PANM 14

Ladislav Lukšan; Ctirad Matonoha; Jan Vlček

Primal interior point method for generalized minimax functions

In: Jan Chleboun and Petr Přikryl and Karel Segeth and Tomáš Vejchodský (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 1-6, 2008. Institute of Mathematics AS CR, Prague, 2008. pp. 138–149.

Persistent URL: http://dml.cz/dmlcz/702867

PRIMAL INTERIOR POINT METHOD FOR GENERALIZED MINIMAX FUNCTIONS*

Ladislav Lukšan, Ctirad Matonoha, Jan Vlček

Introduction

Generalized minimax optimization covers many practical problems, e.g., l_1 and l_∞ approximation or classic minimax optimization. In this contribution, we summarize new results described in our previous works [1]–[4], which can be downloaded from `http://www.cs.cas.cz/luksan/reports.html`. In these works, a connection with the current research and additional references are shown.

Definition 1 *We say that $F(x)$ is a generalized minimax function if*

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

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 $F_i(x)$, $1 \le i \le m$, 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 $h(z)$ is twice continuously differentiable and convex satisfying

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

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 $f_{ij}(x)$, $1 \le i \le m$, $1 \le j \le n_i$, 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 the first and second-order derivatives on $\text{conv}\mathcal{L}(\overline{F})$: 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 $x \in \text{conv}\mathcal{L}(\overline{F})$.

*This work was supported by the Grant Agency of the Czech Academy of sciences, project No. IAA1030405, the Grant Agency of the Czech Republic, project No. 201-06-P397, and the institutional research plan No. AV0Z10300504.

Unconstrained minimization of function $F(x)$ is equivalent to the nonlinear programming problem: Minimize the function

$$h(z_1, \dots, z_m)$$

with constraints

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

(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 this problem have the form

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

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

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

Nonlinear programming problem can be solved by using the primal interior point method. For this reason we apply the Newton minimization method to the sequence of barrier functions

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

assuming $0 < \mu \le \overline{\mu}$ and $\mu \to 0$, where $\varphi : (0, \infty) \to R$ is a barrier which satisfies the following conditions.

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.

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 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 proposed, for example barrier

$$\varphi(t) \;=\; \log(t^{-1} + 1), \qquad t \in (0, \infty),$$

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

$$\begin{aligned} \varphi(t) &= -\log t, & 0 < t \le 1, \\ \varphi(t) &= -(t^{-1} - 4\, t^{-1/2} + 3), & t > 1, \end{aligned}$$

which is bounded from below ($\underline{\varphi} = -3$). Both these barriers satisfy Condition 1 and Condition 2.

Iterative determination of the minimax vector

The necessary conditions for (x, z) to be the minimizer of the barrier function have the form

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

and

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

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 use the Newton method whose iteration step can be written in the form

$$\begin{bmatrix} W(x,z) & -A_1(x)v_1(x,z) & \dots & -A_m(x)v_m(x,z) \\ -v_1^T(x,z)A_1^T(x) & h_{11}(z) + e_1^T v_1(x,z) & \dots & h_{1m}(z) \\ \dots & \dots & \dots & \dots \\ -v_m^T(x,z)A_m^T(x) & h_{m1}(z) & \dots & h_{mm}(z) + e_m^T v_m(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z_1 \\ \dots \\ \Delta z_m \end{bmatrix}$$

$$= -\begin{bmatrix} \sum_{i=1}^m A_i(x)u_i(x,z) \\ h_1(z) - e_1^T u_1(x,z) \\ \dots \\ h_m(z) - e_m^T u_m(x,z) \end{bmatrix},$$

where

$$W(x, z) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \nabla^2 f_{ij}(x)u_{ij}(x, z) + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \nabla f_{ij}(x)v_{ij}(x, z)(\nabla f_{ij}(x))^T,$$

