Chapter 11
Numerical Solution of Generalized Minimax Problems

Ladislav Lukšan, Ctirad Matonoha, and Jan Vlček

Abstract This contribution contains the description and investigation of three numerical methods for solving generalized minimax problems. These problems consists in the minimization of nonsmooth functions which are compositions of special smooth convex functions with maxima of smooth functions. The most important functions of this type are the sums of maxima of smooth functions. Section 11.2 is devoted to primal interior point methods which use solutions of nonlinear equations for obtaining minimax vectors. Section 11.3 contains investigation of smoothing methods, based on using exponential smoothing terms. Section 11.4 contains 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.

11.1 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 $F(\boldsymbol{x}) = \max_{1 \le k \le m} f_k(\boldsymbol{x})$, or problems where the function to be minimized is a nonsmooth norm, e.g. $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_\infty$, $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_\infty$, $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_1$, $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_1$ with $f(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})]^T$ and $f_+(\boldsymbol{x}) = [\max\{f_1(\boldsymbol{x}), 0\}, \ldots, \max\{f_m(\boldsymbol{x}), 0\}]^T$, 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

L. Lukšan (✉) · C. Matonoha · J. Vlček
Institute of Computer Science, The Czech Academy of Sciences, Prague, Czech Republic
e-mail: luksan@cs.cas.cz; matonoha@cs.cas.cz; vlcek@cs.cas.cz

A. M. Bagirov et al. (eds.), *Numerical Nonsmooth Optimization*,
https://doi.org/10.1007/978-3-030-34910-3_11

minimax problem consists in the use of the function

$$F(\boldsymbol{x}) = \max_{1 \le k \le \overline{k}} \boldsymbol{p}_k^T \boldsymbol{f}(\boldsymbol{x}), \tag{11.1}$$

where $\boldsymbol{p}_k \in \mathbb{R}^m$, $1 \le k \le \overline{k}$, and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth mapping. This function is a special case of composite nonsmooth functions of the form $F(\boldsymbol{x}) = f_0(\boldsymbol{x}) + \max_{1 \le k \le \overline{k}}(\boldsymbol{p}_k^T \boldsymbol{f}(\boldsymbol{x}) + b_k)$, where $f_0 : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function [8, Section 14.1].

Remark 11.1 We can express all above mentioned minimax problems and nonsmooth norms in form (11.1).

(a) Setting $\boldsymbol{p}_k = \boldsymbol{e}_k$, where $\boldsymbol{e}_k$ is the k-th column of an identity matrix and $\overline{k} = m$, we obtain $F(\boldsymbol{x}) = \max_{1 \le k \le m} f_k(\boldsymbol{x})$ (the classical minimax).
(b) Setting $\boldsymbol{p}_k = \boldsymbol{e}_k$, $\boldsymbol{p}_{m+k} = -\boldsymbol{e}_k$ and $\overline{k} = 2m$, we obtain $F(\boldsymbol{x}) = \max_{1 \le k \le m} \max\{f_k(\boldsymbol{x}), -f_k(\boldsymbol{x})\} = \|f(\boldsymbol{x})\|_\infty$.
(c) Setting $\boldsymbol{p}_k = \boldsymbol{e}_k$, $\boldsymbol{p}_{m+1} = \boldsymbol{0}$ and $\overline{k} = m + 1$, we obtain $F(\boldsymbol{x}) = \max\{\max_{1 \le k \le m} f_k(\boldsymbol{x}), 0\} = \|f(\boldsymbol{x})_+\|_\infty$.
(d) If $\overline{k} = 2^m$ and $\boldsymbol{p}_k$, $1 \le k \le 2^m$, are mutually different vectors whose elements are either 1 or -1, we obtain $F(\boldsymbol{x}) = \sum_{k=1}^m \max\{f_k(\boldsymbol{x}), -f_k(\boldsymbol{x})\} = \|f(\boldsymbol{x})\|_1$.
(e) If $\overline{k} = 2^m$ and $\boldsymbol{p}_k$, $1 \le k \le 2^m$, are mutually different vectors whose elements are either 1 or 0, we obtain $F(\boldsymbol{x}) = \sum_{k=1}^m \max\{f_k(\boldsymbol{x}), 0\} = \|f_+(\boldsymbol{x})\|_1$.

Remark 11.2 Since the mapping $\boldsymbol{f}(\boldsymbol{x})$ is continuously differentiable, the function (11.1) is Lipschitz. Thus, if the point $\boldsymbol{x} \in \mathbb{R}^n$ is a local minimum of $F(\boldsymbol{x})$, then $\boldsymbol{0} \in \partial F(\boldsymbol{x})$ [25, Theorem 3.2.5] holds. According to [25, Theorem 3.2.13], one has

$$\partial F(\boldsymbol{x}) = (\nabla \boldsymbol{f}(\boldsymbol{x}))^T \operatorname{conv}\left\{\boldsymbol{p}_k : k \in \bar{\mathcal{I}}(\boldsymbol{x})\right\},$$

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

$$\sum_{k=1}^{\overline{k}} \lambda_k = 1 \quad \text{and} \quad \sum_{k=1}^{\overline{k}} \lambda_k J^T(\boldsymbol{x}) \boldsymbol{p}_k = \boldsymbol{0},$$

where $J(\boldsymbol{x})$ is a Jacobian matrix of the mapping $\boldsymbol{f}(\boldsymbol{x})$.

Remark 11.3 It is clear that a minimum of function (11.1) is a solution of a nonlinear programming problem consisting in minimization of a function $\tilde{F} : \mathbb{R}^{n+1} \to \mathbb{R}$, where $\tilde{F}(\boldsymbol{x}, z) = z$, on the set

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

Denoting $c_k(\boldsymbol{x}, z) = \boldsymbol{p}_k^T \boldsymbol{f}(\boldsymbol{x}) - z$ and $\boldsymbol{a}_k = \nabla c_k(\boldsymbol{x}, z)$, $1 \le k \le \overline{k}$, we obtain $\boldsymbol{a}_k = [\boldsymbol{p}_k^T J(\boldsymbol{x}), -1]^T$ and $\boldsymbol{g} = \nabla \tilde{F}(\boldsymbol{x}, z) = [\boldsymbol{0}^T, 1]^T$, so the necessary KKT conditions can be written in the form

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

$\lambda_k(\boldsymbol{p}_k^T \boldsymbol{f}(\boldsymbol{x}) - z) = 0$, where $\lambda_k \ge 0$, $1 \le k \le \overline{k}$, are the Lagrange multipliers and $z = F(\boldsymbol{x})$. Thus, we obtain the same necessary conditions for an extremum as in Remark 11.2.

From the examples given in Remark 11.1 it follows that composite nondifferentiable functions are not suitable for representation of the functions $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_1$ and $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_1$ because in this case the expression on the right-hand side of (11.1) contains 2^m elements with vectors $\boldsymbol{p}_k$, $1 \le k \le 2^m$. In the subsequent considerations, we will choose a somewhat different approach. We will consider generalized minimax functions established in [5] and [23].

Definition 11.1 We say that $F : \mathbb{R}^n \to \mathbb{R}$ is a generalized minimax function if

$$F(\boldsymbol{x}) = h(F_1(\boldsymbol{x}), \ldots, F_m(\boldsymbol{x})), \quad F_k(\boldsymbol{x}) = \max_{1 \le l \le m_k} f_{kl}(\boldsymbol{x}), \quad 1 \le k \le m, \tag{11.2}$$

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

Assumption 11.1 *Functions f_{kl}, $1 \le k \le m$, $1 \le l \le m_k$, are bounded from below on $\mathbb{R}^n$, so that there exists a constant $\underline{F} \in \mathbb{R}$ such that $f_{kl}(\boldsymbol{x}) \ge \underline{F}$, $1 \le k \le m$, $1 \le l \le m_k$, for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Assumption 11.2 *Functions F_k, $1 \le k \le m$, are bounded from below on $\mathbb{R}^n$, so that there exist constants $\underline{F}_k \in \mathbb{R}$ such that $F_k(\boldsymbol{x}) \ge \underline{F}_k$, $1 \le k \le m$, for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Assumption 11.3 *The function h is twice continuously differentiable and convex satisfying*

$$0 < \underline{h}_k \le \frac{\partial}{\partial z_k} h(\boldsymbol{z}) \le \overline{h}_k, \quad 1 \le k \le m, \tag{11.3}$$

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

Assumption 11.4 *Functions $f_{kl}(\boldsymbol{x})$, $1 \le k \le m$, $1 \le l \le m_k$, are twice continuously differentiable on the convex hull of the level set*

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

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

Remark 11.4 The conditions imposed on the function $h(\boldsymbol{z})$ are relatively strong but many important nonsmooth functions satisfy them.

(1) Let $h : \mathbb{R} \to \mathbb{R}$ be an identity mapping, so $h(z) = z$ and $h'(z) = 1 > 0$. Then setting $\overline{k} = 1, m_1 = \overline{l}$ and

$$F(\boldsymbol{x}) = h(F_1(\boldsymbol{x})) = F_1(\boldsymbol{x}) = \max_{1 \le l \le \overline{l}} \boldsymbol{p}_l^T \boldsymbol{f}(\boldsymbol{x})$$

(since $f_{1l} = \boldsymbol{p}_l^T \boldsymbol{f}(\boldsymbol{x})$), we obtain the composite nonsmooth function (11.1) and therefore the functions $F(\boldsymbol{x}) = \max_{1 \le k \le m} f_k(\boldsymbol{x})$, $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_\infty$, $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_\infty$.

(2) Let $h : \mathbb{R}^m \to \mathbb{R}$, where $h(\boldsymbol{z}) = z_1 + \cdots + z_m$, so $\frac{\partial}{\partial z_k} h(\boldsymbol{z}) = 1 > 0, 1 \le k \le m$. Then function (11.2) has the form

$$F(\boldsymbol{x}) = \sum_{k=1}^{m} F_k(\boldsymbol{x}) = \sum_{k=1}^{m} \max_{1 \le l \le m_k} f_{kl}(\boldsymbol{x}) \tag{11.4}$$

(the sum of maxima). If $m_k = 2$ and $F_k(\boldsymbol{x}) = \max\{f_k(\boldsymbol{x}), -f_k(\boldsymbol{x})\}$, we obtain the function $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_1$. If $m_k = 2$ and $F_k(\boldsymbol{x}) = \max\{f_k(\boldsymbol{x}), 0\}$, we obtain the function $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_1$. It follows that the expression of functions $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_1$ and $F(\boldsymbol{x}) = \|f_+(\boldsymbol{x})\|_1$ by (11.2) 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 11.1(d)–(e).

Remark 11.5 Since the functions $F_k(\boldsymbol{x})$, $1 \le k \le m$, are regular [25, Theorem 3.2.13], the function $h(\boldsymbol{z})$ is continuously differentiable, and $h_k = \frac{\partial}{\partial z_k} h(\boldsymbol{z}) > 0$, one can write [25, Theorem 3.2.9]

$$\partial F(\boldsymbol{x}) = \operatorname{conv} \sum_{k=1}^{m} h_k \partial F_k(\boldsymbol{x}) = \sum_{k=1}^{m} h_k \partial F_k(\boldsymbol{x})$$

$$= \sum_{k=1}^{m} h_k \operatorname{conv}\{\boldsymbol{g}_{kl}(\boldsymbol{x}) : l \in \bar{\mathcal{I}}_k(\boldsymbol{x})\},$$

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

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

where for $1 \le k \le m$ it holds $\lambda_{kl} \ge 0$, $\lambda_{kl}(F_k(\boldsymbol{x}) - f_{kl}(\boldsymbol{x})) = 0$, $1 \le l \le m_k$, 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

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

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

$$\sum_{k=1}^{m} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl} = \boldsymbol{0}, \quad \sum_{l=1}^{m_k} u_{kl} = h_k, \quad h_k = \frac{\partial}{\partial z_k} h(z), \tag{11.5}$$

$$u_{kl} \ge 0, \quad F_k(\boldsymbol{x}) - f_{kl}(\boldsymbol{x}) \ge 0, \quad u_{kl}(F_k(\boldsymbol{x}) - f_{kl}(\boldsymbol{x})) = 0. \tag{11.6}$$

Remark 11.6 Unconstrained minimization of function (11.2) is equivalent to the nonlinear programming problem

$$\begin{cases} \text{minimize} & \tilde{F}(\boldsymbol{x}, z) = h(z) \\ \text{subject to} & f_{kl}(\boldsymbol{x}) \le z_k, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \end{cases} \tag{11.7}$$

The condition (11.3) is sufficient for satisfying equalities $z_k = F_k(\boldsymbol{x})$, $1 \le k \le m$, at the minimum point. Denoting $c_{kl}(\boldsymbol{x}, z) = f_{kl}(\boldsymbol{x}) - z_k$ and $\boldsymbol{a}_{kl}(\boldsymbol{x}, z) = \nabla c_{kl}(\boldsymbol{x}, z)$, $1 \le k \le m$, $1 \le l \le m_k$, we obtain $\boldsymbol{a}_{kl}(\boldsymbol{x}, z) = [\boldsymbol{g}_{kl}^T(\boldsymbol{x}), -\boldsymbol{e}_k^T]^T$, where $\boldsymbol{g}_{kl}(\boldsymbol{x})$ is a gradient of $f_{kl}(\boldsymbol{x})$ in $\boldsymbol{x}$ and $\boldsymbol{e}_k$ is the k-th column of the unit matrix of order m. Thus, the necessary first-order (KKT) conditions have the form

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl} = \boldsymbol{0}, \quad \sum_{l=1}^{m_k} u_{kl} = h_k, \quad h_k = \frac{\partial}{\partial z_k} h(z), \tag{11.8}$$

$$u_{kl} \ge 0, \quad z_k - f_{kl}(\boldsymbol{x}) \ge 0, \quad u_{kl}(z_k - f_{kl}(\boldsymbol{x})) = 0, \tag{11.9}$$

where u_{kl}, $1 \le k \le m$, $1 \le l \le m_k$, are Lagrange multipliers and $z_k = F_k(\boldsymbol{x})$. So we obtain the same necessary conditions for an extremum as in Remark 11.5.

