differentiation, variable metric updates, and a barrier parameter decrease. Using standard weak assumptions, we prove that this algorithm is globally convergent if a bounded barrier is used. Then, using stronger assumptions, we prove that it is globally convergent also for the logarithmic barrier. Finally, we present results of computational experiments confirming the efficiency of the primal interior point method for special cases of generalized minimax problems.

Keywords: unconstrained optimization, large-scale optimization, nonsmooth optimization, generalized minimax optimization, interior-point methods, modified Newton methods, variable metric methods, global convergence, computational experiments

Classification: 49K35, 90C06, 90C47, 90C51

1. INTRODUCTION

Functions which we need to minimize are often nonsmooth since they contain absolute values or pointwise maxima of smooth functions. Typical examples are the norms $\|f(x)\|_1$ and $\|f(x)\|_\infty$ of a smooth mapping $f : R^n \to R^m$ (see [4]). Generalizations of these functions are composite non-smooth functions of the form

$$F(x) = \max_{1\le i\le l} p_i^T f(x), \tag{1}$$

where $p_i \in R^m$, $1 \le i \le l$, and $f : R^n \to R^m$ is a smooth mapping (see [5]). In this way we can express nonsmooth functions $\max_{1\le i\le m} f_i(x)$, $\|f(x)\|_\infty$, $\|f_+(x)\|_\infty$, $\|f(x)\|_1$, $\|f_+(x)\|_1$, where $f_+(x) = [\max(f_1(x), 0), \dots, \max(f_m(x), 0)]^T$, by a suitable choice of the matrix $P = [p_1, \dots, p_l]$.

In this contribution we focus our attention on a different class of structured non-smooth functions, the so-called generalized minimax functions (see [18] and references therein) defined by the following way.

Definition 1.1. We say that F is a generalized minimax function if

$$F(x) = h(F_1(x), \dots, F_m(x)), \quad F_i(x) = \max_{1 \le j \le n_i} f_{ij}(x), \quad 1 \le i \le m, \tag{2}$$

where $h : R^m \to R$ and $f_{ij} : R^n \to R$, $1 \le i \le m$, $1 \le j \le n_i$, are smooth functions satisfying the following assumptions.

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

Assumption 2. Function h is twice continuously differentiable and convex satisfying

$$\partial h(z)/\partial z_i \ge \underline{h}_i > 0, \quad 1 \le i \le m, \tag{3}$$

for every $z \in Z = \{z \in R^m : z_i \ge \underline{F}_i,\ 1 \le i \le m\}$ (vector $z \in R^m$ will be called the minimax vector).

Assumption 3. Functions f_{ij}, $1 \le i \le m$, $1 \le j \le n_i$, are twice continuously differentiable on the convex hull of the level set

$$\mathcal{L}(\overline{F}) = \{x \in R^n : F_i(x) \le \overline{F},\ 1 \le i \le m\}$$

for a sufficiently large upper bound $\overline{F}$ and they have bounded first and second-order derivatives on $\operatorname{conv}\mathcal{L}(\overline{F})$: there are $\overline{g}$ and $\overline{G}$ such that $\|\nabla f_{ij}(x)\| \le \overline{g}$ and $\|\nabla^2 f_{ij}(x)\| \le \overline{G}$ for all $1 \le i \le m$, $1 \le j \le n_i$ and $x \in \operatorname{conv}\mathcal{L}(\overline{F})$.

Sometimes, we use the following stronger assumption instead of Assumption 1.

Assumption 4. Functions f_{ij}, $1 \le i \le m$, $1 \le j \le n_i$, are bounded from below on R^n: there is $\underline{F} \in R$ such that $f_{ij}(x) \ge \underline{F}$, $1 \le i \le m$, $1 \le j \le n_i$, for all $x \in R^n$. Note that Assumption 4 implies Assumption 1.

Conditions imposed on the function $h(z)$ are relatively strong, but many functions satisfy them, e. g., the sum of maxima

$$h(z) = \sum_{i=1}^{m} z_i.$$

It is clear that we can express all the nonsmooth functions mentioned above in this way. Since $|f_i(x)| = \max(f_i(x), -f_i(x))$, function (2) covers the case when

$$F(x) = h(|f_1(x)|, \dots, |f_m(x)|).$$

The expression of functions $\|f(x)\|_1$, $\|f_+(x)\|_1$ by (2) is much easier in comparison with (1), since the matrix P contains 2^m columns in these cases.

Unconstrained minimization of function (2) is equivalent to the nonlinear programming problem: Minimize the function

$$h(z_1, \dots, z_m) \tag{4}$$

with constraints

$$f_{ij}(x) \le z_i, \quad 1 \le i \le m, \quad 1 \le j \le n_i \tag{5}$$

(conditions $\partial h(z)/\partial z_i \ge \underline{h}_i > 0$, $1 \le i \le m$, for $z \in Z$ are sufficient for satisfying equalities $z_i = F_i(x)$, $1 \le i \le m$, at the minimum point). The necessary first-order (KKT) conditions for a solution of (4)−(5) have the form

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} u_{ij} \nabla f_{ij}(x) = 0, \quad \sum_{j=1}^{n_i} u_{ij} = \frac{\partial h(z)}{\partial z_i}, \quad 1 \le i \le m, \tag{6}$$

$$u_{ij} \ge 0, \quad z_i - f_{ij}(x) \ge 0, \quad u_{ij}(z_i - f_{ij}(x)) = 0, \quad 1 \le i \le m, \quad 1 \le j \le n_i, \tag{7}$$

where u_{ij}, $1 \le i \le m$, $1 \le j \le n_i$, are Lagrange multipliers.

Nonlinear programming problem (4)−(5) can be solved by using the primal interior point method. For this reason we apply the Newton minimization method to the barrier function

$$B_\mu(x, z) = h(z) + \mu \sum_{i=1}^{m} \sum_{j=1}^{n_i} \varphi(z_i - f_{ij}(x)), \quad 0 < \mu \le \overline{\mu}, \tag{8}$$

assuming $\mu \to 0$, where $\varphi : (0, \infty) \to R$ is a barrier which satisfies the following condition.

Condition 1. $\varphi(t)$, $t \in (0, \infty)$, is a twice continuously differentiable function such that $\varphi(t)$ is decreasing, strictly convex, with $\lim_{t \to 0} \varphi(t) = \infty$, $\varphi'(t)$ is increasing, strictly concave, with $\lim_{t \to \infty} \varphi'(t) = 0$, and $-t\varphi'(t)$ is bounded from above.

The following additional condition is useful for studying the global convergence.

Condition 2. $\varphi(t)$, $t \in (0, \infty)$, is bounded from below: there is $\underline{\varphi} \le 0$ such that $\varphi(t) \ge \underline{\varphi}$ for all $t \in (0, \infty)$ (the non-positive value $\underline{\varphi} \le 0$ was chosen to simplify proofs in Section 5 and Section 6).

The most known and frequently used logarithmic barrier $\varphi(t) = \log t^{-1} = -\log t$ satisfies Condition 1, but does not satisfy Condition 2, since $\log t \to \infty$ as $t \to \infty$. Therefore, additional barriers have been studied (see [13] and references therein). As examples, we can introduce the function

$$\varphi(t) = \log(t^{-1} + 1), \quad t \in (0, \infty), \tag{9}$$

which is positive ($\underline{\varphi} = 0$), or the function

$$\varphi(t) = -\log t, \qquad 0 < t \le 1, \tag{10}$$

$$\varphi(t) = -(t^{-1} - 4\,t^{-1/2} + 3), \qquad t > 1, \tag{11}$$

which is bounded from below ($\underline{\varphi} = -3$). Both functions satisfy Condition 1 and Condition 2. Note that (8) implies

$$B_\mu(x, z) \ge h(z) + \overline{m}\,\overline{\mu}\,\underline{\varphi}, \quad \overline{m} = \sum_{i=1}^{m} n_i, \tag{12}$$

if Condition 2 holds.

A primal interior point method is based on the fact that it is easy to find a vector $z \in R^m$ satisfying constraints (5). Hence, it is not necessary to introduce slack variables, add equality constraints, use a penalty function and iterate the Lagrangian multipliers. In the subsequent sections, we describe two approaches which differ in the determination of the minimax vector $z \in R^m$ and the Algorithm which implements the second approach. We use the notation

$$A_{ij}(x) = \nabla f_{ij}(x), \quad G_{ij}(x) = \nabla^2 f_{ij}(x),$$

for $1 \le i \le m$, $1 \le j \le n_i$, and focus our attention on the problems whose structure allows us to use a sparse matrix technique.

The paper is organized as follows. In Section 2, we derive basic equations of the Newton method applied to the nonlinear KKT system of the interior point subproblem. Section 3 contains a description of the primal interior-point method (i. e. interior point method that uses explicitly computed approximations of Lagrange multipliers instead of their updates). In Section 4, we introduce the basic algorithm and give more details concerning its implementation covering numerical differentiation, variable metric updates, and a barrier parameter decrease. In Section 5 and Section 6, we study theoretical properties of the primal interior-point method. Using standard weak assumptions, we prove that this method is globally convergent if a bounded barrier is used. Then, using stronger assumptions, we prove that it is globally convergent also for the logarithmic barrier. Finally, in Section 7 we present results of computational experiments confirming the efficiency of the primal interior point method for special cases of generalized minimax problems.

