Institute of Computer Science
Academy of Sciences of the Czech Republic

Numerical solution of generalized minimax problems

L.Lukšan, C.Matonoha, J.Vlček

Technical report No. 1255

January 2018

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

Numerical solution of generalized minimax problems

L.Lukšan, C.Matonoha, J.Vlček

Technical report No. 1255

January 2018

Abstract:

This contribution contains the description and investigation of four numerical methods for solving generalized minimax problems, which consists in the minimization of functions which are compositions of special smooth convex functions with maxima of smooth functions (the most important problem of this type is the sum of maxima of smooth functions). Section 1 is introductory. In Section 2, we study recursive quadratic programming methods. This section also contains the description of the dual method for solving corresponding quadratic programming problems. Section 3 is devoted to primal interior points methods which use solutions of nonlinear equations for obtaining minimax vectors. Section 4 contains investigation of smoothing methods, based on using exponential smoothing terms. Section 5 contains a short description of primal-dual interior point methods based on transformation of generalized minimax problems to general nonlinear programming problems. Finally the last section contains results of numerical experiments.

Keywords:
Numerical optimization, nonlinear approximation, nonsmooth optimization, generalized minimax problems, recursive quadratic programming methods, interior point methods, smoothing methods, algorithms, numerical experiments.

Content

1 Generalized minimax problems

In many practical problems we need to minimize functions that contain absolute values or pointwise maxima of smooth functions. Such functions are nonsmooth but they often have a special structure enabling the use of special methods that are more efficient than methods for minimization of general nonsmooth functions. The classical minimax problem, where $F(x) = \max_{1\le k\le m} f_k(x)$, or problems where the function to be minimized is a nonsmooth norm, e.g. $F(x) = \|f(x)\|_\infty$, $F(x) = \|f_+(x)\|_\infty$, $F(x) = \|f(x)\|_1$, $F(x) = \|f_+(x)\|_1$ with $f(x) = [f_1(x), \dots, f_m(x)]^T$ and $f_+(x) = [\max(f_1(x), 0), \dots, \max(f_m(x), 0)]^T$, are typical examples. Such functions can be considered as special cases of more general functions, so it is possible to formulate more general theories and construct more general numerical methods. One possibility for generalization of the classical minimax problem consists in the use of the function

$$F(x) = \max_{1\le k\le \overline{k}} p_k^T f(x), \tag{1}$$

where $p_k \in R^m$, $1 \le k \le \overline{k}$, and $f : R^n \to R^m$ is a smooth mapping. This function is a special case of composite nonsmooth functions of the form $F(x) = f_0(x) + \max_{1\le k\le \overline{k}}(p_k^T f(x) + b_k)$, where $f_0 : R^n \to R$ is a continuously differentiable function [9, Section 14.1].

Remark 1. *We can express all above mentioned minimax problems and nonsmooth norms in form (1).*

(a) Setting $p_k = e_k$, where e_k is the k-th column of a unit matrix and $\overline{k} = m$, we obtain $F(x) = \max_{1\le k\le m} f_k(x)$ (the classical minimax).

(b) Setting $p_k = e_k$, $p_{m+k} = -e_k$ and $\overline{k} = 2m$, we obtain $F(x) = \max_{1\le k\le m} \max(f_k(x), -f_k(x)) = \|f(x)\|_\infty$.

(c) Setting $p_k = e_k$, $p_{m+1} = 0$ and $\overline{k} = m + 1$, we obtain $F(x) = \max(\max_{1\le k\le m} f_k(x), 0) = \|f(x)_+\|_\infty$.

(d) If $\overline{k} = 2^m$ and p_k, $1 \le k \le 2^m$, are mutually different vectors whose elements are either 1 or -1, we can write $F(x) = \sum_{k=1}^m \max(f_k(x), -f_k(x)) = \|f(x)\|_1$.

(e) If $\overline{k} = 2^m$ and p_k, $1 \le k \le 2^m$, are mutually different vectors whose elements are either 1 or 0, we can write $F(x) = \sum_{k=1}^m \max(f_k(x), 0) = \|f_+(x)\|_1$.

Remark 2. *Since the mapping $f(x)$ is continuously differentiable, the function (1) is Lipschitz. Thus, if the point $x \in R^n$ is a local minimum of $F(x)$, then $0 \in \partial F(x)$ [31, Theorem 3.2.5] holds. According to [31, Theorem 3.2.13], one has*

$$\partial F(x) = (\nabla f(x))^T \mathrm{conv}\left\{p_k : k \in \bar{I}(x)\right\},$$

where $\bar{I}(x) = \{k \in \{1, \dots, \overline{k}\} : p_k^T f(x) = F(x)\}$. Thus, if the point $x \in R^n$ is a local minimum of $F(x)$, then multipliers $\lambda_k \ge 0$, $1 \le k \le \overline{k}$, exist, such that $\lambda_k(p_k^T f(x) - F(x)) = 0$, $1 \le k \le \overline{k}$,

$$\sum_{k=1}^{\overline{k}} \lambda_k = 1 \quad \textit{and} \quad \sum_{k=1}^{\overline{k}} \lambda_k J(x)^T p_k = 0,$$

where $J(x)$ is a Jacobian matrix of the mapping $f(x)$.

Remark 3. *It is clear that a minimum of function (1) is a solution of a nonlinear programming problem consisting in minimization of a function $\tilde{F} : R^{n+1} \to R$, where $\tilde{F}(x, z) = z$, on the set*

$$\mathcal{C} = \{(x, z) \in R^{n+1} : p_k^T f(x) \le z,\ 1 \le k \le \overline{k}\}.$$

Obviously, $a_k = \nabla c_k(x,z) = (p_k^T J(x), -1)$, $1 \le k \le \overline{k}$, *and* $g_k = \nabla \tilde{F}(x,z) = (0,1)$, *so the necessary KKT conditions can be written in the form*

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \sum_{k=1}^{\overline{k}} \begin{bmatrix} J^T(x)p_k \\ -1 \end{bmatrix} \lambda_k = 0,$$

$\lambda_k(p_k^T f(x) - z) = 0$, *where* $\lambda_k \ge 0$ *are the Lagrange multipliers and* $z = F(x)$. *Thus, we obtain the same necessary conditions for an extremum as in Remark 2.*

From the examples given in Remark 1 it follows that composite nondifferentiable functions are not suitable for representation of functions $F(x) = \|f(x)\|_1$ and $F(x) = \|f_+(x)\|_1$ because in this case the expression on the right-hand side of (1) contains 2^m elements with vectors p_k, $1 \le k \le 2^m$. In the subsequent considerations, we will choose a somewhat different approach. We will consider generalized minimax functions established in [6] and [26].

Definition 1. *We say that* $F : R^n \to R$ *is a generalized minimax function if*

$$F(x) = h(F_1(x), \dots, F_m(x)), \quad F_k(x) = \max_{1 \le l \le m_k} f_{kl}(x), \quad 1 \le k \le m, \tag{2}$$

where $h : R^m \to R$ *and* $f_{kl} : R^n \to R$, $1 \le k \le m$, $1 \le l \le m_k$, *are smooth functions satisfying the following assumptions.*

Assumption X1a. *Functions* f_{kl}, $1 \le k \le m$, $1 \le l \le m_k$, *are bounded from below on* R^n, *so that there exists a constant* $\underline{F} \in R$ *such that* $f_{kl}(x) \ge \underline{F}$, $1 \le k \le m$, $1 \le l \le m_k$, *for all* $x \in R^n$.

Assumption X1b. *Functions* F_k, $1 \le k \le m$, *are bounded from below on* R^n, *so that there exist constants* $\underline{F}_k \in R$ *such that* $F_k(x) \ge \underline{F}_k$, $1 \le k \le m$, *for all* $x \in R^n$.

Assumption X2. *The function* h *is twice continuously differentiable and convex satisfying*

$$0 < \underline{h}_k \le \partial h(z)/\partial z_k \le \overline{h}_k, \qquad 1 \le k \le m, \tag{3}$$

for every $z \in Z = \{z \in R^m : z_k \ge \underline{F}_k,\ 1 \le k \le m\}$ *(vector* $z \in R^m$ *is called the minimax vector).*

Assumption X3. *Functions* $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, *are twice continuously differentiable on the convex hull of the level set*

$$\mathcal{D}_F(\overline{F}) = \{x \in R^n : F_k(x) \le \overline{F},\ 1 \le k \le m\}$$

for a sufficiently large upper bound $\overline{F}$ *and subsequently, constants* $\overline{g}$ *and* $\overline{G}$ *exist such that* $\|g_{kl}(x)\| \le \overline{g}$ *and* $\|G_{kl}(x)\| \le \overline{G}$ *for all* $1 \le k \le m$, $1 \le l \le m_k$, *and* $x \in \mathrm{conv}\mathcal{D}_F(\overline{F})$, *where* $g_{kl}(x) = \nabla f_{kl}(x)$ *and* $G_{kl}(x) = \nabla^2 f_{kl}(x)$.

Remark 4. *The conditions imposed on the function* $h(z)$ *are relatively strong but many important nonsmooth functions satisfy them.*

(1) Let $h : R \to R$ *be an identity mapping, so* $h(z) = z$ *and* $h'(z) = 1 > 0$. *Then setting* $\overline{k} = 1$, $m_1 = \overline{l}$ *and* $F(x) = h(F_1(x)) = F_1(x) = \max_{1 \le l \le \overline{l}} p_l^T f(x)$ *($f_{1l} = p_l^T f(x)$), we obtain composite nonsmooth function (1) and therefore functions* $F(x) = \max_{1 \le k \le m} f_k(x)$, $F(x) = \|f(x)\|_\infty$, $F(x) = \|f_+(x)\|_\infty$.

(2) Let $h : R^m \to R$, *where* $h(z) = z_1 + \dots + z_m$, *so* $\partial h(z)/\partial z_k = 1 > 0$, $1 \le k \le m$. *Then function (2) has the form*

$$F(x) = \sum_{k=1}^{m} F_k(x) = \sum_{k=1}^{m} \max_{1 \le l \le m_k} f_{kl}(x) \tag{4}$$

(the sum of maxima). If $m_k = 2$ *and* $F_k(x) = \max(f_k(x), -f_k(x))$, *we obtain the function* $F(x) = \|f(x)\|_1$. *If* $m_k = 2$ *and* $F_k(x) = \max(f_k(x), 0)$, *we obtain the function* $F(x) = \|f_+(x)\|_1$. *It follows that the expression of functions* $F(x) = \|f(x)\|_1$ *and* $F(x) = \|f_+(x)\|_1$ *by (2) contains only* m *summands and each summand is a maximum of two function values. Thus, this approach is much more economic than the use of formulas stated in Remark 1 (d)-(e).*

Remark 5. *Since the functions $F_k(x)$, $1 \le k \le m$, are regular [31, Theorem 3.2.13], the function $h(z)$ is continuously differentiable, and $h_k = \partial h(z)/\partial z_k > 0$, one can write [31, Theorem 3.2.9]*

$$\partial F(x) = \operatorname{conv} \sum_{k=1}^{m} h_k \partial F_k(x) = \sum_{k=1}^{m} h_k \partial F_k(x) = \sum_{k=1}^{m} h_k \operatorname{conv}\{g_{kl} : l \in \bar{I}_k(x)\},$$

where $\bar{I}_k(x) = \{l : 1 \le l \le m_k, f_{kl}(x) = F_k(x)\}$. Thus, one has

$$\partial F(x) = \sum_{k=1}^{m} h_k \sum_{l=1}^{m_k} \lambda_{kl} g_{kl},$$

where for $1 \le k \le m$ it holds $\lambda_{kl} \ge 0$, $\lambda_{kl}(F_k(x) - f_{kl}(x)) = 0$, $1 \le l \le m_k$, and $\sum_{l=1}^{m_k} \lambda_{kl} = 1$. Setting $u_{kl} = h_k \lambda_{kl}$, $1 \le k \le m$, $1 \le l \le m_k$, we can write

$$\partial F(x) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} u_{kl} g_{kl},$$

where for $1 \le k \le m$ it holds $u_{kl} \ge 0$, $u_{kl}(F_k(x) - f_{kl}(x)) = 0$, $1 \le l \le m_k$, and $\sum_{l=1}^{m_k} u_{kl} = h_k$. If a point $x \in R^n$ is a minimum of a function $F(x)$, then $0 \in \partial F(x)$, so there exist multipliers u_{kl}, $1 \le k \le m$, $1 \le l \le m_k$, such that

$$\sum_{k=1}^{m} \sum_{l=1}^{m_k} g_{kl}(x) u_{kl} = 0, \quad \sum_{l=1}^{m_k} u_{kl} = h_k, \quad h_k = \frac{\partial h(z)}{\partial z_k}, \quad 1 \le k \le m, \tag{5}$$

$$u_{kl} \ge 0, \quad F_k - f_{kl}(x) \ge 0, \quad u_{kl}(F_k - f_{kl}(x)) = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \tag{6}$$

Remark 6. *Unconstrained minimization of function (2) is equivalent to the nonlinear programming problem*

$$\text{minimize} \quad \tilde{F}(x, z) = h(z) \quad \text{subject to} \quad f_{kl}(x) \le z_k, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \tag{7}$$

Condition (3) is sufficient for satisfying equalities $z_k = F_k(x)$, $1 \le k \le m$, at the minimum point. Denote $a_{kl}(x, z)$ gradients of functions $c_{kl}(x, z) = f_{kl}(x) - z_k$. Obviously, $a_{kl}(x, z) = (g_{kl}(x), -e_k)$, $1 \le k \le m$, $1 \le l \le m_k$, where $g_{kl}(x)$ is a gradient of $f_{kl}(x)$ in x and e_k is the k-th column of a unit matrix of order m. Thus, the necessary first-order (KKT) conditions have the form

$$g(x, u) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} g_{kl}(x) u_{kl} = 0, \quad \sum_{l=1}^{m_k} u_{kl} = h_k, \quad h_k = \frac{\partial h(z)}{\partial z_k}, \quad 1 \le k \le m, \tag{8}$$

$$u_{kl} \ge 0, \quad z_k - f_{kl}(x) \ge 0, \quad u_{kl}(z_k - f_{kl}(x)) = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k, \tag{9}$$

where u_{kl} are Lagrange multipliers and $z_k = F_k(x)$. So we obtain the same necessary conditions for an extremum as in Remark 5.

Remark 7. *A classical minimax problem*

$$F(x) = \max_{1 \le k \le m} f_k(x), \tag{10}$$

can be replaced with an equivalent nonlinear programming problem

$$\text{minimize} \quad \tilde{F}(x, z) = z \quad \text{subject to} \quad f_k(x) \le z, \quad 1 \le k \le m, \tag{11}$$

and necessary KKT conditions have the form

$$\sum_{k=1}^{m} g_k(x) u_k = 0, \qquad \sum_{k=1}^{m} u_k = 1, \tag{12}$$

$$u_k \ge 0, \quad z - f_k(x) \ge 0, \quad u_k(z - f_k(x)) = 0, \quad 1 \le k \le m. \tag{13}$$

Remark 8. *Minimization of the sum of absolute values*

$$F(x) = \sum_{k=1}^{m} |f_k(x)| = \sum_{k=1}^{m} \max(f_k^+(x), f_k^-(x)), \qquad f_k^+(x) = f_k(x), \qquad f_k^-(x) = -f_k(x) \tag{14}$$

can be replaced with an equivalent nonlinear programming problem

$$\text{minimize} \quad \tilde{F}(x,z) = \sum_{k=1}^{m} z_k \quad \text{subject to} \quad -z_k \le f_k(x) \le z_k \tag{15}$$

(there are two constraints $c_k^-(x) = z_k - f_k(x) \ge 0$ and $c_k^+(x) = z_k + f_k(x) \ge 0$ for each index $1 \le k \le m$) and necessary KKT conditions have the form

$$\sum_{k=1}^{m} g_k(x)(u_k^+ - u_k^-) = 0, \quad u_k^+ + u_k^- = 1, \quad 1 \le k \le m, \tag{16}$$

$$u_k^+ \ge 0, \quad z_k - f_k(x) \ge 0, \quad u_k^+(z_k - f_k(x)) = 0, \quad 1 \le k \le m, \tag{17}$$

$$u_k^- \ge 0, \quad z_k + f_k(x) \ge 0, \quad u_k^-(z_k + f_k(x)) = 0, \quad 1 \le k \le m. \tag{18}$$

If we set $u_k = u_k^+ - u_k^-$ and use the equality $u_k^+ + u_k^- = 1$, we obtain $u_k^+ = (1+u_k)/2$, $u_k^- = (1-u_k)/2$. From conditions $u_k^+ \ge 0$, $u_k^- \ge 0$ the inequalities $-1 \le u_k \le 1$, or $|u_k| \le 1$, follow. The condition $u_k^+ + u_k^- = 1$ implies that the numbers u_k^+, u_k^- cannot be simultaneously zero, so either $z_k = f_k(x)$ or $z_k = -f_k(x)$, that is $z_k = |f_k(x)|$. If $f_k(x) \ne 0$, it cannot simultaneously hold $z_k = f_k(x)$ and $z_k = -f_k(x)$, so the numbers u_k^+, u_k^- cannot be simultaneously nonzero. Then either $u_k = u_k^+ = 1$ and $z_k = f_k(x)$ or $u_k = -u_k^- = -1$ and $z_k = -f_k(x)$, that is $u_k = f_k(x)/|f_k(x)|$. Thus, the necessary KKT conditions have the form

$$\sum_{k=1}^{m} g_k(x) u_k = 0, \quad z_k = |f_k(x)|, \quad |u_k| \le 1, \quad \text{and} \quad u_k = \frac{f_k(x)}{|f_k(x)|}, \quad \text{if} \quad |f_k(x)| > 0. \tag{19}$$

Remark 9. *Minimization of the sum of absolute values can also be reformulated so that more slack variables are used. We obtain the problem*

$$\text{minimize} \quad \tilde{F}(x,z) = \sum_{k=1}^{m} (z_k^+ + z_k^-) \quad \text{subject to} \quad f_k(x) = z_k^+ - z_k^-, \quad z_k^+ \ge 0, \quad z_k^- \ge 0, \tag{20}$$

where $1 \le k \le m$. This problem contains m general equality constraints and $2m$ simple bounds for $2m$ slack variables.

In the subsequent considerations, we will restrict ourselves to functions of the form (4), the sums of maxima that include most cases important for applications. In this case, it holds

$$h(z) = \sum_{k=1}^{m} z_k, \qquad \nabla h(z) = \tilde{e}, \qquad \nabla^2 h(z) = 0, \tag{21}$$

where $\tilde{e} \in R^m$ is a vector with unit elements. The case when $h(z)$ is a general function satisfying Assumption X2 is studied in [26]. For simplicity, we will often use the notation $\mathrm{vec}(a,b)$ instead of $[a^T, b^T]^T \in R^{n+m}$.

2 Recursive quadratic programming methods

2.1 Basic properties

Suppose the function $h(z)$ is of form (21). In this case the necessary KKT conditions are of form (8)–(9), where $\partial h(z)/\partial z_k = 1$, $1 \le k \le m$. If we linearize these conditions in a neighborhood of a point $x \in R^n$, we can write for $d \in R^n$

$$\sum_{k=1}^{m}\sum_{l=1}^{m_k}(g_{kl}(x) + G_{kl}(x)d)u_{kl} = 0, \quad \sum_{l=1}^{m_k} u_{kl} = 1, \quad 1 \le k \le m,$$

$$u_{kl} \ge 0, \quad f_{kl}(x) + g_{kl}^T(x)d - z_k \le 0, \quad u_{kl}(f_{kl}(x) + g_{kl}^T(x)d - z_k) = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k.$$

But these are the necessary KKT conditions for solving a quadratic programming problem: minimize a quadratic function

$$Q(d,z) = \sum_{k=1}^{m} z_k + \frac{1}{2}d^T G d, \qquad G = \sum_{k=1}^{m}\sum_{l=1}^{m_k} G_{kl}(x)u_{kl} \tag{22}$$

on the set

$$\mathcal{C} = \{(d,z) \in R^{n+m} : f_{kl}(x) + g_{kl}^T(x)d \le z_k, \quad 1 \le k \le m, \quad 1 \le l \le m_k\}. \tag{23}$$

Note that coefficients u_{kl}, $1 \le k \le m$, $1 \le l \le m_k$, in (22) are old Lagrange multipliers. New Lagrange multipliers along with new values of variables z_k, $1 \le k \le m$, are determined by solving quadratic programming problem (22)–(23).

For simplification, we will omit the argument x and use the notation

$$f_k = \begin{bmatrix} f_{k1}(x) \\ \dots \\ f_{km_k}(x) \end{bmatrix}, \qquad u_k = \begin{bmatrix} u_{k1} \\ \dots \\ u_{km_k} \end{bmatrix}, \qquad \tilde{e}_k = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix},$$

$A_k = [g_{k1}(x), \dots, g_{km_k}(x)]$. Problem (22)–(23) will be written in the form

$$\text{minimize} \quad Q(d,z) = \sum_{k=1}^{m} z_k + \frac{1}{2}d^T G d \quad \text{subject to} \quad f_k + A_k^T d \le z_k \tilde{e}_k, \quad 1 \le k \le m, \tag{24}$$

from where the necessary KKT conditions

$$Gd + \sum_{k=1}^{m} A_k u_k = 0, \quad \tilde{e}_k^T u_k = 1, \quad 1 \le k \le m, \tag{25}$$

$$u_k \ge 0, \quad f_k + A_k^T d - z_k\tilde{e}_k \le 0, \quad u_k^T(f_k + A_k^T d - z_k\tilde{e}_k) = 0, \quad 1 \le k \le m, \tag{26}$$

follow. Note that from (25)–(26) we have

$$z_k = u_k^T(f_k + A_k^T d), \quad 1 \le k \le m. \tag{27}$$

Quadratic programming problem (24) is convex, so there exists a dual problem stated in [32, Theorem 12.14]. We will use the notation

$$f = \begin{bmatrix} f_1 \\ \dots \\ f_m \end{bmatrix}, \qquad u = \begin{bmatrix} u_1 \\ \dots \\ u_m \end{bmatrix}, \qquad v = \begin{bmatrix} v_1 \\ \dots \\ v_m \end{bmatrix}, \qquad w = \begin{bmatrix} w_1 \\ \dots \\ w_m \end{bmatrix}, \qquad z = \begin{bmatrix} z_1 \\ \dots \\ z_m \end{bmatrix},$$

$A = [A_1, \dots, A_m]$, during deriving a dual quadratic programming problem. Obviously, $f \in R^{\overline{m}}$, $u \in R^{\overline{m}}$, $v \in R^{\overline{m}}$, and $w \in R^m$, $z \in R^m$, where $\overline{m} = \sum_{k=1}^{m} m_k$. The following theorem states the dual problem for special quadratic programming problem (24).

Theorem 1. *Consider quadratic programming problem (24) with positive definite matrix G (so this problem is convex). Then the dual problem can be written in the form*

$$\text{minimize} \quad \tilde{Q}(u) = \frac{1}{2}u^T A^T H A u - f^T u \quad \text{subject to} \quad u_k \geq 0, \quad \tilde{e}_k^T u_k = 1, \quad 1 \leq k \leq m, \tag{28}$$

where $H = G^{-1}$. Problem (28) is convex as well and the dual problem to this problem is primal problem (24). If the pair $(\mathrm{vec}(d, z), u)$ is a KKT pair of the primal problem, then the pair $(u, \mathrm{vec}(v, w))$, where $v_k = -(A_k^T d + f_k - z_k \tilde{e}_k)$ and $w_k = z_k$, $1 \leq k \leq m$, is a KKT pair of the dual problem. If the pair $(u, \mathrm{vec}(v, w))$ is a KKT pair of the dual problem, then the pair $(\mathrm{vec}(d, z), u)$, where $d = -HAu$ and $z_k = w_k$, $1 \leq k \leq m$, is a KKT pair of the primal problem and $A_k^T d + f_k - z_k \tilde{e}_k = -v_k$, $1 \leq k \leq m$.