Remark 11.7 A classical minimax problem

$$F(\boldsymbol{x}) = \max_{1 \le k \le m} f_k(\boldsymbol{x}) \tag{11.10}$$

can be replaced with an equivalent nonlinear programming problem

$$\begin{cases} \text{minimize} & \tilde{F}(\boldsymbol{x}, z) = z \\ \text{subject to} & f_k(\boldsymbol{x}) \le z, \quad 1 \le k \le m, \end{cases} \tag{11.11}$$

and the necessary KKT conditions have the form

$$\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x}) u_k = \boldsymbol{0}, \quad \sum_{k=1}^{m} u_k = 1, \tag{11.12}$$

$$u_k \ge 0, \quad z - f_k(\boldsymbol{x}) \ge 0, \quad u_k(z - f_k(\boldsymbol{x})) = 0, \quad 1 \le k \le m. \tag{11.13}$$

Remark 11.8 Minimization of the sum of absolute values

$$F(\boldsymbol{x}) = \sum_{k=1}^{m} |f_k(\boldsymbol{x})| = \sum_{k=1}^{m} \max\{f_k^+(\boldsymbol{x}), f_k^-(\boldsymbol{x})\}, \tag{11.14}$$

where

$$f_k^+(\boldsymbol{x}) = f_k(\boldsymbol{x}), \quad f_k^-(\boldsymbol{x}) = -f_k(\boldsymbol{x})$$

can be replaced with an equivalent nonlinear programming problem

$$\begin{cases} \text{minimize} & \tilde{F}(\boldsymbol{x}, \boldsymbol{z}) = \sum\limits_{k=1}^{m} z_k \\ \text{subject to} & -z_k \le f_k(\boldsymbol{x}) \le z_k, \end{cases} \tag{11.15}$$

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

$$\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x})(u_k^+ - u_k^-) = \boldsymbol{0}, \quad u_k^+ + u_k^- = 1, \tag{11.16}$$

$$u_k^+ \ge 0, \quad z_k - f_k(\boldsymbol{x}) \ge 0, \quad u_k^+(z_k - f_k(\boldsymbol{x})) = 0, \tag{11.17}$$

$$u_k^- \ge 0, \quad z_k + f_k(\boldsymbol{x}) \ge 0, \quad u_k^-(z_k + f_k(\boldsymbol{x})) = 0, \tag{11.18}$$

where $1 \le k \le m$. 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 $z_k = f_k(\boldsymbol{x})$ or $z_k = -f_k(\boldsymbol{x})$, that is $z_k = |f_k(\boldsymbol{x})|$. If $f_k(\boldsymbol{x}) \ne 0$, it cannot simultaneously hold $z_k = f_k(\boldsymbol{x})$ and $z_k = -f_k(\boldsymbol{x})$, so the numbers u_k^+, u_k^- cannot be simultaneously nonzero. Then either $u_k = u_k^+ = 1$ and $z_k = f_k(\boldsymbol{x})$ or $u_k = -u_k^- = -1$ and $z_k = -f_k(\boldsymbol{x})$, that is $u_k = f_k(\boldsymbol{x})/|f_k(\boldsymbol{x})|$. Thus, the necessary KKT conditions have the form

$$\sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x}) u_k = \boldsymbol{0}, \quad z_k = |f_k(\boldsymbol{x})|,$$

$$|u_k| \le 1, \quad u_k = \frac{f_k(\boldsymbol{x})}{|f_k(\boldsymbol{x})|}, \quad \text{if} \quad |f_k(\boldsymbol{x})| > 0. \tag{11.19}$$

Remark 11.9 Minimization of the sum of absolute values can also be reformulated so that more slack variables are used. We obtain the problem

$$\begin{cases} \text{minimize} & \tilde{F}(\boldsymbol{x}, \boldsymbol{z}) = \sum\limits_{k=1}^{m} (z_k^+ + z_k^-) \\ \text{subject to} & f_k(\boldsymbol{x}) = z_k^+ - z_k^-, \quad z_k^+ \ge 0, \quad z_k^- \ge 0, \end{cases} \tag{11.20}$$

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 (11.4), the sums of maxima that include most cases important for applications. In this case, it holds

$$h(\boldsymbol{z}) = \sum_{k=1}^{m} z_k, \quad \nabla h(\boldsymbol{z}) = \tilde{\boldsymbol{e}}, \quad \nabla^2 h(\boldsymbol{z}) = 0, \tag{11.21}$$

where $\tilde{\boldsymbol{e}} \in \mathbb{R}^m$ is a vector with unit elements. The case when $h(\boldsymbol{z})$ is a general function satisfying Assumption 11.3 is studied in [23].

11.2 Primal Interior Point Methods

11.2.1 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 $0 < \mu \le$

$\overline{\mu} < \infty$, is successively minimized on $\mathbb{R}^n$ (without any constraints), where $\mu \downarrow 0$. Applying this approach on the problem (11.7), we obtain a barrier function

$$B_\mu(\boldsymbol{x}, \boldsymbol{z}) = h(\boldsymbol{z}) + \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \varphi(z_k - f_{kl}(\boldsymbol{x})), \quad 0 < \mu \leq \overline{\mu}, \tag{11.22}$$

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

Assumption 11.5 *Function $\varphi(t)$, $t \in (0, \infty)$, is twice continuously differentiable, decreasing, and strictly convex, with $\lim_{t \downarrow 0} \varphi(t) = \infty$. Function $\varphi'(t)$ is increasing and strictly concave such that $\lim_{t \uparrow \infty} \varphi'(t) = 0$. 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*

$$-t\varphi'(t) \geq \underline{c} \tag{11.23}$$

and

$$\varphi'(t)\varphi'''(t) - \varphi''(t)^2 > 0. \tag{11.24}$$

Remark 11.10 A logarithmic barrier function

$$\varphi(t) = \log t^{-1} = -\log t, \tag{11.25}$$

is most frequently used. It satisfies Assumption 11.5 with $\underline{c} = 1$ and $\tau = \infty$ but it is not bounded from below since $\log t \uparrow \infty$ for $t \uparrow \infty$. For that reason, barriers bounded from below are sometimes used, e.g. a function

$$\varphi(t) = \log(t^{-1} + \tau^{-1}) = -\log \frac{t\tau}{t + \tau}, \tag{11.26}$$

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

$$\varphi(t) = -\log t, \quad 0 < t \leq \tau, \qquad \varphi(t) = at^{-2} + bt^{-1} + c, \quad t \geq \tau, \tag{11.27}$$

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

$$\varphi(t) = -\log t, \quad 0 < t \leq \tau, \qquad \varphi(t) = at^{-1} + bt^{-1/2} + c, \quad t \geq \tau, \tag{11.28}$$

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 11.5 [23] (the proof of this statement is trivial for logarithmic barrier (11.25)).

Even if bounded from below barriers (11.26)–(11.28) have more advantageous theoretical properties (Assumption 11.1 can be replaced with a weaker Assump-

tion 11.2 and the proof of Lemma 11.2 below is much simpler, see [23]), algorithms using logarithmic barrier (11.26) are usually more efficient. Therefore, we will only deal with methods using the logarithmic barrier $\varphi(t) = -\log t$ in the subsequent considerations.

11.2.2 Iterative Determination of a Minimax Vector

Suppose the function $h(z)$ is of form (11.21). Using the logarithmic barrier $\varphi(t) = -\log t$, function (11.22) can be written as

$$B_\mu(\boldsymbol{x}, z) = \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k - f_{kl}(\boldsymbol{x})), \quad 0 < \mu \le \overline{\mu}. \tag{11.29}$$

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

$$\begin{aligned} u_{kl}(\boldsymbol{x}, z) &= \frac{\mu}{z_k - f_{kl}(\boldsymbol{x})} \ge 0, \\ v_{kl}(\boldsymbol{x}, z) &= \frac{\mu}{(z_k - f_{kl}(\boldsymbol{x}))^2} = \frac{1}{\mu} u_{kl}^2(\boldsymbol{x}, z) \ge 0, \end{aligned} \tag{11.30}$$

and

$$\boldsymbol{u}_k(\boldsymbol{x}, z) = \begin{bmatrix} u_{k1}(\boldsymbol{x}, z) \\ \vdots \\ u_{km_k}(\boldsymbol{x}, z) \end{bmatrix}, \quad \boldsymbol{v}_k(\boldsymbol{x}, z) = \begin{bmatrix} v_{k1}(\boldsymbol{x}, z) \\ \vdots \\ v_{km_k}(\boldsymbol{x}, z) \end{bmatrix}, \quad \tilde{\boldsymbol{e}}_k = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Denoting by $\boldsymbol{g}(\boldsymbol{x}, z)$ the gradient of the function $B_\mu(\boldsymbol{x}, z)$ and $\gamma_k(\boldsymbol{x}, z) = \frac{\partial}{\partial z_k} B_\mu(\boldsymbol{x}, z)$, the necessary conditions for an extremum of the barrier function (11.22) can be written in the form

$$\boldsymbol{g}(\boldsymbol{x}, z) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x}, z) = \sum_{k=1}^{m} A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}, z) = \boldsymbol{0}, \tag{11.31}$$

$$\gamma_k(\boldsymbol{x}, z) = 1 - \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}, z) = 1 - \tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}, z) = 0, \quad 1 \le k \le m, \tag{11.32}$$

where $A_k(\boldsymbol{x}) = [\boldsymbol{g}_{k1}(\boldsymbol{x}), \ldots, \boldsymbol{g}_{km_k}(\boldsymbol{x})]$, which is a system of $n + m$ nonlinear equations for unknown vectors $\boldsymbol{x}$ 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 (11.31) and (11.32)) are computed. Denoting

$$G(\boldsymbol{x},z)=\sum_{k=1}^{m}\sum_{l=1}^{m_k}G_{kl}(\boldsymbol{x})u_{kl}(\boldsymbol{x},z), \tag{11.33}$$

the Hessian matrix of the Lagrange function and setting

$$U_k(\boldsymbol{x},z)=\mathrm{diag}[u_{k1}(\boldsymbol{x},z),\ldots,u_{km_k}(\boldsymbol{x},z)],$$
$$V_k(\boldsymbol{x},z)=\mathrm{diag}[v_{k1}(\boldsymbol{x},z),\ldots,v_{km_k}(\boldsymbol{x},z)]=\frac{1}{\mu}U_k^2(\boldsymbol{x},z),$$

we can write

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x},z)&=\sum_{k=1}^{m}\sum_{l=1}^{m_k}G_{kl}(\boldsymbol{x})u_{kl}(\boldsymbol{x},z)+\sum_{k=1}^{m}\sum_{l=1}^{m_k}\boldsymbol{g}_{kl}(\boldsymbol{x})v_{kl}(\boldsymbol{x},z)\boldsymbol{g}_{kl}^T(\boldsymbol{x})\\&=G(\boldsymbol{x},z)+\sum_{k=1}^{m}A_k(\boldsymbol{x})V_k(\boldsymbol{x},z)A_k^T(\boldsymbol{x}),\end{aligned} \tag{11.34}$$

$$\frac{\partial}{\partial z_k}\boldsymbol{g}(\boldsymbol{x},z)=-\sum_{l=1}^{m_k}\boldsymbol{g}_{kl}(\boldsymbol{x})v_{kl}(\boldsymbol{x},z)=-A_k(\boldsymbol{x})\boldsymbol{v}_k(\boldsymbol{x},z), \tag{11.35}$$

$$\frac{\partial}{\partial\boldsymbol{x}}\gamma_k(\boldsymbol{x},z)=-\sum_{l=1}^{m_k}v_{kl}(\boldsymbol{x},z)\boldsymbol{g}_{kl}^T(\boldsymbol{x})=-\boldsymbol{v}_k^T(\boldsymbol{x},z)A_k^T(\boldsymbol{x}), \tag{11.36}$$

$$\frac{\partial}{\partial z_k}\gamma_k(\boldsymbol{x},z)=\sum_{l=1}^{m_k}v_{kl}(\boldsymbol{x},z)=\tilde{\boldsymbol{e}}_k^T\boldsymbol{v}_k(\boldsymbol{x},z). \tag{11.37}$$