2. ITERATIVE DETERMINATION OF THE MINIMAX VECTOR

The necessary conditions for (x, z) to be a minimum of function (8) have the form

$$\nabla_x B_\mu(x,z) = -\mu \sum_{i=1}^{m} \sum_{j=1}^{n_i} A_{ij}(x) \varphi'(z_i - f_{ij}(x)) = 0 \tag{13}$$

and

$$\frac{\partial B_\mu(x,z)}{\partial z_i} = h_i'(z) + \mu \sum_{j=1}^{n_i} \varphi'(z_i - f_{ij}(x)) = 0, \quad 1 \le i \le m, \tag{14}$$

where $h_i'(z) = \partial h(z)/\partial z_i$, $1 \le i \le m$. For solving this system of $n + m$ nonlinear equations we can use the Newton method whose iteration step can be written in the form

$$\begin{bmatrix} W(x,z) & -A_1(x)v_1(x,z) & \dots & -A_m(x)v_m(x,z) \\ -v_1^T(x,z)A_1^T(x) & h_{11}''(z) + e_1^T v_1(x,z) & \dots & h_{1m}''(z) \\ \dots & \dots & \dots & \dots \\ -v_m^T(x,z)A_m^T(x) & h_{m1}''(z) & \dots & h_{mm}''(z) + e_m^T v_m(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z_1 \\ \dots \\ \Delta z_m \end{bmatrix}$$

$$= - \begin{bmatrix} \sum_{i=1}^m A_i(x)u_i(x,z) \\ h_1'(z) - e_1^T u_1(x,z) \\ \dots \\ h_m'(z) - e_m^T u_m(x,z) \end{bmatrix}, \tag{15}$$

where

$$\begin{aligned} W(x,z) &= \sum_{i=1}^m \sum_{j=1}^{n_i} G_{ij}(x)u_{ij}(x,z) + \sum_{i=1}^m \sum_{j=1}^{n_i} A_{ij}(x)v_{ij}(x,z)A_{ij}^T(x) \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} G_{ij}(x)u_{ij}(x,z) + \sum_{i=1}^m A_i(x)V_i(x,z)A_i^T(x), \end{aligned}$$

$$u_{ij}(x,z) = -\mu\varphi'(z_i - f_{ij}(x)), \quad v_{ij}(x,z) = \mu\varphi''(z_i - f_{ij}(x)), \quad h_{ij}''(z) = \frac{\partial^2 h(z)}{\partial z_i \partial z_j},$$

(note that $u_{ij}(x,z) > 0$, $v_{ij}(x,z) > 0$ by Condition 1) for $1 \le i \le m$, $1 \le j \le n_i$, and where $A_i(x) = [A_{i1}(x), \dots, A_{in_i}(x)]$, $V_i(x,z) = \mathrm{diag}(v_{i1}(x,z), \dots, v_{in_i}(x,z))$,

$$u_i(x,z) = \begin{bmatrix} u_{i1}(x,z) \\ \dots \\ u_{in_i}(x,z) \end{bmatrix}, \quad v_i(x,z) = \begin{bmatrix} v_{i1}(x,z) \\ \dots \\ v_{in_i}(x,z) \end{bmatrix}, \quad e_i = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix},$$

for $1 \le i \le m$. This formula can be easily verified by the differentiation of (13) and (14) by x and z. Setting

$$C(x,z) = [A_1(x)v_1(x,z), \dots, A_m(x)v_m(x,z)], \quad g(x,z) = \sum_{i=1}^m A_i(x)u_i(x,z),$$

$$\Delta z = \begin{bmatrix} \Delta z_1 \\ \dots \\ \Delta z_m \end{bmatrix}, \quad c(x,z) = \begin{bmatrix} h_1'(z) - e_1^T u_1(x,z) \\ \dots \\ h_m'(z) - e_m^T u_m(x,z) \end{bmatrix},$$

$$H(z) = \nabla^2 h(z), \quad V(x,z) = \mathrm{diag}(e_1^T v_1(x,z), \dots, e_m^T v_m(x,z)),$$

we can rewrite equation (15) in the form

$$\begin{bmatrix} W(x,z) & -C(x,z) \\ -C^T(x,z) & H(z) + V(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} g(x,z) \\ c(x,z) \end{bmatrix}. \tag{16}$$

Now let us have a large-scale partially separable problem (i. e. the number of variables n is large and the functions $f_{ij}(x)$, $1 \le i \le m$, $1 \le j \le n_i$, depend on a small number of variables). Then we can assume that the matrix $W(x,z)$ is sparse and it can be efficiently decomposed. Two cases will be investigated.

First, if m is small (for example in the minimax problems, where $m = 1$), we use the fact that

$$\begin{bmatrix} W & -C \\ -C^T & H + V \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} W^{-1} - W^{-1}C(C^TW^{-1}C - H - V)^{-1}C^TW^{-1} & -W^{-1}C(C^TW^{-1}C - H - V)^{-1} \\ -(C^TW^{-1}C - H - V)^{-1}C^TW^{-1} & -(C^TW^{-1}C - H - V)^{-1} \end{bmatrix}.$$

We assume that matrix W is nonsingular, since it is perturbed by the Gill–Murray decomposition procedure [6] if it is singular (see Step 4 of Algorithm 1). The solution is determined from the formulas

$$\Delta z = (C^TW^{-1}C - H - V)^{-1}(C^TW^{-1}g + c), \tag{17}$$

$$\Delta x = W^{-1}(C\Delta z - g). \tag{18}$$

In this case, we need to decompose the large sparse matrix W of order n and the small dense matrix $C^TW^{-1}C - H - V$ of order m.

In the second case we assume that the numbers n_i, $1 \le i \le m$, are small and the matrix $H(z)$ is diagonal (as in the sums of absolute values) so the matrix

$$W(x, z) - C(x, z)D^{-1}(x, z)C^T(x, z), \quad D(x, z) = H(z) + V(x, z),$$

is sparse (matrix D is positive definite, since $H(z)$ is positive semidefinite by Assumption 2 and diagonal matrix $V(x, z)$ has positive diagonal elements). Then we can use the fact that

$$\begin{bmatrix} W & -C \\ -C^T & D \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (W - CD^{-1}C^T)^{-1} & (W - CD^{-1}C^T)^{-1}CD^{-1} \\ D^{-1}C^T(W - CD^{-1}C^T)^{-1} & D^{-1} + D^{-1}C^T(W - CD^{-1}C^T)^{-1}CD^{-1} \end{bmatrix}.$$

The solution is determined from the formulas

$$\Delta x = -(W - CD^{-1}C^T)^{-1}(g + CD^{-1}c), \tag{19}$$

$$\Delta z = D^{-1}(C^T\Delta x - c). \tag{20}$$

In this case, we need to decompose the large sparse matrix $W - CD^{-1}C^T$ of order n. The inversion of the diagonal matrix D of order m is trivial.

In every step of the primal interior point method with the iterative determination of the minimax vector we know the value of the parameter μ and the vectors $x \in R^n$, $z \in R^m$ such that $z_i > F_i(x)$, $1 \le i \le m$. Using (17) − (18) or (19) − (20), we determine direction vectors Δx, Δz and select a step-size α in such a way that

$$B_\mu(x + \alpha\Delta x, z + \alpha\Delta z) < B_\mu(x, z) \tag{21}$$

and $z_i + \alpha\Delta z_i > F_i(x + \alpha\Delta x)$, $1 \le i \le m$. Finally, we set $x^+ = x + \alpha\Delta x$, $z^+ = z + \alpha\Delta z$ and determine a new value $\mu^+ \le \mu$.

Inequality (21) is satisfied for sufficiently small values of the step-size α, if the matrix of system (16) is positive definite.

Theorem 2.1. Let the matrix $G = \sum_{i=1}^m \sum_{j=1}^{n_i} G_{ij}(x)u_{ij}(x, z)$ be positive definite. Then the matrix of system (16) is positive definite.

Proof. The matrix of equation (16) is positive definite if and only if the matrix $D = H + V$ as well as its Schur complement $W - CD^{-1}C^T$ are both positive definite. The matrix $D = H + V$ is positive definite since H is positive semidefinite and V is positive definite. Now we use the fact that the matrix $V^{-1} - D^{-1}$ is positive semidefinite, since the matrix $H = D - V$ is positive semidefinite (see [12]). Thus $v^T(W - CD^{-1}C^T)v \geq v^T(W - CV^{-1}C^T)v \ \forall v \in R^n$ so it suffices to prove that the matrix $W - CV^{-1}C^T$ is positive definite. But

$$W - CV^{-1}C^T = G + \sum_{i=1}^{m} \left(A_i V_i A_i^T - A_i V_i e_i (e_i^T V_i e_i)^{-1} (A_i V_i e_i)^T \right),$$

the matrices $A_i V_i A_i^T - A_i V_i e_i (e_i^T V_i e_i)^{-1} (A_i V_i e_i)^T$, $1 \leq i \leq m$, are positive semidefinite by the Schwarz inequality and the matrix G is positive definite by assumption. □

3. DIRECT DETERMINATION OF THE MINIMAX VECTOR

Minimization of the barrier function can be considered as the two-level optimization

$$z(x;\mu) = \arg \min_{z \in R^m} B_\mu(x,z), \tag{22}$$

$$x^* = \arg \min_{x \in R^n} B(x;\mu), \quad B(x;\mu) \triangleq B_\mu(x, z(x;\mu)). \tag{23}$$

Equation (22) serves for the determination of the optimal vector $z(x;\mu) \in R^m$ corresponding to a given vector $x \in R^n$. Assuming x fixed, function $B_\mu(x,z)$ is strictly convex (as a function of vector z), since it is a sum of convex function $h(z)$ and strictly convex functions $\mu\varphi(z_i - f_{ij}(x))$, $1 \leq i \leq m$, $1 \leq j \leq n_i$. As a stationary point, its minimum is the solution of the set of equations (14). We prove existence and uniqueness of this solution for the logarithmic barrier, for which $\varphi'(t) = -1/t$.