Proof. The Lagrange function of problem (24) has the form

$$L(d, z, u) = \sum_{k=1}^{m} z_k + \frac{1}{2} d^T G d + \sum_{k=1}^{m} u_k^T (f_k + A_k^T d - z_k \tilde{e}_k) \tag{29}$$

and its gradient is

$$g(d, z, u) = \begin{bmatrix} Gd + \sum_{k=1}^{m} A_k u_k \\ 1 - \tilde{e}_1^T u_1 \\ \cdots \\ 1 - \tilde{e}_m^T u_m \end{bmatrix}. \tag{30}$$

By [32, Theorem 12.14], the dual problem consists in maximizing the Lagrange function $L(d, z, u)$ on the set of constraints $u \geq 0$ and $g(d, z, u) = 0$. Substituting (30) into the equation $g(d, z, u) = 0$, we obtain

$$d = -H \sum_{k=1}^{m} A_k u_k = -HAu, \qquad \tilde{e}_k^T u_k = 1, \quad 1 \leq k \leq m, \tag{31}$$

which after substituting into (29) gives

$$\begin{aligned} L(d, z, u) &= \sum_{k=1}^{m} z_k + \frac{1}{2} d^T G d + u^T (f + A^T d) - \sum_{k=1}^{m} z_k = \frac{1}{2} u^T A^T H A u + u^T f - u^T A^T H A u \\ &= -\frac{1}{2} u^T A^T H A u + f^T u, \end{aligned}$$

so maximization of the Lagrange function $L(d, z, u)$ is equivalent to minimization of the function $\tilde{Q}(u)$. Since the matrix H is positive definite, problem (28) is convex and we can set up the dual problem consisting in maximizing the Lagrange function

$$\tilde{L}(u, v, w) = \frac{1}{2} u^T A^T H A u - f^T u - \sum_{k=1}^{m} v_k^T u_k + \sum_{k=1}^{m} w_k (\tilde{e}_k^T u_k - 1) \tag{32}$$

on the set of constraints $\tilde{g}(u, v, w) = 0$. That is,

$$A_k^T H A u - f_k - v_k + w_k \tilde{e}_k = 0, \quad 1 \leq k \leq m, \tag{33}$$

and

$$u_k \geq 0, \quad \tilde{e}_k^T u_k = 1, \quad v_k \geq 0, \quad v_k^T u_k = 0, \quad 1 \leq k \leq m.$$

If we set $d = -HAu$, and substitute this relation into (33), we obtain

$$-v_k = f_k + A_k^T d - w_k \tilde{e}_k, \quad 1 \leq k \leq m, \tag{34}$$

which along with $u_k^T v_k = 0$ and $u_k^T \tilde{e}_k = 1$ gives $w_k = u_k^T(f_k + A_k^T d)$. Thus, it holds $w_k = z_k$, $1 \le k \le m$, by (27). If we substitute these equalities along with $d = -HAu$ into (32), we can write

$$\begin{aligned}\tilde{L}(u,v,w) &= \frac{1}{2}d^T G d - \sum_{k=1}^{m} f_k^T u_k + \sum_{k=1}^{m}(f_k + A_k^T d - z_k\tilde{e}_k)^T u_k = \frac{1}{2}d^T G d + \sum_{k=1}^{m} u_k^T A_k^T d - \sum_{k=1}^{m} z_k \\ &= \frac{1}{2}d^T G d - d^T G d - \sum_{k=1}^{m} z_k = -\left(\sum_{k=1}^{m} z_k + \frac{1}{2}d^T G d\right),\end{aligned}$$

so maximization of the Lagrange function $\tilde{L}(u,v,w)$ is equivalent to minimization of the function $Q(d,z)$. □

Remark 10. *Note that by (28) and (31), it holds that*

$$\tilde{Q}(u) = \frac{1}{2}d^T G d - f^T u, \tag{35}$$

so

$$Q(d,z) - \tilde{Q}(u) = \sum_{k=1}^{m} z_k + f^T u. \tag{36}$$

The following theorem, which is a generalization of a similar theorem given in [15], shows that the solution of quadratic programming problem (22)–(23) is a descent direction for the objective function $F(x)$.

Theorem 2. *Let Assumption X3 be satisfied and let vectors $d \in R^n$, $z \in R^m$ be a solution of quadratic programming problem (22)–(23) with the positive definite matrix G and a corresponding vector of Lagrange multipliers $u \in R^{\overline{m}}$. If $d = 0$, then the pair $(\mathrm{vec}(x,z), u)$ is the KKT pair of problem (7). If $d \neq 0$, then $F'(x,d) = d^T g(x,u) < 0$, where $F'(x,d)$ is a directional derivative of function (4) along a vector d at a point x and $g(x,u)$ is a vector given by (8). If $\kappa(G) \le 1/\varepsilon_0^2$, where $\kappa(G)$ is a spectral condition number of G, then $d^T g(x,u) \le -\varepsilon_0 \|d\| \|g(x,u)\|$ and for an arbitrary number $0 < \varepsilon_1 < 1/2$ there exists a steplength $0 < \overline{\alpha} \le 1$ such that*

$$F(x + \alpha d) - F(x) \le \varepsilon_1 \alpha d^T g(x,u) \tag{37}$$

if $0 < \alpha \le \overline{\alpha}$.

Proof.

(a) If $d = 0$, then conditions (25)–(26) are equivalent to conditions (28)–(29). Thus, if a pair $(\mathrm{vec}(0,z), u)$ is the KKT pair of problem (24), then a pair $(\mathrm{vec}(x,z), u)$ is the KKT pair of problem (7).

(b) Function (4) is a sum of maxima of differentiable functions, so it is regular by [31, Theorem 3.2.13] and there exists a directional derivative

$$F'(x,d) = \lim_{\alpha \downarrow 0} \frac{F(x+\alpha d) - F(x)}{\alpha} = \sum_{k=1}^{m} \lim_{\alpha \downarrow 0} \frac{F_k(x+\alpha d) - F_k(x)}{\alpha}.$$

Let $0 < \alpha \le 1$ and l_k be indices such that $f_{kl_k}(x + \alpha d) = F_k(x + \alpha d)$, $1 \le k \le m$. Then by Assumption X3 it holds that

$$f_{kl_k}(x + \alpha d) \le f_{kl_k}(x) + \alpha g_{kl_k}^T(x) d + \frac{1}{2}\alpha^2 \overline{G} \|d\|^2, \quad 1 \le k \le m.$$

Using inequality $0 < \alpha \le 1$ and relations (25)–(26), we obtain

$$\begin{aligned} f_{kl_k} + \alpha\, g_{kl_k}^T d &\le f_{kl_k} + \alpha(z_k - f_{kl_k}) = \alpha z_k + (1-\alpha) f_{kl_k} \le \alpha z_k + (1-\alpha) F_k \\ &= F_k + \alpha(z_k - F_k) = F_k + \alpha\, u_k^T(z_k \tilde{e}_k - F_k \tilde{e}_k) \le F_k + \alpha\, u_k^T(z_k \tilde{e}_k - f_k) \\ &= F_k + \alpha\, u_k^T(z_k \tilde{e}_k - f_k - A_k^T d) + \alpha\, u_k^T A_k^T d = F_k + \alpha\, u_k^T A_k^T d. \end{aligned}$$

Thus, we can write

$$\frac{F_k(x+\alpha d)-F_k(x)}{\alpha}=\frac{f_{kl_k}(x+\alpha d)-F_k(x)}{\alpha}\le d^TA_ku_k+\frac{1}{2}\alpha\overline{G}\|d\|^2, \tag{38}$$

so

$$F'(x,d)=\sum_{k=1}^m\lim_{\alpha\downarrow 0}\frac{F_k(x+\alpha d)-F_k(x)}{\alpha}\le\sum_{k=1}^m d^TA_ku_k=d^TAu=d^Tg(x,u).$$

Since $Gd=-Au=-g(x,u)$, see (25), and matrix G is positive definite, we have $F'(x,d)=d^Tg(x,u)=-d^TGd<0$.

(c) If $\kappa(G)\le 1/\varepsilon_0^2$, then $d^Tg(x,u)\le-\varepsilon_0\|d\|\|g(x,u)\|$, see [32, Section 3.2]. Since $d=-G^{-1}g(x,u)$, it holds $\|d\|\le\|g(x,u)\|/\underline{G}$, which along with the previous inequality gives

$$\|d\|^2\le\frac{1}{\underline{G}}\|d\|\|g(x,u)\|\le-\frac{1}{\varepsilon_0\underline{G}}d^Tg(x,u).$$

Using (38) we obtain

$$\begin{aligned}\frac{F(x+\alpha d)-F(x)}{\alpha} &= \sum_{k=1}^m\frac{F_k(x+\alpha d)-F_k(x)}{\alpha}=\sum_{k=1}^m\left(d^TA_ku_k+\frac{1}{2}\alpha\overline{G}\|d\|^2\right)\\ &= d^Tg(x,u)+\frac{m}{2}\alpha\overline{G}\|d\|^2\le\left(1-\alpha\frac{m\overline{G}}{2\varepsilon_0\underline{G}}\right)d^Tg(x,u),\end{aligned}$$

so (37) holds if

$$1-\alpha\frac{m\overline{G}}{2\varepsilon_0\underline{G}}\ge\varepsilon_1\quad\Rightarrow\quad\alpha\le\frac{2\varepsilon_0(1-\varepsilon_1)\underline{G}}{m\overline{G}}\triangleq\overline{\alpha}.$$

□

Remark 11. *A number $0<\alpha\le 1$ satisfying (37) can be determined using the Armijo steplength selection [32, Section 3.1]. Then α is a first term meeting (37) in the sequence α_j, $j\in N$, such that $\alpha_1=1$ and $\underline{\beta}\alpha_j\le\alpha_{j+1}\le\overline{\beta}\alpha_j$, where $0<\underline{\beta}\le\overline{\beta}<1$. At most $\mathrm{int}(\log\overline{\alpha}/\log\overline{\beta}+1)$ steps is used, where $\mathrm{int}(t)$ is the largest integer such that $\mathrm{int}(t)\le t$ and $\alpha\ge\underline{\beta}\overline{\alpha}$. Substituting this inequality into (37) we obtain*

$$\begin{aligned}F(x+\alpha d)-F(x) &\le \varepsilon_1\underline{\beta}\,\overline{\alpha}d^Tg(x,u)\le-\varepsilon_0\varepsilon_1\underline{\beta}\,\overline{\alpha}\|d\|\|g(x,u)\|\le-\frac{\varepsilon_0\varepsilon_1\underline{\beta}\,\overline{\alpha}}{\overline{G}}\|g(x,u)\|^2\\ &\triangleq -c\|g(x,u)\|^2.\end{aligned}\tag{39}$$

2.2 Solving special quadratic programming problems

In this section we will deal with dual methods for solving quadratic programming problems of form (22)–(23). We restrict ourselves to classical minimax problems. Theoretical considerations are practically the same in the case of the sum of maxima but a formal description of algorithms is significantly more complicated. Thus, we will consider the primal problem

$$\text{minimize}\quad Q(d,z)=z+\frac{1}{2}d^TGd,\quad\text{subject to}\quad f_k+g_k^Td\le z,\quad 1\le k\le m,\tag{40}$$

and the dual problem

$$\text{minimize}\quad\tilde{Q}(u)=\frac{1}{2}u^TA^THAu-f^Tu\quad\text{subject to}\quad u\ge 0,\quad\tilde{e}^Tu=1.\tag{41}$$

Note that by (31) we have

$$d = -HAu, \tag{42}$$

which after substituting into (41) gives

$$\tilde{Q}(u) = \frac{1}{2} d^T G d - f^T u. \tag{43}$$

The solution of primal problem (40) can be obtained by an efficient method for solving dual problem (41) as it is described in [19]. Let $K \subset I = \{1, \dots m\}$ be a set of indices such that $u_k = 0$ if $k \notin K$ and $v_k = 0$ if $k \in K$, where $-v_k = f_k + g_k^T d - z$, $k \in K$, are the values of constraints of the primal problem (v_k, $k \in K$, are the Lagrange multipliers of the dual problem). To simplify notation, we denote $u = [u_k, k \in K]$, $v = [v_k, k \in K]$ vectors, whose elements are Lagrange multipliers with indices belonging to K, so dimensions of these vectors are equal to the number of indices in K (similar notation we use for vectors f, $\tilde{e}$ and for columns of matrix A). Note that multipliers u_k and v_k, $k \notin K$, exist, but they are not elements of the vectors u and v. Using (31), (34), and $\tilde{e}^T u = 1$, we can determine elements u_k, $k \in K$, of a vector u and a variable z. Setting $v_k = 0$, $k \in K$, we obtain

$$v = -A^T d - f + \tilde{e} z = A^T H A u - f + \tilde{e} z = 0, \tag{44}$$

so

$$u = (A^T H A)^{-1}(f - \tilde{e} z), \tag{45}$$

and since

$$\tilde{e}^T u = \tilde{e}^T (A^T H A)^{-1}(f - \tilde{e} z) = 1,$$

we obtain

$$z = \frac{\tilde{e}^T (A^T H A)^{-1} f - 1}{\tilde{e}^T (A^T H A)^{-1} \tilde{e}}. \tag{46}$$

Definition 2. *The set of indices $K \subset I$ such that $u_k = 0$, $k \notin K$, and $v_k = 0$, $k \in K$, is called the set of active constraint indices (active set for short) of the primal problem. If $u_k \geq 0$, $k \in K$, then we say that K is an acceptable active set of the primal problem.*

Remark 12. *Formulas (45)–(46) cannot be used if A has linearly dependent columns. It may happen even if the Jacobian matrix $[A^T, -\tilde{e}]$ has full rank. In order not to investigate this singular case separately, we will use matrices*

$$\tilde{A} = \begin{bmatrix} A \\ -\tilde{e}^T \end{bmatrix}, \qquad \tilde{H} = \begin{bmatrix} H & 0 \\ 0 & \mu \end{bmatrix}, \tag{47}$$

where $\mu > 0$. Then $\tilde{A}^T \tilde{H} \tilde{A} = A^T H A + \mu \tilde{e} \tilde{e}^T$, so by (44) it holds

$$\tilde{A}^T \tilde{H} \tilde{A} u = f + (z - \mu)\tilde{e} \tag{48}$$

and formulas (45)–(46) can be written in the form

$$u = Cf - (z - \mu)p, \qquad z = \mu + \frac{p^T f - 1}{p^T e}, \tag{49}$$

where $C = (\tilde{A}^T \tilde{H} \tilde{A})^{-1}$ and $p = C\tilde{e}$. The value $\mu > 0$ should be comparable with elements of matrix H. The choice $\mu = 1$ is usually suitable.

The active constraint method for solving dual problem (41) introduced in [19] is based on generating a sequence of acceptable active sets of primal problem (40). An initial acceptable active set is determined in the way that we choose an arbitrary index $k \in I$ and set $K = \{k\}$, so $u_k = 1$ and $u_l = 0$ if $l \in I$ and $l \neq k$. At each step we first test if the necessary (in a convex case also sufficient) KKT conditions are satisfied. If $v_l < 0$ for some index $l \notin K$, then we try to remove the active constraint of the dual problem by considering the set $K^+ = K \cup \{l\}$. However, this set need not be acceptable (it may hold $u_k < 0$ for some index $k \in K^+$).

Therefore, we need to remove some active constraints of the primal problem in advance, that is, to construct an acceptable set $\bar{K} \subset K$ such that the set $K^+ = \bar{K} \cup \{l\}$ was acceptable as well. For this reason we will change the constraints of the primal problem with index l into $-v_l(\lambda) = A_l^T d + f_l - z + (1-\lambda)v_l \le 0$ (parameter λ is introduced as an argument), so $-v_l(0) = A_l^T d + f_l - z + v_l = 0$ and $u_k(0) \ge 0$ if $k \in K \cup \{l\}$. In the subsequent considerations we will use the notation $a_l = g_l$.

Lemma 1. *Let K be an acceptable active set of primal problem (40) and $v_l < 0$. Suppose that a vector $\tilde{a}_l = [a_l^T, -1]^T$ is not a linear combination of columns of matrix $\tilde{A}$ and denote $p = C\tilde{e}$, $q_l = C\tilde{A}^T\tilde{H}\tilde{a}_l$, $\beta_l = 1 - \tilde{e}^T q_l$, $\gamma_l = \beta_l / \tilde{e}^T p$, and $\delta_l = \tilde{a}_l^T(\tilde{H} - \tilde{H}\tilde{A}C\tilde{A}^T\tilde{H})\tilde{a}_l = \tilde{a}_l^T\tilde{H}(\tilde{a}_l - \tilde{A}q_l)$, where $C = (\tilde{A}^T\tilde{H}\tilde{A})^{-1}$, so $\delta_l > 0$. Then*

$$u(\lambda) = u(0) - \alpha(q_l + \gamma_l p), \qquad u_l(\lambda) = u_l(0) + \alpha, \qquad z(\lambda) = z(0) + \alpha\gamma_l, \tag{50}$$

where $\alpha = -\lambda v_l / (\beta_l\gamma_l - \delta_l)$.

Proof. Using relation (48) augmented by the equation with index l, one can write

$$\begin{bmatrix} \tilde{A}^T\tilde{H}\tilde{A} & \tilde{A}^T\tilde{H}\tilde{a}_l \\ \tilde{a}_l^T\tilde{H}\tilde{A} & \tilde{a}_l^T\tilde{H}\tilde{a}_l \end{bmatrix} \begin{bmatrix} u(\lambda) \\ u_l(\lambda) \end{bmatrix} = \begin{bmatrix} f - (z(\lambda) - \mu)\tilde{e} \\ f_l + (1-\lambda)v_l - (z(\lambda) - \mu) \end{bmatrix}. \tag{51}$$

Subtracting equations for $u(0)$ and $u_l(0)$ we obtain

$$\begin{bmatrix} \tilde{A}^T\tilde{H}\tilde{A} & \tilde{A}^T\tilde{H}\tilde{a}_l \\ \tilde{a}_l^T\tilde{H}\tilde{A} & \tilde{a}_l^T\tilde{H}\tilde{a}_l \end{bmatrix} \begin{bmatrix} u(\lambda) - u(0) \\ u_l(\lambda) - u_l(0) \end{bmatrix} = - \begin{bmatrix} (z(\lambda) - z(0))\tilde{e} \\ \lambda v_l + (z(\lambda) - z(0)) \end{bmatrix}.$$

The inverse matrix of this system can be expressed by the inverse matrix $C = (\tilde{A}^T\tilde{H}\tilde{A})^{-1}$, which gives

$$\begin{aligned} \begin{bmatrix} u(\lambda) - u(0) \\ u_l(\lambda) - u_l(0) \end{bmatrix} &= - \begin{bmatrix} C + \dfrac{q_l q_l^T}{\delta_l} & -\dfrac{q_l}{\delta_l} \\ -\dfrac{q_l^T}{\delta_l} & \dfrac{1}{\delta_l} \end{bmatrix} \begin{bmatrix} (z(\lambda) - z(0))\tilde{e} \\ \lambda v_l + (z(\lambda) - z(0)) \end{bmatrix} \\ &= - \begin{bmatrix} (p - \dfrac{\beta_l}{\delta_l} q_l)(z(\lambda) - z(0)) - \dfrac{\lambda v_l}{\delta_l} q_l \\ \dfrac{\beta_l}{\delta_l}(z(\lambda) - z(0)) + \dfrac{\lambda v_l}{\delta_l} \end{bmatrix}. \end{aligned} \tag{52}$$

Since $\tilde{e}^T u(\lambda) + u_l(\lambda) = 1$ for $\lambda \ge 0$, we can write

$$\tilde{e}^T(u(\lambda) - u(0)) + (u_l(\lambda) - u_l(0)) = 0, \tag{53}$$

which along with (52) gives

$$\begin{bmatrix} \tilde{e}^T & 1 \end{bmatrix} \begin{bmatrix} u(\lambda) - u(0) \\ u_l(\lambda) - u_l(0) \end{bmatrix} = - \left(\tilde{e}^T p + \frac{\beta_l^2}{\delta_l} \right) (z(\lambda) - z(0)) - \frac{\beta_l}{\delta_l}\lambda v_l = 0,$$

that is

$$z(\lambda) - z(0) = -\frac{\beta_l}{\delta_l} \frac{\delta_l}{\delta_l \tilde{e}^T p + \beta_l^2} \lambda v_l = -\gamma_l \frac{\lambda v_l}{\delta_l + \beta_l\gamma_l} = \alpha\gamma_l, \tag{54}$$

which is the last equality in (50). Substituting (54) into (52) and performing formal arrangements we obtain the remaining equalities in (50). □

Remark 13. *Using (47) we obtain*

$$\begin{bmatrix} \tilde{A}^T\tilde{H}\tilde{A} & \tilde{A}^T\tilde{H}\tilde{a}_l \\ \tilde{a}_l^T\tilde{H}\tilde{A} & \tilde{a}_l^T\tilde{H}\tilde{a}_l \end{bmatrix} = \begin{bmatrix} A^T H A & A^T H a_l \\ a_l^T H A & a_l^T H a_l \end{bmatrix} + \mu \begin{bmatrix} \tilde{e}\tilde{e}^T & \tilde{e} \\ \tilde{e}^T & 1 \end{bmatrix}, \tag{55}$$

so

$$q_l = C\tilde{A}^T\tilde{H}\tilde{a}_l = CA^THa_l + \mu p, \tag{56}$$

$$\delta_l = \tilde{a}_l^T\tilde{H}(\tilde{a}_l - \tilde{A}q_l) = a_l^TH(a_l - Aq_l) + \mu\beta_l, \tag{57}$$

Note that $\beta_l\gamma_l + \delta_l = \beta_l^2/\tilde{e}^TC\tilde{e} + \delta_l = 0$ *if and only if* $\beta_l = \gamma_l = \delta_l = 0$.

Lemma 2. *Let the assumptions of Lemma 1 be satisfied. Then*

$$Q(d(\lambda), z(\lambda)) = Q(d(0), z(0)) + \frac{1}{2}\alpha(\beta_l\gamma_l + \delta_l)(u_l(\lambda) + u_l(0)). \tag{58}$$

$$\tilde{Q}(u(\lambda), u_l(\lambda)) = \tilde{Q}(u(0), u_l(0)) + \frac{1}{2}\alpha^2(\beta_l\gamma_l + \delta_l) + \alpha v_l. \tag{59}$$

Proof.

(a) Using (42) and (50) we can write

$$\begin{aligned} d(\lambda) - d(0) &= -HA(u(\lambda) - u(0)) - Ha_l(u_l(\lambda) - u_l(0)) = \alpha H(A(q_l + \gamma_l p) - a_l), \\ d(0) &= -H(Au(0) + a_l u_l(0)) \end{aligned}$$

and by (55)–(56) it holds

$$\begin{aligned} A^THAq_l &= \tilde{A}^T\tilde{H}\tilde{A}q_l - \mu\tilde{e}\tilde{e}^Tq_l = A^THa_l + \mu\tilde{e} - \mu\tilde{e}\tilde{e}^Tq_l = A^THa_l + \mu\beta_l\tilde{e}, \\ A^THAp &= \tilde{A}^T\tilde{H}\tilde{A}p - \mu\tilde{e}\tilde{e}^Tp = \tilde{e} - \mu\tilde{e}\tilde{e}^Tp, \end{aligned}$$

because $\tilde{A}^T\tilde{H}\tilde{A} = C^{-1}$. Using these equalities and formulas (56)–(57) we obtain

$$\begin{aligned} A^TH(A(q_l + \gamma_l p) - a_l) &= A^THa_l + \mu\beta_l\tilde{e} + \gamma_l\tilde{e} - \mu\gamma_l\tilde{e}\tilde{e}^Tp - A^THa_l = \gamma_l\tilde{e}, \\ -a_l^TH(A(q_l + \gamma_l p) - a_l) &= \delta_l - \mu\beta_l - \gamma_l a_l^THAC\tilde{e} = \delta_l - \mu\beta_l - \gamma_l\tilde{e}^Tq_l + \mu\gamma_l\tilde{e}^Tp = \delta_l - \gamma_l\tilde{e}^Tq_l, \end{aligned}$$

which after substitution gives

$$\begin{aligned} (d(\lambda) - d(0))^TGd(0) &= -\alpha(A(q_l + \gamma_l p) - a_l)^TH(Au(0) + a_l u_l(0)) \\ &= -\alpha\gamma_l\tilde{e}^Tu(0) + \alpha(\delta_l - \gamma_l\tilde{e}^Tq_l)u_l(0) \\ &= -\alpha\gamma_l(1 - u_l(0)) + \alpha(\delta_l - \gamma_l(1 - \beta_l))u_l(0) \\ &= -\alpha\gamma_l + \alpha(\delta_l + \beta_l\gamma_l)u_l(0), \qquad (60) \\ (d(\lambda) - d(0))^TG(d(\lambda) - d(0)) &= \alpha^2(A(q_l + \gamma_l p) - a_l)^THA(q_l + \gamma_l p) - a_l) \\ &= \alpha^2\gamma_l\tilde{e}^T(q_l + \gamma_l p) + \alpha^2(\delta_l - \gamma_l\tilde{e}^Tq_l) \\ &= \alpha^2(\delta_l + \beta_l\gamma_l) \qquad (61) \end{aligned}$$

(because $\tilde{e}^Tu(\lambda) + u_l(\lambda) = 1$ for $\lambda \geq 0$). Since $z(\lambda) - z(0) = \alpha\gamma_l$, using (60)–(61) we can write

$$\begin{aligned} Q(d(\lambda), z(\lambda)) &= Q(d(0), z(0)) + z(\lambda) - z(0) + (d(\lambda) - d(0))Gd(0) \\ &+ \frac{1}{2}(d(\lambda) - d(0))^TG(d(\lambda) - d(0)) \\ &= Q(d(0)), z(0)) + \alpha\gamma_l - \alpha\gamma_l + \alpha(\delta_l + \beta_l\gamma_l)u_l(0) + \frac{1}{2}\alpha^2(\delta_l + \beta_l\gamma_l) \\ &= Q(d(0)), z(0)) + \frac{1}{2}\alpha(\delta_l + \beta_l\gamma_l)(u(\lambda) + u_l(0)), \end{aligned}$$

because $\alpha = u_l(\lambda) - u_l(0))$.