$$u_{ij}(x, z) = -\mu\varphi'(z_i - f_{ij}(x)), \quad v_{ij}(x, z) = \mu\varphi''(z_i - f_{ij}(x)),$$

$$h_{ij}(z) = \frac{\partial^2 h(z)}{\partial z_i \partial z_j}, \quad 1 \le i \le m, \quad 1 \le j \le n_i,$$

and where $A_i(x) = [\nabla f_{i1}(x), \dots, \nabla f_{in_i}(x)]$,

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

This formula can be easily verified by the differentiation of the KKT conditions by vectors x and z. Setting

$$C(x, z) = [A_1(x)v_1(x, z), \dots, A_m(x)v_m(x, z)],$$

$$g(x,z) = \sum_{i=1}^{m} A_i(x) u_i(x,z),$$

$$\Delta z = \begin{bmatrix} \Delta z_1 \\ \dots \\ \Delta z_m \end{bmatrix}, \quad c(x,z) = \begin{bmatrix} h_1(z) - e_1^T u_1(x,z) \\ \dots \\ h_m(z) - e_m^T u_m(x,z) \end{bmatrix},$$

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

we can rewrite the Newton system in the form

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

Now, let us have a problem, which is large-scale (the number of variables n is large), but partially separable (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. If m is small (for example in the classic minimax problems, where $m = 1$), we use the fact that

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

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

The solution is determined from the formulas

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

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

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). Denoting $D = H(z) + V(x,z)$, the matrix

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

is sparse and 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

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

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

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 the Newton system, we determine direction vectors Δx, Δz and select a step-size α in such a way that

$$B_\mu(x + \alpha\Delta x, z + \alpha\Delta z) < B_\mu(x, z)$$

and $z_i^+ > F_i(x^+)$, $1 \le i \le m$. Finally, we set $x^+ = x + \alpha\Delta x$, $z^+ = z + \alpha\Delta z$ and determine a new value $\mu^+ < \mu$. The above inequality is satisfied for sufficiently small values of the step-size α, if the matrix of the Newton system is positive definite.

Theorem 1. *Let the matrix $G = \sum_{i=1}^m \sum_{j=1}^{n_i} \nabla^2 f_{ij} u_{ij}$ be positive definite. Then the matrix of the Newton system is positive definite.*

Direct determination of the minimax vector

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

$$z(x) = \arg\min_{z \in Z} B_\mu(x, z),$$

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

where Z is the set used in Assumption 2. The first equation serves for the determination of the optimal vector $z(x) \in R^m$ 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 $\mu\varphi(z_i - f_{ij}(x))$, $1 \le i \le m$, $1 \le j \le n_i$. As a stationary point, its minimum is uniquely determined by the KKT conditions. The following theorem holds for the logarithmic barrier.

Theorem 2. *The system of equations*

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

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

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

where

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

and where $\underline{h}_i > 0$ are bounds used in Assumption 2 and $\overline{h_i} = h_i(\overline{z}_1, \ldots, \overline{z}_m)$.

Similar results can be obtained for other barriers as well. Using barrier

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

we get equations

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

and inequalities for $z_i(x;\mu)$ with bound

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

The system of nonlinear equations 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, this system is decomposed on m scalar equations, which can be efficiently solved by robust methods. If we are able to find a solution of the nonlinear system 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. *One has*

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

$$\nabla^2 B(x;\mu) = W(x, z(x)) - C(x, z(x)) D(x, z(x))^{-1} C^T(x, z(x)),$$

where $W(x, z(x))$, $C(x, z(x))$, $H(z(x))$, $V(x, z(x))$ are matrices introduced in the previous section and $D(x, z(x)) = H(z(x)) + V(x, z(x))$. If the matrix $H(z(x))$ is diagonal, we can express the Hessian matrix in the form

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

where $A_i(x)$, $V_i(x, z(x))$, $1 \le i \le m$, and $G(x, z(x))$ are matrices introduced in the previous section.