Theorem 3.1. The system of equations

$$h_i'(z) - \sum_{j=1}^{n_i} \frac{\mu}{z_i - f_{ij}(x)} = 0, \quad h_i'(z) = \frac{\partial h(z)}{\partial z_i}, \quad 1 \leq i \leq m, \tag{24}$$

with $x \in R^n$ fixed, has the unique solution $z(x;\mu) \in Z \subset R^m$ such that

$$F_i(x) < \underline{z}_i \leq z_i(x;\mu) \leq \overline{z}_i, \quad 1 \leq i \leq m, \tag{25}$$

with

$$\underline{z}_i = F_i(x) + \mu/\overline{h}_i, \quad \overline{z}_i = F_i(x) + n_i\mu/\underline{h}_i,$$

where $\underline{h}_i > 0$ are bounds used in (3) and $\overline{h_i} = h_i(\overline{z}_1, \ldots, \overline{z}_m)$.

Proof. Let $\overline{z}_i = F_i(x) + n_i\mu/\underline{h}_i$, $\overline{h_i} = h_i(\overline{z}_1, \ldots, \overline{z}_m)$, $\underline{z}_i = F_i(x) + \mu/\overline{h}_i$ for $1 \leq i \leq m$. If (24) holds, then

$$\underline{h}_i - \frac{n_i\mu}{z_i - F_i(x)} \leq 0 \quad \Rightarrow \quad z_i - F_i(x) \leq n_i\mu/\underline{h}_i$$

and

$$\overline{h}_i - \frac{\mu}{z_i - F_i(x)} \geq 0 \quad \Rightarrow \quad z_i - F_i(x) \geq \mu/\underline{h}_i,$$

which proves (25). Choosing an arbitrary (sufficiently small) number $\varepsilon > 0$, the function $B_\mu(x,z)$ attains its minimum on the compact set

$$Z_\varepsilon(x;\mu) = \{z \in R^m : \ \underline{z}_i - \varepsilon\mu/\overline{h}_i \leq z_i \leq \overline{z}_i + \varepsilon n_i \mu/\underline{h}_i, \ 1 \leq i \leq m\} \subset \text{int } Z,$$

since it is continuous on int Z. Now we will show that this minimum cannot lie on the boundary of $Z_\varepsilon(x;\mu)$. It is clear that for every point of this boundary there is at least one index $1 \leq i \leq m$ such that either $z_i = \underline{z}_i - \varepsilon\mu/\overline{h}_i$ or $z_i = \overline{z}_i + \varepsilon n_i\mu/\underline{h}_i$ holds. If $z_i = \underline{z}_i - \varepsilon\mu/\overline{h}_i$, then

$$\begin{aligned}\frac{\partial B_\mu(x,z)}{\partial z_i} &= h_i'(z) - \sum_{j=1}^{n_i} \frac{\mu}{z_i - f_{ij}(x)} \leq \overline{h}_i - \frac{\mu}{\underline{z}_i - \varepsilon\mu/\overline{h}_i - F_i(x)} \\ &= \overline{h}_i - \frac{\mu}{(1-\varepsilon)\mu/\overline{h}_i} = -\frac{\varepsilon \overline{h}_i}{1-\varepsilon} < 0,\end{aligned}$$

so a small increase of the variable z_i can decrease the function value of $B_\mu(x,z)$. If $z_i = \overline{z}_i + \varepsilon n_i\mu/\underline{h}_i$, then

$$\begin{aligned}\frac{\partial B_\mu(x,z)}{\partial z_i} &= h_i'(z) - \sum_{j=1}^{n_i} \frac{\mu}{z_i - f_{ij}(x)} \geq \underline{h}_i - \frac{n_i\mu}{\overline{z}_i + \varepsilon n_i\mu/\underline{h}_i - F_i(x)} \\ &= \underline{h}_i - \frac{n_i\mu}{(1+\varepsilon)n_i\mu/\underline{h}_i} = \frac{\varepsilon \underline{h}_i}{1+\varepsilon} > 0,\end{aligned}$$

so a small decrease of the variable z_i can decrease the function value of $B_\mu(x,z)$. The above considerations imply that the minimum of the function $B_\mu(x,z)$ is an interior point of the set $Z_\varepsilon(x;\mu)$ and since $B_\mu(x,z)$ is continuously differentiable on $Z_\varepsilon(x;\mu)$, necessary conditions (24) have to be satisfied. Since the number $\varepsilon > 0$ can be chosen arbitrarily, the solution satisfies inequalities $F_i(x) < \underline{z}_i \leq z_i(x;\mu) \leq \overline{z}_i$, $1 \leq i \leq m$. The uniqueness of this solution follows from the strict convexity of $B_\mu(x,z)$. □

Similar results can be obtained for other barriers as well. Using barrier (9), we get equations

$$h_i'(z) - \sum_{j=1}^{n_i} \frac{\mu}{(z_i - f_{ij}(x))(z_i - f_{ij}(x) + 1)} = 0, \quad 1 \leq i \leq m,$$

and inequalities of the form (25) with bounds

$$\underline{z}_i = F_i(x) + \frac{2\mu/\overline{h}_i}{1 + \sqrt{1 + 4\mu/\overline{h}_i}}, \quad \overline{z}_i = F_i(x) + \frac{2n_i\mu/\underline{h}_i}{1 + \sqrt{1 + 4n_i\mu/\underline{h}_i}},$$

see [13] (where also a bounded barrier similar as (10) – (11) is investigated).

System of equations (14) can be solved by the Newton method started, e. g., from the point z such that $z_i = \overline{z}_i$, $1 \le i \le m$. If the Hessian matrix of the function $h(z)$ is diagonal, then system (14) is decomposed on m scalar equations, which can be efficiently solved, e. g. by methods described in [9], [10] (see [13]).

If we are able to find a solution of system (14) for an arbitrary vector $x \in R^n$, we can restrict our attention to the unconstrained minimization of the function $B(x;\mu) = B_\mu(x, z(x;\mu))$, which has n variables. It is suitable to know the gradient and the Hessian matrix of this function.

Theorem 3.2. One has

$$\nabla B(x;\mu) = \sum_{i=1}^{m} A_i(x) u_i(x;\mu) = A(x)u(x;\mu), \tag{26}$$

where $A(x) = [A_1(x), \dots, A_m(x)]$, $u(x;\mu) = [u_1^T(x;\mu), \dots, u_m^T(x;\mu)]^T$, and also

$$\nabla^2 B(x;\mu) = W(x;\mu) - C(x;\mu)\left(H(z(x;\mu)) + V(x;\mu)\right)^{-1} C^T(x;\mu), \tag{27}$$

where $W(x;\mu) = W(x, z(x;\mu))$, $C(x;\mu) = C(x, z(x;\mu))$, $V(x;\mu) = V(x, z(x;\mu))$, and $u_i(x;\mu) = u_i(x, z(x;\mu))$, $1 \le i \le m$ (see the previous section). If the matrix $H(z(x;\mu))$ is diagonal, we can express (27) in the form

$$\begin{aligned} \nabla^2 B(x;\mu) &= G(x;\mu) + \sum_{i=1}^{m} A_i(x) V_i(x;\mu) A_i^T(x) \\ &- \sum_{i=1}^{m} \frac{A_i(x) V_i(x;\mu) e_i e_i^T V_i(x;\mu) A_i^T(x)}{\partial^2 h(z(x;\mu))/\partial z_i^2 + e_i^T V_i(x;\mu) e_i}, \end{aligned} \tag{28}$$

where $G(x;\mu) = G(x, z(x;\mu))$ and $V_i(x;\mu) = V_i(x, z(x;\mu))$, $1 \le i \le m$ (see the previous section).

Proof. Differentiating the function

$$B(x;\mu) = h(z(x;\mu)) + \mu \sum_{i=1}^{m} \sum_{j=1}^{n_i} \varphi(z_i(x;\mu) - f_{ij}(x)), \tag{29}$$

we obtain

$$\begin{aligned} \nabla B(x;\mu) &= \sum_{i=1}^{m} \frac{\partial h(z(x;\mu))}{\partial z_i} \frac{\partial z_i(x;\mu)}{\partial x} - \sum_{i=1}^{m} \sum_{j=1}^{n_i} u_{ij}(x;\mu) \left(\frac{\partial z_i(x;\mu)}{\partial x} - \frac{\partial f_{ij}(x)}{\partial x} \right) \\ &= \sum_{i=1}^{m} \frac{\partial z_i(x;\mu)}{\partial x} \left(\frac{\partial h(z(x;\mu))}{\partial z_i} - \sum_{j=1}^{n_i} u_{ij}(x;\mu) \right) + \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{\partial f_{ij}(x)}{\partial x} u_{ij}(x;\mu) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} A_{ij}(x) u_{ij}(x;\mu) = \sum_{i=1}^{m} A_i(x) u_i(x;\mu) \end{aligned}$$

by (14), where

$$u_{ij}(x;\mu) = -\mu\varphi'(z_i(x;\mu) - f_{ij}(x)), \quad 1 \le i \le m, \quad 1 \le j \le n_i. \tag{30}$$

Formula (27) can be derived by an additional differentiation of relations (14) and (26) using (30). A simpler way is based on the use of formula (19). Since (14) implies $c(x, z(x;\mu)) = 0$, we can substitute $c = 0$ into (19) to obtain the relation

$$\Delta x = -\left(W(x,z) - C(x,z)\,(H(z) + V(x,z))^{-1}\,C^T(x,z)\right)^{-1} g(x,z)$$

with $z = z(x;\mu)$, which confirms a validity of formula (27) (more details are given in [13]). $\square$

To determine the Hessian matrix inverse, we can use relations (17)–(18) which, after substitution $c(x, z(x;\mu)) = 0$, give

$$\begin{aligned}(\nabla^2 B(x;\mu))^{-1} \quad = \quad & W^{-1}(x;\mu) - W^{-1}(x;\mu)C(x;\mu) \\ & \left(C^T(x;\mu)W^{-1}(x;\mu)C(x;\mu) - H(z(x;\mu)) - V(x;\mu)\right)^{-1} \\ & C^T(x;\mu)W^{-1}(x;\mu).\end{aligned} \tag{31}$$

If system (14) is not solved with a sufficient precision, we use (19) − (20) rather than (27) and (17) − (18) rather than (31), where the actual vector $c(x, z(x;\mu)) \neq 0$ is substituted.