(b) Using (51), (53), and (55) we obtain

$$\begin{aligned}(d(\lambda)-d(0))^T G d(0) &= \begin{bmatrix} u(\lambda)-u(0) \\ u_l(\lambda)-u_l(0) \end{bmatrix}^T \begin{bmatrix} A^T H A & A^T H a_l \\ a_l^T H A & a_l^T H a_l \end{bmatrix} \begin{bmatrix} u(0) \\ u_l(0) \end{bmatrix} \\ &= \begin{bmatrix} u(\lambda)-u(0) \\ u_l(\lambda)-u_l(0) \end{bmatrix}^T \begin{bmatrix} f-z(0)\tilde{e} \\ f_l+v_l-z(0) \end{bmatrix} \\ &= (u(\lambda)-u(0))^T f + (u_l(\lambda)-u_l(0))f_l + (u_l(\lambda)-u_l(0))v_l,\end{aligned}$$

which along with (43) and (50) gives

$$\begin{aligned}\tilde{Q}(u(\lambda),u_l(\lambda)) &= \tilde{Q}(u(0),u_l(0)) + (d(\lambda)-d(0))Gd(0) + \frac{1}{2}(d(\lambda)-d(0))^T G(d(\lambda)-d(0)) \\ &\qquad -f^T(u(\lambda)-u(0)) - f_l(u_l(\lambda)-u_l(0)) \\ &= \tilde{Q}(u(0),u_l(0)) + \frac{1}{2}(d(\lambda)-d(0))^T G(d(\lambda)-d(0)) + \alpha v_l,\end{aligned}$$

so (59) holds by (61).

□

Remark 14. *Denote* $\tilde{I} = \{k \in I \cap K : e_k^T(q_l + \gamma_l p) > 0\}$ *and set*

$$\alpha_1 = -\frac{v_l}{\beta_l\gamma_l + \delta_l}, \qquad \alpha_2 = \frac{u_j(0)}{e_j^T(q_l+\gamma_l p)} \triangleq \min_{k\in\tilde{I}} \frac{u_k(0)}{e_k^T(q_l+\gamma_l p)}, \tag{62}$$

where $\alpha_1 = \infty$ *if* $\beta_l\gamma_l + \delta_l = 0$ *and* $\alpha_2 = \infty$ *if* $\tilde{I} = \emptyset$. *Let* $\alpha = \min(\alpha_1, \alpha_2)$. *Three cases can occur.*

(1) If $\alpha = \alpha_1 = \infty$, *the dual problem has no optimal solution because* $\beta_l\gamma_l + \delta_l = 0$ *and* $v_l < 0$. *Thus,* $\tilde{Q}(u(\alpha), u_l(\alpha)) \to -\infty$ *by (59) if* $\alpha \to \infty$. *The primal problem has no feasible solution in this case.*

(2) If $\alpha = \alpha_1 < \infty$, *so* $\beta_l\gamma_l + \delta_l > 0$, *we can set* $K^+ = K \cup \{l\}$. *This corresponds to adding the constraint with index* l *into a set of active constraints of the primal problem.*

(3) If $\alpha = \alpha_2 < \alpha_1$, *we need to remove the constraint with index* j *(formula (62)) from the set of active constraints of the primal problem.*

In cases (2) and (3), it is necessary to update representation of the linear variety defined by new active constraints.

Remark 15. *If* $\alpha = \alpha_2 < \alpha_1$, *we cannot add index* l *into a set* K. *In this case, we need to remove index* j *from the set* K *and create a set* $\bar{K}_1 = \bar{K}_0 \setminus \{j\}$, *where* $\bar{K}_0 = K$, *which is possible because* $u_j = 0$. *At the same time, we need to multiply the value* v_l *by* $1 - \alpha/\alpha_1$ *(if* $\alpha_1 = \infty$, *then the value* v_l *is unchanged). Performing these adjustments we can try to add index* l *into the set* $\bar{K}_1$. *Repeating these procedure we obtain a sequence of sets* $K = \bar{K}_0 \supset \bar{K}_1 \supset \cdots \supset \bar{K}_p$. *Since the number of constraints of the primal problem is finite, then there exists a set* $\bar{K} = \bar{K}_p$, *where* $p \geq 0$, *such that either* $\alpha = \infty$ *(no solution of problem (40) exists) or* $\alpha = \alpha_1$, *so a set* $K^+ = \bar{K} \cup \{l\}$ *is an acceptable active set of the primal problem.*

Lemma 3. *Let a solution of problem (40) exist and* K, $K^+ = \bar{K} \cup \{l\}$ *be the sets mentioned in Remark 15. Let* d *and* d^+ *be direction vectors given by (42), where the vectors* u *and* u^+ *correspond to acceptable sets* K *and* K^+. *Then* $Q(d^+, z^+) > Q(d, z)$.

Proof. Denote by $\bar{d}_i$, $0 \leq i \leq p$, direction vectors given by (42), where the vectors $\bar{u}_i$, $0 \leq i \leq p$, correspond to the sets $\bar{K}_i$, $0 \leq i \leq p$. Then $Q(\bar{d}_i, \bar{z}_i) \geq Q(\bar{d}_{i-1}, \bar{z}_{i-1})$, $1 \leq i \leq p$, holds by Lemma 2. Since $\alpha = \alpha_1$ holds in the last step determined by the set $\bar{K} = \bar{K}_p$, so $\beta_l\gamma_l + \delta_l > 0$ and $\alpha > 0$, and since $\bar{u}_l(0) \geq 0$, we can write $Q(d^+, z^+) > Q(\bar{d}, \bar{z}) = Q(\bar{d}_p, \bar{z}_p) \geq Q(\bar{d}_0, \bar{z}_0) = Q(d, z)$. □

Algorithm 1. *Dual method of active constraints*

Step 1 *Choose an arbitrary index $1 \le l \le m$ (e.g. $l = 1$) and a number μ (e.g. $\mu = 1$). Set $K := \{l\}$, $u := [1]$, $\tilde{e} := [1]$, $A := [a_l]$, $R := [a_l^T H a_l + \mu]^{1/2}$. Compute a number $z := f_l - a_l^T H a_l$. Set $v_l := 0$ and $u_k := 0$ for $k \notin K$.*

Step 2 *Compute a vector $d := -HAu$ (formula (42)), set $v_k := z - (a_k^T d + f_k)$ for $k \notin K$ and determine an index $l \notin K$ such that $v_l = \min_{k \notin K} v_k$. If $v_l \ge 0$, terminate the computation (a pair $(d, z) \in R^{n+1}$ is a solution of primal problem (40) and a vector u is a solution of dual problem (41)).*

Step 3 *Determine the vector p by solving the system of equations $R^T R p = \tilde{e}$ and the vector q_l by solving the system of equations $R^T R q_l = \tilde{A}^T \tilde{H} \tilde{a}_l$. Set $\beta_l := 1 - \tilde{e}^T q_l$, $\gamma_l := \beta_l / \tilde{e}^T p$, $\delta_l := \tilde{a}_l^T \tilde{H} (\tilde{a}_l - \tilde{A} q_l)$ (Remark 16). Compute numbers α_1, α_2 defined in Remark 14 and set $\alpha := \min(\alpha_1, \alpha_2)$. If $\alpha = \infty$, terminate the computation (the primal problem has no feasible solution and the dual problem has no optimal solution). If $\alpha < \infty$, set $u := u - \alpha(q_l + \gamma_l p)$, $u_l := u_l + \alpha$, $z := z + \alpha \gamma_l$, $v_l := (1 - \alpha/\alpha_1) v_l$.*

Step 4 *If $\alpha = \alpha_1$, set $K := K \cup \{l\}$, $u := u^+$, $f := f^+$, $\tilde{e} := \tilde{e}^+$, $A := A^+$, $R := R^+$, where $u^+ = [u^T, u_l]^T$, $f^+ = [f^T, f_l]^T$, $\tilde{e}^+ = [\tilde{e}^T, 1]^T$ and A^+, R^+ are matrices defined in Remark 16. Go to Step 2.*

Step 5 *If $\alpha \ne \alpha_1$, set $K := K \setminus \{j\}$, $u := u^-$, $f := f^-$, $\tilde{e} := \tilde{e}^-$, $A := A^-$, $R := R^-$, where $j \in K$ is an index determined by (62), vectors u^-, f^-, $\tilde{e}^-$ result from vectors u, f, $\tilde{e}$ by removing the element with index j and A^-, R^- are matrices defined in Remark 16. Go to Step 3.*

Remark 16. *The vector q_l and the number δ_l used in Step 3 of Algorithm 1 can be computed so that we solve two systems of equations $R^T r_l = \tilde{A}^T \tilde{H} \tilde{a}_l = A^T H a_l + \mu \tilde{e}$ and $R q_l = r_l$ with triangular matrices R^T and R and set $\delta_l = \rho_l^2$ where $\rho_l^2 = a_l^T H a_l - r_l^T r_l$. Then, in Step 4 it holds that*

$$A^+ = [A, a_l], \qquad R^+ = \begin{bmatrix} R & r_l \\ 0 & \rho_l \end{bmatrix}.$$

In Step 5 we determine a permutation matrix Π such that $A\Pi = [A^-, a_j]$ and $R\Pi$ is an upper Hessenberg matrix. Furthermore, we determine an orthogonal matrix Q such that the matrix $QR\Pi$ is upper triangular. Then

$$QR\Pi = \begin{bmatrix} R^- & r_j \\ 0 & \rho_j \end{bmatrix}$$

holds. Derivation of these relations can be found in [19].

Theorem 3. *After a finite number of steps of Algorithm 1, either a solution of problems (40) and (41) is found or the fact that these problems have no solution is detected.*

Proof. Algorithm 1 generates a sequence of active sets K_{j_i}, $j_i \in N$, of the primal problem, where the set K_{j_1}, $j_1 = 1$, is acceptable. Suppose that the set K_{j_i}, $j_i \in N$, is acceptable. By Remark 15 and Lemma 3, after at most m steps, we either find out that problems (40) and (41) have no solution or obtain an acceptable set $K_{j_{i+1}}$ where $j_{i+1} - j_i \le m$ and where $Q(d_{j_{i+1}}, z_{j_{i+1}}) > Q(d_{j_i}, z_{j_i})$. Thus, the sets $K_{j_{i+1}}$ and K_{j_i} are different and since the number of different subsets of the set $\{1, \ldots, m\}$ is finite, the computation must terminate after a finite number of steps. □

3 Primal interior point methods

3.1 Barriers and barrier functions

Primal interior point methods for equality constraint minimization problems are based on adding a barrier term containing constraint functions to the minimized function. A resulting barrier function, depending

on a barrier parameter $0 < \mu \leq \overline{\mu} < \infty$, is successively minimized on R^n (without any constraints), where $\mu \to 0$. Applying this approach on problem (7), we obtain a barrier function

$$B_\mu(x,z) = h(z) + \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \varphi(z_k - f_{kl}(x)), \quad 0 < \mu \leq \overline{\mu}, \tag{63}$$

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

Assumption B1. *Function $\varphi(t)$, $t \in (0,\infty)$, is twice continuously differentiable, decreasing, and strictly convex, with $\lim_{t\to 0} \varphi(t) = \infty$. Function $\varphi'(t)$ is increasing and strictly concave such that $\lim_{t\to\infty} \varphi'(t) = 0$. For $t \in (0,\infty)$ it holds $-t\varphi'(t) \leq 1$, $t^2\varphi''(t) \leq 1$. There exist numbers $\tau > 0$ and $\underline{c} > 0$ such that for $t < \tau$ it holds*

$$-t\varphi'(t) \geq \underline{c} \tag{64}$$

and

$$\varphi'(t)\varphi'''(t) - \varphi''(t)^2 > 0. \tag{65}$$

Remark 17. *A logarithmic barrier function*

$$\varphi(t) = \log t^{-1} = -\log t, \tag{66}$$

is most frequently used. It satisfies Assumption B1 with $\underline{c} = 1$ and $\tau = \infty$ but it is not bounded from below since $\log t \to \infty$ for $t \to \infty$. For that reason, barriers bounded from below are sometimes used, e.g. a function

$$\varphi(t) = \log(t^{-1} + \tau^{-1}) = -\log \frac{t\tau}{t+\tau}, \tag{67}$$

which is bounded from below by number $\underline{\varphi} = -\log\tau$, or a function

$$\varphi(t) = -\log t, \quad 0 < t \leq \tau, \qquad \varphi(t) = at^{-2} + bt^{-1} + c, \quad t \geq \tau, \tag{68}$$

which is bounded from below by number $\underline{\varphi} = c = -\log\tau - 3/2$, or a function

$$\varphi(t) = -\log t, \quad 0 < t \leq \tau, \qquad \varphi(t) = at^{-1} + bt^{-1/2} + c, \quad t \geq \tau, \tag{69}$$

which is bounded from below by number $\underline{\varphi} = c = -\log\tau - 3$. Coefficients a, b, c are chosen so that function $\varphi(t)$ as well as its first and second derivatives are continuous in $t = \tau$. All these barriers satisfy Assumption B1 [26] (the proof of this statement is trivial for logarithmic barrier (66)).

Even if bounded from below barriers (67)-(69) have more advantageous theoretical properties (Assumption X1a can be replaced with a weaker Assumption X1b), algorithms using logarithmic barrier (67) are usually more efficient. Therefore, we will only deal with methods using the logarithmic barrier $\varphi(t) = -\log t$ in the subsequent considerations.

3.2 Iterative determination of a minimax vector

Suppose the function $h(z)$ is of form (21). Using the logarithmic barrier $\varphi(t) = -\log t$, function (63) can be written as

$$B_\mu(x,z) = \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k - f_{kl}(x)), \quad 0 < \mu \leq \overline{\mu}. \tag{70}$$

Further, we will denote $g_{kl}(x)$ and $G_{kl}(x)$ gradients and Hessian matrices of functions $f_{kl}(x)$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, and set

$$u_{kl}(x,z) = \frac{\mu}{z_k - f_{kl}(x)} \geq 0, \qquad v_{kl}(x,z) = \frac{\mu}{(z_k - f_{kl}(x))^2} = \frac{1}{\mu} u_{kl}^2(x,z) \geq 0, \tag{71}$$

$$u_k(x,z) = \begin{bmatrix} u_{k1}(x,z) \\ \dots \\ u_{km_k}(x,z) \end{bmatrix}, \qquad v_k(x,z) = \begin{bmatrix} v_{k1}(x,z) \\ \dots \\ v_{km_k}(x,z) \end{bmatrix}, \qquad \tilde{e}_k = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}.$$

Denoting by $g(x,z)$ the gradient of the function $B_\mu(x,z)$ and $\gamma_k(x,z) = \partial B_\mu(x,z)/\partial z_k$, the necessary conditions for an extremum of barrier function (63) can be written in the form

$$g(x,z) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x)u_{kl}(x,z) = \sum_{k=1}^{m} A_k(x)u_k(x,z) = 0, \tag{72}$$

$$\gamma_k(x,z) = 1 - \sum_{l=1}^{m_k} u_{kl}(x,z) = 1 - \tilde{e}_k^T u_k(x,z) = 0, \qquad 1 \le k \le m, \tag{73}$$

where $A_k(x) = [g_{k1}(x), \dots, g_{km_k}(x)]$, which is a system of $n+m$ nonlinear equations for unknown vectors x and z. These equations can be solved by the Newton method. In this case, the second derivatives of the Lagrange function (which are the first derivatives of expressions (72) and (73)) are computed. Denoting

$$G(x,z) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} G_{kl}(x)u_{kl}(x,z) \tag{74}$$

the Hessian matrix of the Lagrange function and setting

$$\begin{aligned} U_k(x,z) &= \operatorname{diag}(u_{k1}(x,z), \dots, u_{km_k}(x,z)), \\ V_k(x,z) &= \operatorname{diag}(v_{k1}(x,z), \dots, v_{km_k}(x,z)) = \frac{1}{\mu}U_k^2(x,z), \end{aligned}$$

we can write

$$\begin{aligned} \frac{\partial g(x,z)}{\partial x} &= \sum_{k=1}^{m}\sum_{l=1}^{m_k} G_{kl}(x)u_{kl}(x,z) + \sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x)v_{kl}(x,z)g_{kl}^T(x) \\ &= G(x,z) + \sum_{k=1}^{m} A_k(x)V_k(x,z)A_k^T(x), \end{aligned} \tag{75}$$

$$\frac{\partial g(x,z)}{\partial z_k} = -\sum_{l=1}^{m_k} g_{kl}(x)v_{kl}(x,z) = -A_k(x)v_k(x,z), \tag{76}$$

$$\frac{\partial \gamma_k(x,z)}{\partial x} = -\sum_{l=1}^{m_k} v_{kl}(x,z)g_{kl}^T(x) = -v_k^T(x,z)A_k^T(x), \tag{77}$$

$$\frac{\partial \gamma_k(x,z)}{\partial z_k} = \sum_{l=1}^{m_k} v_{kl}(x,z) = \tilde{e}_k^T v_k(x,z). \tag{78}$$

Using these formulas we obtain a system of linear equations describing a step of the Newton method

$$\begin{bmatrix} W(x,z) & -A_1(x)v_1(x,z) & \dots & -A_m(x)v_m(x,z) \\ -v_1^T(x,z)A_1^T(x) & \tilde{e}_1^T v_1(x,z) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -v_m^T(x,z)A_m^T(x) & 0 & \dots & \tilde{e}_m^T v_m(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z_1 \\ \dots \\ \Delta z_m \end{bmatrix} = - \begin{bmatrix} g_k(x,z) \\ \gamma_1(x,z) \\ \dots \\ \gamma_m(x,z) \end{bmatrix}, \tag{79}$$

where

$$W(x,z) = G(x,z) + \sum_{k=1}^{m} A_k(x)V_k(x,z)A_k^T(x). \tag{80}$$

Setting

$$C(x,z) = [A_1(x)v_1(x,z), \dots, A_m(x)v_m(x,z)], \qquad D(x,z) = \operatorname{diag}(\tilde{e}_1^T v_1(x,z), \dots, \tilde{e}_m^T v_m(x,z))$$

and $\gamma(x,z) = [\gamma_1(x,z), \ldots, \gamma_m(x,z)]^T$, a step of the Newton method can be written in the form

$$\begin{bmatrix} W(x,z) & -C(x,z) \\ -C^T(x,z) & D(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} g(x,z) \\ \gamma(x,z) \end{bmatrix}. \tag{81}$$

The diagonal matrix $D(x,z)$ is positive definite since it has positive diagonal elements.

Remark 18. *If the number m is small (as in case of a classical minimax problem, where $m = 1$), we will use the expression*

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

We suppose that the matrix W is regular (otherwise, it can be regularized e.g. by the Gill-Murray decomposition [11]). Then, a solution of system of equations (81) can be computed by

$$\Delta z = (C^TW^{-1}C - D)^{-1}(C^TW^{-1}g + \gamma), \tag{82}$$

$$\Delta x = W^{-1}(C\Delta z - g). \tag{83}$$

In this case, a large matrix W of order n, which is sparse if $G(x,z)$ is sparse, and a small dense matrix $C^TW^{-1}C - D$ of order m are decomposed.

Remark 19. *If the numbers m_k, $1 \le k \le m$, are small (as in case of a sum of sbsolute values, where $m_k = 2$, $1 \le k \le m$), the matrix $W(x,z) - C(x,z)D^{-1}(x,z)C^T(x,z)$ is sparse. Thus, we can use the expression*

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

Then, a solution of system of equations (81) can be computed by

$$\Delta x = -(W - CD^{-1}C^T)^{-1}(g + CD^{-1}\gamma), \tag{84}$$

$$\Delta z = D^{-1}(C^T\Delta x - \gamma). \tag{85}$$

In this case, a large matrix $W - CD^{-1}C^T$ of order n, which is usually sparse if $G(x,z)$ is sparse, is decomposed. The inverse of the diagonal matrix D of order m makes no problem.

During iterative determination of a minimax vector we know a value of the parameter μ and vectors $x \in R^n$, $z \in R^m$ such that $z_k > F_k(x)$, $1 \le k \le m$. Using formulas (82)–(83) or (84)–(85) we determine direction vectors Δx, Δz. Then, we choose a steplength α so that

$$B_\mu(x + \alpha\Delta x, z + \alpha\Delta z) < B_\mu(x,z) \tag{86}$$

and $z_k + \alpha\Delta z_k > F_k(x + \alpha\Delta x)$, $1 \le k \le m$. Finally, we set $x_+ = x + \alpha\Delta x$, $z_+ = z + \alpha\Delta z$ and determine a new value $\mu_+ < \mu$. If the matrix of system of equations (81) is positive definite, inequality (86) is satisfied for a sufficiently small value of the steplength α.

Theorem 4. *Let the matrix $G(x,z)$ given by (74) be positive definite. Then the matrix of system of equations (81) is positive definite.*

Proof. The matrix of system of equations (81) is positive definite if and only if the matrix D and its Schur complement $W - CD^{-1}C^T$ are positive definite [8, Theorem 2.5.6]. The matrix D is positive definite since it has positive diagonal elements. Further, it holds

$$W - CD^{-1}C^T = G + \sum_{k=1}^{m} \left(A_kV_kA_k^T - A_kV_k\tilde{e}_k(\tilde{e}_k^TV_k\tilde{e}_k)^{-1}(A_kV_k\tilde{e}_k)^T \right),$$

matrices $A_kV_kA_k^T - A_kV_k\tilde{e}_k(\tilde{e}_k^TV_k\tilde{e}_k)^{-1}(A_kV_k\tilde{e}_k)^T$, $1 \le k \le m$, are positive semidefinite due to the Schwarz inequality and the matrix G is positive definite by the assumption. □

3.3 Direct determination of a minimax vector

Now we will show how to solve system of equations (72)–(73) by direct determination of a minimax vector using two-level optimization

$$z(x;\mu) = \arg\min_{z\in R^m} B_\mu(x,z), \tag{87}$$

$$x^* = \arg\min_{x\in R^n} B(x;\mu), \quad B(x;\mu) \stackrel{\Delta}{=} B_\mu(x, z(x;\mu)). \tag{88}$$

Problem (87) serves for determination of an optimal vector $z(x;\mu) \in R^m$. Let $\tilde{B}_\mu(z) = B_\mu(x,z)$ for a fixed chosen vector $x \in R^n$. The function $\tilde{B}_\mu(z)$ is strictly convex (as a function of a vector z), since it is a sum of convex function (21) and strictly convex functions $-\mu\log(z_k - f_{kl}(x))$, $1 \le k \le m$, $1 \le l \le m_k$. A minimum of the function $\tilde{B}_\mu(z)$ is its stationary point, so it is a solution of system of equations (73) with Lagrange multipliers (71). The following theorem shows that this solution exists and is unique.