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

$$\begin{bmatrix}W(\boldsymbol{x},z) & -A_1(\boldsymbol{x})\boldsymbol{v}_1(\boldsymbol{x},z) & \cdots & -A_m(\boldsymbol{x})\boldsymbol{v}_m(\boldsymbol{x},z)\\ -\boldsymbol{v}_1^T(\boldsymbol{x},z)A_1^T(\boldsymbol{x}) & \tilde{\boldsymbol{e}}_1^T\boldsymbol{v}_1(\boldsymbol{x},z) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ -\boldsymbol{v}_m^T(\boldsymbol{x},z)A_m^T(\boldsymbol{x}) & 0 & \cdots & \tilde{\boldsymbol{e}}_m^T\boldsymbol{v}_m(\boldsymbol{x},z)\end{bmatrix}\begin{bmatrix}\Delta\boldsymbol{x}\\ \Delta z_1\\ \vdots\\ \Delta z_m\end{bmatrix} \tag{11.38}$$
$$=-\begin{bmatrix}\boldsymbol{g}(\boldsymbol{x},z)\\ \gamma_1(\boldsymbol{x},z)\\ \vdots\\ \gamma_m(\boldsymbol{x},z)\end{bmatrix},$$

where

$$W(\boldsymbol{x},z)=G(\boldsymbol{x},z)+\sum_{k=1}^{m}A_k(\boldsymbol{x})V_k(\boldsymbol{x},z)A_k^T(\boldsymbol{x}). \tag{11.39}$$

Setting

$$C(\boldsymbol{x},z)=[A_1(\boldsymbol{x})\boldsymbol{v}_1(\boldsymbol{x},z),\ldots,A_m(\boldsymbol{x})\boldsymbol{v}_m(\boldsymbol{x},z)],$$
$$D(\boldsymbol{x},z)=\mathrm{diag}[\tilde{\boldsymbol{e}}_1^T\boldsymbol{v}_1(\boldsymbol{x},z),\ldots,\tilde{\boldsymbol{e}}_m^T\boldsymbol{v}_m(\boldsymbol{x},z)]$$

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

$$\begin{bmatrix} W(\boldsymbol{x},z) & -C(\boldsymbol{x},z)\\ -C^T(\boldsymbol{x},z) & D(\boldsymbol{x},z)\end{bmatrix}\begin{bmatrix}\Delta\boldsymbol{x}\\ \Delta z\end{bmatrix}=-\begin{bmatrix}\boldsymbol{g}(\boldsymbol{x},z)\\ \boldsymbol{\gamma}(\boldsymbol{x},z)\end{bmatrix}. \tag{11.40}$$

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

During iterative determination of a minimax vector we know a value of the parameter μ and vectors $\boldsymbol{x}\in\mathbb{R}^n$, $z\in\mathbb{R}^m$ such that $z_k>F_k(\boldsymbol{x})$, $1\le k\le m$. Using formula (11.40) we determine direction vectors $\Delta\boldsymbol{x}$, Δz. Then, we choose a step length α so that

$$B_\mu(\boldsymbol{x}+\alpha\Delta\boldsymbol{x},z+\alpha\Delta z)<B_\mu(\boldsymbol{x},z) \tag{11.41}$$

and $z_k+\alpha\Delta z_k>F_k(\boldsymbol{x}+\alpha\Delta\boldsymbol{x})$, $1\le k\le m$. Finally, we set $\boldsymbol{x}_+=\boldsymbol{x}+\alpha\Delta\boldsymbol{x}$, $z_+=z+\alpha\Delta z$ and determine a new value $\mu_+<\mu$. If the matrix of system of equations (11.40) is positive definite, inequality (11.41) is satisfied for a sufficiently small value of the step length α.

Theorem 11.1 *Let the matrix $G(\boldsymbol{x},z)$ given by (11.33) be positive definite. Then the matrix of system of equations (11.40) is positive definite.*

Proof The matrix of system of equations (11.40) is positive definite if and only if the matrix D and its Schur complement $W-CD^{-1}C^T$ are positive definite [7, Theorem 2.5.6]. The matrix D is positive definite since it has positive diagonal elements. Further, it holds

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

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

11.2.3 Direct Determination of a Minimax Vector

Now we will show how to solve system of equations (11.31)–(11.32) by direct determination of a minimax vector using two-level optimization

$$z(\boldsymbol{x};\mu) = \operatorname*{argmin}_{z\in\mathbb{R}^m} B_\mu(\boldsymbol{x},z), \tag{11.42}$$

and

$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \hat{B}(\boldsymbol{x};\mu), \quad \hat{B}(\boldsymbol{x};\mu) \stackrel{\Delta}{=} B_\mu(\boldsymbol{x},z(\boldsymbol{x};\mu)). \tag{11.43}$$

The problem (11.42) serves for determination of an optimal vector $z(\boldsymbol{x};\mu)\in\mathbb{R}^m$. Let $\tilde{B}_\mu(z) = B_\mu(\boldsymbol{x},z)$ for a fixed chosen vector $\boldsymbol{x}\in\mathbb{R}^n$. The function $\tilde{B}_\mu(z)$ is strictly convex (as a function of a vector z), since it is a sum of convex function (11.21) and strictly convex functions $-\mu\log(z_k - f_{kl}(\boldsymbol{x}))$, $1\le k\le m$, $1\le l\le m_k$. A minimum of the function $\tilde{B}_\mu(z)$ is its stationary point, so it is a solution of system of equations (11.32) with Lagrange multipliers (11.30). The following theorem shows that this solution exists and is unique.

Theorem 11.2 *The function $\tilde{B}_\mu(z) : (F(\boldsymbol{x}),\infty)\to\mathbb{R}$ has a unique stationary point which is its global minimum. This stationary point is characterized by a system of equations $\boldsymbol{\gamma}(\boldsymbol{x},z)=\boldsymbol{0}$, or*

$$1 - \tilde{\boldsymbol{e}}_k^T\boldsymbol{u}_k = 1 - \sum_{l=1}^{m_k}\frac{\mu}{z_k - f_{kl}(\boldsymbol{x})} = 0, \quad 1\le k\le m, \tag{11.44}$$

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

$$F_k(\boldsymbol{x}) < F_k(\boldsymbol{x}) + \mu < z_k(\boldsymbol{x};\mu) < F_k(\boldsymbol{x}) + m_k\mu \tag{11.45}$$

for $1\le k\le m$.

Proof Definition 11.1 implies $f_{kl}(\boldsymbol{x})\le F_k(\boldsymbol{x})$, $1\le k\le m$, $1\le l\le m_k$, where the equality occurs for at least one index l.

(a) If (11.44) holds, then we can write

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

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

which proves inequalities (11.45).

(b) Since

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

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

and the function $\gamma_k(\boldsymbol{x}, z_k)$ is continuous and decreasing in $F_k(\boldsymbol{x}) + \mu < z_k(\boldsymbol{x}; \mu) < F_k(\boldsymbol{x}) + m_k$ by (11.37), the equation $\gamma_k(\boldsymbol{x}, z_k) = 0$ has a unique solution in this interval. Since the function $\tilde{B}_\mu(\boldsymbol{z})$ is convex this solution corresponds to its global minimum.

□

System (11.44) is a system of m scalar equations with localization inequalities (11.45). These scalar equations can be efficiently solved by robust methods described e.g. in [14] and [15] (details are stated in [22]). Suppose that $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x}; \mu)$ and denote

$$\hat{B}(\boldsymbol{x}; \mu) = \sum_{k=1}^{m} z_k(\boldsymbol{x}; \mu) - \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k(\boldsymbol{x}; \mu) - f_{kl}(\boldsymbol{x})). \tag{11.46}$$

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

Theorem 11.3 *Consider the barrier function (11.46). Then*

$$\nabla \hat{B}(\boldsymbol{x}; \mu) = \sum_{k=1}^{m} A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}; \mu), \tag{11.47}$$

$$\begin{aligned} \nabla^2 \hat{B}(\boldsymbol{x}; \mu) &= W(\boldsymbol{x}; \mu) - C(\boldsymbol{x}; \mu) D^{-1}(\boldsymbol{x}; \mu) C^T(\boldsymbol{x}; \mu) \\ &= G(\boldsymbol{x}; \mu) + \sum_{k=1}^{m} A_k(\boldsymbol{x}) V_k(\boldsymbol{x}; \mu) A_k^T(\boldsymbol{x}) \\ &\quad - \sum_{k=1}^{m} \frac{A_k(\boldsymbol{x}) V_k(\boldsymbol{x}; \mu) \tilde{\boldsymbol{e}}_k \tilde{\boldsymbol{e}}_k^T V_k(\boldsymbol{x}; \mu) A_k^T(\boldsymbol{x})}{\tilde{\boldsymbol{e}}_k^T V_k(\boldsymbol{x}; \mu) \tilde{\boldsymbol{e}}_k}, \end{aligned} \tag{11.48}$$

where $G(\boldsymbol{x}; \mu) = G(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{x}; \mu))$ *and* $W(\boldsymbol{x}; \mu)$, $C(\boldsymbol{x}; \mu)$, $D(\boldsymbol{x}; \mu)$, $U_k(\boldsymbol{x}; \mu)$, $V_k(\boldsymbol{x}; \mu) = U_k^2(\boldsymbol{x}; \mu)/\mu$, $1 \le k \le m$, *are obtained by the same substitution. A solution of equation*

$$\nabla^2 \hat{B}(\boldsymbol{x}; \mu) \Delta \boldsymbol{x} = -\nabla \hat{B}(\boldsymbol{x}; \mu) \tag{11.49}$$

is identical with $\Delta \boldsymbol{x}$ given by (11.40), where $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x}; \mu)$ (so $\boldsymbol{\gamma}(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{x}; \mu)) = \boldsymbol{0}$).

Proof Differentiating the barrier function (11.46) and using (11.32) we obtain

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

where

$$u_{kl}(\boldsymbol{x}; \mu) = \frac{\mu}{z_k(\boldsymbol{x}; \mu) - f_{kl}(\boldsymbol{x})}, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \tag{11.50}$$

Formula (11.48) can be obtained by additional differentiation of relations (11.32) and (11.47) using (11.50). A simpler way is based on using (11.40). Since (11.32) implies $\boldsymbol{\gamma}(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{x}; \mu)) = \boldsymbol{0}$, we can substitute $\boldsymbol{\gamma} = \boldsymbol{0}$ into (11.40), which yields the equation

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

where $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x}; \mu)$, that confirms validity of formulas (11.48) and (11.49) (details can be found in [22]). □

Remark 11.11 To determine an inverse of the Hessian matrix, one can use a Woodbury formula [7, Theorem 12.1.4] which gives

$$\begin{aligned}(\nabla^2 \hat{B}(\boldsymbol{x}; \mu))^{-1} = {} & W^{-1}(\boldsymbol{x}; \mu) - W^{-1}(\boldsymbol{x}; \mu) C(\boldsymbol{x}; \mu) \\ & \left(C^T(\boldsymbol{x}; \mu) W^{-1}(\boldsymbol{x}; \mu) C(\boldsymbol{x}; \mu) - D(\boldsymbol{x}; \mu) \right)^{-1} \\ & C^T(\boldsymbol{x}; \mu) W^{-1}(\boldsymbol{x}; \mu).\end{aligned} \tag{11.51}$$

If the matrix $\nabla^2 \hat{B}(\boldsymbol{x}; \mu)$ is not positive definite, it can be replaced by a matrix $LL^T = \nabla^2 \hat{B}(\boldsymbol{x}; \mu) + E$, obtained by the Gill–Murray decomposition [10]. Note that it is more advantageous to use system of linear equations (11.40) instead of (11.49) for determination of a direction vector $\Delta \boldsymbol{x}$ because the system of nonlinear equations (11.44) is solved with prescribed finite precision, and thus a vector $\boldsymbol{\gamma}(\boldsymbol{x}, \boldsymbol{z})$, defined by (11.32), need not be zero.