In every step of the primal interior point method with the direct determination of the minimax vector we know the value of the parameter μ and the vector $x \in R^n$. Solving system (14) we determine the vector $z(x;\mu)$, using the Hessian matrix (27) or its inverse (31) we determine a direction vector Δx and select a step-size α in such a way that

$$B_\mu(x + \alpha\Delta x, z(x + \alpha\Delta x;\mu)) < B_\mu(x, z(x;\mu)) \tag{32}$$

(the vector $z(x + \alpha\Delta x;\mu)$ is obtained as a solution of system (14), in which x is replaced by $x + \alpha\Delta x$). Finally, we set $x^+ = x + \alpha\Delta x$ and determine a new value $\mu^+ \le \mu$. Conditions for the direction vector Δx to be descent are the same as in Theorem 2.1. It suffices when the matrix $G(x;\mu)$ is positive definite.

4. IMPLEMENTATION

In this section, we restrict our attention on the direct determination of the minimax vector. There are two possibilities, the line search implementation or the trust-region implementation. The first one was used in [13] for large-scale minimax optimization and the second one in [14] for large-scale l_1 optimization. These papers contain all necessary details concerning both implementations. Here we briefly describe the line search implementation realized by the following algorithm. To prove the global convergence, the direction vector $d = \Delta x$ is modified in such a way that

$$-g^T d \ge \varepsilon_0 \|g\|\|d\|, \quad \underline{c}\|g\| \le \|d\| \le \overline{c}\|g\|, \tag{33}$$

where $g = A(x)u(x;\mu)$ and ε_0, $\underline{c}$, $\overline{c}$ are suitable constants (see Step 6 of Algorithm 1).

Algorithm 1.

Data: Termination parameter $\underline{\varepsilon} > 0$, precision for the nonlinear equation solver $\underline{\delta} > 0$, bounds for the barrier parameter $0 < \underline{\mu} < \overline{\mu}$, 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 $\overline{\Delta} > 0$, symptom of direction determination $\mathcal{D}$ ($\mathcal{D} = 1$ or $\mathcal{D} = 2$).

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

Step 1: *Initiation.* Set $\mu = \overline{\mu}$. If $\mathcal{D} = 1$, determine the sparsity pattern of matrix $W = W(x; \mu)$ from the sparsity pattern of matrix $A(x)$ and carry out a symbolic decomposition of W. If $\mathcal{D} = 2$, determine the sparsity pattern of matrices $W = W(x; \mu)$ and $C = C(x; \mu)$ from the sparsity pattern of matrix $A(x)$ and carry out a symbolic decomposition of matrix $W - CD^{-1}C^T$. Compute values $f_{ij}(x)$, $1 \le i \le m$, $1 \le j \le n_i$, $F_i(x) = \max_{1 \le j \le n_i} f_{ij}(x)$, $1 \le i \le m$, and $F(x) = h(F_1(x), \ldots, F_m(x))$. Set $k := 0$ (iteration count) and $r := 0$ (restart indicator).

Step 2: *Termination.* Solve nonlinear equations (14) with precision $\underline{\delta}$ to obtain vectors $z(x; \mu)$ and $u(x; \mu)$. Compute matrix $A := A(x)$ and vector $g := g(x; \mu) = A(x)u(x; \mu)$. If $\mu \le \underline{\mu}$ and $\|g\| \le \underline{\varepsilon}$, then terminate the computation. Otherwise set $k := k + 1$.

Step 3: *Approximation of the Hessian matrix.* Set $G = G(x; \mu)$ or compute an approximation G of the Hessian matrix $G(x; \mu)$ by using either gradient differences or variable metric updates (more details are given below).

Step 4: *Direction determination.* If $\mathcal{D} = 1$, determine vector $d = \Delta x$ from (17)–(18) by using the Gill–Murray decomposition of matrix W. If $\mathcal{D} = 2$, determine vector $d = \Delta x$ from (19)–(20) by using the Gill–Murray decomposition of matrix $W - CD^{-1}C^T$.

Step 5: *Restart.* If $r = 0$ and (33) does not hold, select a positive definite diagonal matrix $\tilde{D}$, set $G = \tilde{D}$, $r := 1$ and go to Step 4 (more details are given in [13]). If $r = 1$ and (33) does not hold, set $d := -g$ (the steepest descent direction). Set $r := 0$.

Step 6: *Step-length selection.* Define the maximum step-length $\overline{\alpha} = \min(1, \overline{\Delta}/\|d\|)$. Find a minimum integer $l \ge 0$ such that $B(x + \beta^l \overline{\alpha} d; \mu) \le B(x; \mu) + \varepsilon_1 \beta^l \overline{\alpha} g^T d$ (note that nonlinear equations (14) has to be solved at all points $x + \beta^j \overline{\alpha} d$, $0 \le j \le l$). Set $x := x + \beta^l \overline{\alpha} d$. Compute values $f_{ij}(x)$, $1 \le i \le m$, $1 \le j \le n_i$, $F_i(x) = \max_{1 \le j \le n_i} f_{ij}(x)$, $1 \le i \le m$, and $F(x) = h(F_1(x), \ldots, F_m(x))$.

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

In Step 3 of Algorithm 1 we assume that $G = G(x;\mu)$, where $G(x;\mu)$ is either given analytically or determined by using automatic differentiation, see [7]. In practical computations, G is frequently an approximation of $G(x;\mu)$ obtained by using either gradient differences or variable metric updates. In the first case, G is computed by differences $A(x+\delta w_j)u(x;\mu) - A(x)u(x;\mu)$ for a suitable set of vectors w_j, $j = 1,2,\dots,\underline{n}$, where $\underline{n} \ll n$ if G is sparse. Determination of vectors w_j, $j = 1,2,\dots,\underline{n}$, is equivalent to a graph coloring problem, see [3]. The corresponding code is proposed in [2]. In the second case, G is defined by the expression

$$G = \sum_{i=1}^{m}\sum_{j=1}^{n_i} u_{ij}(x;\mu)G_{ij}, \tag{34}$$

where approximations G_{ij} of $\nabla^2 f_{ij}(x)$ are computed by using variable metric updates described in [8]. In our implementation we use safeguarded scaled BFGS updates. Let $R^n_{ij} \subset R^n$, $1 \le i \le m$, $1 \le j \le n_i$, be subspaces defined by independent variables of functions f_{ij} and Z_{ij} be matrices whose columns form canonical orthonormal bases in these subspaces (they are columns of the unit matrix of order n). Then we can define reduced approximations of the Hessian matrices $\tilde{G}_{ij} = Z_{ij}^T G_{ij} Z_{ij}$, $1 \le i \le m$, $1 \le j \le n_i$. New reduced approximations of the Hessian matrices, used in the next iteration, are computed by the formulas

$$\begin{aligned} \tilde{G}^+_{ij} &= \frac{1}{\tilde{\gamma}_{ij}}\left(\tilde{G}_{ij} - \frac{\tilde{G}_{ij}\tilde{s}_{ij}\tilde{s}^T_{ij}\tilde{G}_{ij}}{\tilde{s}^T_{ij}\tilde{G}_{ij}\tilde{s}_{ij}}\right) + \frac{\tilde{y}_{ij}\tilde{y}^T_{ij}}{\tilde{s}^T_{ij}\tilde{y}_{ij}}, \quad && \tilde{s}^T_{ij}\tilde{y}_{ij} > 0, \\ \tilde{G}^+_{ij} &= \tilde{G}_{ij}, && \tilde{s}^T_{ij}\tilde{y}_{ij} \le 0, \end{aligned}$$

where

$$\tilde{s}_{ij} = Z^T_{ij}(x^+ - x), \quad \tilde{y}_{ij} = Z^T_{ij}(\nabla f_{ij}(x^+) - \nabla f_{ij}(x)), \quad 1 \le i \le m, \quad 1 \le j \le n_i,$$

and where either $\tilde{\gamma}_{ij} = 1$ or $\tilde{\gamma}_{ij} = \tilde{s}^T_{ij}\tilde{G}_{ij}\tilde{s}_{ij}/\tilde{s}^T_{ij}\tilde{y}_{ij}$ (we denote by + quantities from the next iteration). The particular choice of $\tilde{\gamma}_{ij}$ is determined by the controlled scaling strategy described in [15]. In the first iteration we set $\tilde{G}_{ij} = I_{ij}$, where I_{ij} are unit matrices of suitable orders. Finally, $G^+_{ij} = Z_{ij}\tilde{G}^+_{ij}Z^T_{ij}$, $1 \le i \le m$, $1 \le j \le n_i$.