Theorem 5. *The function $\tilde{B}_\mu(z) : (F(x),\infty) \to R$ has a unique stationary point which is its global minimum. This stationary point is characterized by a system of equations $\gamma(x,z) = 0$, or*

$$1 - \tilde{e}_k^T u_k = 1 - \sum_{l=1}^{m_k} \frac{\mu}{z_k - f_{kl}(x)} = 0, \qquad 1 \le k \le m, \tag{89}$$

which has a unique solution $z(x;\mu) \in Z \subset R^m$ such that

$$F_k(x) < F_k(x) + \mu < z_k(x;\mu) < F_k(x) + m_k\mu \tag{90}$$

for $1 \le k \le m$.

Proof. Definition 1 implies $f_{kl}(x) \le F_k(x)$, $1 \le k \le m$, $1 \le l \le m_k$, where the equality occurs for at least one index l.

(a) If (89) holds, then we can write

$$1 = \sum_{l=1}^{m_k} \frac{\mu}{z_k - f_{kl}(x)} > \frac{\mu}{z_k - F_k(x)} \quad \Leftrightarrow \quad z_k - F_k(x) > \mu,$$

$$1 = \sum_{l=1}^{m_k} \frac{\mu}{z_k - f_{kl}(x)} < \frac{m_k\mu}{z_k - F_k(x)} \quad \Leftrightarrow \quad z_k - F_k(x) < m_k\mu,$$

which proves inequalities (90).

(b) Since

$$\gamma_k(x, F+\mu) = 1 - \sum_{l=1}^{m_k} \frac{\mu}{\mu + F_k(x) - f_{kl}(x)} < 1 - \frac{\mu}{\mu} = 0,$$

$$\gamma_k(x, F+m_k\mu) = 1 - \sum_{l=1}^{m_k} \frac{\mu}{m_k\mu + F_k(x) - f_{kl}(x)} > 1 - \frac{m_k\mu}{m_k\mu} = 0,$$

and the function $\gamma_k(x, z_k)$ is continuous and decreasing in $F_k(x) + \mu < z_k(x;\mu) < F_k(x) + m_k$ by (78), the equation $\gamma_k(x, z_k) = 0$ has a unique solution in this interval. Since the function $\tilde{B}_\mu(z)$ is convex, this solution corresponds to its global minimum.

□

System (89) is a system of m scalar equations with localization inequalities (90). These scalar equations can be efficiently solved by robust methods described e.g. in [16] and [17] (details are stated in [25]). Suppose that $z = z(x;\mu)$ and denote

$$B(x;\mu) = \sum_{k=1}^{m} z_k(x;\mu) - \mu \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)). \tag{91}$$

To find a minimum of $B_\mu(x,z)$ in R^{n+m}, it suffices to minimize $B(x;\mu)$ in R^n.

Theorem 6. *Consider barrier function (91). Then*

$$\nabla B(x;\mu) = \sum_{k=1}^{m} A_k(x) u_k(x;\mu), \tag{92}$$

$$\begin{aligned} \nabla^2 B(x;\mu) &= W(x;\mu) - C(x;\mu) D^{-1}(x;\mu) C^T(x;\mu) \\ &= G(x;\mu) + \sum_{k=1}^{m} A_k(x) V_k(x;\mu) A_k^T(x) - \sum_{k=1}^{m} \frac{A_k(x) V_k(x;\mu) \tilde{e}_k \tilde{e}_k^T V_k(x;\mu) A_k^T(x)}{\tilde{e}_k^T V_k(x;\mu) \tilde{e}_k}, \end{aligned} \tag{93}$$

where $W(x;\mu) = W(x, z(x;\mu))$, $G(x;\mu) = G(x, z(x;\mu))$, $C(x;\mu) = C(x, z(x;\mu))$, $D(x;\mu) = D(x, z(x;\mu))$ and $U_k(x;\mu) = U_k(x, z(x;\mu))$, $V_k(x;\mu) = V_k(x, z(x;\mu)) = U_k^2(x;\mu)/\mu$, $1 \le k \le m$. A solution of equation

$$\nabla^2 B(x;\mu) \Delta x = -\nabla B(x;\mu) \tag{94}$$

is identical with a vector Δx given by (84), where $z = z(x;\mu)$ (so $\gamma(x, z(x;\mu)) = 0$).

Proof. Differentiating barrier function (91) and using (73) we obtain

$$\begin{aligned} \nabla B(x;\mu) &= \sum_{k=1}^{m} \frac{\partial z_k(x;\mu)}{\partial x} - \sum_{k=1}^{m}\sum_{l=1}^{m_k} u_{kl}(x;\mu) \left(\frac{\partial z_k(x;\mu)}{\partial x} - \frac{\partial f_{kl}(x)}{\partial x} \right) \\ &= \sum_{k=1}^{m} \frac{\partial z_k(x;\mu)}{\partial x} \left(1 - \sum_{l=1}^{m_k} u_{kl}(x;\mu) \right) + \sum_{k=1}^{m}\sum_{l=1}^{m_k} \frac{\partial f_{kl}(x)}{\partial x} u_{kl}(x;\mu) \\ &= \sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x) u_{kl}(x;\mu) = \sum_{k=1}^{m} A_k(x) u_k(x;\mu), \end{aligned}$$

where

$$u_{kl}(x;\mu) = \frac{\mu}{z_k(x;\mu) - f_{kl}(x)}, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \tag{95}$$

Formula (93) can be obtained by additional differentiation of relations (73) and (92) using (95). A simpler way is based on using (84). Since (73) implies $\gamma(x, z(x;\mu)) = 0$, we can substitute $\gamma = 0$ into (84), which yields the equation

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

where $z = z(x;\mu)$, that confirms validity of formulas (93) and (94) (details can be found in [25]). □

Remark 20. *To determine an inverse of the Hessian matrix, one can use a Woodbury formula [8, Theorem 12.1.4] which gives*

$$\begin{aligned} (\nabla^2 B(x;\mu))^{-1} &= W^{-1}(x;\mu) - W^{-1}(x;\mu) C(x;\mu) \\ &\quad \left(C^T(x;\mu) W^{-1}(x;\mu) C(x;\mu) - D(x;\mu) \right)^{-1} \\ &\quad C^T(x;\mu) W^{-1}(x;\mu). \end{aligned} \tag{96}$$

If the matrix $\nabla^2 B(x;\mu)$ is not positive definite, it can be replaced by a matrix $LL^T = \nabla^2 B(x;\mu) + E$, obtained by the Gill-Murray decomposition [11]. Note that it is more advantageous to use system of linear equations (81) instead of (94) for determination of a direction vector Δx because the system of nonlinear equations (89) is solved with prescribed finite precision, and thus a vector $\gamma(x,z)$, defined by (73), need not be zero.

From

$$V_k(x;\mu) = \frac{1}{\mu} U_k^2(x;\mu), \qquad u_k(x;\mu) \geq 0, \qquad \tilde{e}_k^T u_k(x;\mu) = 1, \qquad 1 \leq k \leq m,$$

it follows that $\|V_k(x;\mu)\| \to \infty$ if $\mu \to 0$, so Hessian matrix (93) may be ill-conditioned if the value μ is very small. From this reason, we use a lower bound $\underline{\mu} > 0$ for μ.

Theorem 7. *Let Assumption X3 be satisfied and let $\mu \geq \underline{\mu} > 0$. If the matrix $G(x;\mu)$ is uniformly positive definite (i.e. there exists a constant $\underline{G}$ such that $v^T G(x;\mu) v \geq \underline{G}\|v\|^2$), there exists a number $\overline{\kappa} \geq 1$ such that $\kappa(\nabla^2 B(x;\mu)) \leq \overline{\kappa}$.*

Proof.

(a) Using (71), (93), and Assumption X3, we obtain

$$\begin{aligned} \|\nabla^2 B(x;\mu)\| &\leq \left\| G(x;\mu) + \sum_{k=1}^{m} A_k(x) V_k(x;\mu) A_k^T(x) \right\| \\ &\leq \sum_{k=1}^{m} \sum_{l=1}^{m_k} \left(|G_{kl}(x) u_{kl}(x,\mu)| + \frac{1}{\mu} \left| u_{kl}^2(x;\mu) g_{kl}(x) g_{kl}^T(x) \right| \right) \\ &\leq \frac{\overline{m}}{\mu} \left(\overline{\mu}\, \overline{G} + \overline{g}^2 \right) \triangleq \frac{\overline{c}}{\mu} \leq \frac{\overline{c}}{\underline{\mu}} \end{aligned} \tag{97}$$

because $0 \leq u_{kl}(x;\mu) \leq \tilde{e}_k^T u_k(x;\mu) = 1$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, by (89).

(b) As in the proof of Theorem 4, for an arbitrary vector $v \in R^n$ it holds

$$\begin{aligned} v^T \nabla^2 B(x;\mu) v &= v^T (W(x;\mu) - C(x;\mu) D^{-1}(x;\mu) C^T(x;\mu)) v \\ &= v^T G(x;\mu) v + v^T A_k(x) V_k(x;\mu) A_k^T(x) v - \frac{v^T A_k(x) V_k(x;\mu) \tilde{e}_k \tilde{e}_k^T V(x;\mu) A_k^T(x) v}{\tilde{e}_k^T V_k(x;\mu) \tilde{e}_k} \\ &\geq v^T G(x;\mu) v \geq \underline{G} \|v\|^2, \end{aligned}$$

so $\underline{\lambda}(\nabla^2 B(x;\mu)) \geq \underline{G}$.

(c) Since (a) implies $\overline{\lambda}(\nabla^2 B(x;\mu)) = \|\nabla^2 B(x;\mu)\| \leq \overline{c}/\underline{\mu}$, using (b) we can write

$$\kappa(\nabla^2 B(x;\mu)) = \frac{\overline{\lambda}(\nabla^2 B(x;\mu))}{\underline{\lambda}(\nabla^2 B(x;\mu))} \leq \frac{\overline{c}}{\underline{\mu}\, \underline{G}} \triangleq \overline{\kappa}. \tag{98}$$

□

Remark 21. *If there exists a number $\overline{\kappa} > 0$ such that $\kappa(\nabla^2 B(x_i;\mu_i)) \leq \overline{\kappa}$, $i \in N$, the direction vector Δx_i, given by solving a system of equations $\nabla^2 B(x_i;\mu_i)\Delta x_i = -\nabla B(x_i;\mu_i)$, satisfies the condition*

$$(\Delta x_i)^T g(x_i;\mu_i) \leq -\varepsilon_0 \|\Delta x_i\| \|g(x_i;\mu_i)\|, \qquad i \in N, \tag{99}$$

where $\varepsilon_0 = 1/\sqrt{\overline{\kappa}}$. Then, for arbitrary numbers $0 < \varepsilon_1 \leq \varepsilon_2 < 1$ one can find a steplength parameter $\alpha_i > 0$ such that for $x_{i+1} = x_i + \alpha_i \Delta x_i$ it holds

$$\varepsilon_1 \leq \frac{B(x_{i+1};\mu_i) - B(x_i;\mu_i)}{\alpha_i (\Delta x_i)^T g(x_i;\mu_i)} \leq \varepsilon_2, \tag{100}$$

so there exists a number $c > 0$ such that (see [32, Section 3.2])

$$B(x_{i+1};\mu_i) - B(x_i;\mu_i) \le -c\|g(x_i;\mu_i)\|^2, \quad i \in N. \tag{101}$$

If Assumption X3 is not satisfied, then only $(\Delta x_i)^T g(x_i;\mu_i) < 0$ holds (because the matrix $\nabla^2 B(x;\mu)$ is positive definite by Theorem 4) and

$$B(x_{i+1};\mu_i) - B(x_i;\mu_i) \le 0, \quad i \in N. \tag{102}$$

3.4 Implementation

Remark 22. *In (80), it is assumed that $G(x,z)$ is the Hessian matrix of the Lagrange function. Direct computation of the matrix $G(x;\mu) = G(x,z(x;\mu))$ is usually difficult (one can use automatic differentiation as described in [14]). Thus, various approximations $G \approx G(x;\mu)$ are mostly used.*

- *The matrix $G \approx G(x;\mu)$ can be determined using differences*

$$Gw_j = \frac{A(x+\delta w_j)u(x;\mu) - A(x)u(x;\mu)}{\delta}, \quad 1 \le j \le \overline{k}.$$

  The vectors w_j, $1 \le j \le \overline{k}$, are chosen so that the number of them is as small as possible [4], [35].

- *The matrix $G \approx G(x;\mu)$ can be determined using the variable metric methods [27]. The vectors*

$$d = x_+ - x, \qquad y = A(x_+)u(x_+;\mu) - A(x)u(x_+;\mu)$$

  are used for an update of G.

- *If the problem is separable (i.e. $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, are functions of a small number $n_{kl} = O(1)$ of variables), one can set as in [13]*

$$G = \sum_{k=1}^{m} \sum_{l=1}^{m_k} Z_{kl} \hat{G}_{kl} Z_{kl}^T u_{kl}(x,z),$$

  where the reduced Hessian matrices $\hat{G}_{kl}$ are updated using the reduced vectors $\hat{d}_{kl} = Z_{kl}^T(x_+ - x)$ and $\hat{y}_{kl} = Z_{kl}(g_{kl}(x_+) - g_{kl}(x))$.

Remark 23. *The matrix $G \approx G(x;\mu)$ obtained by the approach stated in Remark 22 can be ill-conditioned so condition (99) (with a chosen value $\varepsilon_0 > 0$) may not be satisfied. In this case it is possible to restart the iteration process and set $G = I$. Then $\overline{G} = 1$ and $\underline{G} = 1$ in (97) and (98), so it is a higher probability of fulfilment of condition (99). If the choice $G = I$ does not satisfy (99), we set $\Delta x = -g(x;\mu)$ (a steepest descent direction).*

An update of μ is an important part of interior point methods. Above all, $\mu \to 0$ must hold, which is a main property of interior point methods. Moreover, rounding errors may cause that $z_k(x;\mu) = F_k(x)$ when the value μ is small (because $F_k(x) < z_k(x;\mu) \le F_k(x) + m_k\mu$ and $F_k(x) + m_k\mu \to F_k(x)$ if $\mu \to 0$), which leads to a breakdown (division by $z_k(x;\mu) - F_k(x) = 0$) when computing $1/(z_k(x;\mu) - F_k(x))$. Therefore, we need to use a lower bound $\underline{\mu}$ for a barrier parameter (e.g. $\underline{\mu} = 10^{-8}$ when computing in double precision).

The efficiency of interior point methods also depends on the way of decreasing the value of a barrier parameter. The following heuristic procedures proved successful in practice, where $g(x_i;\mu_i) = A(x_i)u(x_i;\mu_i)$ and $\underline{g}$ is a suitable constant.

Procedure A

Phase 1 If $\|g(x_i;\mu_i)\| \geq \underline{g}$, then $\mu_{i+1} = \mu_i$ (the value of a barrier parameter is unchanged).

Phase 2 If $\|g(x_i;\mu_i)\| < \underline{g}$, then

$$\mu_{i+1} = \max\left(\tilde{\mu}_{i+1},\, \underline{\mu},\, 10\,\varepsilon_M |F(x_{i+1})|\right), \tag{103}$$

where $F(x_{i+1}) = F_1(x_{i+1}) + \cdots + F_m(x_{i+1})$, ε_M is a machine precision, and

$$\tilde{\mu}_{i+1} = \min\left[\max(\lambda\mu_i,\, \mu_i/(\sigma\mu_i + 1)),\, \max(\|g(x_i;\mu_i)\|^2,\, 10^{-2k})\right]. \tag{104}$$

The values $\underline{\mu} = 10^{-8}$, $\lambda = 0.85$, and $\sigma = 100$ are usually used.

Procedure B

Phase 1 If $\|g(x_i;\mu_i)\|^2 \geq \vartheta\mu_i$, then $\mu_{i+1} = \mu_i$ (the value of a barrier parameter is unchanged).

Phase 2 If $\|g(x_i;\mu_i)\|^2 < \vartheta\mu_i$, then

$$\mu_{i+1} = \max(\underline{\mu}, \|g_i(x_i;\mu_i)\|^2). \tag{105}$$

The values $\underline{\mu} = 10^{-8}$ and $\vartheta = 0.1$ are usually used.

The choice of $\underline{g}$ in Procedure A is not critical. We can set $\underline{g} = \infty$ but a lower value is sometimes more advantageous. Formula (104) requires several comments. The first argument of the minimum controls the decreasing speed of the value of a barrier parameter which is linear (a geometric sequence) for small i (the term $\lambda\mu_i$) and sublinear (a harmonic sequence) for large i (the term $\mu_i/(\sigma\mu_i + 1)$). Thus, the second argument ensuring that the value μ is small in a neighborhood of a desired solution is mainly important for large i. This situation may appear if the gradient norm $\|g(x_i;\mu_i)\|$ is small even if x_i is far from a solution. The idea of Procedure B proceeds from the fact that a barrier function $B(x;\mu)$ should be minimized with a sufficient precision for a given value of a parameter μ.

The considerations up to now are summarized in the following algorithm which supposes that the matrix $A(x)$ is sparse. If it is dense, the algorithm is simplified because there is no symbolic decomposition.

Algorithm 2. *Primal interior point method*

Data. *A tolerance for the gradient norm of the Lagrange function $\underline{\varepsilon} > 0$. A precision for determination of a minimax vector $\underline{\delta} > 0$. Bounds for a barrier parameter $0 < \underline{\mu} < \overline{\mu}$. Coefficients for decrease of a barrier parameter $0 < \lambda < 1$, $\sigma > 1$ (or $0 < \vartheta < 1$). A tolerance for a uniform descent $\varepsilon_0 > 0$. A tolerance for a steplength selection $\varepsilon_1 > 0$. A maximum steplength $\overline{\Delta} > 0$.*

Input. *A sparsity pattern of the matrix $A(x) = [A_1(x), \ldots, A_m(x)]$. A starting point $x \in R^n$.*

Step 1 Initiation. *Choose $\mu \leq \overline{\mu}$. Determine a sparse structure of the matrix $W = W(x;\mu)$ from the sparse structure of the matrix $A(x)$ and perform a symbolic decomposition of the matrix W (described in [2, Section 1.7.4]). Compute values $f_{kl}(x)$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, values $F_k(x) = \max_{1\leq l\leq m_k} f_{kl}(x)$, $1 \leq k \leq m$, and the value of objective function (4). Set $r = 0$ (restart indicator).*

Step 2 Termination. *Solve nonlinear equations (89) with precision $\underline{\delta}$ to obtain a minimax variable $z(x;\mu)$ and a vector of Lagrange multipliers $u(x;\mu)$. Determine a matrix $A = A(x)$ and a vector $g = g(x;\mu) = A(x)u(x;\mu)$. If $\mu \leq \underline{\mu}$ and $\|g\| \leq \underline{\varepsilon}$, terminate the computation.*

Step 3 Hessian matrix approximation. *Set $G = G(x;\mu)$ or compute an approximation G of the Hessian matrix $G(x;\mu)$ using gradient differences or using quasi-Newton updates (Remark 22).*

Step 4 Direction determination. *Determine a matrix $\nabla^2 B(x;\mu)$ by (93) and a vector Δx by solving equations (94) with the right-hand side defined by (92).*

Step 5 Restart. *If $r = 0$ and (99) does not hold (where $s = \Delta x$), set $G = I$, $r = 1$ and go to Step 4. If $r = 1$ and (99) does not hold, set $\Delta x = -g$. Set $r = 0$.*

Step 6 Steplength selection. *Determine a steplength $\alpha > 0$ satisfying inequalities (100) (for a barrier function $B(x;\mu)$ defined by (91)) and $\alpha \le \overline{\Delta}/\|\Delta x\|$. Note that nonlinear equations (89) are solved at the point $x + \alpha\Delta x$. Set $x := x + \alpha\Delta x$. Compute values $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, values $F_k(x) = \max_{1\le l\le m_k} f_{kl}(x)$, $1 \le k \le m$, and the value of objective function (4).*

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

The values $\underline{\varepsilon} = 10^{-6}$, $\underline{\delta} = 10^{-6}$, $\underline{\mu} = 10^{-8}$, $\overline{\mu} = 1$, $\lambda = 0.85$, $\sigma = 100$, $\vartheta = 0.1$, $\varepsilon_0 = 10^{-8}$, $\varepsilon_1 = 10^{-4}$, and $\overline{\Delta} = 1000$ were used in our numerical experiments.

3.5 Global convergence

Now we prove the global convergence of the method realized by Algorithm 2.

Lemma 4. *Let vectors $z_k(x;\mu)$, $1 \le k \le m$, be solutions of equations (89). Then*

$$\frac{\partial z_k(x;\mu)}{\partial \mu} > 0, \quad 1 \le k \le m, \qquad \frac{\partial B(x;\mu)}{\partial \mu} = -\sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)).$$

Proof. Differentiating (89) with respect to μ, one can write for $1 \le k \le m$

$$-\sum_{l=1}^{m_k} \frac{1}{z_k(x;\mu) - f_{kl}(x)} + \sum_{l=1}^{m_k} \frac{\mu}{(z_k(x;\mu) - f_{kl}(x))^2} \frac{\partial z_k(x;\mu)}{\partial \mu} = 0,$$

which after multiplication of μ together with (71) and (89) gives

$$\frac{\partial z_k(x;\mu)}{\partial \mu} = \left(\sum_{l=1}^{m_k} \frac{\mu^2}{(z_k(x;\mu) - f_{kl}(x))^2}\right)^{-1} = \left(\sum_{l=1}^{m_k} u_{kl}^2(x;\mu)\right)^{-1} > 0.$$

Differentiating a function

$$B(x;\mu) = \sum_{k=1}^{m} z_k(x;\mu) - \mu \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)) \tag{106}$$

and using (89) we obtain

$$\begin{aligned}
\frac{\partial B(x;\mu)}{\partial \mu} &= \sum_{k=1}^{m} \frac{\partial z_k(x;\mu)}{\partial \mu} - \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)) - \sum_{k=1}^{m}\sum_{l=1}^{m_k} \frac{\mu}{z_k(x;\mu) - f_{kl}(x)} \frac{\partial z_k(x;\mu)}{\partial \mu} \\
&= \frac{\partial z_k(x;\mu)}{\partial \mu} \sum_{k=1}^{m} \left(1 - \sum_{l=1}^{m_k} \frac{\mu}{z_k(x;\mu) - f_{kl}(x)}\right) - \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)) \\
&= -\sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)).
\end{aligned}$$

□

Lemma 5. *Let Assumption X1a be satisfied. Let x_i and μ_i, $i \in N$, be the sequences generated by Algorithm 2. Then the sequences $B(x_i;\mu_i)$, $z(x_i;\mu_i)$, and $F(x_i)$, $i \in N$, are bounded. Moreover, there exists a constant $L \ge 0$ such that for $i \in N$ it holds*

$$B(x_{i+1};\mu_{i+1}) \le B(x_{i+1};\mu_i) + L(\mu_i - \mu_{i+1}). \tag{107}$$

Proof.

(a) We first prove boundedness from below. Using (106) and Assumption X1a, one can write

$$\begin{aligned} B(x;\mu) - \underline{F} &= \sum_{k=1}^{m} z_k(x;\mu) - \underline{F} - \mu \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x;\mu) - f_{kl}(x)) \\ &\geq \sum_{k=1}^{m} \left(z_k(x;\mu) - \underline{F} - m_k\mu \log(z_k(x;\mu) - \underline{F})\right). \end{aligned}$$

A convex function $\psi(t) = t - m\mu\log(t)$ has a unique minimum at the point $t = m\mu$ because $\psi'(m\mu) = 1 - m\mu/m\mu = 0$. Thus, it holds

$$\begin{aligned} B(x;\mu) &\geq \underline{F} + \sum_{k=1}^{m}(m_k\mu - m_k\mu\log(m_k\mu)) \geq \underline{F} + \sum_{k=1}^{m} \min\left(0,\, m_k\overline{\mu}(1 - \log(m_k\overline{\mu})\right) \\ &\geq \underline{F} + \sum_{k=1}^{m} \min\left(0,\, m_k\overline{\mu}(1 - \log(2m_k\overline{\mu})\right) \triangleq \underline{B}. \end{aligned}$$

Boundedness from below of sequences $z(x_i;\mu_i)$ and $F(x_i)$, $i \in N$, follows from inequalities (90) and Assumption X1a.