From

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

it follows that $\|V_k(\boldsymbol{x};\mu)\|\uparrow\infty$ if $\mu\downarrow 0$, so Hessian matrix (11.48) 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 11.4 *Let Assumption 11.4 be satisfied and $\mu\geq\underline{\mu}>0$. If $G(\boldsymbol{x};\mu)$ is uniformly positive definite (if a constant $\underline{G}$ exists such that $\boldsymbol{v}^TG(\boldsymbol{x};\mu)\boldsymbol{v}\geq\underline{G}\|\boldsymbol{v}\|^2$), then there is a number $\overline{\kappa}\geq 1$ such that $\kappa(\nabla^2\hat{B}(\boldsymbol{x};\mu))\leq\overline{\kappa}$.*

Proof

(a) Using (11.30), (11.48), and Assumption 11.4, we obtain

$$\begin{aligned}\|\nabla^2\hat{B}(\boldsymbol{x};\mu)\|&\leq\left\|G(\boldsymbol{x};\mu)+\sum_{k=1}^{m}A_k(\boldsymbol{x})V_k(\boldsymbol{x};\mu)A_k^T(\boldsymbol{x})\right\|\\&\leq\sum_{k=1}^{m}\sum_{l=1}^{m_k}\left(|G_{kl}(\boldsymbol{x})u_{kl}(\boldsymbol{x},\mu)|+\frac{1}{\mu}\left|u_{kl}^2(\boldsymbol{x};\mu)\boldsymbol{g}_{kl}(\boldsymbol{x})\boldsymbol{g}_{kl}^T(\boldsymbol{x})\right|\right)\\&\leq\frac{\overline{m}}{\mu}\left(\overline{\mu}\,\overline{G}+\overline{g}^2\right)\triangleq\frac{\overline{c}}{\mu}\leq\frac{\overline{c}}{\underline{\mu}},\end{aligned}\tag{11.52}$$

because $0\leq u_{kl}(\boldsymbol{x};\mu)\leq\tilde{\boldsymbol{e}}_k^T\boldsymbol{u}_k(\boldsymbol{x};\mu)=1$, $1\leq k\leq m$, $1\leq l\leq m_k$, by (11.44).

(b) From the proof of Theorem 11.1 it follows that the matrix $\nabla^2\hat{B}(\boldsymbol{x};\mu)-G(\boldsymbol{x};\mu)$ is positive semidefinite. Therefore,

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

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

$$\kappa(\nabla^2\hat{B}(\boldsymbol{x};\mu))=\frac{\overline{\lambda}(\nabla^2\hat{B}(\boldsymbol{x};\mu))}{\underline{\lambda}(\nabla^2\hat{B}(\boldsymbol{x};\mu))}\leq\frac{\overline{c}}{\underline{\mu}\,\underline{G}}\triangleq\overline{\kappa}.\tag{11.53}$$

□

Remark 11.12 If there exists a number $\overline{\kappa}>0$ such that $\kappa(\nabla^2\hat{B}(\boldsymbol{x}_i;\mu_i))\leq\overline{\kappa}$, $i\in\mathbb{N}$, the direction vector $\Delta\boldsymbol{x}_i$, given by solving a system of equations $\nabla^2\hat{B}(\boldsymbol{x}_i;\mu_i)\Delta\boldsymbol{x}_i=-\nabla\hat{B}(\boldsymbol{x}_i;\mu_i)$, satisfies the condition

$$(\Delta\boldsymbol{x}_i)^T\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\leq-\varepsilon_0\|\Delta\boldsymbol{x}_i\|\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|,\quad i\in\mathbb{N},\tag{11.54}$$

where $\varepsilon_0=1/\sqrt{\overline{\kappa}}$ and $\boldsymbol{g}(\boldsymbol{x};\mu)=\nabla\hat{B}(\boldsymbol{x};\mu)$. Then, for arbitrary numbers $0<\varepsilon_1\leq\varepsilon_2<1$ one can find a step length parameter $\alpha_i>0$ such that for $\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\alpha_i\Delta\boldsymbol{x}_i$

it holds

$$\varepsilon_1 \leq \frac{\hat{B}(\boldsymbol{x}_{i+1};\mu_i) - \hat{B}(\boldsymbol{x}_i;\mu_i)}{\alpha_i(\Delta\boldsymbol{x}_i)^T\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)} \leq \varepsilon_2, \tag{11.55}$$

so there exists a number $c > 0$ such that (see [26, Section 3.2])

$$\hat{B}(\boldsymbol{x}_{i+1};\mu_i) - \hat{B}(\boldsymbol{x}_i;\mu_i) \leq -c\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|^2, \quad i \in \mathbb{N}. \tag{11.56}$$

If Assumption 11.4 is not satisfied, then only $(\Delta\boldsymbol{x}_i)^T\boldsymbol{g}(\boldsymbol{x}_i;\mu_i) < 0$ holds (because the matrix $\nabla^2\hat{B}(\boldsymbol{x};\mu)$ is positive definite by Theorem 11.1) and

$$\hat{B}(\boldsymbol{x}_{i+1};\mu_i) - \hat{B}(\boldsymbol{x}_i;\mu_i) \leq 0, \quad i \in \mathbb{N}. \tag{11.57}$$

11.2.4 Implementation

Remark 11.13 In (11.39), it is assumed that $G(\boldsymbol{x},\boldsymbol{z})$ is the Hessian matrix of the Lagrange function. Direct computation of the matrix $G(\boldsymbol{x};\mu) = G(\boldsymbol{x},\boldsymbol{z}(\boldsymbol{x};\mu))$ is usually difficult (one can use automatic differentiation as described in [13]). Thus, various approximations $G \approx G(\boldsymbol{x};\mu)$ are mostly used.

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

$$G\boldsymbol{w}_j = \frac{1}{\delta}\left(\sum_{k=1}^{m} A_k(\boldsymbol{x}+\delta\boldsymbol{w}_j)\boldsymbol{u}_k(\boldsymbol{x};\mu) - \sum_{k=1}^{m} A(\boldsymbol{x})\boldsymbol{u}_k(\boldsymbol{x};\mu)\right).$$

  The vectors $\boldsymbol{w}_j$, $1 \leq j \leq \overline{k}$, are chosen so that the number of them is as small as possible [4, 27].
- The matrix $G \approx G(\boldsymbol{x};\mu)$ can be determined using the variable metric methods [17]. The vectors

$$\boldsymbol{d} = \boldsymbol{x}_+ - \boldsymbol{x}, \quad \boldsymbol{y} = \sum_{k=1}^{m} A_k(\boldsymbol{x}_+)\boldsymbol{u}_k(\boldsymbol{x}_+;\mu) - \sum_{k=1}^{m} A_k(\boldsymbol{x})\boldsymbol{u}_k(\boldsymbol{x}_+;\mu)$$

  are used for an update of G.
- If the problem is separable (i.e. $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, are functions of a small number $n_{kl} = O(1)$ of variables), one can set as in [12]

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

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

Remark 11.14 The matrix $G \approx G(\boldsymbol{x}; \mu)$ obtained by the approach stated in Remark 11.13 can be ill-conditioned so condition (11.54) (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 (11.52) and (11.53), so it is a higher probability of fulfilment of condition (11.54). If the choice $G = I$ does not satisfy (11.54), we set $\Delta\boldsymbol{x} = -\boldsymbol{g}(\boldsymbol{x}; \mu)$ (a steepest descent direction).

An update of μ is an important part of interior point methods. Above all, $\mu \downarrow 0$ must hold, which is a main property of interior point methods. Moreover, rounding errors may cause that $z_k(\boldsymbol{x}; \mu) = F_k(\boldsymbol{x})$ when the value μ is small (because $F_k(\boldsymbol{x}) < z_k(\boldsymbol{x}; \mu) \leq F_k(\boldsymbol{x}) + m_k\mu$ and $F_k(\boldsymbol{x}) + m_k\mu \to F_k(\boldsymbol{x})$ if $\mu \downarrow 0$), which leads to a breakdown (division by $z_k(\boldsymbol{x}; \mu) - F_k(\boldsymbol{x}) = 0$) when computing $1/(z_k(\boldsymbol{x}; \mu) - F_k(\boldsymbol{x}))$. 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 $\underline{g}$ is a suitable constant.

Procedure A

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

Phase 2. If $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| < \underline{g}$, then

$$\mu_{i+1} = \max\left\{\tilde{\mu}_{i+1},\ \underline{\mu},\ 10\,\varepsilon_M|F(\boldsymbol{x}_{i+1})|\right\}, \tag{11.58}$$

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

$$\tilde{\mu}_{i+1} = \min\left\{\max\{\lambda\mu_i,\ \mu_i/(\sigma\mu_i + 1)\}, \max\{\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2,\ 10^{-2k}\}\right\}. \tag{11.59}$$

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

Procedure B

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

Phase 2. If $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2 < \vartheta\mu_i$, then

$$\mu_{i+1} = \max\{\underline{\mu}, \|\boldsymbol{g}_i(\boldsymbol{x}_i; \mu_i)\|^2\}. \tag{11.60}$$

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 (11.59) 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 $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|$ is small even if $\boldsymbol{x}_i$ is far from a solution. The idea of Procedure B proceeds from the fact that a barrier function $\hat{B}(\boldsymbol{x}; \mu)$ should be minimized with a sufficient precision for a given value of a parameter μ.

The considerations up to now are summarized in Algorithm 11.1 introduced in the Appendix. This algorithm supposes that the matrix $A(\boldsymbol{x})$ is sparse. If it is dense, the algorithm is simplified because there is no symbolic decomposition.

11.2.5 Global Convergence

Now we prove the global convergence of the method realized by Algorithm 11.1.

Lemma 11.1 *Let numbers $z_k(\boldsymbol{x}; \mu)$, $1 \le k \le m$, be solutions of Eq. (11.44). Then*

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

Proof Differentiating (11.44) with respect to μ, one can write for $1 \le k \le m$

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

which after multiplication of μ together with (11.30) and (11.44) gives

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

Differentiating the function

$$\hat{B}(\boldsymbol{x}; \mu) = \sum_{k=1}^{m} z_k(\boldsymbol{x}; \mu) - \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k(\boldsymbol{x}; \mu) - f_{kl}(\boldsymbol{x})) \tag{11.61}$$

and using (11.44) we obtain

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

□

Lemma 11.2 *Let Assumption 11.1 be satisfied. Let* $\{\boldsymbol{x}_i\}$ *and* $\{\mu_i\}$, $i \in \mathbb{N}$, *be the sequences generated by Algorithm 11.1. Then the sequences* $\{\hat{B}(\boldsymbol{x}_i;\mu_i)\}$, $\{z(\boldsymbol{x}_i;\mu_i)\}$, *and* $\{F(\boldsymbol{x}_i)\}$, $i \in \mathbb{N}$, *are bounded. Moreover, there exists a constant* $L \geq 0$ *such that for* $i \in \mathbb{N}$ *it holds*

$$\hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) \leq \hat{B}(\boldsymbol{x}_{i+1};\mu_i) + L(\mu_i - \mu_{i+1}). \tag{11.62}$$

Proof

(a) We first prove boundedness from below. Using (11.61) and Assumption 11.1, one can write

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

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

$$\hat{B}(\boldsymbol{x};\mu) \geq \underline{F} + \sum_{k=1}^{m}(m_k\mu - m_k\mu\log(m_k\mu))$$

$$\geq \underline{F} + \sum_{k=1}^{m} \min\{0,\ m_k \overline{\mu}(1 - \log(m_k \overline{\mu})\}$$

$$\geq \underline{F} + \sum_{k=1}^{m} \min\{0,\ m_k \overline{\mu}(1 - \log(2m_k \overline{\mu})\} \stackrel{\Delta}{=} \underline{B}.$$

Boundedness from below of sequences $\{z(\boldsymbol{x}_i; \mu_i)\}$ and $\{F(\boldsymbol{x}_i)\}$, $i \in \mathbb{N}$, follows from inequalities (11.45) and Assumption 11.1.