Restart in Step 5 of Algorithm 1 assures that the direction vectors are uniformly descent and gradient-related ((33) holds). If Assumptions 1–3 are satisfied, then the Armijo line search (Step 6 of Algorithm 1) guarantees that a constant c exists such that

$$B(x_{k+1};\mu_k) - B(x_k;\mu_k) \le -c\|g(x_k;\mu_k)\|^2 \quad \forall k \in N, \tag{35}$$

see [5] (note that $g(x_k;\mu_k) = \nabla B(x_k;\mu_k)$ by Theorem 3.2). If only Assumptions 1–2 hold, the Armijo line search implies weaker inequality

$$B(x_{k+1};\mu_k) - B(x_k;\mu_k) \le 0 \quad \forall k \in N. \tag{36}$$

Restarts are sometimes used when $G_k = G(x_k;\mu_k)$, since G_k can be indefinite in this case. If G_k is determined using partitioned variable metric (safeguarded BFGS) updates, then G_k is positive definite and restarts are unnecessary. More details concerning restarts are given in [13].

A very important part of Algorithm 1 is the barrier parameter update. 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 $z_i(x;\mu) = F_i(x)$ when μ is too small (since $F_i(x) < z_i(x;\mu) \le \overline{z}_i(x;\mu)$ and $\overline{z}_i(x;\mu) \to F_i(x)$ as $\mu \to 0$ for all barriers mentioned in Section 1), which leads to a breakdown (division by $z_i(x;\mu) - F_i(x) = 0$ in computation of $\varphi'(z_i(x;\mu) - F_i(x))$). Thus a lower bound $\underline{\mu}$ for the barrier parameter has to be used (we recommend the value $\underline{\mu} = 10^{-10}$ in a double precision arithmetic).

Algorithm 1 is also sensitive to the way in which the barrier parameter decreases. Denoting by $s_{ij}(x;\mu) = z_i(x;\mu) - f_{ij}(x)$, $1 \le i \le m$, $1 \le j \le n_i$, slack variables, we can see from (30) that $u_{ij}(x;\mu)s_{ij}(x;\mu) = \mu$, $1 \le i \le m$, $1 \le j \le n_i$, if the logarithmic barrier is used. In this case, interior-point methods assume that μ decreases linearly (see [19]). We have tested various possibilities for the barrier parameter update including simple geometric sequences, which proved to be unsuitable. Better results were obtained by the following two heuristic procedures, where $g(x_k;\mu_k) = A(x_k)u(x_k;\mu_k)$ and $\underline{g}$ is a suitable constant.

Procedure A.

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

Phase 2: If $\|g(x_k;\mu_k)\| < \underline{g}$, we set

$$\mu_{k+1} = \max\left(\tilde{\mu}_{k+1},\, \underline{\mu},\, 10\,\varepsilon_M |F(x_{k+1})|\right), \tag{37}$$

where $F(x_{k+1}) = h\left(F_1(x_{k+1}), \dots, F_m(x_{k+1})\right)$, ε_M is the machine precision, and

$$\tilde{\mu}_{k+1} = \min\left[\max(\lambda\mu_k,\, \mu_k/(\sigma\mu_k + 1)),\, \max(\|g(x_k;\mu_k)\|^2,\, 10^{-2k})\right]. \tag{38}$$

The values $\underline{\mu} = 10^{-10}$, $\lambda = 0.85$, and $\sigma = 100$ are chosen as defaults.

Procedure B.

Phase 1: If $\|g(x_k;\mu_k)\|^2 \ge \rho\mu_k$, we set $\mu_{k+1} = \mu_k$, i. e., the barrier parameter is not changed.

Phase 2: If $\|g(x_k;\mu_k)\|^2 < \rho\mu_k$, we set

$$\mu_{k+1} = \max(\underline{\mu}, \|g_k(x_k;\mu_k)\|^2). \tag{39}$$

The values $\underline{\mu} = 10^{-10}$ and $\rho = 0.1$ are chosen as defaults.

The choice of $\underline{g}$ in Procedure A is not critical. We can set $\underline{g} = \infty$ but a lower value is sometimes more suitable. Formula (38) 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 $\mu_k/(\sigma\mu_k+1)$). 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 $\|g(x_k;\mu_k)\|$ is small, even if x_k is far from the solution. The idea of Procedure B follows from the requirement that $B(x;\mu)$ should be sufficiently minimized for a current value of μ. Thus the parameter μ_k is changed only if $\|g(x_k;\mu_k)\|$ is sufficiently small.

5. GLOBAL CONVERGENCE FOR BOUNDED BARRIERS

In this section, we first assume that function $\varphi(t)$ is bounded from below, $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$ and all computations are exact. We will investigate an infinite sequence $\{x_k\}_1^\infty$ generated by Algorithm 1.

Lemma 5.1. Let Assumption 1, Assumption 2, Condition 1, Condition 2 be satisfied. Let $\{x_k\}_1^\infty$ and $\{\mu_k\}_1^\infty$ be sequences generated by Algorithm 1. Then sequences $\{B(x_k;\mu_k)\}_1^\infty$, $\{z(x_k;\mu_k)\}_1^\infty$, and $\{F(x_k)\}_1^\infty$ are bounded. Moreover, there is $L \ge 0$ such that

$$B(x_{k+1};\mu_{k+1}) \le B(x_{k+1};\mu_k) + L(\mu_k - \mu_{k+1}) \quad \forall k \in N. \tag{40}$$

Proof. (a) Since function $\varphi(t)$ is bounded from below, Assumption 1, Assumption 2, Condition 2 and (12) imply that

$$B(x;\mu) \ge h(z(x;\mu)) + \overline{m}\,\overline{\mu}\,\underline{\varphi} \ge h(\underline{F}_1,\dots,\underline{F}_m) + \overline{m}\,\overline{\mu}\,\underline{\varphi} \triangleq \underline{B}.$$

Furthermore, using (25), we obtain $z_i \ge F_i(x) \ge \underline{F}_i$, $1 \le i \le m$, and the boundedness from below is proved.

(b) Differentiating function (29) and using (14) one has

$$\begin{aligned}\frac{\partial B(x;\mu)}{\partial \mu} &= \sum_{i=1}^m \frac{\partial h(z(x;\mu))}{\partial z_i}\frac{\partial z_i(x;\mu)}{\partial \mu} + \mu \sum_{i=1}^m \sum_{j=1}^{n_i} \varphi'(z_i(x;\mu) - f_{ij}(x))\frac{\partial z_i(x;\mu)}{\partial \mu} \\ &\quad + \sum_{i=1}^m \sum_{j=1}^{n_i} \varphi\left(z_i(x;\mu) - f_{ij}(x)\right) = \sum_{i=1}^m \sum_{j=1}^{n_i} \varphi\left(z_i(x;\mu) - f_{ij}(x)\right) \ge \overline{m}\,\underline{\varphi}.\end{aligned}$$

(c) Using the mean value theorem and (b), we obtain

$$\begin{aligned}B(x_{k+1};\mu_{k+1}) - B(x_{k+1};\mu_k) &= \sum_{i=1}^m \sum_{j=1}^{n_i} \varphi\left(z_i(x_{k+1};\tilde{\mu}_k) - f_{ij}(x)\right)(\mu_{k+1} - \mu_k) \\ &\le \overline{m}\,\underline{\varphi}\,(\mu_{k+1} - \mu_k) \triangleq L(\mu_k - \mu_{k+1})\end{aligned}$$

which together with (36) gives $B(x_{k+1};\mu_{k+1}) \le B(x_k;\mu_k) + L(\mu_k - \mu_{k+1})\ \forall k \in N$. Thus

$$B(x_k;\mu_k) \le B(x_1;\mu_1) + L(\mu_1 - \mu_k) \le B(x_1;\mu_1) + L\mu_1 \triangleq \overline{B} \quad \forall k \in N.$$

Furthermore, using (12), Assumption 2, Condition 2 and (a), one has

$$\begin{aligned}\overline{B} &\ge B(x_k;\mu_k) \ge h(z(x_k;\mu_k)) + \overline{m}\,\overline{\mu}\,\underline{\varphi} \ge \underline{B} + \sum_{i=1}^{m} \underline{h}_i(z_i(x_k;\mu_k) - \underline{F}_i)\\ &\ge \underline{B} + \underline{h}_i(z_i(x_k;\mu_k) - \underline{F}_i), \quad 1 \le i \le m,\end{aligned}$$

which gives $F_i(x_k) \le z_i(x_k;\mu_k) \le (\overline{B} - \underline{B})/\underline{h}_i + \underline{F}_i$ for $1 \le i \le m$ and $k \in N$. Thus the boundedness from above is proved. □

The assertion of Lemma 5.1 does not depend on bounds $\overline{g}$ and $\overline{G}$, since we do not use Assumption 3. Thus an upper bound $\overline{F}$ (independent of $\overline{g}$ and $\overline{G}$) exists such that $F(x_k) \le \overline{F}$ for all $k \in N$. This bound can be used for the definition of the level set in Assumption 3.

Lemma 5.2. Let assumptions of Lemma 5.1 and Assumption 3 be satisfied. Then the values $\{\mu_k\}_1^\infty$, generated by Algorithm 1, form a non-increasing sequence such that $\mu_k \to 0$.