(b) Now we prove boundedness from above. Similarly as in (a) we can write

$$B(x;\mu) - \underline{F} \geq \sum_{k=1}^{m} \frac{z_k(x;\mu) - \underline{F}}{2} + \sum_{k=1}^{m} \left(\frac{z_k(x;\mu) - \underline{F}}{2} - m_k\mu\log(z(x;\mu) - \underline{F})\right).$$

A convex function $t/2 - m\mu\log(t)$ has a unique minimum at the point $t = 2m\mu$. Thus, it holds

$$B(x;\mu) \geq \sum_{k=1}^{m} \frac{z_k(x;\mu) - \underline{F}}{2} + \underline{F} + \sum_{k=1}^{m} \min\left(0,\, m\overline{\mu}(1 - \log(2m_k\overline{\mu})\right) = \sum_{k=1}^{m} \frac{z_k(x;\mu) - \underline{F}}{2} + \underline{B}$$

or

$$\sum_{k=1}^{m}(z_k(x;\mu) - \underline{F}) \leq 2(B(x;\mu) - \underline{B}). \tag{108}$$

Using the mean value theorem and Lemma 4, we obtain

$$\begin{aligned} B(x_{i+1};\mu_{i+1}) - B(x_{i+1};\mu_i) &= \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x_{i+1};\tilde{\mu}_i) - f_{kl}(x_{i+1}))(\mu_i - \mu_{i+1}) \\ &\leq \sum_{k=1}^{m}\sum_{l=1}^{m_k} \log(z_k(x_{i+1};\mu_i) - f_{kl}(x_{i+1}))(\mu_i - \mu_{i+1}) \\ &\leq \sum_{k=1}^{m} m_k \log(z_k(x_{i+1};\mu_i) - \underline{F})(\mu_i - \mu_{i+1}), \end{aligned} \tag{109}$$

where $0 < \mu_{i+1} \leq \tilde{\mu}_i \leq \mu_i$. Since $\log(t) \leq t/e$ (where $e = \exp(1)$) for $t > 0$, we can write using inequalities (108), (109), (90)

$$\begin{aligned} B(x_{i+1};\mu_{i+1}) - \underline{B} &\leq B(x_{i+1};\mu_i) - \underline{B} + \sum_{k=1}^{m} m_k \log(z_k(x_{i+1};\mu_i) - \underline{F})(\mu_i - \mu_{i+1}) \\ &\leq B(x_{i+1};\mu_i) - \underline{B} + e^{-1}\sum_{k=1}^{m} m_k(z_k(x_{i+1};\mu_i) - \underline{F})(\mu_i - \mu_{i+1}) \\ &\leq B(x_{i+1};\mu_i) - \underline{B} + 2e^{-1}\overline{m}(B(x_{i+1};\mu_i) - \underline{B})(\mu_i - \mu_{i+1}) \\ &= (1 + \lambda\delta_i)(B(x_{i+1};\mu_i) - \underline{B}) \leq (1 + \lambda\delta_i)(B(x_i;\mu_i) - \underline{B}), \end{aligned}$$

where $\lambda = 2\overline{m}/e$ and $\delta_i = \mu_i - \mu_{i+1}$. Therefore,

$$B(x_{i+1};\mu_{i+1}) - \underline{B} \le \prod_{j=1}^{i}(1+\lambda\delta_j)(B(x_1;\mu_1) - \underline{B}) \le \prod_{i=1}^{\infty}(1+\lambda\delta_i)(B(x_1;\mu_1) - \underline{B}) \tag{110}$$

and since

$$\sum_{i=1}^{\infty}\lambda\delta_i = \lambda\sum_{i=1}^{\infty}(\mu_i - \mu_{i+1}) = \lambda(\overline{\mu} - \lim_{i\to\infty}\mu_i) \le \lambda\overline{\mu},$$

the expression on the right-hand side of (110) is finite. Thus, the sequence $B(x_i;\mu_i)$, $i \in N$, is bounded from above and the sequences $z(x_i;\mu_i)$ and $F(x_i)$, $i \in N$, are bounded from above as well by (108) and (90).

(c) Finally, we prove formula (107). Using (109) and (90) we obtain

$$\begin{aligned} B(x_{i+1};\mu_{i+1}) - B(x_{i+1};\mu_i) &\le \sum_{k=1}^{m} m_k \log(z_k(x_{i+1};\mu_i) - \underline{F})(\mu_i - \mu_{i+1}) \\ &\le \sum_{k=1}^{m} m_k \log(F_k(x_{i+1}) + m_k\mu_i - \underline{F})(\mu_i - \mu_{i+1}) \\ &\le \sum_{k=1}^{m} m_k \log(\overline{F} + m_k\overline{\mu} - \underline{F})(\mu_i - \mu_{i+1}) \triangleq L(\mu_i - \mu_{i+1}) \end{aligned}$$

(the existence of a constant $\overline{F}$ follows from boundedness of a sequence $F(x_i)$, $i \in N$), which together with (102) gives $B(x_{i+1};\mu_{i+1}) \le B(x_i;\mu_i) + L(\mu_i - \mu_{i+1})$, $i \in N$. Thus, it holds

$$B(x_i;\mu_i) \le B(x_1;\mu_1) + L(\mu_1 - \mu_i) \le B(x_1;\mu_1) + L\overline{\mu} \triangleq \overline{B}, \qquad i \in N. \tag{111}$$

□

The upper bounds $\overline{g}$ and $\overline{G}$ are not used in Lemma 5, so Assumption X3 may not be satisfied. Thus, there exists an upper bound $\overline{F}$ (independent of $\overline{g}$ and $\overline{G}$) such that $F(x_i) \le \overline{F}$ for all $i \in N$. This upper bound can be used in definition of a set $\mathcal{D}_F(\overline{F})$ in Assumption X3.

Lemma 6. *Let Assumption X3 and the assumptions of Lemma 5 be satisfied. Then, if we use Procedure A or Procedure B for an update of parameter μ, the values μ_i, $i \in N$, form a non-decreasing sequence such that $\mu_i \to 0$.*

Proof. The value of parameter μ is unchanged in the first phase of Procedure A or Procedure B. Since a function $B(x;\mu)$ is continuous, bounded from below by Lemma 5, and since inequality (101) is satisfied (with $\mu_i = \mu$), it holds $\|g(x_i;\mu)\| \to 0$ if phase 1 contains an infinite number of subsequent iterative steps [32, Section 3.2]. Thus, there exists a step (with index i) belonging to the first phase such that either $\|g(x_i;\mu)\| < \underline{g}$ in Procedure A or $\|g(x_i;\mu)\|^2 < \vartheta\mu$ in Procedure B. However, this is in contradiction with the definition of the first phase. Thus, there exists an infinite number of steps belonging to the second phase, where the value of parameter μ is decreased so that $\mu_i \to 0$. □

Theorem 8. *Let assumptions of Lemma 6 be satisfied. Consider a sequence x_i, $i \in N$, generated by Algorithm 2, where $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$. Then*

$$\lim_{i\to\infty}\sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x_i)u_{kl}(x_i;\mu_i) = 0, \quad \sum_{l=1}^{m_k} u_{kl}(x_i;\mu_i) = 1,$$

$$z_k(x_i;\mu_i) - f_{kl}(x_i) \ge 0, \quad u_{kl}(x_i;\mu_i) \ge 0,$$

$$\lim_{i\to\infty} u_{kl}(x_i;\mu_i)(z_k(x_i;\mu_i) - f_{kl}(x_i)) = 0$$

for $1 \le k \le m$ and $1 \le l \le m_k$.

Proof.

(a) Equalities $\tilde{e}_k^T u_k(x_i;\mu_i) = 1$, $1 \le k \le m$, are satisfied by (89) because $\underline{\delta} = 0$. Inequalities $z_k(x_i;\mu_i) - f_{kl}(x_i) \ge 0$ and $u_{kl}(x_i;\mu_i) \ge 0$ follow from formulas (90) and statement (95).

(b) Relations (101) and (107) yield

$$\begin{aligned} B(x_{i+1};\mu_{i+1}) - B(x_i;\mu_i) &= (B(x_{i+1};\mu_{i+1}) - B(x_{i+1};\mu_i)) + (B(x_{i+1};\mu_i) - B(x_i;\mu_i)) \\ &\le L\,(\mu_i - \mu_{i+1}) - c\,\|g(x_i;\mu_i)\|^2 \end{aligned}$$

and since $\lim_{i\to\infty} \mu_i = 0$ (Lemma 6), we can write by (111) that

$$\begin{aligned} \underline{B} &\le \lim_{i\to\infty} B(x_{i+1};\mu_{i+1}) \le B(x_1;\mu_1) + L\sum_{i=1}^{\infty}(\mu_i - \mu_{i+1}) - c\sum_{i=1}^{\infty}\|g(x_i;\mu_i)\|^2 \\ &\le B(x_1;\mu_1) + L\overline{\mu} - c\sum_{i=1}^{\infty}\|g(x_i;\mu_i)\|^2 = \overline{B} - c\sum_{i=1}^{\infty}\|g(x_i;\mu_i)\|^2. \end{aligned}$$

Thus, it holds

$$\sum_{i=1}^{\infty}\|g(x_i;\mu_i)\|^2 \le \frac{1}{c}(\overline{B} - \underline{B}) < \infty,$$

which gives $g(x_i;\mu_i) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x_i)u_{kl}(x_i;\mu_i) \to 0$.

(c) Let indices $1 \le k \le m$ and $1 \le l \le m_k$ are chosen arbitrarily. Using (95) and Lemma 6 we obtain

$$u_{kl}(x_i;\mu_i)(z_k(x_i;\mu_i) - f_{kl}(x_i)) = \frac{\mu_i}{z_k(x_i;\mu_i) - f_{kl}(x_i)}(z_k(x_i;\mu_i) - f_{kl}(x_i)) = \mu_i \to 0.$$

□

Corollary 1. *Let the assumptions of Theorem 8 be satisfied. Then, every cluster point $x \in R^n$ of a sequence x_i, $i \in N$, satisfies necessary KKT conditions (8)-(9) where z and u (with elements z_k and u_{kl}, $1 \le k \le m$, $1 \le l \le m_k$) are cluster points of sequences $z(x_i;\mu_i)$ and $u(x_i;\mu_i)$, $i \in N$.*

Now we will suppose that the values $\underline{\delta}$, $\underline{\varepsilon}$, and $\underline{\mu}$ are nonzero and show how a precise solution of the system of KKT equations will be after termination of computation.

Theorem 9. *Let the assumptions of Lemma 6 be satisfied. Consider a sequence x_i, $i \in N$, generated by Algorithm 2. Then, if the values $\underline{\delta} > 0$, $\underline{\varepsilon} > 0$, and $\underline{\mu} > 0$ are chosen arbitrarily, there exists an index $i \ge 1$ such that*

$$\|g(x_i;\mu_i)\| \le \underline{\varepsilon}, \quad \left|1 - \sum_{l=1}^{m_k} u_{kl}(x_i;\mu_i)\right| \le \underline{\delta},$$

$$z_k(x_i;\mu_i) - f_{kl}(x_i) \ge 0, \quad u_{kl}(x_i;\mu_i) \ge 0,$$

$$u_{kl}(x_i;\mu_i)(z_k(x_i;\mu_i) - f_{kl}(x_i)) \le \underline{\mu}$$

for $1 \le k \le m$ and $1 \le l \le m_k$.

Proof. Inequality $|1 - \tilde{e}_k^T u_k(x_i;\mu_i)| \le \underline{\delta}$ follows immediately from the fact that the equation $\tilde{e}_k^T u_k(x_i;\mu_i) = 1$, $1 \le k \le m$, is solved with precision $\underline{\delta}$. Inequalities $z_k(x_i;\mu_i) - f_{kl}(x_i) \ge 0$, $u_{kl}(x_i;\mu_i) \ge 0$ follow from formulas (90) and statement (95) as in the proof of Theorem 8. Since $\mu_i \to 0$ and $g(x_i;\mu_i) \to 0$ by Lemma 6 and Theorem 8, there exists an index $i \ge 1$ such that $\mu_i \le \underline{\mu}$ and $\|g(x_i;\mu_i)\| \le \underline{\varepsilon}$. Using (95) we obtain

$$u_{kl}(x_i;\mu_i)(z_k(x_i;\mu_i) - f_{kl}(x_i)) = \frac{\mu_i}{z_k(x_i;\mu_i) - f_{kl}(x_i)}(z_k(x_i;\mu_i) - f_{kl}(x_i)) = \mu_i \le \underline{\mu}.$$

□

3.6 Special cases

Both the simplest and most widely considered generalized minimax problem is the classical minimax problem (10), when $m = 1$ in (4) (in this case we write m, z, u, v, U, V instead of m_1, z_1, u_1, v_1, U_1, V_1). For solving a classical minimax problem one can use Algorithm 2, where a major part of computation is very simplified. System of equations (79) is of order $n+1$ and has the form

$$\begin{bmatrix} G(x,z) + A(x)V(x,z)A^T(x) & -A(x)V(x,z)\tilde{e} \\ -\tilde{e}^T V(x,z)A^T(x) & \tilde{e}^T V(x,z)\tilde{e} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} g(x,z) \\ \gamma(x,z) \end{bmatrix}, \tag{112}$$

where $g(x,z) = A(x)u(x,z)$, $\gamma(x,z) = 1 - \tilde{e}^T u(x,z)$, $V(x,z) = U^2(x,z)/\mu = \mathrm{diag}(u_1^2(x,z), \ldots, u_m^2(x,z))/\mu$, and $u_k(x,z) = \mu/(z - f_k(x)$, $1 \le k \le m$. System of equations (89) is reduced to one nonlinear equation

$$1 - \tilde{e}^T u(x,z) = 1 - \sum_{k=1}^{m} \frac{\mu}{z - f_k(x)} = 0, \tag{113}$$

whose solution $z(x;\mu)$ lies in the interval $F(x) + \mu \le z(x;\mu) \le F(x) + m\mu$. To find this solution by robust methods from [16], [17] is not difficult. A barrier function has the form

$$B(x;\mu) = z(x;\mu) - \mu \sum_{k=1}^{m} \log(z(x;\mu) - f_k(x)) \tag{114}$$

with

$$\begin{aligned} \nabla B(x;\mu) &= A(x)u(x;\mu), \\ \nabla^2 B(x;\mu) &= G(x;\mu) + A(x)V(x;\mu)A^T(x) - \frac{A(x)V(x;\mu)\tilde{e}\tilde{e}^T V(x;\mu)A^T(x)}{\tilde{e}^T V(x;\mu)\tilde{e}}. \end{aligned}$$

If we write system (112) in the form

$$\begin{bmatrix} W(x,z) & -c(x,z) \\ -c^T(x,z) & \delta(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} g(x,z) \\ \gamma(x,z) \end{bmatrix}$$

where $W(x,z) = G(x,z) + A(x)V(x,z)A^T(x)$, $c(x,z) = A(x)V(x,z)\tilde{e}$ and $\delta(x,z) = \tilde{e}^T V(x,z)\tilde{e}$, then

$$\nabla^2 B(x;\mu) = W(x;\mu) - \frac{c(x;\mu)c^T(x;\mu)}{\delta(x;\mu)}.$$

Since

$$\begin{bmatrix} W & -c \\ -c^T & \delta \end{bmatrix}^{-1} = \begin{bmatrix} W^{-1} - W^{-1}c\,\omega^{-1}c^T H^{-1} & -W^{-1}c\,\omega^{-1} \\ -\omega^{-1}c^T W^{-1} & -\omega^{-1} \end{bmatrix},$$

where $\omega = c^T W^{-1} c - \delta$, we can write

$$\begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = - \begin{bmatrix} W & -c \\ -c^T & \delta \end{bmatrix}^{-1} \begin{bmatrix} g \\ \gamma \end{bmatrix} = \begin{bmatrix} W^{-1}(c\,\Delta z - g) \\ \Delta z \end{bmatrix},$$

where

$$\Delta z = \omega^{-1}(c^T W^{-1} g + \gamma).$$

The matrix W is sparse if the matrix $A(x)$ has sparse columns. If the matrix W is not positive definite, we can use the Gill-Murray decomposition

$$W + E = LL^T, \tag{115}$$

where E is a positive semidefinite diagonal matrix. Then we solve the equations

$$LL^T p = g, \quad LL^T q = c \tag{116}$$

and set

$$\Delta z = \frac{c^T p + \gamma}{c^T q - \delta}, \quad \Delta x = q\,\Delta z - p. \tag{117}$$

If we solve the classical minimax problem, Algorithm 2 must be somewhat modified. In Step 2, we solve only equation (113) instead of the system of equations (89). In Step 4, we determine a vector Δx by solving equations (116) and using relations (117). In Step 4, we use the barrier function (114) (nonlinear equation (113) must be solved at the point $x + \alpha\Delta x$).

Minimization of a sum of absolute values, i.e., minimization of the function

$$F(x) = \sum_{k=1}^{m} |f_k(x)| = \sum_{k=1}^{m} \max(f_k^+(x), f_k^-(x)), \qquad f_k^+(x) = f_k(x), \qquad f_k^-(x) = -f_k(x),$$

is another important generalized minimax problem. In this case, a barrier function has the form

$$\begin{aligned} B_\mu(x,z) &= \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \log(z_k - f_k^+(x)) - \mu \sum_{k=1}^{m} \log(z_k - f_k^-(x)) \\ &= \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \log(z_k - f_k(x)) - \mu \sum_{k=1}^{m} \log(z_k + f_k(x)) \\ &= \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \log(z_k^2 - f_k^2(x)), \end{aligned} \tag{118}$$

where $z_k > |f_k(x)|$, $1 \le k \le m$. Differentiating $B_\mu(x,z)$ with respect to x and z we obtain the necessary conditions for an extremum

$$\sum_{k=1}^{m} \frac{2\mu f_k(x)}{z_k^2 - f_k^2(x)}\, g_k(x) = \sum_{k=1}^{m} u_k(x,z_k)\, g_k(x) = 0, \qquad u_k(x,z_k) = \frac{2\mu f_k(x)}{z_k^2 - f_k^2(x)} \tag{119}$$

and

$$1 - \frac{2\mu z_k}{z_k^2 - f_k^2(x)} = 1 - u_k(x,z_k)\frac{z_k}{f_k(x)} = 0 \quad \Rightarrow \quad u_k(x,z_k) = \frac{f_k(x)}{z_k}, \quad 1 \le k \le m. \tag{120}$$

Denoting $A(x) = [g_1(x), \ldots, g_m(x)]$,

$$f(x) = \begin{bmatrix} f_1(x) \\ \ldots \\ f_m(x) \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ \ldots \\ z_m \end{bmatrix}, \quad u(x,z) = \begin{bmatrix} u_1(x,z_1) \\ \ldots \\ u_m(x,z_m) \end{bmatrix} \tag{121}$$

and $Z = \mathrm{diag}(z_1, \ldots, z_m)$, one can write

$$A(x)u(x,z) = 0, \qquad u(x,z) = Z^{-1} f(x). \tag{122}$$

Let $\tilde{B}_\mu(z) = B_\mu(x,z)$ for a fixed chosen vector $x \in R^n$. The function $\tilde{B}_\mu(z)$ is convex for $z_k > |f_k(x)|$, $1 \le k \le m$, because it is a sum of convex functions. Thus, a stationary point of $\tilde{B}_\mu(z)$ exists and it is its global minimum. Differentiating $\tilde{B}_\mu(z)$ with respect to z we obtain quadratic equations

$$\frac{2\mu z_k(x;\mu)}{z_k^2(x;\mu) - f_k^2(x)} = 1 \quad \Leftrightarrow \quad z_k^2(x;\mu) - f_k^2(x) = 2\mu z_k(x;\mu), \quad 1 \le k \le m, \tag{123}$$

defining its unique stationary point, that have a solution

$$z_k(x;\mu) = \mu + \sqrt{\mu^2 + f_k^2(x)}, \quad 1 \le k \le m, \tag{124}$$

(the second solutions of quadratic equations (123) do not satisfy the condition $z_k > |f_k(x)|$, so the obtained vector z does not belong to a domain of $\tilde{B}_\mu(z)$). Using (120) and (124) we obtain

$$u_k(x;\mu) = u_k(x, z(x;\mu)) = \frac{f_k(x)}{z_k(x;\mu)} = \frac{f_k(x)}{\mu + \sqrt{\mu^2 + f_k^2(x)}}, \quad 1 \le k \le m, \tag{125}$$

and

$$\begin{aligned} B(x;\mu) &= B(x, z(x;\mu)) = \sum_{k=1}^{m} z_k(x;\mu) - \mu \sum_{k=1}^{m} \log(z_k^2(x;\mu) - f_k^2(x)) \\ &= \sum_{k=1}^{m} z_k(x;\mu) - \mu \sum_{k=1}^{m} \log(2\mu z_k(x;\mu)) \\ &= \sum_{i=1}^{m} [z_i(x;\mu) - \mu \log(z_i(x;\mu))] - \mu m \log(2\mu). \end{aligned} \tag{126}$$

Theorem 10. *Consider barrier function (126). Then*

$$\nabla B(x;\mu) = A(x)u(x;\mu) \tag{127}$$

and

$$\nabla^2 B(x;\mu) = W(x;\mu) = G(x;\mu) + A(x)V(x;\mu)A^T(x), \tag{128}$$

where

$$G(x;\mu) = \sum_{k=1}^{m} G_k(x) u_k(x;\mu), \tag{129}$$

$G_k(x)$ are the Hessian matrices of functions $f_k(x)$, $1 \le k \le m$, $V(x;\mu) = \operatorname{diag}(v_1(x;\mu), \ldots, v_m(x;\mu))$, and

$$v_k(x;\mu) = \frac{2\mu}{z_k^2(x;\mu) + f_k^2(x)}, \qquad 1 \le k \le m. \tag{130}$$

Proof. Differentiating (126) and using (123) and (119) we can write

$$\begin{aligned} \nabla B(x;\mu) &= \sum_{k=1}^{m} \nabla z_k(x;\mu) - 2\mu \sum_{i=1}^{m} \frac{z_i(x;\mu)\nabla z_k(x;\mu) - f_k(x)g_k(x)}{z_k^2(x;\mu) - f_k^2(x)} \\ &= \sum_{k=1}^{m} \left(1 - \frac{2\mu z_k(x;\mu)}{z_k^2(x;\mu) - f_k^2(x)}\right) \nabla z_k(x;\mu) + \sum_{k=1}^{m} \frac{2\mu f_k(x) g_k(x)}{z_k^2(x;\mu) - f_k^2(x)} \\ &= \sum_{k=1}^{m} u_k(x;\mu) g_k(x) = A(x)u(x;\mu). \end{aligned}$$

Differentiating (123) we obtain

$$\frac{\nabla z_k(x;\mu)}{z_k^2(x;\mu) - f_k^2(x)} - \frac{2z_k(x;\mu)(z_k(x;\mu)\nabla z_k(x;\mu) - f_k(x)g_k(x))}{(z_k^2(x;\mu) - f_k^2(x))^2} = 0$$

for $1 \le k \le m$, which after arrangements gives

$$\nabla z_k(x;\mu) = \frac{2z_k(x;\mu) f_k(x) g_k(x)}{z_k^2(x;\mu) + f_k^2(x)} \tag{131}$$

for $1 \le k \le m$. Thus, by (125), (131), (123), and (127) it holds

$$\begin{aligned}
\nabla u_k(x;\mu) &= \nabla\left(\frac{f_k(x)}{z_k(x;\mu)}\right) = \frac{z_k(x;\mu)g_k(x) - f_k(x)\nabla z_k(x;\mu)}{z_k^2(x;\mu)} \\
&= \left(1 - \frac{2f_k^2(x)}{z_k^2(x;\mu)+f_k^2(x)}\right)\frac{g_k(x)}{z_k(x;\mu)} = \frac{z_k^2(x;\mu)-f_k^2(x)}{z_k^2(x;\mu)+f_k^2(x)}\,\frac{g_k(x)}{z_k(x;\mu)} \\
&= \frac{2\mu}{z_k^2(x;\mu)+f_k^2(x)}g_k(x) = v_k(x;\mu)g_k(x).
\end{aligned}$$

Differentiating (127) and using the previous expression we obtain

$$\begin{aligned}
\nabla^2 B(x;\mu) &= \nabla\sum_{k=1}^m u_k(x;\mu)g_k(x) = \sum_{k=1}^m u_k(x;\mu)G_k(x) + \sum_{k=1}^m \nabla u_k(x;\mu)g_k^T(x) \\
&= \sum_{k=1}^m u_k(x;\mu)G_k(x) + \sum_{k=1}^m v_k(x;\mu)g_k(x)g_k^T(x),
\end{aligned}$$

which is equation (128). □

A vector $\Delta x \in R^n$ is determined by solving the equation

$$\nabla^2 B(x;\mu)\Delta x = -g(x;\mu), \tag{132}$$

where $g(x;\mu) = \nabla B(x;\mu) \ne 0$. From (132) it follows

$$\begin{aligned}
(\Delta x)^T g(x;\mu) &= -(\Delta x)^T\nabla^2 B(x;\mu)\Delta x = -(\Delta x)^T G(x;\mu)\Delta x - (\Delta x)^T A(x)V(x;\mu)A^T(x)\Delta x \\
&\le -(\Delta x)^T G(x;\mu)\Delta x,
\end{aligned}$$

so if a matrix $G(x;\mu)$ is positive definite, a matrix $\nabla B(x;\mu)$ is positive definite as well (since a diagonal matrix $V(x;\mu)$ is positive definite by (130)) and $(\Delta x)^T g(x;\mu) < 0$ holds (a direction vector Δx is descent for a function $B(x;\mu)$).