To determine the Hessian matrix inverse, we can use the relation obtained by the decomposition of the Newton system described in the previous section. Using substitution $c(x, z(x)) = 0$ we get

$$
\begin{aligned}
(\nabla^2 B_\mu(x))^{-1} &= W(x,z(x))^{-1} - W(x,z(x))^{-1}C(x,z(x)) \\
&\quad \Big(C^T(x,z(x))\,W^{-1}(x,z(x))\,C(x,z(x)) - H(z(x)) - V(x,z(x))\Big)^{-1} \\
&\quad C^T(x,z(x))\,W(x,z(x))^{-1}.
\end{aligned}
$$

If the nonlinear system is not solved with a sufficient precision, we rather use the Newton system from the previous section, 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 the nonlinear system we determine the vector $z(x)$. Using the Hessian matrix or its inverse, we determine a direction vector Δx and select a step-size α in such a way that

$$B_\mu(x+\alpha\Delta x, z(x+\alpha\Delta x;\mu)) < B_\mu(x, z(x;\mu))$$

(the vector $z(x+\alpha\Delta x;\mu)$ is obtained as a solution of the nonlinear system, in which x is replaced by $x+\alpha\Delta x$). Finally, we set $x^+ = x+\alpha\Delta x$ and determine a new value $\mu^+ < \mu$. Conditions for the direction vector Δx to be descent are the same as in Theorem 1. It suffices when the matrix $G(x, z(x))$ is positive definite.

Now, we describe an algorithm, in which the direction vector $d = \Delta x$ is determined in such a way that

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

(uniform descent) where $g = A(x)u(x;\mu)$.

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 $\overline{\Delta} > 0$, way 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 $f_{ij}(x)$, $1 \leq i \leq m$, $1 \leq j \leq n_i$, $F_i(x) = \max_{1\leq j\leq n_i} f_{ij}(x)$, $1 \leq i \leq m$, and $F(x) = h(F_1(x), \ldots, F_m(x))$. Set $k := 0$ (iteration count) and $r := 0$ (restart indicator).

Step 2: *Termination.* Solving the nonlinear system with precision $\underline{\delta}$, 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.

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

Step 5: *Restart.* If $r = 0$ and the direction vector is not uniformly descent, select a positive definite diagonal matrix $\tilde{D}$, set $G = \tilde{D}$, $r := 1$ and go to Step 4. If $r = 1$ and the direction vector is not uniformly descent, 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 $B(x+\beta^l\overline{\alpha}d;\mu) \le B(x;\mu) + \varepsilon_1\beta^l\overline{\alpha}g^Td$ (the nonlinear system 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 $F(x) = h(F_1(x), \dots, F_m(x))$.

Step 7: *Barrier parameter update.* Determine a new value of the barrier parameter $\mu \ge \underline{\mu}$ by Procedure A or Procedure B. Go to Step 2.

Procedure A.

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

$$\mu_{k+1} = \max\left(\tilde{\mu}_{k+1},\ \underline{\mu}\right),$$

where

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

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

Procedure B.

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

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

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

Global convergence for bounded barriers

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. Proofs of all assertions are given in [4].

Lemma 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 $L \geq 0$ such that*

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

Lemma 2. *Let assumptions of Lemma 1 and Assumption 3 be satisfied. Then the values $\{\mu_k\}_1^\infty$, generated by Algorithm 1, form a non-increasing sequence such that $\mu_k \to 0$.*

Theorem 4. *Let assumptions of Lemma 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*

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

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

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

pro $1 \leq i \leq m$ a $1 \leq j \leq n_i$.