P r o o f. In Phase 1, the value of μ is fixed. Since the function $B(x;\mu)$ is continuous, bounded from below by Lemma 5.1, and since (35) (with $\mu_k = \mu$) holds, it can be proved (see [5]) that $\|g(x_k;\mu)\| \to 0$ if Phase 1 contains an infinite number of consecutive steps. Thus a step (with index l) belonging to Phase 1 exists such that either $\|g(x_l;\mu)\| < \underline{g}$ in Procedure A or $\|g(x_l;\mu)\|^2 < \rho\mu$ in Procedure B. This is a contradiction with the definition of Phase 1. □

Theorem 5.3. Let assumptions of Lemma 5.1 and Assumption 3 be satisfied. Consider a sequence $\{x_k\}_1^\infty$ generated by Algorithm 1 (with $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$). Then

$$\lim_{k\to\infty} \sum_{i=1}^{m}\sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k)\nabla f_{ij}(x_k) = 0, \quad \sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k) = h_i'(z(x_k;\mu_k)),$$

$$u_{ij}(x_k;\mu_k) \ge 0, \quad z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0,$$

$$\lim_{k\to\infty} u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) = 0$$

for $1 \le i \le m$ and $1 \le j \le n_i$.

P r o o f. (a) Equalities $e^T u_i(x_k;\mu_k) = h_i'(z(x_k;\mu_k))$, $1 \le i \le m$, hold since $\underline{\delta} = 0$. Inequalities $u_{ij}(x_k;\mu_k) \ge 0$ and $z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0$ follow from (30) and (25).

(b) Since (35) and (40) hold, we can write

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

which (since $\lim_{k\to\infty}\mu_k = 0$ by Lemma 5.2) implies

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

where $\underline{B}$ is a lower bound defined in the proof of Lemma 5.1. Thus one has

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

which implies that $g(x_k;\mu_k) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k)\nabla f_{ij}(x_k) \to 0$.

(c) Let $1 \le i \le m$ and $1 \le j \le n_i$ be chosen arbitrarily. Using the definition of $u_{ij}(x_k;\mu_k)$ and boundedness of $t\varphi'(t)$, we obtain

$$\begin{aligned} u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) &= -\mu_k\varphi'(z_i(x_k;\mu_k) - f_{ij}(x_k))(z_i(x_k;\mu_k) - f_{ij}(x_k)) \\ &\le \overline{c}\,\mu_k \to 0 \end{aligned}$$

by Lemma 5.2 ($\overline{c}$ is an upper bound for $-t\varphi'(t)$). □

Corollary 5.4. Let assumptions of Theorem 5.3 hold. Then every cluster point $x \in R^n$ of the sequence $\{x_k\}_1^\infty$ satisfies KKT conditions (6)−(7), where z and u (with elements z_i and u_{ij}, $1 \le i \le m$, $1 \le j \le n_i$) are cluster points of sequences $\{z(x_k;\mu_k)\}_1^\infty$ and $\{u(x_k;\mu_k)\}_1^\infty$.

Now, assuming that the values $\underline{\delta}$, $\underline{\varepsilon}$, $\underline{\mu}$ are nonzero, we can prove the following theorem informing us about the precision obtained, when Algorithm 1 terminates.

Theorem 5.5. Consider the sequence $\{x_k\}_1^\infty$ generated by Algorithm 1. Let assumptions of Lemma 5.1 and Assumption 3 hold. Then, choosing $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, $\underline{\mu} > 0$ arbitrarily, there is an index $k \ge 1$ such that

$$\|g(x_k;\mu_k)\| \le \underline{\varepsilon}, \quad |h_i'(z(x_k;\mu_k)) - \sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k)| \le \underline{\delta},$$

$$u_{ij}(x_k;\mu_k) \ge 0, \quad z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0,$$
$$u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) \le \overline{c}\,\overline{\mu}$$

for all $1 \le i \le m$ and $1 \le j \le n_i$ (note that $\overline{c} = 1$ for all barriers mentioned in Section 1).

P r o o f. The inequality $|h_i'(z(x_k;\mu_k)) - e^T u_i(x_k;\mu_k)| \le \underline{\delta}$ follows immediately from the fact that equations $e^T u_i(x_k;\mu_k) = h_i'(z(x_k;\mu_k))$, $1 \le i \le m$, are solved with the precision $\underline{\delta}$. Inequalities $u_{ij}(x_k;\mu_k) \ge 0$, $z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0$ follow from the definition of $u_{ij}(x_k;\mu_k)$ and from (25) as in the proof of Theorem 5.3. Since $\mu_k \to 0$ by Lemma 5.2 and $g(x_k;\mu_k) \to 0$ by Theorem 5.3, there is an index $k \ge 1$ such that $\mu_k \le \underline{\mu}$ and $\|g(x_k;\mu_k)\| \le \underline{\varepsilon}$ (thus Algorithm 1 terminates at the kth iteration). Using the definition of $u_{ij}(x_k;\mu_k)$, we obtain

$$\begin{aligned} u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k)-f_{ij}(x_k)) &= -\mu_k\varphi'(z_i(x_k;\mu_k)-f_{ij}(x_k))(z_i(x_k;\mu_k)-f_{ij}(x_k)) \\ &\le \overline{c}\,\mu_k \le \overline{c}\,\overline{\mu}. \end{aligned}$$

□

6. GLOBAL CONVERGENCE FOR THE LOGARITHMIC BARRIER

In this section, we first assume that $\varphi(t) = -\log t$, $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$ and all computations are exact. We will investigate an infinite sequence $\{x_k\}_1^\infty$ generated by Algorithm 1.

Lemma 6.1. Let Assumptions 2 and 4 be satisfied and $\varphi(t) = -\log t$. Then $B(x;\mu)$ is bounded from below.

P r o o f. Using (8), Assumption 2 (convexity of $h(z)$ and (3)) and Assumption 4, we can write

$$\begin{aligned} B(x;\mu) &= h(z(x;\mu)) - \mu\sum_{i=1}^m\sum_{j=1}^{n_i}\log\left(z_i(x;\mu) - f_{ij}(x)\right) \\ &\ge h(\underline{F}_1,\dots,\underline{F}_m) + \sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}_i\right) - \sum_{i=1}^m n_i\mu\log\left(z_i(x;\mu) - \underline{F}\right) \\ &\ge \underline{H} + \sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}\right) - \sum_{i=1}^m n_i\mu\log\left(z_i(x;\mu) - \underline{F}\right), \end{aligned}$$

where $\underline{H} = h(\underline{F}_1,\dots,\underline{F}_m) - \sum_{i=1}^m \underline{h}_i\left(\underline{F}_i - \underline{F}\right)$. Convex functions $\psi_i(t) = \underline{h}_i t - n_i\mu\log(t)$ have unique minima at the points $t_i = n_i\mu/\underline{h}_i$, $1 \le i \le m$. Thus

$$\begin{aligned} B(x;\mu) &\ge \underline{H} + \sum_{i=1}^m \underline{h}_i\frac{n_i\mu}{\underline{h}_i}\left(1 - \log\left(\frac{n_i\mu}{\underline{h}_i}\right)\right) \\ &\ge \underline{H} + \sum_{i=1}^m \underline{h}_i\min\left(0,\ \frac{n_i\overline{\mu}}{\underline{h}_i}\left(1 - \log\left(\frac{n_i\overline{\mu}}{\underline{h}_i}\right)\right)\right) \\ &\ge \underline{H} + \frac{1}{2}\sum_{i=1}^m \underline{h}_i\min\left(0,\ \frac{2n_i\overline{\mu}}{\underline{h}_i}\left(1 - \log\left(\frac{2n_i\overline{\mu}}{\underline{h}_i}\right)\right)\right) \triangleq \underline{B}. \end{aligned}$$

□

Now we clarify the dependence of $z(x;\mu)$ and $B(x;\mu)$ on the parameter μ.

Lemma 6.2. Let Assumption 2 be satisfied and $z(x;\mu)$ be a solution of linear system (24). Then

$$\frac{\partial B(x;\mu)}{\partial \mu} = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(z_i(x;\mu) - f_{ij}(x)\right).$$

If the Hessian matrix $H(z(x;\mu))$ is diagonal, then $\partial z(x;\mu)/\partial\mu > 0$.

Proof. (a) Differentiating the function

$$B(x;\mu) = h(z(x;\mu)) - \mu \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(z_i(x;\mu) - f_{ij}(x)\right)$$

and using (24), one has

$$\begin{aligned}
\frac{\partial B(x;\mu)}{\partial \mu} &= \sum_{i=1}^{m} \frac{\partial h(z(x;\mu))}{\partial z_i}\frac{\partial z_i(x;\mu)}{\partial \mu} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{\mu}{z_i(x;\mu) - f_{ij}(x)}\frac{\partial z_i(x;\mu)}{\partial \mu} \\
&\quad - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(z_i(x;\mu) - f_{ij}(x)\right) \\
&= -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(z_i(x;\mu) - f_{ij}(x)\right).
\end{aligned}$$

(b) Differentiating equation (24), which has the form

$$\frac{\partial h(z(x;\mu))}{\partial z_i} - \sum_{j=1}^{n_i} \frac{\mu}{z_i(x;\mu) - f_{ij}(x)} = 0,$$

we obtain

$$\sum_{k=1}^{m} \frac{\partial^2 h(z(x;\mu))}{\partial z_i \partial z_k}\frac{\partial z_k(x;\mu)}{\partial \mu} + \sum_{j=1}^{n_i} \frac{\mu}{(z_i(x;\mu) - f_{ij}(x))^2}\frac{\partial z_i(x;\mu)}{\partial \mu} - \sum_{j=1}^{n_i} \frac{1}{z_i(x;\mu) - f_{ij}(x)} = 0,$$

which gives

$$\sum_{k=1}^{m} h''_{ik}(z(x;\mu))\frac{\partial z_k(x;\mu)}{\partial \mu} + \left(\sum_{j=1}^{n_i} v_{ij}(x;\mu)\right)\frac{\partial z_i(x;\mu)}{\partial \mu} = \frac{1}{\mu}\sum_{j=1}^{n_i} u_{ij}(x;\mu) = \frac{1}{\mu} h'_i(z(x;\mu)),$$

or

$$(\mu H(z(x;\mu)) + \mu V(x;\mu))\frac{\partial z(x;\mu)}{\partial \mu} = \frac{\partial h(z(x;\mu))}{\partial z}.$$

If the Hessian matrix $H(z(x;\mu))$ is diagonal, then also $H(z(x;\mu)) + V(x;\mu)$ is diagonal with positive diagonal elements, which together with (3) imply that $\partial z(x;\mu)/\partial\mu > 0$. □

Now we prove that $B(x;\mu)$, $z(x;\mu)$, and $F(x)$ are bounded and $B(x;\mu)$ is a Lipschitz continuous function of μ.