By (130), a norm of a matrix $V(x;\mu)$ is bounded from above if the numbers $f_k^2(x)$, $1 \le k \le m$, are sufficiently positive. If $f_k^2(x)$ tends to zero faster than μ, then the element $v_k(x;\mu)$ may tend to infinity and a matrix $\nabla^2 B(x;\mu)$ (given by (128)) may be ill-conditioned. However, if Assumption X3 is satisfied and if $0 < \underline{\mu} \le \mu \le \overline{\mu}$ holds, one can write

$$\begin{aligned}
\|\nabla^2 B(x;\mu)\| &= \left\|G(x;\mu) + A(x)V(x;\mu)A^T(x)\right\| \\
&\le \left\|\sum_{k=1}^m u_k(x;\mu)G_k(x)\right\| + \left\|\sum_{k=1}^m v_k(x;\mu)g_k(x)g_k^T(x)\right\| \\
&\le m\overline{G} + m\overline{g}^2\|V(x;\mu)\|
\end{aligned}$$

because $|u_k(x;\mu)| \le 1$, $1 \le k \le m$, holds by (125). Since a matrix $V(x;\mu)$ is diagonal, we can write by (130) that

$$\|V(x;\mu)\| = \max_{1\le k\le m}|v_k(x;\mu)| = \max_{1\le k\le m}\left(\frac{2\mu}{z_k^2(x;\mu)+f_k^2(x)}\right). \tag{133}$$

Using (123) and (124) we obtain

$$z_k^2(x;\mu) + f_k^2(x) = 2\mu z_k(x;\mu) = 2\mu\left(\mu + \sqrt{\mu^2 + f_k^2(x)}\right) \ge 4\mu^2$$

for $1 \le k \le m$, which after substitution to (133) gives $\|V(x;\mu)\| \le 1/(2\mu)$. Thus, inequality

$$\|\nabla^2 B(x;\mu)\| \le \frac{\overline{c}}{\mu} \le \frac{\overline{c}}{\underline{\mu}} \tag{134}$$

where $\bar{c} = m(\overline{\mu}\,\overline{G} + \overline{g}^2/2)$ is satisfied.

A slightly modified Algorithm 2 can be used for minimization of a sum of absolute values. However, the problems of this type are characterized by ill-posedness of a matrix $\nabla^2 B(x;\mu)$. Thus, it is more convenient to use trust region methods [22]. In this case, a direction vector Δx is determined by approximate minimization of a quadratic function

$$Q(\Delta x) = \frac{1}{2}(\Delta x)^T \nabla^2 B(x;\mu)\Delta x + g^T(x;\mu)\Delta x$$

on the set $\|\Delta x\| \leq \Delta$, where Δ is a trust region radius. A direction vector Δx serves for determination of a new approximation of the solution x_+. Denoting

$$\rho(\Delta x) = \frac{B(x+\Delta x;\mu) - B(x;\mu)}{Q(\Delta x)},$$

we set $x_+ = x$ if $\rho(\Delta x) < \underline{\rho}$ or $x_+ = x + \Delta x$ if $\rho(\Delta x) \geq \underline{\rho}$. The trust region radius is updated so that $\beta_1 \leq \Delta_+ \leq \beta_2$ if $\rho(\Delta x) < \underline{\rho}$ or $\Delta_+ \geq \Delta$ if $\rho(\Delta x) \geq \underline{\rho}$ where $0 < \beta_1 \leq \beta_2 < 1$. More details can be found in [5] and [22].

4 Smoothing methods

4.1 Basic properties

Similarly as in Section 2.1 we will restrict ourselves to sums of maxima, where a function $h : R^n \to R^m$ is a sum of its arguments, so (4) holds. Smoothing methods for minimization of sums of maxima replace function (4) by a smoothing function

$$S(x;\mu) = \sum_{k=1}^{m} S_k(x;\mu), \tag{135}$$

where

$$S_k(x;\mu) = \mu \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(x)}{\mu}\right) = F_k(x) + \mu \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(x) - F_k(x)}{\mu}\right), \tag{136}$$

depending on a smoothing parameter $0 < \mu \leq \overline{\mu}$, which is successively minimized on R^n with $\mu \to 0$. Since $f_{kl}(x) \leq F_k(x)$, $1 \leq l \leq m_k$, and the equality arises for at least one index, at least one exponential function on the right-hand side of (136) has the value 1, so the logarithm is positive. Thus, it holds

$$F_k(x) \leq S_k(x;\mu) \leq F_k(x) + \mu \log m_k, \quad 1 \leq k \leq m \quad \Rightarrow \quad F(x) \leq S(x;\mu) \leq F(x) + \mu \sum_{k=1}^{m} \log m_k, \tag{137}$$

so $S(x;\mu) \to F(x)$ if $\mu \to 0$.

Remark 24. *Similarly as in Section 3.2 we will denote $g_{kl}(x)$ and $G_{kl}(x)$ the gradients and Hessian matrices of functions $f_{kl}(x)$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, and*

$$u_k(x;\mu) = \begin{bmatrix} u_{k1}(x;\mu) \\ \dots \\ u_{km_k}(x;\mu) \end{bmatrix}, \qquad \tilde{e}_k = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix},$$

where

$$u_{kl}(x;\mu) = \frac{\exp(f_{kl}(x)/\mu)}{\sum_{l=1}^{m_k} \exp(f_{kl}(x)/\mu)} = \frac{\exp((f_{kl}(x) - F_k(x))/\mu)}{\sum_{l=1}^{m_k} \exp((f_{kl}(x) - F_k(x))/\mu)}. \tag{138}$$

Thus, it holds $u_{kl}(x;\mu) \geq 0$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, *and*

$$\tilde{e}_k^T u_k(x;\mu) = \sum_{l=1}^{m_k} u_{kl}(x;\mu) = 1. \tag{139}$$

Further, we denote $A_k(x) = J_k^T(x) = [g_{k1}(x), \ldots, g_{km_k}(x)]$ *and* $U_k(x;\mu) = \mathrm{diag}(u_{k1}(x;\mu), \ldots, u_{km_k}(x;\mu))$ *for* $1 \leq k \leq m$.

Theorem 11. *Consider smoothing function (135). Then*

$$\nabla S(x;\mu) = g(x;\mu) \tag{140}$$

and

$$\begin{aligned} \nabla^2 S(x;\mu) &= G(x;\mu) + \frac{1}{\mu}\sum_{k=1}^{m} A_k(x)U_k(x;\mu)A_k^T(x) - \frac{1}{\mu}\sum_{k=1}^{m} A_k(x)u_k(x;\mu)(A_k(x)u_k(x;\mu))^T \\ &= G(x;\mu) + \frac{1}{\mu}A(x)U(x;\mu)A^T(x) - \frac{1}{\mu}C(x;\mu)C(x;\mu)^T \end{aligned} \tag{141}$$

where $g(x;\mu) = \sum_{k=1}^{m} A_k(x)u_k(x;\mu) = A(x)u(x)$ *and*

$$G(x;\mu) = \sum_{k=1}^{m} G_k(x)u_k(x;\mu), \qquad A(x) = [A_1(x), \ldots, A_m(x)],$$

$$U(x;\mu) = \mathrm{diag}(U_1(x;\mu), \ldots, U_m(x;\mu)), \qquad C(x;\mu) = [A_1(x)u_1(x;\mu), \ldots, A_m(x)u_m(x;\mu)].$$

Proof. Obviously,

$$\nabla S(x;\mu) = \sum_{k=1}^{m} \nabla S_k(x;\mu), \qquad \nabla^2 S(x;\mu) = \sum_{k=1}^{m} \nabla^2 S_k(x;\mu).$$

Differentiating functions (136) and using (138) we obtain

$$\begin{aligned} \nabla S_k(x;\mu) &= \frac{\mu}{\sum_{l=1}^{m_k}\exp(f_{kl}(x)/\mu)} \sum_{l=1}^{m_k} \frac{1}{\mu}\exp(f_{kl}(x)/\mu)g_{kl}(x) \\ &= \sum_{l=1}^{m_k} g_{kl}(x)u_{kl}(x;\mu) = A_k(x)u_k(x;\mu). \end{aligned} \tag{142}$$

Adding up these expressions yields (140). Further, it holds

$$\begin{aligned} \nabla u_{kl}(x;\mu) &= \frac{1}{\mu}\frac{\exp(f_{kl}(x)/\mu)g_{kl}(x)}{\sum_{l=1}^{m_k}\exp(f_{kl}(x)/\mu)} - \frac{\exp(f_{kl}(x)/\mu)}{\left(\sum_{l=1}^{m_k}\exp(f_{kl}(x)/\mu)\right)^2}\sum_{l=1}^{m_k}\frac{1}{\mu}\exp(f_{kl}(x)/\mu)g_{kl}(x) \\ &= \frac{1}{\mu}u_{kl}(x;\mu)g_{kl}(x) - \frac{1}{\mu}u_{kl}(x;\mu)\sum_{l=1}^{m_k} u_{kl}(x;\mu)g_{kl}(x). \end{aligned} \tag{143}$$

Differentiating (142) and using (143) we obtain

$$\begin{aligned} \nabla^2 S_k(x;\mu) &= \sum_{l=1}^{m_k} G_{kl}(x)u_{kl}(x;\mu) + \sum_{l=1}^{m_k} g_{kl}(x)\nabla u_{kl}(x;\mu) \\ &= G_k(x;\mu) + \frac{1}{\mu}\sum_{l=1}^{m_k} g_{kl}(x)u_{kl}(x;\mu)g_{kl}^T(x) - \frac{1}{\mu}\sum_{l=1}^{m_k} g_{kl}(x)u_{kl}(x;\mu)\left(\sum_{l=1}^{m_k} g_{kl}(x)u_{kl}(x;\mu)\right)^T \\ &= G_k(x;\mu) + \frac{1}{\mu}A_k(x)U_k(x;\mu)A_k^T(x) - \frac{1}{\mu}A_k(x)u_k(x;\mu)(A_k(x)u_k(x;\mu))^T, \end{aligned}$$

where $G_k(x;\mu) = \sum_{l=1}^{m_k} G_{kl}(x)u_{kl}(x;\mu)$. Adding up these expressions yields (141). □

Remark 25. *Note that using (141) and the Schwarz inequality we obtain*

$$\begin{aligned} v^T\nabla^2 S(x;\mu)v &= v^T G(x;\mu)v + \frac{1}{\mu}\sum_{k=1}^{m}\left(v^T A_k(x)U_k(x;\mu)A_k^T(x)v - \sum_{k=1}^{m}\frac{(v^T A_k(x)U_k(x;\mu)\tilde{e}_k)^2}{\tilde{e}_k^T U_k(x;\mu)\tilde{e}_k}\right) \\ &\geq v^T G(x;\mu)v \end{aligned}$$

because $\tilde{e}_k^T U_k(x;\mu)\tilde{e}_k = \tilde{e}_k^T u_k(x;\mu) = 1$*, so the Hessian matrix* $\nabla^2 S(x;\mu)$ *is positive definite if the matrix* $G(x;\mu)$ *is positive definite.*

Using Theorem 11, a step of the Newton method can be written in the form $x_+ = x + \alpha\Delta x$ where

$$\nabla^2 S(x;\mu)\Delta x = -\nabla S(x;\mu),$$

or

$$\left(W(x;\mu) - \frac{1}{\mu}C(x;\mu)C^T(x;\mu)\right)\Delta x = -g(x;\mu), \tag{144}$$

where

$$W(x;\mu) = G(x;\mu) + \frac{1}{\mu}A(x)U(x;\mu)A^T(x), \qquad g(x;\mu) = A(x)u(x;\mu). \tag{145}$$

A matrix W in (145) has the same structure as a matrix W in (93) and, by Theorem 11, smoothing function (135) has similar properties as barrier function (91). Thus, one can use an algorithm that is analogous to Algorithm 2 and considerations stated in Remark 21, where $S(x;\mu)$ and $\nabla^2 S(x;\mu)$ are used instead of $B(x;\mu)$ and $\nabla^2 B(x;\mu)$. It means that

$$S(x_{i+1};\mu_i) - S(x_i;\mu_i) \leq -c\|g(x_i;\mu_i)\|^2 \quad \forall i \in N \tag{146}$$

if Assumption X3 is satisfied and

$$S(x_{i+1};\mu_i) - S(x_i;\mu_i) \leq 0 \quad \forall i \in N \tag{147}$$

in remaining cases.

Algorithm 3. *Smoothing method*

Data. *A tolerance for the gradient norm of the smoothing function* $\underline{\varepsilon} > 0$*. Bounds for a smoothing parameter* $0 < \underline{\mu} < \overline{\mu}$*. Coefficients for decrease of a smoothing parameter* $0 < \lambda < 1$*,* $\sigma > 1$ *(or* $0 < \vartheta < 1$*). A tolerance for a uniform descent* $\varepsilon_0 > 0$*. A tolerance for a steplength selection* $\varepsilon_1 > 0$*. A maximum steplength* $\overline{\Delta} > 0$*.*

Input. *A sparsity pattern of the matrix* $A(x) = [A_1(x), \ldots, A_m(x)]$*. A starting point* $x \in R^n$*.*

Step 1 Initiation. *Choose* $\mu \leq \overline{\mu}$*. Determine a sparse structure of the matrix* $W = W(x;\mu)$ *from the sparse structure of the matrix* $A(x)$ *and perform a symbolic decomposition of the matrix* W *(described in [2, Section 1.7.4]). Compute values* $f_{kl}(x)$*,* $1 \leq k \leq m$*,* $1 \leq l \leq m_k$*, values* $F_k(x) = \max_{1\leq l\leq m_k} f_{kl}(x)$*,* $1 \leq k \leq m$*, and the value of objective function (4). Set* $r = 0$ *(restart indicator).*

Step 2 Termination. *Determine a vector of smoothing multipliers* $u(x;\mu)$ *by (138). Determine a matrix* $A = A(x)$ *and a vector* $g = g(x;\mu) = A(x)u(x;\mu)$*. If* $\mu \leq \underline{\mu}$ *and* $\|g\| \leq \underline{\varepsilon}$*, terminate the computation.*

Step 3 Hessian matrix approximation. *Set* $G = G(x;\mu)$ *or compute an approximation* G *of the Hessian matrix* $G(x;\mu)$ *using gradient differences or using quasi-Newton updates (Remark 22).*

Step 4 Direction determination. *Determine a matrix* W *by (145) and a vector* Δx *by (144) using the Gill-Murray decomposition of a matrix* W*.*

Step 5 Restart. *If $r = 0$ and (99) does not hold (where $s = \Delta x$), set $G = I$, $r = 1$ and go to Step 4. If $r = 1$ and (99) does not hold, set $\Delta x = -g$. Set $r = 0$.*

Step 6 Steplength selection. *Determine a steplength $\alpha > 0$ satisfying inequalities (100) (for a smoothing function $S(x;\mu)$) and $\alpha = \overline{\Delta}/\|\Delta x\|$. Set $x := x + \alpha\Delta x$. Compute values $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, values $F_k(x) = \max_{1\le l\le m_k} f_{kl}(x)$, $1 \le k \le m$, and the value of the objective function (4).*

Step 7 Smoothing parameter update. *Determine a new value of the smoothing parameter $\mu \ge \underline{\mu}$ using Procedure A or Procedure B. Go to Step 2.*

Algorithm 3 differs from Algorithm 2 in that a nonlinear equation $\tilde{e}^T u(x;\mu) = 1$ need not be solved in Step 2 (because (139) follows from (138)), equations (144)–(145) instead of (116)–(117) are used in Step 4, and a barrier function $B(x;\mu)$ is replaced with a smoothing function $S(x;\mu)$ in Step 6. Note that the parameter μ in (135) has different meaning than the same parameter in (91), so we could use another procedure for its update in Step 7. However, it is becoming apparent that using Procedure A or Procedure B is very efficient. On the other hand, it must be noted that using exponential functions in Algorithm 3 has certain disadvantages. Computation of the values of exponential functions is more time consuming than performing standard arithmetic operations and underflow may also happen (i.e. replacing nonzero values by zero values) if the value of a parameter μ is very small.

The values $\underline{\varepsilon} = 10^{-6}$, $\underline{\mu} = 10^{-6}$, $\overline{\mu} = 1$, $\lambda = 0.85$, $\sigma = 100$, $\vartheta = 0.1$, $\varepsilon_0 = 10^{-8}$, $\varepsilon_1 = 10^{-4}$, and $\overline{\Delta} = 1000$ were used in our numerical experiments.

4.2 Global convergence

Now we prove the global convergence of the smoothing method realized by Algorithm 3.

Lemma 7. *Choose a fixed vector $x \in R^n$. Then functions $S_k(x;\mu) : (0,\infty) \to R$, $1 \le k \le m$, are nondecreasing convex functions of $\mu > 0$ and*

$$0 \le \log \underline{m}_k \le \frac{\partial S_k(x;\mu)}{\partial \mu} \le \log m_k, \tag{148}$$

where $\underline{m}_k$ is a number of active functions (for which $f_{kl}(x) = F_k(x)$) and

$$\frac{\partial S_k(x;\mu)}{\partial \mu} = \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(x) - F_k(x)}{\mu}\right) - \sum_{l=1}^{m_k} \left(\frac{f_{kl}(x) - F_k(x)}{\mu}\right) u_{kl}(x;\mu). \tag{149}$$

Proof. Denoting $\varphi_{kl}(x;\mu) = (f_{kl}(x) - F_k(x))/\mu \le 0$, $1 \le k \le m$, so

$$\varphi'_{kl}(x;\mu) \triangleq \frac{\partial \varphi_{kl}(x;\mu)}{\partial \mu} = -\frac{\varphi_{kl}(x;\mu)}{\mu} \ge 0,$$

we can write by (136) that

$$S_k(x;\mu) = F_k(x) + \mu \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(x;\mu)$$

and

$$\begin{aligned} \frac{\partial S_k(x;\mu)}{\partial \mu} &= \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(x;\mu) + \mu \frac{\sum_{l=1}^{m_k} \varphi'_{kl}(x;\mu) \exp \varphi_{kl}(x;\mu)}{\sum_{l=1}^{m_k} \exp \varphi_{kl}(x;\mu)} \\ &= \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(x;\mu) - \sum_{l=1}^{m_k} \varphi_{kl}(x;\mu) u_{kl}(x;\mu) \ge 0 \end{aligned} \tag{150}$$

because $\varphi_{kl}(x;\mu) \le 0$, $u_{kl}(x;\mu) \ge 0$, $1 \le k \le m$, and $\varphi_{kl}(x;\mu) = 0$ holds for at least one index. Thus, functions $S_k(x;\mu)$, $1 \le k \le m$, are nondecreasing. Differentiating (138) with respect to μ we obtain

$$\begin{aligned}\frac{\partial u_{kl}(x;\mu)}{\partial \mu} &= -\frac{1}{\mu}\frac{\varphi_{kl}(x;\mu)\exp\varphi_{kl}(x;\mu)}{\sum_{l=1}^{m_k}\exp\varphi_{kl}(x;\mu)} + \frac{1}{\mu}\frac{\exp\varphi_{kl}(x;\mu)}{\sum_{l=1}^{k} m\exp\varphi_{kl}(x;\mu)}\frac{\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)\exp\varphi_{kl}(x;\mu)}{\sum_{l=1}^{m_k}\exp\varphi_{kl}(x;\mu)} \\ &= -\frac{1}{\mu}\varphi_{kl}(x;\mu)u_{kl}(x;\mu) + \frac{1}{\mu}u_{kl}(x;\mu)\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu). \end{aligned} \tag{151}$$

Differentiating (150) with respect to μ and using equations (139) and (151) we can write

$$\begin{aligned}\frac{\partial^2 S_k(x;\mu)}{\partial \mu^2} &= -\frac{1}{\mu}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu) + \frac{1}{\mu}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu) - \frac{1}{\mu}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)\frac{\partial u_{kl}(x;\mu)}{\partial \mu} \\ &= -\frac{1}{\mu}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)\frac{\partial u_{kl}(x;\mu)}{\partial \mu} \\ &= \frac{1}{\mu^2}\left(\sum_{l=1}^{m_k}\varphi_{kl}^2(x;\mu)u_{kl}(x;\mu)\right)\left(\sum_{l=1}^{m_k}u_{kl}(x;\mu)\right) - \frac{1}{\mu^2}\left(\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu)\right)^2 \ge 0 \end{aligned}$$

because

$$\left(\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu)\right)^2 = \left(\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)\sqrt{u_{kl}(x;\mu)}\sqrt{u_{kl}(x;\mu)}\right)^2 \le \sum_{l=1}^{m_k}\varphi_{kl}^2(x;\mu)u_{kl}(x;\mu)\sum_{l=1}^{m_k}u_{kl}(x;\mu)$$

holds by the Schwarz inequality. Thus, functions $S_k(x;\mu)$, $1 \le k \le m$, are convex, so their derivatives $\partial S_k(x;\mu)/\partial\mu$ are nondecreasing. Obviously, it holds

$$\begin{aligned}\lim_{\mu\to 0}\frac{\partial S_k(x;\mu)}{\partial \mu} &= \lim_{\mu\to 0}\log\sum_{l=1}^{m_k}\exp\varphi_{kl}(x;\mu) - \lim_{\mu\to 0}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu) \\ &= \log\underline{m}_k - \frac{1}{\underline{m}_k}\lim_{\mu\to 0}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)\exp\varphi_{kl}(x;\mu) = \log\underline{m}_k \end{aligned}$$

because $\varphi_{kl}(x;\mu) = 0$ if $f_{kl}(x) = F_k(x)$ and $\lim_{\mu\to 0}\varphi_{kl}(x;\mu) = -\infty$, $\lim_{\mu\to 0}\varphi_{kl}(x;\mu)\exp\varphi_{kl}(x;\mu) = 0$ if $f_{kl}(x) < F_k(x)$. Similarly, it holds

$$\lim_{\mu\to\infty}\frac{\partial S_k(x;\mu)}{\partial \mu} = \lim_{\mu\to\infty}\log\sum_{l=1}^{m_k}\exp\varphi_{kl}(x;\mu) - \lim_{\mu\to\infty}\sum_{l=1}^{m_k}\varphi_{kl}(x;\mu)u_{kl}(x;\mu) = \log m$$

because $\lim_{\mu\to\infty}\varphi_{kl}(x;\mu) = 0$ and $\lim_{\mu\to\infty}|u_{kl}(x;\mu)| \le 1$ for $1 \le k \le m$. □

Lemma 8. *Let Assumptions X1b and X3 be satisfied. Then the values μ_i, $i \in N$, generated by Algorithm 3, create a nonincreasing sequence such that $\mu_i \to 0$.*

Proof. Lemma 8 is a direct consequence of Lemma 6 because the same procedures for an update of a parameter μ are used and (146) holds. □

Theorem 12. *Let the assumptions of Lemma 8 be satisfied. Consider a sequence x_i $i \in N$, generated by Algorithm 3, where $\underline{\varepsilon} = \underline{\mu} = 0$. Then*

$$\lim_{i\to\infty}\sum_{k=1}^{m}\sum_{l=1}^{m_k}u_{kl}(x_i;\mu_i)g_{kl}(x_i) = 0, \qquad \sum_{l=1}^{m_k}u_{kl}(x_i;\mu_i) = 1$$

and

$$F_k(x_i) - f_{kl}(x_i) \ge 0, \quad u_{kl}(x_i;\mu_i) \ge 0, \quad \lim_{i\to\infty}u_{kl}(x_i;\mu_i)(F_k(x_i) - f_{kl}(x_i)) = 0$$

for $1 \le k \le m$ and $1 \le l \le m_k$.