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

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

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

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

$$= \sum_{k=1}^{m} \frac{z_k(\boldsymbol{x}; \mu) - \underline{F}}{2} + \underline{B}$$

or

$$\sum_{k=1}^{m} (z_k(\boldsymbol{x}; \mu) - \underline{F}) \leq 2(\hat{B}(\boldsymbol{x}; \mu) - \underline{B}). \tag{11.63}$$

Using the mean value theorem and Lemma 11.1, we obtain

$$\hat{B}(\boldsymbol{x}_{i+1}; \mu_{i+1}) - \hat{B}(\boldsymbol{x}_{i+1}; \mu_i)$$

$$= \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k(\boldsymbol{x}_{i+1}; \tilde{\mu}_i) - f_{kl}(\boldsymbol{x}_{i+1}))(\mu_i - \mu_{i+1})$$

$$\leq \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(z_k(\boldsymbol{x}_{i+1}; \mu_i) - f_{kl}(\boldsymbol{x}_{i+1}))(\mu_i - \mu_{i+1})$$

$$\leq \sum_{k=1}^{m} m_k \log(z_k(\boldsymbol{x}_{i+1}; \mu_i) - \underline{F})(\mu_i - \mu_{i+1}), \tag{11.64}$$

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 (11.63), (11.64), and (11.45)

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

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

$$\begin{aligned}
\hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) - \underline{B} &\leq \prod_{j=1}^{i}(1 + \lambda\delta_j)(\hat{B}(\boldsymbol{x}_1;\mu_1) - \underline{B}) \\
&\leq \prod_{i=1}^{\infty}(1 + \lambda\delta_i)(\hat{B}(\boldsymbol{x}_1;\mu_1) - \underline{B}) \qquad (11.65)
\end{aligned}$$

and since

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

the expression on the right-hand side of (11.65) is finite. Thus, the sequence $\{\hat{B}(\boldsymbol{x}_i;\mu_i)\}$, $i \in \mathbb{N}$, is bounded from above and the sequences $\{z(\boldsymbol{x}_i;\mu_i)\}$ and $\{F(\boldsymbol{x}_i)\}$, $i \in \mathbb{N}$, are bounded from above as well by (11.63) and (11.45).

(c) Finally, we prove formula (11.62). Using (11.64) and (11.45) we obtain

$$\begin{aligned}
&\hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) - \hat{B}(\boldsymbol{x}_{i+1};\mu_i) \\
&\qquad \leq \sum_{k=1}^{m} m_k \log(z_k(\boldsymbol{x}_{i+1};\mu_i) - \underline{F})(\mu_i - \mu_{i+1})
\end{aligned}$$

$$\leq \sum_{k=1}^{m} m_k \log(F_k(\boldsymbol{x}_{i+1}) + m_k\mu_i - \underline{F})(\mu_i - \mu_{i+1})$$

$$\leq \sum_{k=1}^{m} m_k \log(\overline{F} + m_k\overline{\mu} - \underline{F})(\mu_i - \mu_{i+1})$$

$$\stackrel{\Delta}{=} L(\mu_i - \mu_{i+1})$$

(the existence of a constant $\overline{F}$ follows from boundedness of a sequence $\{F(\boldsymbol{x}_i)\}$, $i \in \mathbb{N}$), which together with (11.57) gives $\hat{B}(\boldsymbol{x}_{i+1}; \mu_{i+1}) \leq \hat{B}(\boldsymbol{x}_i; \mu_i) + L(\mu_i - \mu_{i+1})$, $i \in \mathbb{N}$. Thus, it holds

$$\hat{B}(\boldsymbol{x}_i; \mu_i) \leq \hat{B}(\boldsymbol{x}_1; \mu_1) + L(\mu_1 - \mu_i) \leq \hat{B}(\boldsymbol{x}_1; \mu_1) + L\overline{\mu} \stackrel{\Delta}{=} \overline{B}, \quad i \in N. \tag{11.66}$$

□

The upper bounds $\overline{g}$ and $\overline{G}$ are not used in Lemma 11.2, so Assumption 11.4 may not be satisfied. Thus, there exists an upper bound $\overline{F}$ (independent of $\overline{g}$ and $\overline{G}$) such that $F(\boldsymbol{x}_i) \leq \overline{F}$ for all $i \in \mathbb{N}$. This upper bound can be used in definition of a set $\mathcal{D}_F(\overline{F})$ in Assumption 11.4.

Lemma 11.3 *Let Assumption 11.4 and the assumptions of Lemma 11.2 be satisfied. Then, if we use Procedure A or Procedure B for an update of parameter μ, the values $\{\mu_i\}$, $i \in \mathbb{N}$, form a non-decreasing sequence such that $\mu_i \downarrow 0$.*

Proof The value of parameter μ is unchanged in the first phase of Procedure A or Procedure B. Since a function $\hat{B}(\boldsymbol{x}; \mu)$ is continuous, bounded from below by Lemma 11.2, and since inequality (11.56) is satisfied (with $\mu_i = \mu$), it holds $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu)\| \downarrow 0$ if phase 1 contains an infinite number of subsequent iterative steps [26, Section 3.2]. Thus, there exists a step (with index i) belonging to the first phase such that either $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu)\| < \underline{g}$ in Procedure A or $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu)\|^2 < \vartheta\mu$ 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 $\mu_i \downarrow 0$. □

Theorem 11.5 *Let assumptions of Lemma 11.3 be satisfied. Consider a sequence $\{\boldsymbol{x}_i\}$, $i \in \mathbb{N}$, generated by Algorithm 11.1, where $\underline{\delta} = \underline{\varepsilon} = \underline{\mu} = 0$. Then*

$$\lim_{i\uparrow\infty} \sum_{k=1}^{m} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}_i) u_{kl}(\boldsymbol{x}_i; \mu_i) = \boldsymbol{0}, \quad \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}_i; \mu_i) = 1,$$

$$z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i) \geq 0, \quad u_{kl}(\boldsymbol{x}_i; \mu_i) \geq 0,$$

$$\lim_{i\uparrow\infty} u_{kl}(\boldsymbol{x}_i; \mu_i)(z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i)) = 0$$

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

Proof

(a) Equalities $\tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i;\mu_i) = 1$, $1 \le k \le m$, are satisfied by (11.44) because $\underline{\delta} = 0$. Inequalities $z_k(\boldsymbol{x}_i;\mu_i) - f_{kl}(\boldsymbol{x}_i) \ge 0$ and $u_{kl}(\boldsymbol{x}_i;\mu_i) \ge 0$ follow from formulas (11.45) and statement (11.50).

(b) Relations (11.56) and (11.62) yield

$$\begin{aligned}\hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) - \hat{B}(\boldsymbol{x}_i;\mu_i) &= (\hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) - \hat{B}(\boldsymbol{x}_{i+1};\mu_i)) \\ &\quad + (\hat{B}(\boldsymbol{x}_{i+1};\mu_i) - \hat{B}(\boldsymbol{x}_i;\mu_i)) \\ &\le L\,(\mu_i - \mu_{i+1}) - c\,\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|^2\end{aligned}$$

and since $\lim_{i\uparrow\infty} \mu_i = 0$ (Lemma 11.3), we can write by (11.66) that

$$\begin{aligned}\underline{B} &\le \lim_{i\uparrow\infty} \hat{B}(\boldsymbol{x}_{i+1};\mu_{i+1}) \\ &\le \hat{B}(\boldsymbol{x}_1;\mu_1) + L\sum_{i=1}^{\infty}(\mu_i - \mu_{i+1}) - c\sum_{i=1}^{\infty}\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|^2 \\ &\le \hat{B}(\boldsymbol{x}_1;\mu_1) + L\overline{\mu} - c\sum_{i=1}^{\infty}\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|^2 = \overline{B} - c\sum_{i=1}^{\infty}\|\boldsymbol{g}(\boldsymbol{x}_i;\mu_i)\|^2.\end{aligned}$$

Thus, it holds

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

which gives $\boldsymbol{g}(\boldsymbol{x}_i;\mu_i) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}_i)u_{kl}(\boldsymbol{x}_i;\mu_i) \downarrow \boldsymbol{0}$.

(c) Let indices $1 \le k \le m$ and $1 \le l \le m_k$ are chosen arbitrarily. Using (11.50) and Lemma 11.3 we obtain

$$u_{kl}(\boldsymbol{x}_i;\mu_i)(z_k(\boldsymbol{x}_i;\mu_i) - f_{kl}(\boldsymbol{x}_i)) = \frac{\mu_i(z_k(\boldsymbol{x}_i;\mu_i) - f_{kl}(\boldsymbol{x}_i))}{z_k(\boldsymbol{x}_i;\mu_i) - f_{kl}(\boldsymbol{x}_i)} = \mu_i \downarrow 0.$$

□

Corollary 11.1 *Let the assumptions of Theorem 11.5 be satisfied. Then, every cluster point $\boldsymbol{x} \in \mathbb{R}^n$ of a sequence $\{\boldsymbol{x}_i\}$, $i \in \mathbb{N}$, satisfies necessary KKT conditions (11.8)–(11.9) where $\boldsymbol{z}$ and $\boldsymbol{u}$ (with elements z_k and u_{kl}, $1 \le k \le m$, $1 \le l \le m_k$) are cluster points of sequences $\{\boldsymbol{z}(\boldsymbol{x}_i;\mu_i)\}$ and $\{\boldsymbol{u}(\boldsymbol{x}_i;\mu_i)\}$, $i \in \mathbb{N}$.*

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 11.6 *Let the assumptions of Lemma 11.3 be satisfied. Consider a sequence $\{\boldsymbol{x}_i\}$, $i \in \mathbb{N}$, generated by Algorithm 11.1. Then, if the values $\underline{\delta} > 0$,*

$\underline{\varepsilon} > 0$, and $\underline{\mu} > 0$ are chosen arbitrarily, there exists an index $i \geq 1$ such that

$$\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \leq \underline{\varepsilon}, \quad \left|1 - \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}_i; \mu_i)\right| \leq \underline{\delta},$$
$$z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i) \geq 0, \quad u_{kl}(\boldsymbol{x}_i; \mu_i) \geq 0,$$
$$u_{kl}(\boldsymbol{x}_i; \mu_i)(z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i)) \leq \underline{\mu},$$

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

Proof Inequality $|1 - \tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i; \mu_i)| \leq \underline{\delta}$ follows immediately from the fact that the equation $\tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i; \mu_i) = 1$, $1 \leq k \leq m$, is solved with precision $\underline{\delta}$. Inequalities $z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i) \geq 0$, $u_{kl}(\boldsymbol{x}_i; \mu_i) \geq 0$ follow from formulas (11.45) and statement (11.50) as in the proof of Theorem 11.5. Since $\mu_i \downarrow 0$ and $\boldsymbol{g}(\boldsymbol{x}_i; \mu_i) \downarrow \boldsymbol{0}$ by Lemma 11.3 and Theorem 11.5, there exists an index $i \geq 1$ such that $\mu_i \leq \underline{\mu}$ and $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \leq \underline{\varepsilon}$. Using (11.50) we obtain

$$u_{kl}(\boldsymbol{x}_i; \mu_i)(z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i)) = \frac{\mu_i (z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i))}{z_k(\boldsymbol{x}_i; \mu_i) - f_{kl}(\boldsymbol{x}_i)} = \mu_i \leq \underline{\mu}.$$

□

Theorem 11.5 is a standard global convergence result. If the stopping parameters $\underline{\delta}, \underline{\varepsilon}, \underline{\mu}$ are zero, the sequence of generated points converges to the point satisfying the KKT conditions for the equivalent nonlinear programming problem. Theorem 11.6 determines a precision of the obtained solution if the stopping parameters are nonzero.

11.2.6 Special Cases

Both the simplest and most widely considered generalized minimax problem is the classical minimax problem (11.10), when $m = 1$ in (11.4) (in this case we write m, z, $\boldsymbol{u}$, $\boldsymbol{v}$, U, V, A instead of m_1, z_1, $\boldsymbol{u}_1$, $\boldsymbol{v}_1$, U_1, V_1, A_1). For solving a classical minimax problem one can use Algorithm 11.1, where a major part of computation is very simplified. System of equations (11.38) is of order $n + 1$ and has the form

$$\begin{bmatrix} G(\boldsymbol{x}, z) + A(\boldsymbol{x})V(\boldsymbol{x}, z)A^T(\boldsymbol{x}) & -A(\boldsymbol{x})V(\boldsymbol{x}, z)\tilde{\boldsymbol{e}} \\ -\tilde{\boldsymbol{e}}^T V(\boldsymbol{x}, z)A^T(\boldsymbol{x}) & \tilde{\boldsymbol{e}}^T V(\boldsymbol{x}, z)\tilde{\boldsymbol{e}} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta z \end{bmatrix} = - \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x}, z) \\ \gamma(\boldsymbol{x}, z) \end{bmatrix}, \tag{11.67}$$

where $\boldsymbol{g}(\boldsymbol{x}, z) = A(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}, z)$, $\gamma(\boldsymbol{x}, z) = 1 - \tilde{\boldsymbol{e}}^T \boldsymbol{u}(\boldsymbol{x}, z)$, $V(\boldsymbol{x}, z) = U^2(\boldsymbol{x}, z)/\mu = \mathrm{diag}[u_1^2(\boldsymbol{x}, z), \ldots, u_m^2(\boldsymbol{x}, z)]/\mu$, and $u_k(\boldsymbol{x}, z) = \mu/(z - f_k(\boldsymbol{x}))$, $1 \leq k \leq m$. System

of equations (11.44) is reduced to one nonlinear equation

$$1 - \tilde{\boldsymbol{e}}^T \boldsymbol{u}(\boldsymbol{x}, z) = 1 - \sum_{k=1}^{m} \frac{\mu}{z - f_k(\boldsymbol{x})} = 0, \tag{11.68}$$

whose solution $z(\boldsymbol{x}; \mu)$ lies in the interval $F(\boldsymbol{x}) + \mu \le z(\boldsymbol{x}; \mu) \le F(\boldsymbol{x}) + m\mu$. To find this solution by robust methods from [14, 15] is not difficult. A barrier function has the form

$$\hat{B}(\boldsymbol{x}; \mu) = z(\boldsymbol{x}; \mu) - \mu \sum_{k=1}^{m} \log(z(\boldsymbol{x}; \mu) - f_k(\boldsymbol{x})) \tag{11.69}$$

with $\nabla \hat{B}(\boldsymbol{x}; \mu) = A(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}; \mu)$ and

$$\nabla^2 \hat{B}(\boldsymbol{x}; \mu) = G(\boldsymbol{x}; \mu) + A(\boldsymbol{x})V(\boldsymbol{x}; \mu)A^T(\boldsymbol{x}) - \frac{A(\boldsymbol{x})V(\boldsymbol{x}; \mu)\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}^T V(\boldsymbol{x}; \mu)A^T(\boldsymbol{x})}{\tilde{\boldsymbol{e}}^T V(\boldsymbol{x}; \mu)\tilde{\boldsymbol{e}}}.$$