Lemma 6.3. Let assumptions of Lemma 6.1 be satisfied and let the Hessian matrix $H(z(x;\mu))$ be diagonal. Let $\{x_k\}_1^\infty$ and $\{\mu_k\}_1^\infty$ be sequences generated by Algorithm 1. Then sequences $\{B(x_k;\mu_k)\}_1^\infty$, $\{z(x_k;\mu_k)\}_1^\infty$, and $\{F(x_k)\}_1^\infty$ are bounded. Moreover, there is $L \geq 0$ such that

$$B(x_{k+1};\mu_{k+1}) \leq B(x_{k+1};\mu_k) + L(\mu_k - \mu_{k+1}) \quad \forall k \in N. \tag{41}$$

P r o o f. Boundedness from below simply follows from Assumption 1, inequalities (25) and Lemma 6.1.
(a) As in the proof of Lemma 6.1, we can write

$$\begin{aligned} B(x;\mu) &\geq \underline{H} + \frac{1}{2}\sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}\right) \\ &\quad + \frac{1}{2}\sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}\right) - \sum_{i=1}^m n_i\mu \log\left(z_i(x;\mu) - \underline{F}\right) \\ &\geq \underline{H} + \frac{1}{2}\sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}\right) + \frac{1}{2}\sum_{i=1}^m \underline{h}_i \frac{2n_i\mu}{\underline{h}_i}\left(1 - \log\left(\frac{2n_i\mu}{\underline{h}_i}\right)\right) \\ &\geq \underline{H} + \frac{1}{2}\sum_{i=1}^m \underline{h}_i\left(z_i(x;\mu) - \underline{F}\right) + \frac{1}{2}\sum_{i=1}^m \underline{h}_i \min\left(0, \frac{2n_i\overline{\mu}}{\underline{h}_i}\left(1 - \log\left(\frac{2n_i\overline{\mu}}{\underline{h}_i}\right)\right)\right) \\ &\geq \underline{B} + \frac{\underline{h}}{2}\sum_{i=1}^m \left(z_i(x;\mu) - \underline{F}\right), \end{aligned}$$

where $\underline{h} = \min(\underline{h}_1, \ldots, \underline{h}_m)$. Thus

$$\sum_{i=1}^m \left(z_i(x;\mu) - \underline{F}\right) \leq \frac{2}{\underline{h}}\left(B(x;\mu) - \underline{B}\right). \tag{42}$$

(b) Using the mean value theorem and the first part of Lemma 6.2, we obtain

$$\begin{aligned} B(x_{k+1};\mu_{k+1}) - B(x_{k+1};\mu_k) &= \sum_{i=1}^m \sum_{j=1}^{n_i} \log\left(z_i(x_{k+1};\tilde{\mu}_k) - f_{ij}(x)\right)(\mu_k - \mu_{k+1}) \\ &\leq \sum_{i=1}^m n_i \log\left(z_i(x_{k+1};\tilde{\mu}_k) - \underline{F}\right)(\mu_k - \mu_{k+1}) \\ &\leq \frac{\overline{n}}{e}\sum_{i=1}^m \left(z_i(x_{k+1};\tilde{\mu}_k) - \underline{F}\right)(\mu_k - \mu_{k+1}), \end{aligned} \tag{43}$$

where $\mu_{k+1} \leq \tilde{\mu}_k \leq \mu_k$ and $\overline{n} = \max(n_1, \ldots, n_m)$. The last inequality follows from the relation $\log t \leq t/e$ (where $e = \exp(1)$), which holds for all $t > 0$. But $z_i(x_{k+1};\tilde{\mu}_k) \leq z_i(x_{k+1};\mu_k)$ by the second part of Lemma 6.2. Thus using (42), we can write

$$\begin{aligned} B(x_{k+1};\mu_{k+1}) &\leq B(x_{k+1};\mu_k) + \frac{\overline{n}}{e}\sum_{i=1}^{m}\left(z_i(x_{k+1};\mu_k) - \underline{F}\right)(\mu_k - \mu_{k+1}) \\ &\leq B(x_{k+1};\mu_k) + \frac{2\overline{n}}{e\underline{h}}(B(x_{k+1};\mu_k) - \underline{B})(\mu_k - \mu_{k+1}), \end{aligned}$$

which using (36) implies

$$B(x_{k+1};\mu_{k+1}) - \underline{B} \leq (1+\lambda\delta_k)(B(x_{k+1};\mu_k) - \underline{B}) \leq (1+\lambda\delta_k)(B(x_k;\mu_k) - \underline{B}),$$

where $\lambda = 2\overline{n}/(e\underline{h})$ and $\delta_k = \mu_k - \mu_{k+1}$. Then

$$\begin{aligned} B(x_{k+1};\mu_{k+1}) - \underline{B} &\leq \prod_{i=1}^{k}(1+\lambda\delta_i)(B(x_1;\mu_1) - \underline{B}) \\ &\leq \prod_{i=1}^{\infty}(1+\lambda\delta_i)(B(x_1;\mu_1) - \underline{B}) \end{aligned}$$

and since

$$\sum_{i=1}^{\infty}\lambda\delta_i \leq \lambda(\overline{\mu} - \lim_{k\to\infty}\mu_k) \leq \lambda\overline{\mu},$$

the above product is finite. This together with (25) and (42) proves that sequences $\{B(x_k;\mu_k)\}_1^\infty$, $\{z(x_k;\mu_k)\}_1^\infty$, and $\{F(x_k)\}_1^\infty$ are bounded from above.

(c) Using (43) and (25), we can write

$$\begin{aligned} B(x_{k+1};\mu_{k+1}) - B(x_{k+1};\mu_k) &\leq \sum_{i=1}^{m} n_i \log\left(z_i(x_{k+1};\tilde{\mu}_k) - \underline{F}\right)(\mu_k - \mu_{k+1}) \\ &\leq \sum_{i=1}^{m} n_i \log\left(\overline{F} + \frac{n_i\overline{\mu}}{\underline{h}_i} - \underline{F}\right)(\mu_k - \mu_{k+1}) \\ &\triangleq L(\mu_k - \mu_{k+1}), \end{aligned} \tag{44}$$

for all $k \in N$, where existence of $\overline{F}$ follows from boundedness of $\{F(x_k)\}_1^\infty$. □

The assertion of Lemma 6.3 does not depend on bounds $\overline{g}$ and $\overline{G}$, since we do not use Assumption 3. Thus an upper bound $\overline{F}$ (independent of $\overline{g}$ and $\overline{G}$) exists such that $F(x_k) \leq \overline{F}$ for all $k \in N$. This bound can be used for the definition of the level set in Assumption 3.

Lemma 6.4. Let assumptions of Lemma 6.1 and Assumption 3 be satisfied. Then the values $\{\mu_k\}_1^\infty$, generated by Algorithm 1, form a non-increasing sequence such that $\mu_k \to 0$.

Proof. The same as the proof of Lemma 5.2 (using Lemma 6.1 instead of Lemma 5.1). □

Theorem 6.5. Let assumptions of Lemma 6.3 and Assumption 3 be satisfied. Consider a sequence $\{x_k\}_1^\infty$ generated by Algorithm 1 (with $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$). Then

$$\lim_{k\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k)\nabla f_{ij}(x_k) = 0, \quad \sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k) = h_i'(z(x_k;\mu_k)),$$

$$u_{ij}(x_k;\mu_k) \ge 0, \quad z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0,$$

$$\lim_{k\to\infty} u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) = 0$$

for $1 \le i \le m$ and $1 \le j \le n_i$.

Proof. The same as the proof of Theorem 5.3 (using Lemma 6.1, Lemma 6.3 and Lemma 6.4 instead of Lemma 5.1). □

Corollary 6.6. Let assumptions of Theorem 6.5 hold. Then every cluster point $x \in R^n$ of the sequence $\{x_k\}_1^\infty$ satisfies KKT conditions (6)–(7), where z and u (with elements z_i and u_{ij}, $1 \le i \le m$, $1 \le j \le n_i$) are cluster points of sequences $\{z(x_k;\mu_k)\}_1^\infty$ and $\{u(x_k;\mu_k)\}_1^\infty$.

Now, assuming that the values $\underline{\delta}$, $\underline{\varepsilon}$, $\underline{\mu}$ are nonzero, we can prove the following theorem informing us about the precision obtained, when Algorithm 1 terminates.

Theorem 6.7. Consider the sequence $\{x_k\}_1^\infty$ generated by Algorithm 1. Let assumptions of Lemma 6.3 and Assumption 3 hold. Then, choosing $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, $\underline{\mu} > 0$ arbitrarily, there is an index $k \ge 1$ such that

$$\|g(x_k;\mu_k)\| \le \underline{\varepsilon}, \quad |h_i'(z(x_k;\mu_k)) - \sum_{j=1}^{n_i} u_{ij}(x_k;\mu_k)| \le \underline{\delta},$$

$$u_{ij}(x_k;\mu_k) \ge 0, \quad z_i(x_k;\mu_k) - f_{ij}(x_k) \ge 0,$$

$$u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) \le \underline{\mu}$$

for all $1 \le i \le m$ and $1 \le j \le n_i$.