Proof.

(a) Equations $\tilde{e}_k^T u_k(x_i;\mu_i) = 1$ for $1 \le k \le m$ follow from (139). Inequalities $F_k(x_i) - f_{kl}(x_i) \ge 0$ and $u_{kl}(x_i;\mu_i) \ge 0$ for $1 \le k \le m$ and $1 \le l \le m_k$ follow from (4) and (138).

(b) Since $S_k(x;\mu)$ are nondecreasing functions of the parameter μ by Lemma 7 and (146) holds, we can write

$$\begin{aligned} \underline{F} &\le \sum_{k=1}^{m} F_k(x_{i+1}) \le S(x_{i+1};\mu_{i+1}) \le S(x_{i+1};\mu_i) \le S(x_i;\mu_i) - c\|g(x_i;\mu_i)\|^2 \\ &\le S(x_1;\mu_1) - c\sum_{j=1}^{i} \|g(x_j;\mu_j)\|^2, \end{aligned}$$

where $\underline{F} = \sum_{k=1}^{m} \underline{F}_k$ and $\underline{F}_k$, $1 \le k \le m$, are lower bounds from Assumption X1b. Thus, it holds

$$\underline{F} \le \lim_{i\to\infty} S(x_{i+1};\mu_{i+1}) \le S(x_1;\mu_1) - c\sum_{i=1}^{\infty} \|g(x_i;\mu_i)\|^2,$$

or

$$\sum_{i=1}^{\infty} \|g(x_i;\mu_i)\|^2 \le \frac{1}{c}(S(x_1;\mu_1) - \underline{F}),$$

so $\|g(x_i;\mu_i)\| \to 0$, which together with inequalities $0 \le u_{kl}(x_i;\mu_i) \le 1$, $1 \le k \le m$, $1 \le l \le m_k$, gives $\lim_{i\to\infty} u_{kl}(x_i;\mu_i) g_{kl}(x_i) = 0$.

(c) Let indices $1 \le k \le m$ and $1 \le l \le m_k$ be chosen arbitrarily. Using (138) we get

$$\begin{aligned} 0 &\le u_{kl}(x_i;\mu_i)(F_k(x_i) - f_{kl}(x_i)) = -\mu_i \frac{\varphi_{kl}(x_i;\mu_i)\exp\varphi_{kl}(x_i;\mu_i)}{\sum_{l=1}^{m_k}\exp\varphi_{kl}(x_i;\mu_i)} \\ &\le -\mu_i\varphi_{kl}(x_i;\mu_i)\exp\varphi_{kl}(x_i;\mu_i) \le \frac{\mu_i}{e}, \end{aligned}$$

where $\varphi_{kl}(x_i;\mu_i)$, $1 \le k \le m$, $1 \le l \le m_k$, are functions used in the proof of Lemma 7, because

$$\sum_{l=1}^{m_k} \exp\varphi_{kl}(x_i;\mu_i) \ge 1$$

and the function $t\exp t$ attains its minimal value $-1/e$ at the point $t = -1$. Since $\mu_i \to 0$, we obtain $u_{kl}(x_i;\mu_i)(F_k(x_i) - f_{kl}(x_i)) \to 0$.

□

Corollary 2. *Let the assumptions of Theorem 12 be satisfied. Then every cluster point $x \in R^n$ of a sequence x_i, $i \in N$, satisfies the necessary KKT conditions (5)–(6), where u (with elements u_k, $1 \le k \le m$) is a cluster point of a sequence $u(x_i;\mu_i)$, $i \in N$.*

Now we will suppose that the values $\underline{\varepsilon}$ and $\underline{\mu}$ are nonzero and show how a precise solution of the system of KKT equations will be after termination of computation of Algorithm 3.

Theorem 13. *Let the assumptions of Theorem 8 be satisfied and let x_i, $i \in N$, be a sequence generated by Algorithm 3. Then, if the values $\underline{\varepsilon} > 0$ and $\underline{\mu} > 0$ are chosen arbitrarily, there exists an index $i \ge 1$ such that*

$$\|g(x_i;\mu_i)\| \le \underline{\varepsilon}, \qquad \tilde{e}_k^T u_k(x_i;\mu_i) = 1, \quad 1 \le k \le m,$$

and

$$F_k(x_i) - f_{kl}(x_i) \ge 0, \quad u_{kl}(x_i;\mu_i) \ge 0, \quad u_{kl}(x_i;\mu_i)(F_k(x_i) - f_{kl}(x_i)) \le \frac{\underline{\mu}}{e}$$

for all $1 \le k \le m$ and $1 \le l \le m_k$.

Proof. Equalities $\tilde{e}_k^T u_k(x_i;\mu_i) = 1$, $1 \le k \le m$, follow from (139). Inequalities $F_k(x_i) - f_{kl}(x_i) \ge 0$ and $u_{kl}(x_i;\mu_i) \ge 0$, $1 \le k \le m$, $1 \le l \le m_k$, follow from (10) and (138). Since $\mu_i \to 0$ holds by Lemma 8 and $\|g(x_i;\mu_i)\| \to 0$ holds by Theorem 12, there exists an index $i \ge 1$ such that $\mu_i \le \underline{\mu}$ and $\|g(x_i;\mu_i)\| \le \underline{\varepsilon}$. By (138), as in the proof of Theorem 12, one can write

$$u_{kl}(x_i;\mu_i)(F_k(x_i) - f_{kl}(x_i)) \le -\mu_i \varphi_{kl}(x_i;\mu_i) \exp \varphi_{kl}(x_i;\mu_i) \le \frac{\mu_i}{e} \le \frac{\underline{\mu}}{e}$$

for $1 \le k \le m$ and $1 \le l \le m_k$. □

4.3 Special cases

Both the simplest and most widely considered generalized minimax problem is the classical minimax problem (10), when $m = 1$ in (4) (in this case we write m and z instead of m_1 and z_1). For solving a classical minimax problem one can use Algorithm 3, where a major part of computation is very simplified. A step of the Newton method can be written in the form $x_+ = x + \alpha\Delta x$ where

$$\nabla^2 S(x;\mu)\Delta x = -\nabla S(x;\mu),$$

or

$$\left(W(x;\mu) - \frac{1}{\mu} g(x;\mu) g^T(x;\mu)\right) \Delta x = -g(x;\mu), \tag{152}$$

where

$$W(x;\mu) = G(x;\mu) + \frac{1}{\mu} A(x) U(x;\mu) A^T(x), \qquad g(x;\mu) = A(x) u(x;\mu). \tag{153}$$

Since

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

holds by the Sherman-Morrison formula, the solution of system of equations (152) can be written in the form

$$\Delta x = \frac{\mu}{g^T W^{-1} g - \mu} W^{-1} g. \tag{154}$$

If a matrix W is not positive definite, it may be replaced with a matrix $LL^T = W + E$ obtained by the Gill-Murray decomposition described in [11]. Then, we solve an equation

$$LL^T p = g \tag{155}$$

and set

$$\Delta x = \frac{\mu}{g^T p - \mu} p. \tag{156}$$

Minimization of a sum of absolute values, i.e., minimization of the function

$$F(x) = \sum_{k=1}^{m} |f_k(x)| = \sum_{k=1}^{m} \max(f_k^+(x), f_k^-(x)), \qquad f_k^+(x) = f_k(x), \qquad f_k^-(x) = -f_k(x),$$

is another important generalized minimax problem. In this case, a smoothing function has the form

$$\begin{aligned} S(x;\mu) &= F(x) + \mu \sum_{k=1}^{m} \log\left(\exp\left(-\frac{|f_k(x)| - f_k^+(x)}{\mu}\right) + \exp\left(-\frac{|f_k(x)| - f_k^-(x)}{\mu}\right)\right) \\ &= \sum_{k=1}^{m} |f_k(x)| + \mu \sum_{k=1}^{m} \log\left(1 + \exp\left(-\frac{2|f_k(x)|}{\mu}\right)\right) \end{aligned}$$

because $f_k^+(x) = |f_k(x)|$ if $f_k(x) \geq 0$ and $f_k^-(x) = |f_k(x)|$ if $f_k(x) \leq 0$, and by Theorem 11 we have

$$\nabla S(x;\mu) = \sum_{k=1}^{m}(g_k^+ u_k^+ + g_k^- u_k^-) = \sum_{k=1}^{m} g_k(u_k^+ - u_k^-) = \sum_{k=1}^{m} g_k u_k = g(x;\mu),$$

$$\nabla^2 S(x;\mu) = \sum_{k=1}^{m} G_k(u_k^+ - u_k^-) + \frac{1}{\mu}\sum_{k=1}^{m} g_k g_k^T(u_k^+ + u_k^-) - \frac{1}{\mu}\sum_{k=1}^{m} g_k g_k^T(u_k^+ - u_k^-)^2 = G(x;\mu) + \frac{1}{\mu}\sum_{k=1}^{m} g_k g_k^T(1-u_k^2)$$

(because $u_k^+ + u_k^- = 1$), where $g_k = g_k(x)$ and

$$u_k = u_k^+ - u_k^- = \frac{\exp\left(-\frac{|f_k(x)|-f_k^+(x)}{\mu}\right) - \exp\left(-\frac{|f_k(x)|-f_k^-(x)}{\mu}\right)}{\exp\left(-\frac{|f_k(x)|-f_k^+(x)}{\mu}\right) + \exp\left(-\frac{|f_k(x)|-f_k^-(x)}{\mu}\right)} = \frac{1-\exp\left(-\frac{2|f_k(x)|}{\mu}\right)}{1+\exp\left(-\frac{2|f_k(x)|}{\mu}\right)}\operatorname{sign}(f_k(x)),$$

$$1 - u_k^2 = \frac{4\exp\left(-\frac{2|f_k(x)|}{\mu}\right)}{\left(1+\exp\left(-\frac{2|f_k(x)|}{\mu}\right)\right)^2},$$

and where $\operatorname{sign}(f_k(x))$ is a sign of a function $f_k(x)$.

5 Primal-dual interior point methods

5.1 Basic properties

Primal interior point methods for solving nonlinear programming problems profit from the simplicity of obtaining and keeping a point in the interior of the feasible set (for generalized minimax problems, it suffices to set $z_k > F_k(x)$, $1 \leq k \leq m$). Minimization of a barrier function without constraints and a direct computation of multipliers u_{kl}, $1 \leq k \leq m$, $1 \leq l \leq m_k$, are basic features of these methods. Primal-dual interior point methods are intended for solving general nonlinear programming problems, where it is usually impossible to assure validity of constraints. These methods guarantee feasibility of points by adding slack variables, which appear in a barrier term added to the objective function. Positivity of the slack variables is assured algorithmically (by a steplength selection). Minimization of a barrier function with equality constraints and an iterative computation of the Lagrange multipliers (dual variables) are the main features of primal-dual interior point methods.

Consider function (4). As is mentioned in the introduction, minimization of this function is equivalent to the nonlinear programming problem

$$\text{minimize} \quad \sum_{k=1}^{m} z_k \quad \text{subject to} \quad f_{kl}(x) \leq z_k, \quad 1 \leq k \leq m, \quad 1 \leq l \leq m_k. \tag{157}$$

Using slack variables $s_{kl} > 0$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, and a barrier function

$$B_\mu(x,z,s) = \sum_{k=1}^{m} z_k - \mu\sum_{k=1}^{m}\sum_{l=1}^{m_k}\log(s_{kl}), \tag{158}$$

a solving of problem (157) can be transformed to a successive solving of problems

$$\text{minimize} \quad B_\mu(x,z,s) \quad \text{subject to} \quad f_{kl}(x) + s_{kl} - z_k = 0, \quad 1 \leq k \leq m, \quad 1 \leq l \leq m_k, \tag{159}$$

where $\mu \to 0$. Necessary conditions for an extremum of problem (159) have the form

$$g(x,u) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} g_{kl}(x)u_{kl} = 0,$$

$$1 - \sum_{l=1}^{m_k} u_{kl} = 0, \quad 1 \le k \le m,$$

$$u_{kl} s_{kl} - \mu = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k,$$

$$f_{kl}(x) + s_{kl} - z_k = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k,$$

which is $n+m+2\bar{m}$ equations for $n+m+2\bar{m}$ unknowns (vectors x, $z = [z_k]$, $s = [s_{kl}]$, $u = [u_{kl}]$, $1 \le k \le m$, $1 \le l \le m_k$), where $\bar{m} = m_1 + \cdots + m_m$. Denote $A(x) = [A_1(x), \ldots, A_m(x)]$, $f = [f_{kl}]$, $S = \operatorname{diag}(s_{kl})$, $U = \operatorname{diag}(u_{kl})$, $1 \le k \le m$, $1 \le l \le m_k$, and

$$E = \begin{bmatrix} \tilde{e}_1 & 0 & \ldots & 0 \\ 0 & \tilde{e}_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \tilde{e}_m \end{bmatrix}, \qquad \tilde{e} = \begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \\ \ldots \\ \tilde{e}_m \end{bmatrix}, \qquad z = \begin{bmatrix} z_1 \\ z_2 \\ \ldots \\ z_m \end{bmatrix}$$

(matrices $A_k(x)$, vectors $\tilde{e}_k$, and numbers z_k, $1 \le k \le m$, are defined in Section 3.2). Applying the Newton method to this system of nonlinear equations, we obtain a system of linear equations for increments (direction vectors) Δx, Δz, Δs, Δu. After arrangement and elimination

$$\Delta s = -U^{-1} S (u + \Delta u) + \mu S^{-1} \tilde{e}, \tag{160}$$

this system has the form

$$\begin{bmatrix} G(x,u) & 0 & A(x) \\ 0 & 0 & -E^T \\ A^T(x) & -E & -U^{-1}S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta u \end{bmatrix} = - \begin{bmatrix} g(x,u) \\ \tilde{e} - E^T u \\ f(x) - Ez + \mu U^{-1} \tilde{e} \end{bmatrix}, \tag{161}$$

where $G(x,u) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} G_{kl}(x) u_{kl}$. Vector $\tilde{e}$ in equation $\tilde{e} - E^T u = 0$ has unit elements, but its dimension is different from the dimension of a vector $\tilde{e}$ in (160).

For solving this linear system, we cannot advantageously use the structure of a generalized minimax problem (substituting $z = F_k(x) = \max_{1 \le l \le m_k} f_{kl}(x)$ we would obtain a nonsmooth problem whose solution is much more difficult). Therefore, we need to deal with a general nonlinear programming problem. To simplify subsequent considerations, we use the notation

$$\tilde{x} = \begin{bmatrix} x \\ z \end{bmatrix}, \qquad \tilde{g}(\tilde{x}, u) = \begin{bmatrix} g(x,u) \\ \tilde{e} - E^T u \end{bmatrix}, \qquad \tilde{G}(\tilde{x}, u) = \begin{bmatrix} G(x,u) & 0 \\ 0 & 0 \end{bmatrix}, \qquad \tilde{A}(\tilde{x}) = \begin{bmatrix} A(x) \\ -E^T \end{bmatrix} \tag{162}$$

and write (161) in the form

$$\begin{bmatrix} \tilde{G}(\tilde{x}, u) & \tilde{A}(\tilde{x}) \\ \tilde{A}^T(\tilde{x}) & -U^{-1}S \end{bmatrix} \begin{bmatrix} \Delta \tilde{x} \\ \Delta u \end{bmatrix} = - \begin{bmatrix} \tilde{g}(\tilde{x}, u) \\ c(\tilde{x}) + \mu U^{-1} \tilde{e} \end{bmatrix}, \tag{163}$$

where $c(\tilde{x}) = f(x) - Ez$. This system of equations is more advantageous against systems (94) and (144) in that its matrix does not depend on the barrier parameter μ, so it is not necessary to use a lower bound $\underline{\mu}$. On the other hand, system (163) has a dimension $n+m+\bar{m}$, while systems (94) and (144) have dimensions n. It would be possible to eliminate the vector Δu, so the resulting system

$$(\tilde{G}(\tilde{x}, u) + \tilde{A}(\tilde{x}) M^{-1} \tilde{A}^T(\tilde{x})) \Delta \tilde{x} = -(\tilde{g}(\tilde{x}, u) + \tilde{A}(\tilde{x})(M^{-1} c(\tilde{x}) + \mu S^{-1} \tilde{e})), \tag{164}$$

where $M = U^{-1} S$, would have dimension $n+m$ (i.e., $n+1$ for classical minimax problems). Nevertheless, as follows from the equation $u_{kl} s_{kl} = \mu$, either $u_{kl} \to 0$ or $s_{kl} \to 0$ if $\mu \to 0$, so some elements of a matrix M^{-1} may tend to infinity, which increases the condition number of system (164). Conversely, the solution of equation (163) is easier if the elements of a matrix M are small (if $M = 0$, we obtain the saddle point system, which can be solved by efficient iterative methods [1], [29]). Therefore, it is advantageous to split the constraints to active with $s_{kl} \le \tilde{\varepsilon} u_{kl}$ (we denote active quantities by $\hat{c}(\tilde{x})$, $\hat{A}(\tilde{x})$, $\hat{s}$, $\Delta\hat{s}$, $\hat{S}$, $\hat{u}$, $\Delta\hat{u}$, $\hat{U}$,

$\hat{M} = \hat{U}^{-1}\hat{S}$) and inactive with $s_{kl} > \tilde{\varepsilon} u_{kl}$ (we denote inactive quantities by $\check{c}(\tilde{x})$, $\check{A}(\tilde{x})$, $\check{s}$, $\Delta\check{s}$, $\check{S}$, $\check{u}$, $\Delta\check{u}$, $\check{U}$, $\check{M} = \check{U}^{-1}\check{S}$). Eliminating inactive equations from (163) we obtain

$$\Delta\check{u} = \check{M}^{-1}(\check{c}(\tilde{x}) + \check{A}(\tilde{x})^T \Delta\tilde{x}) + \mu\check{S}^{-1}e \tag{165}$$

and

$$\begin{bmatrix} \hat{G}(\tilde{x},u) & \hat{A}(\tilde{x}) \\ \hat{A}^T(\tilde{x}) & -\hat{M} \end{bmatrix} \begin{bmatrix} \Delta\tilde{x} \\ \Delta\hat{u} \end{bmatrix} = - \begin{bmatrix} \hat{g}(\tilde{x},u) \\ \hat{c}(\tilde{x}) + \mu\hat{U}^{-1}\tilde{e} \end{bmatrix}, \tag{166}$$

where

$$\begin{aligned} \hat{G}(\tilde{x},u) &= G(\tilde{x},u) + \check{A}(\tilde{x})\check{M}^{-1}\check{A}^T(\tilde{x}), \\ \hat{g}(\tilde{x},u) &= g(\tilde{x},u) + \check{A}(\tilde{x})(\check{M}^{-1}\check{c}(\tilde{x}) + \mu\check{S}^{-1}\tilde{e}), \end{aligned}$$

and $\hat{M} = \hat{U}^{-1}\hat{S}$ is a diagonal matrix of order $\hat{m}$, where $0 \le \hat{m} \le \bar{m}$ is the number of active constraints. Substituting (165) into (160) we can write

$$\Delta\hat{s} = -\hat{M}(\hat{u} + \Delta\hat{u}) + \mu\hat{U}^{-1}\tilde{e}, \qquad \Delta\check{s} = -(\check{c} + \check{A}^T\Delta\tilde{x} + \check{s}). \tag{167}$$

The matrix of the linear system (166) is symmetric, but indefinite, so its Choleski decomposition cannot be determined. In this case, we use either dense [3] or sparse [7] Bunch-Parlett decomposition for solving this system. System (166) (especially if it is large and sparse) can be efficiently solved by iterative conjugate gradient method with indefinite preconditioner [20]. If the vectors $\Delta\tilde{x}$ and $\Delta\hat{u}$ are solutions of system (166), we determine vector $\Delta\check{u}$ by (165) and vectors $\Delta\hat{s}$, $\Delta\check{s}$ by (167).

Having vectors $\Delta\tilde{x}$, Δs, Δu, we need to determine a steplength $\alpha > 0$ and set

$$\tilde{x}_+ = \tilde{x} + \alpha\Delta\tilde{x}, \qquad s_+ = s(\alpha), \qquad u_+ = u(\alpha), \tag{168}$$

where $s(\alpha)$ and $u(\alpha)$ are vector functions such that $s(\alpha) > 0$, $s'(0) = \Delta s$ and $u(\alpha) > 0$, $u'(0) = \Delta u$. This step is not trivial, because we need to decrease both the value of the barrier function $\tilde{B}_\mu(\tilde{x}, s) = B_\mu(x, z, s)$ and the norm of constraints $\|c(\tilde{x})\|$, and also to assure positivity of vectors s and u. We can do this in several different ways: using either the augmented Lagrange function [20], [21] or a bi-criterial filter [10], [37] or a special algorithm [12], [18]. In this section, we confine our attention to the augmented Lagrange function which has (for problem (157)) the form

$$P(\alpha) = \tilde{B}_\mu(\tilde{x} + \alpha\Delta\tilde{x}, s(\alpha)) + (u + \Delta u)^T(c(\tilde{x} + \alpha\Delta\tilde{x}) + s(\alpha)) + \frac{\sigma}{2}\|c(\tilde{x} + \alpha\Delta\tilde{x}) + s(\alpha)\|^2, \tag{169}$$

where $\sigma \ge 0$ is a penalty parameter. The following theorem, whose proof is given in [20], holds.

Theorem 14. *Let $s > 0$, $u > 0$ and let vectors $\Delta\tilde{x}$, $\Delta\hat{u}$ be solutions of the linear system*

$$\begin{bmatrix} \hat{G}(\tilde{x},u) & \hat{A}(\tilde{x}) \\ \hat{A}^T(\tilde{x}) & -\hat{M} \end{bmatrix} \begin{bmatrix} \Delta\tilde{x} \\ \Delta\hat{u} \end{bmatrix} + \begin{bmatrix} \hat{g}(\tilde{x},u) \\ \hat{c}(\tilde{x}) + \mu\hat{U}^{-1}\tilde{e} \end{bmatrix} = \begin{bmatrix} r \\ \hat{r} \end{bmatrix}, \tag{170}$$

where r and $\hat{r}$ are residual vectors, and let vectors $\Delta\check{u}$ and Δs be determined by (165) and (167). Then

$$P'(0) = -(\Delta\tilde{x})^T\tilde{G}(\tilde{x},u)\Delta\tilde{x} - (\Delta s)^T M^{-1}\Delta s - \sigma\|c(\tilde{x}) + s\|^2 + (\Delta\tilde{x})^T r + \sigma(\hat{c}(\tilde{x}) + \hat{s})^T\hat{r}. \tag{171}$$

If

$$\sigma > -\frac{(\Delta\tilde{x})^T\tilde{G}(\tilde{x},u)\Delta\tilde{x} + (\Delta s)^T M^{-1}\Delta s}{\|c(\tilde{x}) + s\|^2} \tag{172}$$

and if system (166) is solved in such a way that

$$(\Delta\tilde{x})^T r + \sigma(\hat{c}(\tilde{x}) + \hat{s})^T\hat{r} < (\Delta\tilde{x})^T\tilde{G}(\tilde{x},u)\Delta\tilde{x} + (\Delta s)^T M^{-1}\Delta s + \sigma(\|c(\tilde{x}) + s\|^2), \tag{173}$$

then $P'(0) < 0$.

Inequality (173) is significant only if linear system (166) is solved iteratively and residual vectors r and $\hat{r}$ are nonzero. If these vectors are zero, then (173) follows immediately from (172). Inequality (172) serves for determination of a penalty parameter, which should be as small as possible. If the matrix $\tilde{G}(\tilde{x},u)$ is positive semidefinite, then the right-hand side of (172) is negative and we can choose $\sigma = 0$.

5.2 Implementation

The algorithm of the primal-dual interior point method consists of four basic parts: determination of the matrix $G(x,u)$ or its approximation, solving linear system (166), a steplength selection, and an update of the barrier parameter μ. The matrix $G(x,u)$ has form (74), so its approximation can be determined in the one of the ways introduced in Remark 22.