If we write system (11.67) in the form

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

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

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

Since

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

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

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

where

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

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

$$W + E = LL^T, \tag{11.70}$$

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

$$LL^T \boldsymbol{p} = \boldsymbol{g}, \quad LL^T \boldsymbol{q} = \boldsymbol{c} \tag{11.71}$$

and set

$$\Delta z = \frac{\boldsymbol{c}^T \boldsymbol{p} + \gamma}{\boldsymbol{c}^T \boldsymbol{q} - \delta}, \quad \Delta \boldsymbol{x} = \boldsymbol{q}\, \Delta z - \boldsymbol{p}. \tag{11.72}$$

If we solve the classical minimax problem, Algorithm 11.1 must be somewhat modified. In Step 2, we solve only Eq. (11.68) instead of the system of equations (11.44). In Step 4, we determine a vector $\Delta \boldsymbol{x}$ by solving Eq. (11.71) and using relations (11.72). In Step 4, we use the barrier function (11.69) (the nonlinear equation (11.68) must be solved at the point $\boldsymbol{x} + \alpha \Delta \boldsymbol{x}$).

Minimization of a sum of absolute values, i.e., minimization of the function (11.14) is another important generalized minimax problem. In this case, a barrier function has the form

$$\begin{aligned} B_\mu(\boldsymbol{x}, z) &= \sum_{k=1}^m z_k - \mu \sum_{k=1}^m \log(z_k - f_k(\boldsymbol{x})) - \mu \sum_{k=1}^m \log(z_k + f_k(\boldsymbol{x})) \\ &= \sum_{k=1}^m z_k - \mu \sum_{k=1}^m \log(z_k^2 - f_k^2(\boldsymbol{x})), \end{aligned} \tag{11.73}$$

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

$$\begin{aligned} &\sum_{k=1}^m \frac{2\mu f_k(\boldsymbol{x})}{z_k^2 - f_k^2(\boldsymbol{x})}\, \boldsymbol{g}_k(\boldsymbol{x}) = \sum_{k=1}^m u_k(\boldsymbol{x}, z_k)\, \boldsymbol{g}_k(\boldsymbol{x}) = \boldsymbol{0}, \\ &u_k(\boldsymbol{x}, z_k) = \frac{2\mu f_k(\boldsymbol{x})}{z_k^2 - f_k^2(\boldsymbol{x})} \end{aligned} \tag{11.74}$$

and

$$1 - \frac{2\mu z_k}{z_k^2 - f_k^2(\boldsymbol{x})} = 1 - u_k(\boldsymbol{x}, z_k) \frac{z_k}{f_k(\boldsymbol{x})} = 0 \quad \Rightarrow \quad u_k(\boldsymbol{x}, z_k) = \frac{f_k(\boldsymbol{x})}{z_k}, \tag{11.75}$$

where $\boldsymbol{g}_k(\boldsymbol{x}) = \nabla f_k(\boldsymbol{x})$, $1 \le k \le m$, which corresponds to (11.31)–(11.32). Equations in (11.44) are quadratic of the form

$$\frac{2\mu z_k(\boldsymbol{x};\mu)}{z_k^2(\boldsymbol{x};\mu) - f_k^2(\boldsymbol{x})} = 1 \quad \Leftrightarrow \quad z_k^2(\boldsymbol{x};\mu) - f_k^2(\boldsymbol{x}) = 2\mu z_k(\boldsymbol{x};\mu), \tag{11.76}$$

where $1 \le k \le m$, and their solutions is given by

$$z_k(\boldsymbol{x};\mu) = \mu + \sqrt{\mu^2 + f_k^2(\boldsymbol{x})}, \quad 1 \le k \le m, \tag{11.77}$$

(the second solutions of quadratic equations (11.76) do not satisfy the condition $z_k > |f_k(\boldsymbol{x})|$, so the obtained vector $\boldsymbol{z}$ does not belong to a domain of $\tilde{B}_\mu(z)$). Using (11.75) and (11.77) we obtain

$$u_k(\boldsymbol{x};\mu) = u_k(\boldsymbol{x}, z_k(\boldsymbol{x};\mu)) = \frac{f_k(\boldsymbol{x})}{z_k(\boldsymbol{x};\mu)} = \frac{f_k(\boldsymbol{x})}{\mu + \sqrt{\mu^2 + f_k^2(\boldsymbol{x})}} \tag{11.78}$$

for $1 \le k \le m$ and

$$\begin{aligned}\hat{B}(\boldsymbol{x};\mu) &= B(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{x};\mu)) = \sum_{k=1}^{m} z_k(\boldsymbol{x};\mu) - \mu \sum_{k=1}^{m} \log(z_k^2(\boldsymbol{x};\mu) - f_k^2(\boldsymbol{x})) \\ &= \sum_{k=1}^{m} z_k(\boldsymbol{x};\mu) - \mu \sum_{k=1}^{m} \log(2\mu z_k(\boldsymbol{x};\mu)) \\ &= \sum_{k=1}^{m} \big[z_k(\boldsymbol{x};\mu) - \mu \log(z_k(\boldsymbol{x};\mu))\big] - \mu m \log(2\mu). \end{aligned} \tag{11.79}$$

Using these expressions, we can write formulas (11.47) and (11.48) in the form

$$\nabla \hat{B}(\boldsymbol{x};\mu) = \sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x}) u_k(\boldsymbol{x};\mu) \tag{11.80}$$

and

$$\nabla^2 \hat{B}(\boldsymbol{x};\mu) = W(\boldsymbol{x};\mu) = \sum_{k=1}^{m} G_k(\boldsymbol{x}) u_k(\boldsymbol{x};\mu) + \sum_{k=1}^{m} \boldsymbol{g}_k(\boldsymbol{x}) v_k(\boldsymbol{x};\mu) \boldsymbol{g}_k^T(\boldsymbol{x}), \tag{11.81}$$

where

$$G_k(\boldsymbol{x}) = \nabla^2 f_k(\boldsymbol{x}), \quad v_k(\boldsymbol{x};\mu) = \frac{2\mu}{z_k^2(\boldsymbol{x};\mu) + f_k^2(\boldsymbol{x})}, \quad 1 \le k \le m. \tag{11.82}$$

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

$$\nabla^2 \hat{B}(\boldsymbol{x};\mu)\Delta\boldsymbol{x} = -\boldsymbol{g}(\boldsymbol{x};\mu), \tag{11.83}$$

where $\boldsymbol{g}(\boldsymbol{x};\mu) = \nabla\hat{B}(\boldsymbol{x};\mu) \neq 0$. From (11.83) and (11.81) it follows

$$(\Delta\boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x};\mu) = -(\Delta\boldsymbol{x})^T \nabla^2 \hat{B}(\boldsymbol{x};\mu)\Delta\boldsymbol{x} \le -(\Delta\boldsymbol{x})^T G(\boldsymbol{x};\mu)\Delta\boldsymbol{x},$$

so if a matrix $G(\boldsymbol{x};\mu)$ is positive definite, a matrix $\nabla\hat{B}(\boldsymbol{x};\mu)$ is positive definite as well (since a diagonal matrix $V(\boldsymbol{x};\mu)$ is positive definite by (11.82)) and $(\Delta\boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x};\mu) < 0$ holds (a direction vector $\Delta\boldsymbol{x}$ is descent for a function $\hat{B}(\boldsymbol{x};\mu)$).

If we minimize a sum of absolute values, Algorithm 11.1 needs to be somewhat modified. In Step 2, we solve quadratic equations (11.76) whose solutions are given by (11.77). In Step 4, we determine a vector $\Delta\boldsymbol{x}$ by solving Eq. (11.83), where matrix $\nabla^2\hat{B}(\boldsymbol{x};\mu)$ is given by (11.83). In Step 4, we use the barrier function (11.79).

11.3 Smoothing Methods

11.3.1 Basic Properties

Similarly as in Sect. 11.2.1 we will restrict ourselves to sums of maxima, where a mapping $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ is a sum of its arguments, so (11.4) holds. Smoothing methods for minimization of sums of maxima replace function (11.4) by a smoothing function

$$S(\boldsymbol{x};\mu) = \sum_{k=1}^{m} S_k(\boldsymbol{x};\mu), \tag{11.84}$$

where

$$\begin{aligned} S_k(\boldsymbol{x};\mu) &= \mu \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(\boldsymbol{x})}{\mu}\right) \\ &= F_k(\boldsymbol{x}) + \mu \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(\boldsymbol{x}) - F_k(\boldsymbol{x})}{\mu}\right), \end{aligned} \tag{11.85}$$

depending on a smoothing parameter $0 < \mu \leq \overline{\mu}$, which is successively minimized on $\mathbb{R}^n$ with $\mu \downarrow 0$. Since $f_{kl}(\boldsymbol{x}) \leq F_k(\boldsymbol{x})$, $1 \leq l \leq m_k$, and the equality arises for at least one index, at least one exponential function on the right-hand side of (11.85) has the value 1, so the logarithm is positive. Thus $F_k(\boldsymbol{x}) \leq S_k(\boldsymbol{x}; \mu) \leq F_k(\boldsymbol{x}) + \mu \log m_k$, $1 \leq k \leq m$, hold. Therefore

$$F(\boldsymbol{x}) \leq S(\boldsymbol{x}; \mu) \leq F(\boldsymbol{x}) + \mu \sum_{k=1}^{m} \log m_k, \tag{11.86}$$

so $S(\boldsymbol{x}; \mu) \to F(\boldsymbol{x})$ if $\mu \downarrow 0$.