Proof. The same as the proof of Theorem 5.5 (using Lemma 6.4 and Theorem 6.5 instead of Lemma 5.2 and Theorem 5.3). □

7. SPECIAL CASES AND NUMERICAL EXPERIMENTS

The simplest function of form (2) is the sum

$$F(x) = \sum_{i=1}^{m} F_i(x) = \sum_{i=1}^{m} \max_{1\le j\le n_i} f_{ij}(x). \tag{45}$$

In this case, $\partial h(z)/\partial z_i = 1$, $1 \le i \le m$, for an arbitrary vector z and the matrix $H(z)$ is diagonal. Using the logarithmic barrier, system of equations (14) decomposes on m scalar equations

$$1 - \sum_{j=1}^{n_i} \frac{\mu}{z_i(x;\mu) - f_{ij}(x)} = 0, \qquad 1 \le i \le m, \tag{46}$$

whose solutions lie in the intervals

$$F_i(x) + \mu \le z_i(x;\mu) \le F_i(x) + n_i\mu, \quad 1 \le i \le m,$$

as follows from the proof of Theorem 3.1 substituting $\overline{h}_i = \underline{h}_i = 1$. For $m = 1$ we obtain the classic minimax problem. A primal interior point method for this problem is described in [13]. Table 1, taken from [13], contains a comparison of three implementations of the primal interior point method (P1 uses the logarithmic barrier, P2 uses positive barrier (9), P3 uses bounded barrier (10) – (11)) with the smoothing method SM described in [20], and the primal-dual interior point method DI described in [11]. All these methods were realized as the line-search methods with two modifications: NM denotes the discrete Newton method with the Hessian matrix computed using the differences by the way described in [3] and VM denotes the variable metric method with the partitioned updates described in [8]. The tests were carried out using a collection of 22 test problems introduced in [16] (the source texts can be downloaded from the web page `www.cs.cas.cz/~luksan/test.html` as Test 14). In Table 1, `NIT` denotes the total number of iterations, `NFV` denotes the total number of function evaluations, `NFG` denotes the total number of gradient evaluations, `NR` denotes the total number of restarts, `NL` denotes the number of problems for which the lowest known local minimum was not found, `NF` denotes the number of failures, `NT` denotes the number of problems for which some parameters of the method had to be tuned, and `Time` denotes the total computational time in seconds.

Table 1. Test 14: minimax with 200 variables.

Method	NIT	NFV	NFG	NR	NL	NF	NT	Time
P1-NM	1675	3735	11109	327	-	-	4	1.92
P2-NM	2018	6221	12674	605	-	-	7	2.09
P3-NM	1777	3989	11596	379	1	-	7	2.11
SM-NM	4123	12405	32451	823	-	-	7	9.64
DI-NM	1771	3732	17952	90	1	-	10	6.34
P1-VM	1615	2429	1637	-	-	-	1	1.05
P2-VM	2116	3549	2138	2	-	-	3	1.47
P3-VM	1985	3208	2007	1	-	-	3	1.27
SM-VM	7244	21008	7266	-	1	-	8	9.09
DI-VM	1790	3925	1790	5	1	-	9	4.59

Table 1 indicates that the logarithmic barrier P1 is the best possibility for practical computations (in comparison with P2, P3) even if it needs stronger assumptions to prove its global convergence.

If $n_i = 2$, $1 \le i \le m$, equations (46) are quadratic and their solution has the form

$$z_i(x;\mu) = \mu + \frac{f_{i1}(x) + f_{i2}(x)}{2} + \sqrt{\mu^2 + \left(\frac{f_{i1}(x) - f_{i2}(x)}{2}\right)^2}, \quad 1 \le i \le m. \tag{47}$$

This formula can be used in the case when function $h : R^m \to R$ contains absolute values $F_i(x) = |f_i(x)| = \max(f_i(x), -f_i(x))$. Then $f_{i1}(x) = f_i(x)$ a $f_{i2}(x) = -f_i(x)$, so that

$$z_i(x;\mu) = \mu + \sqrt{\mu^2 + f_i^2(x)}, \quad 1 \le i \le m. \tag{48}$$

The primal interior point method for the sums of absolute values is described in [14]. Table 2 contains a comparison of two realizations of the primal interior point method with the logarithmic barrier (the trust region realization PT and the line-search realization PL) with the primal-dual interior point method DI described in [11] and the bundle variable metric method BM described in [17]. These methods were realized in two modifications: NM denotes the discrete Newton method with the Hessian matrix computed using the differences and VM denotes the variable metric method with the partitioned updates (BM is principally the variable metric method, so it could not be realized as NM). The tests were again carried out using a collection of 22 test problems introduced in [16]. The meaning of the columns is the same as in Table 1.

Table 2. Test 14: sum of absolute values with 200 variables.

Method	NIT	NFV	NFG	NR	NL	NF	NT	Time
PT-NM	3014	3518	27404	1	-	-	4	4.66
PL-NM	2651	12819	22932	3	1	-	6	5.24
DI-NM	5002	7229	42462	328	1	-	13	33.52
PT-VM	3030	3234	3051	-	-	1	1	1.44
PL-VM	2699	3850	2721	-	-	1	2	1.42
DI-VM	7138	14719	14719	9	2	-	9	86.18
BM-VM	34079	34111	34111	22	1	1	11	25.72

Tables 1 and 2 indicate that the primal interior point methods are very suitable for minimization of generalized minimax functions. They are more efficient than special bundle methods and also than general primal-dual interior point methods applied to problem (4) – (5). This is especially caused by the fact that the primal-dual interior point methods require the introduction of an additional slack vector $s \in R^m$ so that the resulting optimization problem contains $n + 2m$ variables x, z, s, which considerably increases the computational time.

ACKNOWLEDGEMENT

This work was supported by the Czech Science Foundation, project No. 201/09/1957, and the institutional research plan No. AV0Z10300504.

(Received December 3, 2009)

REFERENCES

[1] D. P. Bertsekas: Nondifferentiable optimization via approximation. In: Nondifferentiable Optimization (M. L. Balinski and P.Wolfe, eds.), Math. Programming Stud. *3* (1975), 1–25.

[2] T. F. Coleman, B. S. Garbow, and J. J. Moré: Software for estimating sparse Hessian matrices. ACM Trans. Math. Software *11* (1985), 363–377.

[3] T. F. Coleman and J. J. Moré: Estimation of sparse Hessian matrices and graph coloring problems. Math. Programming *28* (1984), 243–270.

[4] V. F. Demyanov and V. N. Malozemov: Introduction to Minimax. Dover Publications, 1990.

[5] R. Fletcher: Practical Methods of Optimization. Second edition. Wiley, New York 1987.

[6] P. E. Gill and W. Murray: Newton type methods for unconstrained and linearly constrained optimization. Math. Programming *7* (1974), 311–350.

[7] A. Griewank: Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. SIAM, Philadelphia 2000.

[8] A. Griewank and P. L. Toint: Partitioned variable metric updates for large-scale structured optimization problems. Numer. Math. *39* (1982), 119–137.

[9] D. Le: Three new rapidly convergent algorithms for finding a zero of a function. SIAM J. Sci. Stat. Computat. *6* (1985), 193–208.

[10] D. Le: An efficient derivative-free method for solving nonlinear equations. ACM Trans. Math. Software *11* (1985), 250–262.

[11] L. Lukšan, C. Matonoha, and J. Vlček: Interior point method for nonlinear nonconvex optimization. Numer. Linear Algebra Appl. *11* (2004), 431–453.

[12] L. Lukšan, C. Matonoha, and J. Vlček: On Lagrange multipliers of trust-region subproblems. BIT Numer. Math. *48* (2008), 763–768.

[13] L. Lukšan, C. Matonoha, and J. Vlček: Primal interior-point method for large sparse minimax optimization. Kybernetika *45* (2009), 841–864.

[14] L. Lukšan, C. Matonoha, and J. Vlček: Trust-region interior-point method for large sparse l_1 optimization. Optimization Methods and Software *22* (2007), 737–753.

[15] L. Lukšan and E. Spedicato: Variable metric methods for unconstrained optimization and nonlinear least squares. J. Comput. Appl. Math. *124* (2000), 61–93.

[16] L. Lukšan and J. Vlček: Sparse and Partially Separable Test Problems for Unconstrained and Equality Constrained Optimization. Technical Report V-767, ICS AS ČR, Prague 1998.

[17] L. Lukšan and J. Vlček: Variable metric method for minimization of partially separable nonsmooth functions. Pacific J. Optim. *2* (2006), 59–70.

[18] G. Di Pillo, L. Grippo, and S. Lucidi: Smooth transformation of the generalized minimax problem. J. Optim. Theory Appl. *95* (1997), 1–24.

[19] J. Vanderbei and D. F. Shanno: An interior point algorithm for nonconvex nonlinear programming. Comput. Optim. Appl. *13* (1999), 231–252.

[20] S. Xu: Smoothing methods for minimax problems. Computational Optim. Appl. *20* (2001), 267–279.

Ladislav Lukšan, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Praha 8 and Technical University of Liberec, Hálkova 6, 461 17 Liberec. Czech Republic.
e-mail: luksan@cs.cas.cz

Ctirad Matonoha, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Praha 8. Czech Republic.
e-mail: matonoha@cs.cas.cz

Jan Vlček, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Praha 8. Czech Republic.
e-mail: vlcek@cs.cas.cz