Corollary 1. *Let assumptions of Theorem 4 hold. Then every cluster point $x \in R^n$ of the sequence $\{x_k\}_1^\infty$ satisfies KKT conditions of the original problem, where z and u (with elements z_i and u_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n_i$) are cluster points of sequences $\{z(x_k;\mu_k)\}_1^\infty$ and $\{u(x_k;\mu_k)\}_1^\infty$.*

Theorem 5. *Consider the sequence $\{x_k\}_1^\infty$ generated by Algorithm 1. Let assumptions of Lemma 1 and Assumption 3 hold. Then, choosing $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, $\underline{\mu} > 0$ arbitrarily, there is an index $k \geq 1$ such that*

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

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

$$u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) \leq \overline{\mu}$$

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

Global convergence for the logarithmic barrier

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. Proofs of all assertions are given in [4].

Lemma 3. *Let Assumption 2, Assumption 4 be satisfied, $\varphi(t) = -\log t$ and 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*

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

Lemma 4. *Let assumptions of Lemma 3 and Assumption 3 be satisfied. Then the values $\{\mu_k\}_1^\infty$, generated by Algorithm 1, form a non-increasing sequence such that $\mu_k \to 0$.*

Theorem 6. *Let assumptions of Lemma 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*

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

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

for $1 \leq i \leq m$ a $1 \leq j \leq n_i$.

Corollary 2. *Let assumptions of Theorem 6 hold. Then every cluster point $x \in R^n$ of the sequence $\{x_k\}_1^\infty$ satisfies KKT conditions of the original problem, where z and u (with elements z_i and u_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n_i$) are cluster points of sequences $\{z(x_k;\mu_k)\}_1^\infty$ and $\{u(x_k;\mu_k)\}_1^\infty$.*

Theorem 7. *Consider the sequence $\{x_k\}_1^\infty$ generated by Algorithm 1. Let assumptions of Lemma 3 and Assumption 3 hold. Then, choosing $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, $\underline{\mu} > 0$ arbitrarily, there is an index $k \geq 1$ such that*

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

$$u_{ij}(x_k;\mu_k) \geq 0, \quad z_i(x_k;\mu_k) - f_{ij}(x_k) \geq 0,$$
$$u_{ij}(x_k;\mu_k)(z_i(x_k;\mu_k) - f_{ij}(x_k)) \leq \overline{\mu}$$

for $1 \leq i \leq m$ a $1 \leq j \leq n_i$.

Special cases and numerical results

The simplest generalized minimax function is the sum

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

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. The nonlinear system decomposes on m scalar equations

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

whose solutions lie in the intervals

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

If $m = 1$, we obtain the classic minimax problems. Numerical experiments for minimax functions were carried out using a collection of 22 test problems (Test 14) described in [5]. The source texts can be downloaded from `http://www.cs.cas.cz/luksan/test.html`. Compared methods: P1–the logarithmic barrier, P2–positive barrier, P3–bounded barrier, SM–smoothing method, DI–primal-dual method. The results:

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

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

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

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

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

Numerical experiments for sums of absolute values were carried out using a collection of 22 test problems (Test 14) described in [5]. The source texts can be downloaded from `http://www.cs.cas.cz/luksan/test.html`. Compared methods: PT–logarithmic barrier and a trust-region realization, PL–logarithmic barrier and a line-search realization, DI–primal-dual method, BM–bundle variable metric method. The results:

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

The above tables demonstrate the high efficiency of Algorithm 1. The use of a minimax structure together with the two-level optimization give much better results than the use of standard nonlinear programming methods applied to the equivalent nonlinear programming problem.

References

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

[2] L. Lukšan, C. Matonoha, J. Vlček: *Primal interior-point method for large sparse minimax optimization.* Technical Report V-941, Inst. Computer Science, Acad. Sci., Czech Rep., 2005.

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

[4] L. Lukšan, C. Matonoha, J. Vlček: *Primal interior point method for minimization of generalized minimax functions.* Technical Report V-1017, Inst. Computer Science, Acad. Sci., Czech Rep., 2007.

[5] L. Lukšan, J. Vlček: *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, Report V-767, Prague, Inst. Computer Science, Acad. Sci., Czech Rep., 1998.