The linear system (166), obtained by determination and subsequent elimination of inactive constraints in the way described in the previous subsection, is solved either directly using the Bunch-Parlett decomposition or iteratively by the conjugate gradient method with the indefinite preconditioner

$$C = \begin{bmatrix} \hat{D} & \hat{A}(\tilde{x}) \\ \hat{A}^T(\tilde{x}) & -\hat{M} \end{bmatrix}, \tag{174}$$

where $\hat{D}$ is a positive definite diagonal matrix that approximates matrix $\hat{G}(\tilde{x},u)$. An iterative process is terminated if residual vectors satisfy conditions (173) and

$$\|r\| \le \omega \|\tilde{g}(\tilde{x},u)\|, \qquad \|\hat{r}\| \le \omega \min(\|\hat{c}(\tilde{x}) + \mu \hat{U}^{-1}\tilde{e}\|, \|\hat{c}(\tilde{x}) + \hat{s}\|),$$

where $0 < \omega < 1$ is a prescribed precision. The directional derivative $P'(0)$ given by (169) should be negative. There are two possibilities how this requirement can be achieved. We either determine the value $\sigma \ge 0$ satisfying inequality (172), which implies $P'(0) < 0$ if (173) holds (Theorem 14), or set $\sigma = 0$ and ignore inequality (173). If $P'(0) \ge 0$, we determine a diagonal matrix $\tilde{D}$ with elements

$$\begin{aligned} \tilde{D}_{jj} &= \underline{\Gamma} \quad \text{if} \quad \frac{\|\tilde{g}\|}{10}|\tilde{G}_{jj}| < \underline{\Gamma}, \\ \tilde{D}_{jj} &= \frac{\|\tilde{g}\|}{10}|\tilde{G}_{jj}| \quad \text{if} \quad \underline{\Gamma} \le \frac{\|\tilde{g}\|}{10}|\tilde{G}_{jj}| \le \overline{\Gamma}, \\ \tilde{D}_{jj} &= \overline{\Gamma} \quad \text{if} \quad \overline{\Gamma} < \frac{\|\tilde{g}\|}{10}|\tilde{G}_{jj}|, \end{aligned} \tag{175}$$

for $1 \le j \le n+m$, where $\tilde{g} = \tilde{g}(\tilde{x},u)$ and $0 < \underline{\Gamma} < \overline{\Gamma}$, set $\tilde{G}(\tilde{x},u) = \tilde{D}$ and restart the iterative process by solving new linear system (166).

We use functions $s(\alpha) = [s_{kl}(\alpha)]$, $u(\alpha) = [u_{kl}(\alpha)]$, where $s_{kl}(\alpha) = s_{kl} + \alpha_{s_{kl}} \Delta s_{kl}$, $u_{kl}(\alpha) = u_{kl} + \alpha_{u_{kl}} \Delta s_{kl}$ and

$$\begin{aligned} \alpha_{s_{kl}} &= \alpha, & \Delta s_{kl} \ge 0, \\ \alpha_{skl} &= \min\left(\alpha, -\gamma \frac{s_{kl}}{\Delta s_{kl}}\right), & \Delta s_{kl} < 0, \\ \alpha_{u_{kl}} &= \alpha, & \Delta u_{kl} \ge 0, \\ \alpha_{u_{kl}} &= \min\left(\alpha, -\gamma \frac{u_{kl}}{\Delta u_{kl}}\right), & \Delta u_{kl} < 0, \end{aligned}$$

when choosing a steplength using the augmented Lagrange function. A parameter $0 < \gamma < 1$ (usually $\gamma = 0.99$) assures the positivity of vectors s_+ and u_+ in (168). A parameter $\alpha > 0$ is chosen to satisfy the inequality $P(\alpha) - P(0) \le \varepsilon_1 \alpha P'(0)$, which is possible because $P'(0) < 0$ and a function $P(\alpha)$ is continuous.

After finishing the iterative step, a barrier parameter μ should be updated. There exist several heuristic procedures for this purpose. The following procedure proposed in [36] seems to be very efficient.

Procedure C

Compute the centrality measure

$$\varrho = \frac{\bar{m} \min_{kl}(s_{kl} u_{kl})}{s^T u},$$

where $\bar{m} = m_1 + \cdots + m_m$ and $1 \le k \le m, 1 \le l \le m_k$. Compute the value

$$\lambda = 0.1 \min\left(0.05\frac{1-\varrho}{\varrho}, 2\right)^3$$

and set $\mu = \lambda s^T u/\bar{m}$.

Algorithm 4. *Primal-dual interior point method*

Data. *A tolerance for the barrier function gradient norm $\underline{\varepsilon} > 0$. A parameter for determination of active constraints $\tilde{\varepsilon} > 0$. A parameter for initiation of the slack variables and the Lagrange multipliers $\delta > 0$. An initial value of the barrier parameter $\overline{\mu} > 0$. Precision for the direction vectors determination $0 \le \omega < 1$. A parameter for the steplength selection $0 < \gamma < 1$. A tolerance for the steplength selection $\varepsilon_1 > 0$. A maximum steplength $\overline{\Delta} > 0$.*

Input. *A sparsity pattern of the matrix $A(x) = [A_1(x), \ldots, A_m(x)]$. A starting point $x \in R^n$.*

Step 1 Initiation. *Compute values $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, and set $F_k(x) = \max_{1 \le l \le m_k} f_{kl}$, $z_k = F_k(x) + \delta$, $1 \le k \le m$. Compute values $c_{kl}(\tilde{x}) = f_{kl}(x) - z_k$, $1 \le k \le m$, $1 \le l \le m_k$, and set $s_{kl} = -c_{kl}(\tilde{x})$, $u_{kl} = \delta$. Set $\mu = \overline{\mu}$ and compute the value of the barrier function $\tilde{B}_\mu(\tilde{x}, s)$.*

Step 2 Termination. *Determine a matrix $\tilde{A}(\tilde{x})$ and a vector $\tilde{g}(\tilde{x}, u) = \tilde{A}(\tilde{x})u$ by (162). If the KKT contitions $\|\tilde{g}(\tilde{x}, u)\| \le \underline{\varepsilon}$, $\|c(\tilde{x}) + s\| \le \underline{\varepsilon}$, and $s^T u \le \underline{\varepsilon}$ are satisfied, terminate the computation.*

Step 3 Hessian matrix approximation. *Set $G = G(x, u)$ or compute an approximation G of the Hessian matrix $G(x, u)$ using gradient differences or utilizing quasi-Newton updates (Remark 22). Determine a parameter $\sigma \ge 0$ by (172) or set $\sigma = 0$. Split the constraints into active if $\hat{s}_{kl} \le \tilde{\varepsilon}\hat{u}_{kl}$ and inactive if $\check{s}_{kl} > \tilde{\varepsilon}\check{u}_{kl}$.*

Step 4 Direction determination. *Determine the matrix $\tilde{G} = \tilde{G}(\tilde{x}, u)$ by (162) (where the Hessian matrix $G(x, u)$ is replaced with its approximation G). Determine vectors $\Delta\tilde{x}$ and $\Delta\hat{u}$ by solving linear system (166), a vector $\Delta\check{u}$ by (165), and a vector Δs by (167). Linear system (166) is solved either directly using the Bunch-Parlett decomposition (we carry out both the symbolic and the numeric decompositions in this step) or iteratively by the conjugate gradient method with indefinite preconditioner (174). Compute the derivative of the augmented Lagrange function by formula (171).*

Step 5 Restart. *If $P'(0) \ge 0$, determine a diagonal matrix $\tilde{D}$ by (175), set $\tilde{G} = \tilde{D}$, $\sigma = 0$, and go to Step 4.*

Step 6 Steplength selection. *Determine a steplength parameter $\alpha > 0$ satisfying inequalities $P(\alpha) - P(0) \le \varepsilon_1 \alpha P'(0)$ and $\alpha \le \overline{\Delta}/\|\Delta x\|$. Determine new vectors $\tilde{x} := \tilde{x} + \alpha\Delta\tilde{x}$, $s := s(\alpha)$, $u := u(\alpha)$ by (168). Compute values $f_{kl}(x)$, $1 \le k \le m$, $1 \le l \le m_k$, and set $c_{kl}(\tilde{x}) = f_{kl}(x) - z_k$, $1 \le k \le m$, $1 \le l \le m_k$. Compute the value of the barrier function $\tilde{B}_\mu(\tilde{x}, s)$.*

Step 7 Barrier parameter update. *Determine a new value of the barrier parameter $\mu \ge \underline{\mu}$ using Procedure C. Go to Step 2.*

The values $\underline{\varepsilon} = 10^{-6}$, $\tilde{\varepsilon} = 0.1$, $\delta = 0.1$, $\omega = 0.9$, $\gamma = 0.99$, $\overline{\mu} = 1$, $\varepsilon_1 = 10^{-4}$, and $\overline{\Delta} = 1000$ were used in our numerical experiments.

6 Numerical experiments

The methods studied in this contribution were tested by using two collections of test problems TEST14 and TEST15 described in [30], which are the parts of the UFO system [28] and can be downloaded from the web page `www.cs.cas.cz/luksan/test.html`. Both these collections contain 22 problems with functions

$f_k(x)$, $1 \leq k \leq m$, $x \in R^n$, where n is an input parameter and $m \geq n$ depends on n (we have used the values $n = 100$ and $n = 1000$ for numerical experiments). Functions $f_k(x)$, $1 \leq k \leq m$, have a sparse structure (the Jacobian matrix of a mapping $f(x)$ is sparse), so sparse matrix decompositions can be used for solving linear equation systems. Since the method described in Section 2.2 does not use sparsity of a quadratic programming problem, it is not comparable with other methods for test problems used. Therefore, it is not presented in the test results.

The tested methods, whose results are reported in Tables 1-5, are denoted by seven letters. The first pair of letters distinguishes the line search methods LS from the trust region methods TR (trust region methods are used only for minimization of the l_1 norm). The second pair of letters gives the problem type: either a classical minimax MX (when a function $F(x)$ has form (10) or $F(x) = \|f(x)\|_\infty$ holds) or a sum of absolute values SA (when $F(x) = \|f(x)\|_1$ holds). Further two letters specify the method used:

`PI` - the primal interior point method (Section 3),

`SM` - the smoothing method (Section 4),

`DI` - the primal-dual interior point method (Section 5).

The last letter denotes the procedure for updating a barrier parameter μ (procedures A and B are described in Section 3.4 and procedure C is described in Section 5.2).

The columns of all tables correspond to two classes of methods. The Newton methods use approximations of the Hessian matrices of the Lagrange function obtained by gradient differences [4] and variable metric methods update approximations of the Hessian matrices of the partial functions by the methods belonging to the Broyden family [13] (Remark 22).

The tables contain total numbers of iterations `NIT`, function evaluations `NFV`, gradient evaluations `NFG`, and also the total computational time, the number of problems with the value $\overline{\Delta}$ decreased and the number of failures (the number of unsolved problems). The decrease of the maximum steplength $\overline{\Delta}$ is used for three reasons. First, too large steps may lead to overflows if arguments of functions (roots, logarithms, exponentials) lie outside of their definition domain. Second, the change of $\overline{\Delta}$ can affect the finding of a suitable (usually global) minimum. Finally, it can prevent from achieving a domain in which the objective function has bad behavior leading to a loss of convergence. The number of times the steplength has decreased is in some sense a symptom of robustness (a lower number corresponds to higher robustness).

Several conclusions can be done from the results stated in these tables.

- For l_1 approximation, it is usually more advantageous to use the trust region methods TR than the line search methods LS.
- The smoothing methods are less efficient than the primal interior point methods. For testing the smoothing methods, we had to use the value $\underline{\mu} = 10^{-6}$, while the primal interior methods work well with the smaller value $\underline{\mu} = 10^{-8}$, which gives more precise results.
- The primal-dual interior point methods are slower than the primal interior point methods, especially for the reason that system of equations (166) is indefinite, so we cannot use the Choleski (or the Gill-Murray [11]) decomposition. If the matrix of linear system (166) is large and sparse, we can use the Bunch-Parlett decomposition [7]. In this case, a large fill-in of new nonzero elements (and thus to overflow of the operational memory or large extension of the computational time) may appear. In this case, we can also use the iterative conjugate gradient method with an indefinite preconditioner [29], however, the ill-conditioned systems can require a large number of iterations and thus a large computational time.
- It cannot be uniquely decided whether Procedure A is better than Procedure B. The Newton methods usually work better with Procedure A while the variable metric methods are more efficient with Procedure B.

- The variable metric methods are usually faster because it is not necessary to determine the elements of the Hessian matrix of the Lagrange function by gradient differences. The Newton methods seem to be more robust (especially in case of l_1 approximation).

References

[1] M.Benzi, G.H.Golub, J.Liesen: Numerical solution of saddle point problems. Acta Numerica 14 (2005) 1-137.

[2] A.Björck: Numerical Methods in Matrix Computations. Springer, New York, 2015.

[3] J.R.Bunch, B.N.Parlett: Direct methods for solving symmetric indefinite systems of linear equations. SIAM J. Numerical Analysis 8 (1971) 639-655.

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

[5] A.R.Conn, N.I.M.Gould, P.L.Toint: Trust-region Methods. SIAM, Phiadelphia, 2000.

[6] G. Di Pillo, L. Grippo, S. Lucidi: Smooth Transformation of the generalized minimax problem. J. of Optimization Theory and Applications 95 (1997) 1-24.

[7] I.S.Duff, N.I.M.Gould, J.K.Reid, K.Turner: The factoruzation of sparse symmetric indefinite metrices. IMA Journal of Numerical Analysis 11 (1991) 181-204.

[8] M.Fiedler: Special Matrices and Their Applications in Numerical Mathematics. Dover Publications, New York, 2008.

[9] R.Fletcher: Practical methods of optimization. Wiley, New York, 1987.

[10] R.Fletcher, S.Leyffer: Nonlinear programming without a penalty function Mathematical Programming 91 (2002) 239-269.

[11] P.E.Gill, W.Murray: Newton type methods for unconstrained and linearly constrained optimization. Mathematical Programming 7 (1974) 311-350.

[12] N.I.M.Gould, P.L.Toint: Nonlinear programming without a penalty function or a filter. Mathematical Programming 122 (2010) 155-196.

[13] A.Griewank, P.L.Toint: Partitioned variable metric updates for large-scale structured optimization problems. Numererische Mathematik 39 (1982) 119-137.

[14] A.Griewank, A.Walther: Evaluating Derivatives. SIAM, Philadelphia, 2008.

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

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

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

[18] X.Liu Y.Yuan: A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization. SIAM J. Optimization 21 (2011) 545-571.

[19] L.Lukšan: Dual method for solving a special problem of quadratic programming as a subproblem at linearly constrained nonlinear minimax approximation. Kybernetika 20 (1984) 445-457.

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

[21] L.Lukšan, C.Matonoha, J.Vlček: Interior-point method for large-scale nonlinear programming. Optimization Methods and Software 20 (2005) 569-582.

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

[23] L.Lukšan, C.Matonoha, J.Vlček: On Lagrange multipliers of trust-region subproblems. BIT Numerical Analysis 48 (2008a) 763-768.

[24] L.Lukšan, C.Matonoha, J.Vlček J.: Algorithm 896: LSA: Algorithms for Large-Scale Optimization. ACM Transactions on Mathematical Software 36 (2009) No. 3.

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

[26] L. Lukšan, C. Matonoha, J. Vlček: Primal interior-point method for minimization of generalized minimax functions. Kybernetika 46 (2010) 697-721.

[27] L.Lukšan, E.Spedicato: Variable metric methods for unconstrained optimization and nonlinear least squares. Journal of Computational and Applied Mathematics 124 (2000) 61-93.

[28] L.Lukšan, M.Tůma, C.Matonoha, J.Vlček J., N.Ramešová, M.Šiška, J.Hartman: UFO 2017. Interactive System for Universal Functional Optimization. Technical Report V-1252. Prague, ICS AS CR 2017.

[29] 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.

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

[31] M.M.Mäkelá, P.Neitaanmäki: Nonsmooth Optimization. World Scientific, London 1992.

[32] J.Nocedal, S.J.Wright: Numerical optimization. Springer-Verlag, New York, 2006.

[33] M.J.D.Powell: On the global convergence of trust region algorithms for unconstrained minimization. Mathematical Programming 29 (1984) 297-303.

[34] P.L.Toint: On sparse and symmetric matrix updating subject to a linear equation. Mathematics of Computation 31 (1977) 954-961.

[35] M.Tůma: A note on direct methods for approximations of sparse Hessian matrices. Applications of Mathematic 33 (1988) 171-176.

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

[37] A.Wachter, L.Biegler: Line search filter methods for nonlinear programming. Motivation and global convergence. SIAM Journal on Computing 16 (2005) 1-31.

[38] Y.Xiao, B.Yu: A truncated aggregate smoothing Newton method for minimax problems. Applied Mathematics and Computation 216 (2010) 1868-1879.

	Newton methods: n=100						Variable metric methods: n=100					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	2232	7265	11575	0.74	4	-	2849	5078	2821	0.32	2	-
LSMXPI-B	2184	5262	9570	0.60	1	-	1567	2899	1589	0.24	1	-
LSMXSM-A	3454	11682	21398	1.29	5	-	4444	12505	4465	1.03	-	-
LSMXSM-B	10241	36891	56399	4.15	3	-	8861	32056	8881	2.21	1	1
LSMXDI-C	1386	2847	14578	0.90	2	-	2627	5373	2627	0.96	3	-
	Newton methods: n=1000						Variable metric methods: n=1000					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	1386	3735	7488	5.58	4	-	3237	12929	3258	5.91	6	-
LSMXPI-B	3153	6885	12989	9.03	4	-	1522	3287	1544	2.68	5	-
LSMXSM-A	10284	30783	82334	54.38	7	-	4221	9519	4242	8.00	8	-
LSMXSM-B	18279	61180	142767	87.76	6	-	13618	54655	13639	45.10	9	1
LSMXDI-C	3796	6677	48204	49.95	6	-	2371	5548	2371	18.89	3	-

Table 1: TEST14 (minimization of maxima) – 22 problems

	Newton methods: n=100						Variable metric methods: n=100					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	2194	5789	10553	0.67	3	-	2890	5049	2912	0.48	1	-
LSMXPI-B	6767	17901	39544	3.79	4	-	1764	3914	1786	0.37	2	-
LSMXSM-A	3500	9926	23568	1.79	7	-	8455	23644	8476	4.69	4	-
LSMXSM-B	15858	48313	92486	8.33	5	-	9546	34376	9566	2.59	9	1
LSMXDI-C	1371	2901	11580	1.12	8	-	2467	5130	2467	1.59	3	-
	Newton methods: n=1000						Variable metric methods: n=1000					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	4110	14633	20299	18.89	4	-	1549	2636	1571	2.51	3	-
LSMXPI-B	6711	31618	29939	30.73	7	-	1992	6403	2013	4.96	4	-
LSMXSM-A	9994	24333	88481	67.45	11	-	6164	15545	6185	29.37	8	-
LSMXSM-B	23948	84127	182604	149.63	8	-	24027	92477	24048	132.08	8	1
LSMXDI-C	3528	9084	26206	49.78	12	-	1932	2845	1932	18.73	5	-

Table 2: TEST14 (l_∞ approximation) – 22 problems

	Newton methods: n=100						Variable metric methods: n=100					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	15525	20272	55506	4.41	1	-	6497	8204	6518	1.37	3	-
LSMXPI-B	7483	17999	27934	3.27	5	-	1764	7598	2488	0.74	2	-
LSMXSM-A	17574	29780	105531	11.09	4	-	9879	15305	9900	5.95	-	-
LSMXSM-B	13446	29249	81938	6.80	9	1	9546	34376	9566	2.59	3	-
LSMXDI-C	980	1402	7356	0.79	1	-	1179	1837	1179	1.06	2	-
	Newton methods: n=1000						Variable metric methods: n=1000					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
LSMXPI-A	10325	15139	32422	39.30	6	-	6484	9904	6502	13.77	2	-
LSMXPI-B	14836	30724	46864	68.70	10	-	7388	15875	7409	19.98	3	-
LSMXSM-A	11722	24882	69643	61.65	10	-	6659	12824	6681	41.55	8	-
LSMXSM-B	13994	31404	86335	78.45	9	1	15125	25984	15147	61.62	10	-
LSMXDI-C	1408	2406	10121	15.63	6	-	2228	3505	2228	35.13	10	-

Table 3: TEST15 (l_∞ approximation) – 22 problems

	Newton methods: n=100						Variable metric methods: n=100					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
TRSAPI-A	2098	2469	10852	0.57	1	-	22710	22903	22731	1.74	1	1
TRSAPI-B	2286	2771	11353	0.56	1	-	22311	22476	22332	1.62	1	1
LSSAPI-A	1647	5545	8795	0.63	5	-	12265	23579	12287	1.37	2	1
LSSAPI-B	1957	7779	10121	0.67	6	-	4695	6217	10608	0.67	3	-
TRSASM-A	2373	2868	19688	0.73	1	-	22668	22918	22689	2.34	2	1
TRSASM-B	3487	4382	28467	1.12	1	-	22022	22244	22044	1.90	2	1
LSSASM-A	1677	4505	16079	0.74	3	-	20025	27369	20047	2.83	4	-
LSSASM-B	2389	8085	23366	1.18	4	-	5656	11637	5678	1.02	2	-
LSSADI-C	4704	13012	33937	4.16	7	1	6547	7012	6547	9.18	8	-
	Newton methods: n=1000						Variable metric methods: n=1000					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
TRSAPI-A	7570	8955	30013	12.54	3	-	24896	25307	24916	16.22	5	1
TRSAPI-B	8555	9913	36282	18.11	6	-	25013	25492	25033	16.64	5	1
LSSAPI-A	7592	19621	46100	15.39	4	-	22277	36610	22298	19.09	7	1
LSSAPI-B	9067	35463	56292	19.14	6	-	16650	35262	16672	14.47	6	1
TRSASM-A	7922	9453	49104	12.66	2	-	26358	26966	26378	26.44	4	1
TRSASM-B	9559	11358	58418	16.39	7	-	24283	24911	24303	17.79	6	1
LSSASM-A	5696	13534	41347	15.28	4	-	20020	30736	20042	23.05	5	1
LSSASM-B	8517	30736	57878	23.60	6	-	18664	28886	18686	18.65	5	1
LSSADI-C	6758	11011	47960	94.78	11	1	13123	14610	13124	295.46	8	2

Table 4: TEST14 (l_1 approximation) – 22 problems

	Newton methods: n=100						Variable metric methods: n=100					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
TRSAPI-A	15882	16327	54025	3.79	6	-	43038	43435	43054	8.40	8	1
TRSAPI-B	3977	4646	14592	1.15	8	-	6526	6958	6548	0.70	7	1
LSSAPI-A	15760	21846	58082	4.24	8	-	39469	58157	39486	6.28	4	1
LSSAPI-B	4592	17050	17778	1.46	5	-	5932	25035	5952	1.48	6	1
TRSASM-A	6310	7210	38094	2.16	6	-	46219	46759	46237	11.68	7	1
TRSASM-B	4452	5340	26841	1.22	5	-	5240	5821	5409	0.99	5	1
LSSASM-A	10098	14801	610511	3.54	5	-	9162	28421	9184	3.65	6	1
LSSASM-B	4528	14477	290379	2.94	8	-	3507	9036	3528	1.27	6	-
LSSADI-C	877	1373	6026	0.84	3	-	15528	15712	15529	14.49	5	1
	Newton methods: n=1000						Variable metric methods: n=1000					
Method	NIT	NFV	NFG	Time	Δ	Fail	NIT	NFV	NFG	Time	Δ	Fail
TRSAPI-A	14828	16249	85433	29.89	7	-	23758	24120	23778	33.22	9	1
TRSAPI-B	8048	9003	39532	17.45	8	-	10488	11044	10506	11.15	9	1
LSSAPI-A	18519	39319	70951	61.04	5	-	27308	44808	27327	36.64	4	1
LSSAPI-B	12405	57969	43189	55.06	7	-	12712	32179	12731	21.48	8	1
TRSASM-A	11496	13335	63214	26.78	8	1	16345	16754	16362	27.76	8	1
TRSASM-B	9564	11006	53413	24.10	5	-	6993	7525	7011	8.23	6	1
LSSASM-A	19317	32966	113121	62.65	8	-	22264	42908	22284	62.46	7	1
LSSASM-B	14331	33572	86739	57.56	6	-	12898	42479	12919	47.05	7	1
LSSADI-C	2093	3681	12616	20.01	3	1	23957	28000	23960	186.92	5	3

Table 5: TEST15 (l_1 approximation) – 22 problems