Remark 11.15 Similarly as in Sect. 11.2.2 we will denote $\boldsymbol{g}_{kl}(\boldsymbol{x})$ and $G_{kl}(x)$ the gradients and Hessian matrices of functions $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, and

$$\boldsymbol{u}_k(\boldsymbol{x}; \mu) = \begin{bmatrix} u_{k1}(\boldsymbol{x}; \mu) \\ \vdots \\ u_{km_k}(\boldsymbol{x}; \mu) \end{bmatrix}, \quad \tilde{\boldsymbol{e}}_k = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

where

$$u_{kl}(\boldsymbol{x}; \mu) = \frac{\exp(f_{kl}(\boldsymbol{x})/\mu)}{\sum_{l=1}^{m_k} \exp(f_{kl}(\boldsymbol{x})/\mu)} = \frac{\exp((f_{kl}(\boldsymbol{x}) - F_k(\boldsymbol{x}))/\mu)}{\sum_{l=1}^{m_k} \exp((f_{kl}(\boldsymbol{x}) - F_k(\boldsymbol{x}))/\mu)}. \tag{11.87}$$

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

$$\tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}; \mu) = \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}; \mu) = 1. \tag{11.88}$$

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

Theorem 11.7 *Consider the smoothing function (11.84). Then*

$$\nabla S(\boldsymbol{x}; \mu) = \boldsymbol{g}(\boldsymbol{x}; \mu) \tag{11.89}$$

and

$$\begin{aligned} \nabla^2 S(\boldsymbol{x}; \mu) &= G(\boldsymbol{x}; \mu) + \frac{1}{\mu} \sum_{k=1}^{m} A_k(\boldsymbol{x}) U_k(\boldsymbol{x}; \mu) A_k^T(\boldsymbol{x}) \\ &\quad - \frac{1}{\mu} \sum_{k=1}^{m} A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}; \mu) (A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}; \mu))^T \\ &= G(\boldsymbol{x}; \mu) + \frac{1}{\mu} A(\boldsymbol{x}) U(\boldsymbol{x}; \mu) A^T(\boldsymbol{x}) - \frac{1}{\mu} C(\boldsymbol{x}; \mu) C(\boldsymbol{x}; \mu)^T \end{aligned} \tag{11.90}$$

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

$$G(\boldsymbol{x};\mu) = \sum_{k=1}^{m} G_k(\boldsymbol{x})\boldsymbol{u}_k(\boldsymbol{x};\mu), \quad A(\boldsymbol{x}) = [A_1(\boldsymbol{x}), \ldots, A_m(\boldsymbol{x})],$$
$$U(\boldsymbol{x};\mu) = \mathrm{diag}[U_1(\boldsymbol{x};\mu), \ldots, U_m(\boldsymbol{x};\mu)],$$
$$C(\boldsymbol{x};\mu) = [A_1(\boldsymbol{x})\boldsymbol{u}_1(\boldsymbol{x};\mu), \ldots, A_m(\boldsymbol{x})\boldsymbol{u}_m(\boldsymbol{x};\mu)].$$

Proof Obviously,

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

Differentiating functions (11.85) and using (11.87) we obtain

$$\begin{aligned}\nabla S_k(\boldsymbol{x};\mu) &= \frac{\mu}{\sum_{l=1}^{m_k} \exp(f_{kl}(\boldsymbol{x})/\mu)} \sum_{l=1}^{m_k} \frac{1}{\mu} \exp(f_{kl}(\boldsymbol{x})/\mu)\boldsymbol{g}_{kl}(\boldsymbol{x}) \\ &= \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x};\mu) = A_k(\boldsymbol{x})\boldsymbol{u}_k(\boldsymbol{x};\mu). \end{aligned} \tag{11.91}$$

Adding up these expressions yields (11.89). Further, it holds

$$\begin{aligned}\nabla u_{kl}(\boldsymbol{x};\mu) &= \frac{1}{\mu} \frac{\exp(f_{kl}(\boldsymbol{x})/\mu)}{\sum_{l=1}^{m_k} \exp(f_{kl}(\boldsymbol{x})/\mu)} \boldsymbol{g}_{kl}(\boldsymbol{x}) \\ &\quad - \frac{\exp(f_{kl}(\boldsymbol{x})/\mu)}{\left(\sum_{l=1}^{m_k} \exp(f_{kl}(\boldsymbol{x})/\mu)\right)^2} \sum_{l=1}^{m_k} \frac{1}{\mu} \exp(f_{kl}(\boldsymbol{x})/\mu)\boldsymbol{g}_{kl}(\boldsymbol{x}) \\ &= \frac{1}{\mu} u_{kl}(\boldsymbol{x};\mu)\boldsymbol{g}_{kl}(\boldsymbol{x}) - \frac{1}{\mu} u_{kl}(\boldsymbol{x};\mu) \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x};\mu)\boldsymbol{g}_{kl}(\boldsymbol{x}). \end{aligned} \tag{11.92}$$

Differentiating (11.91) and using (11.92) we obtain

$$\begin{aligned}\nabla^2 S_k(\boldsymbol{x};\mu) &= \sum_{l=1}^{m_k} G_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x};\mu) + \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) \nabla u_{kl}(\boldsymbol{x};\mu) \\ &= G_k(\boldsymbol{x};\mu) + \frac{1}{\mu} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x};\mu) \boldsymbol{g}_{kl}^T(\boldsymbol{x}) \end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{1}{\mu} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x}; \mu) \left(\sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x}; \mu) \right)^T \\
&= G_k(\boldsymbol{x}; \mu) + \frac{1}{\mu} A_k(\boldsymbol{x}) U_k(\boldsymbol{x}; \mu) A_k^T(\boldsymbol{x}) \\
&\quad - \frac{1}{\mu} A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}; \mu) (A_k(\boldsymbol{x}) \boldsymbol{u}_k(\boldsymbol{x}; \mu))^T,
\end{aligned}
$$

where $G_k(\boldsymbol{x}; \mu) = \sum_{l=1}^{m_k} G_{kl}(\boldsymbol{x}) u_{kl}(\boldsymbol{x}; \mu)$. Adding up these expressions yields (11.90). □

Remark 11.16 Note that using (11.90) and the Schwarz inequality we obtain

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

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

Using Theorem 11.7, a step of the Newton method can be written in the form $\boldsymbol{x}_+ = \boldsymbol{x} + \alpha \Delta \boldsymbol{x}$ where

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

or

$$
\left(W(\boldsymbol{x}; \mu) - \frac{1}{\mu} C(\boldsymbol{x}; \mu) C^T(\boldsymbol{x}; \mu) \right) \Delta \boldsymbol{x} = -\boldsymbol{g}(\boldsymbol{x}; \mu), \tag{11.93}
$$

where

$$
W(\boldsymbol{x}; \mu) = G(\boldsymbol{x}; \mu) + \frac{1}{\mu} A(\boldsymbol{x}) U(\boldsymbol{x}; \mu) A^T(\boldsymbol{x}), \quad \boldsymbol{g}(\boldsymbol{x}; \mu) = A(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x}; \mu). \tag{11.94}
$$

A matrix W in (11.94) has the same structure as a matrix W in (11.48) and, by Theorem 11.7, smoothing function (11.84) has similar properties as the barrier function (11.46). Thus, one can use an algorithm that is analogous to Algorithm 11.1 and considerations stated in Remark 11.12, where $S(\boldsymbol{x}; \mu)$ and $\nabla^2 S(\boldsymbol{x}; \mu)$ are used instead of $\hat{B}(\boldsymbol{x}; \mu)$ and $\nabla^2 \hat{B}(\boldsymbol{x}; \mu)$. It means that

$$
S(\boldsymbol{x}_{i+1}; \mu_i) - S(\boldsymbol{x}_i; \mu_i) \leq -c \|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2 \quad \text{for all } i \in \mathbb{N}, \tag{11.95}
$$

if Assumption 11.4 is satisfied and

$$S(\boldsymbol{x}_{i+1};\mu_i) - S(\boldsymbol{x}_i;\mu_i) \leq 0 \quad \text{for all } i \in \mathbb{N} \tag{11.96}$$

in remaining cases.

The considerations up to now are summarized in Algorithm 11.2 introduced in the Appendix. This algorithm differs from Algorithm 11.1 in that a nonlinear equation $\tilde{\boldsymbol{e}}^T \boldsymbol{u}(\boldsymbol{x};\mu) = 1$ need not be solved in Step 2 (because (11.88) follows from (11.87)), Eq. (11.93)–(11.94) instead of (11.71)–(11.72) are used in Step 4, and a barrier function $\hat{B}(\boldsymbol{x};\mu)$ is replaced with a smoothing function $S(\boldsymbol{x};\mu)$ in Step 6. Note that the parameter μ in (11.84) has different meaning than the same parameter in (11.46), 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 11.2 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.

11.3.2 Global Convergence

Now we prove the global convergence of the smoothing method realized by Algorithm 11.2.

Lemma 11.4 *Choose a fixed vector $\boldsymbol{x} \in \mathbb{R}^n$. Then $S_k(\boldsymbol{x};\mu) : (0,\infty) \to \mathbb{R}$, $1 \leq k \leq m$, are nondecreasing convex functions of $\mu > 0$ and*

$$0 \leq \log \underline{m}_k \leq \frac{\partial}{\partial \mu} S_k(\boldsymbol{x};\mu) \leq \log m_k, \tag{11.97}$$

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

$$\frac{\partial}{\partial \mu} S_k(\boldsymbol{x};\mu) = \log \sum_{l=1}^{m_k} \exp\left(\frac{f_{kl}(\boldsymbol{x}) - F_k(\boldsymbol{x})}{\mu}\right) - \sum_{l=1}^{m_k} \left(\frac{f_{kl}(\boldsymbol{x}) - F_k(\boldsymbol{x})}{\mu}\right) u_{kl}(\boldsymbol{x};\mu). \tag{11.98}$$

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

$$\varphi'_{kl}(\boldsymbol{x};\mu) \stackrel{\Delta}{=} \frac{\partial}{\partial \mu} \varphi_{kl}(\boldsymbol{x};\mu) = -\frac{\varphi_{kl}(\boldsymbol{x};\mu)}{\mu} \geq 0,$$

we can write by (11.85) that

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

and

$$\begin{aligned}\frac{\partial}{\partial \mu} S_k(\boldsymbol{x};\mu) &= \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu) + \mu \frac{\sum_{l=1}^{m_k} \varphi'_{kl}(\boldsymbol{x};\mu) \exp \varphi_{kl}(\boldsymbol{x};\mu)}{\sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu)} \\ &= \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu) - \sum_{l=1}^{m_k} \varphi_{kl}(\boldsymbol{x};\mu) u_{kl}(\boldsymbol{x};\mu) \geq 0, \quad (11.99)\end{aligned}$$

because $\varphi_{kl}(\boldsymbol{x};\mu) \leq 0$, $u_{kl}(\boldsymbol{x};\mu) \geq 0$, $1 \leq k \leq m$, and $\varphi_{kl}(\boldsymbol{x};\mu) = 0$ holds for at least one index. Thus, functions $S_k(\boldsymbol{x};\mu)$, $1 \leq k \leq m$, are nondecreasing. Differentiating (11.87) with respect to μ we obtain

$$\begin{aligned}\frac{\partial}{\partial \mu} u_{kl}(\boldsymbol{x};\mu) &= -\frac{1}{\mu} \frac{\varphi_{kl}(\boldsymbol{x};\mu) \exp \varphi_{kl}(\boldsymbol{x};\mu)}{\sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu)} \\ &\quad + \frac{1}{\mu} \frac{\exp \varphi_{kl}(\boldsymbol{x};\mu)}{\sum_{l=1}^{k} m \exp \varphi_{kl}(\boldsymbol{x};\mu)} \frac{\sum_{l=1}^{m_k} \varphi_{kl}(\boldsymbol{x};\mu) \exp \varphi_{kl}(\boldsymbol{x};\mu)}{\sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu)} \\ &= \frac{1}{\mu} u_{kl}(\boldsymbol{x};\mu) \left(-\varphi_{kl}(\boldsymbol{x};\mu) + \sum_{l=1}^{m_k} \varphi_{kl}(\boldsymbol{x};\mu) u_{kl}(\boldsymbol{x};\mu) \right). \quad (11.100)\end{aligned}$$

Differentiating (11.99) with respect to μ and using Eqs. (11.88) and (11.100) we can write

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

because

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

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

$$\lim_{\mu\downarrow 0} \frac{\partial}{\partial\mu} S_k(\boldsymbol{x};\mu) = \lim_{\mu\downarrow 0} \log \sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x};\mu) - \lim_{\mu\downarrow 0} \sum_{l=1}^{m_k} \varphi_{kl}(\boldsymbol{x};\mu)u_{kl}(\boldsymbol{x};\mu)$$
$$= \log \underline{m}_k - \frac{1}{\underline{m}_k} \lim_{\mu\downarrow 0} \sum_{l=1}^{m_k} \varphi_{kl}(\boldsymbol{x};\mu) \exp \varphi_{kl}(\boldsymbol{x};\mu) = \log \underline{m}_k,$$

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

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

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

Lemma 11.5 *Let Assumptions 11.2 and 11.4 be satisfied. Then the values μ_i, $i \in \mathbb{N}$, generated by Algorithm 11.2, create a nonincreasing sequence such that $\mu_i \downarrow 0$.*

Proof Lemma 11.5 is a direct consequence of Lemma 11.3 because the same procedures for an update of a parameter μ are used and (11.95) holds. □

Theorem 11.8 *Let the assumptions of Lemma 11.5 be satisfied. Consider a sequence $\{\boldsymbol{x}_i\}$ $i \in \mathbb{N}$, generated by Algorithm 11.2, where $\underline{\varepsilon} = \underline{\mu} = 0$. Then*

$$\lim_{i\uparrow\infty} \sum_{k=1}^{m} \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}_i;\mu_i)\boldsymbol{g}_{kl}(\boldsymbol{x}_i) = \boldsymbol{0}, \quad \sum_{l=1}^{m_k} u_{kl}(\boldsymbol{x}_i;\mu_i) = 1$$

and

$$F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i) \geq 0, \quad u_{kl}(\boldsymbol{x}_i;\mu_i) \geq 0, \quad \lim_{i\uparrow\infty} u_{kl}(\boldsymbol{x}_i;\mu_i)(F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i)) = 0$$

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

Proof

(a) Equations $\tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i; \mu_i) = 1$ for $1 \le k \le m$ follow from (11.88). Inequalities $F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i) \ge 0$ and $u_{kl}(\boldsymbol{x}_i; \mu_i) \ge 0$ for $1 \le k \le m$ and $1 \le l \le m_k$ follow from (11.4) and (11.87).

(b) Since $S_k(\boldsymbol{x}; \mu)$ are nondecreasing functions of the parameter μ by Lemma 11.4 and (11.95) holds, we can write

$$\underline{F} \le \sum_{k=1}^{m} F_k(\boldsymbol{x}_{i+1}) \le S(\boldsymbol{x}_{i+1}; \mu_{i+1}) \le S(\boldsymbol{x}_{i+1}; \mu_i)$$

$$\le S(\boldsymbol{x}_i; \mu_i) - c\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2 \le S(\boldsymbol{x}_1; \mu_1) - c \sum_{j=1}^{i} \|\boldsymbol{g}(\boldsymbol{x}_j; \mu_j)\|^2,$$

where $\underline{F} = \sum_{k=1}^{m} \underline{F}_k$ and $\underline{F}_k$, $1 \le k \le m$, are lower bounds from Assumption 11.2. Thus, it holds

$$\underline{F} \le \lim_{i \uparrow \infty} S(\boldsymbol{x}_{i+1}; \mu_{i+1}) \le S(\boldsymbol{x}_1; \mu_1) - c \sum_{i=1}^{\infty} \|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2,$$

or

$$\sum_{i=1}^{\infty} \|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\|^2 \le \frac{1}{c}(S(\boldsymbol{x}_1; \mu_1) - \underline{F}),$$

so $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \downarrow 0$, which together with inequalities $0 \le u_{kl}(\boldsymbol{x}_i; \mu_i) \le 1$, $1 \le k \le m$, $1 \le l \le m_k$, gives $\lim_{i \uparrow \infty} u_{kl}(\boldsymbol{x}_i; \mu_i) \boldsymbol{g}_{kl}(\boldsymbol{x}_i) = \boldsymbol{0}$.

(c) Let indices $1 \le k \le m$ and $1 \le l \le m_k$ be chosen arbitrarily. Using (11.87) we get

$$0 \le u_{kl}(\boldsymbol{x}_i; \mu_i)(F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i)) = -\mu_i \frac{\varphi_{kl}(\boldsymbol{x}_i; \mu_i) \exp \varphi_{kl}(\boldsymbol{x}_i; \mu_i)}{\sum_{l=1}^{m_k} \exp \varphi_{kl}(\boldsymbol{x}_i; \mu_i)}$$

$$\le -\mu_i \varphi_{kl}(\boldsymbol{x}_i; \mu_i) \exp \varphi_{kl}(\boldsymbol{x}_i; \mu_i) \le \frac{\mu_i}{e},$$

where $\varphi_{kl}(\boldsymbol{x}_i; \mu_i)$, $1 \le k \le m$, $1 \le l \le m_k$, are functions used in the proof of Lemma 11.4, because

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

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

□

Corollary 11.2 *Let the assumptions of Theorem 11.8 be satisfied. Then every cluster point $\boldsymbol{x} \in \mathbb{R}^n$ of a sequence $\{\boldsymbol{x}_i\}$, $i \in \mathbb{N}$, satisfies the necessary KKT conditions (11.5) and (11.6), where $\boldsymbol{u}$ (with elements $\boldsymbol{u}_k$, $1 \leq k \leq m$) is a cluster point of a sequence $\{\boldsymbol{u}(\boldsymbol{x}_i; \mu_i)\}$, $i \in \mathbb{N}$.*

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 11.2.

Theorem 11.9 *Let the assumptions of Theorem 11.5 be satisfied and let $\{\boldsymbol{x}_i\}$, $i \in \mathbb{N}$, be a sequence generated by Algorithm 11.2. Then, if the values $\underline{\varepsilon} > 0$ and $\underline{\mu} > 0$ are chosen arbitrarily, there exists an index $i \geq 1$ such that*

$$\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \leq \underline{\varepsilon}, \quad \tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i; \mu_i) = 1, \quad 1 \leq k \leq m,$$

and

$$F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i) \geq 0, \quad u_{kl}(\boldsymbol{x}_i; \mu_i) \geq 0, \quad u_{kl}(\boldsymbol{x}_i; \mu_i)(F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i)) \leq \frac{\underline{\mu}}{e}$$

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

Proof Equalities $\tilde{\boldsymbol{e}}_k^T \boldsymbol{u}_k(\boldsymbol{x}_i; \mu_i) = 1$, $1 \leq k \leq m$, follow from (11.88). Inequalities $F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i) \geq 0$ and $u_{kl}(\boldsymbol{x}_i; \mu_i) \geq 0$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, follow from (11.10) and (11.87). Since $\mu_i \downarrow 0$ holds by Lemma 11.5 and $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \downarrow 0$ holds by Theorem 11.8, there exists an index $i \geq 1$ such that $\mu_i \leq \underline{\mu}$ and $\|\boldsymbol{g}(\boldsymbol{x}_i; \mu_i)\| \leq \underline{\varepsilon}$. By (11.87), as in the proof of Theorem 11.8, one can write

$$u_{kl}(\boldsymbol{x}_i; \mu_i)(F_k(\boldsymbol{x}_i) - f_{kl}(\boldsymbol{x}_i)) \leq -\mu_i \varphi_{kl}(\boldsymbol{x}_i; \mu_i) \exp \varphi_{kl}(\boldsymbol{x}_i; \mu_i) \leq \frac{\mu_i}{e} \leq \frac{\underline{\mu}}{e}$$

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

Theorems 11.8 and 11.9 have the same meaning as Theorems 11.5 and 11.6 introduced in Sect. 11.2.5.

11.3.3 Special Cases

Both the simplest and most widely considered generalized minimax problem is the classical minimax problem (11.10), when $m = 1$ in (11.4) (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 11.2, where a major part of computation is very simplified. A step of the

Newton method can be written in the form $\boldsymbol{x}_+ = \boldsymbol{x} + \alpha\Delta\boldsymbol{x}$ where

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

or

$$\left(W(\boldsymbol{x};\mu) - \frac{1}{\mu}\boldsymbol{g}(\boldsymbol{x};\mu)\boldsymbol{g}^T(\boldsymbol{x};\mu)\right)\Delta\boldsymbol{x} = -\boldsymbol{g}(\boldsymbol{x};\mu), \tag{11.101}$$

where

$$W(\boldsymbol{x};\mu) = G(\boldsymbol{x};\mu) + \frac{1}{\mu}A(\boldsymbol{x})U(\boldsymbol{x};\mu)A^T(\boldsymbol{x}), \quad \boldsymbol{g}(\boldsymbol{x};\mu) = A(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x};\mu). \tag{11.102}$$

Since

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

holds by the Sherman–Morrison formula, the solution of system of equations (11.101) can be written in the form

$$\Delta\boldsymbol{x} = \frac{\mu}{\boldsymbol{g}^T W^{-1}\boldsymbol{g} - \mu}W^{-1}\boldsymbol{g}. \tag{11.103}$$

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 [10]. Then, we solve an equation

$$LL^T\boldsymbol{p} = \boldsymbol{g}, \tag{11.104}$$

and set

$$\Delta\boldsymbol{x} = \frac{\mu}{\boldsymbol{g}^T\boldsymbol{p} - \mu}\boldsymbol{p}. \tag{11.105}$$

Minimization of a sum of absolute values, i.e., minimization of the function (11.14) is another important generalized minimax problem. In this case, a smoothing function has the form

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

because $f_k^+(\boldsymbol{x}) = |f_k(\boldsymbol{x})|$ if $f_k(\boldsymbol{x}) \geq 0$ and $f_k^-(\boldsymbol{x}) = |f_k(\boldsymbol{x})|$ if $f_k(\boldsymbol{x}) \leq 0$, and by Theorem 11.7 we have

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

$$\begin{aligned}\nabla^2 S(\boldsymbol{x};\mu) &= \sum_{k=1}^m G_k (u_k^+ - u_k^-) + \frac{1}{\mu} \sum_{k=1}^m \boldsymbol{g}_k \boldsymbol{g}_k^T (u_k^+ + u_k^-) \\ &\quad - \frac{1}{\mu} \sum_{k=1}^m \boldsymbol{g}_k \boldsymbol{g}_k^T (u_k^+ - u_k^-)^2 = G(\boldsymbol{x};\mu) + \frac{1}{\mu} \sum_{k=1}^m \boldsymbol{g}_k \boldsymbol{g}_k^T (1 - u_k^2),\end{aligned}$$

(because $u_k^+ + u_k^- = 1$), where $\boldsymbol{g}_k = \boldsymbol{g}_k(\boldsymbol{x})$,

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

and

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

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

11.4 Primal-Dual Interior Point Methods

11.4.1 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 $z_k > F_k(\boldsymbol{x})$, $1 \leq k \leq m$). 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 step length 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 (11.4). As is mentioned in the introduction, minimization of this function is equivalent to the nonlinear programming problem

$$\begin{cases} \text{minimize} & \sum\limits_{k=1}^{m} z_k \\ \text{subject to} & f_{kl}(\boldsymbol{x}) \le z_k, \quad 1 \le k \le m, \quad 1 \le l \le m_k. \end{cases} \tag{11.106}$$

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

$$B_\mu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{s}) = \sum_{k=1}^{m} z_k - \mu \sum_{k=1}^{m} \sum_{l=1}^{m_k} \log(s_{kl}), \tag{11.107}$$

a solving of the problem (11.106) can be transformed to a successive solving of problems

$$\begin{cases} \text{minimize} & B_\mu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{s}) \\ \text{subject to} & f_{kl}(\boldsymbol{x}) + s_{kl} - z_k = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k, \end{cases} \tag{11.108}$$

where $\mu \downarrow 0$. Necessary conditions for an extremum of the problem (11.108) have the form

$$\begin{aligned} & \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} \boldsymbol{g}_{kl}(\boldsymbol{x}) u_{kl} = \boldsymbol{0}, \\ & 1 - \sum_{l=1}^{m_k} u_{kl} = 0, \quad 1 \le k \le m, \\ & u_{kl} s_{kl} - \mu = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k, \\ & f_{kl}(\boldsymbol{x}) + s_{kl} - z_k = 0, \quad 1 \le k \le m, \quad 1 \le l \le m_k, \end{aligned}$$

which is $n + m + 2\bar{m}$ equations for $n + m + 2\bar{m}$ unknowns (vectors $\boldsymbol{x}$, $\boldsymbol{z} = [z_k]$, $\boldsymbol{s} = [s_{kl}]$, $\boldsymbol{u} = [u_{kl}]$, $1 \le k \le m$, $1 \le l \le m_k$), where $\bar{m} = m_1 + \cdots + m_m$. Denote

$A(\boldsymbol{x}) = [A_1(\boldsymbol{x}), \ldots, A_m(\boldsymbol{x})]$, $\boldsymbol{f} = [f_{kl}]$, $S = \mathrm{diag}[s_{kl}]$, $U = \mathrm{diag}[u_{kl}]$, $1 \le k \le m$, $1 \le l \le m_k$, and

$$E = \begin{bmatrix} \tilde{\boldsymbol{e}}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{e}}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \tilde{\boldsymbol{e}}_m \end{bmatrix}, \quad \tilde{\boldsymbol{e}} = \begin{bmatrix} \tilde{\boldsymbol{e}}_1 \\ \tilde{\boldsymbol{e}}_2 \\ \vdots \\ \tilde{\boldsymbol{e}}_m \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

(matrices $A_k(\boldsymbol{x})$, vectors $\tilde{\boldsymbol{e}}_k$, and numbers z_k, $1 \le k \le m$, are defined in Sect. 11.2.2). Applying the Newton method to this system of nonlinear equations, we obtain a system of linear equations for increments (direction vectors) $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{z}$, $\Delta\boldsymbol{s}$, $\Delta\boldsymbol{u}$. After arrangement and elimination

$$\Delta\boldsymbol{s} = -U^{-1}S(\boldsymbol{u} + \Delta\boldsymbol{u}) + \mu S^{-1}\tilde{\boldsymbol{e}}, \tag{11.109}$$

this system has the form

$$\begin{bmatrix} G(\boldsymbol{x},\boldsymbol{u}) & \boldsymbol{0} & A(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{0} & -E^T \\ A^T(\boldsymbol{x}) & -E & -U^{-1}S \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{z} \\ \Delta\boldsymbol{u} \end{bmatrix} = - \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}) \\ \tilde{\boldsymbol{e}} - E^T\boldsymbol{u} \\ \boldsymbol{f}(\boldsymbol{x}) - E\boldsymbol{z} + \mu U^{-1}\tilde{\boldsymbol{e}} \end{bmatrix}, \tag{11.110}$$

where $G(\boldsymbol{x},\boldsymbol{u}) = \sum_{k=1}^{m}\sum_{l=1}^{m_k} G_{kl}(\boldsymbol{x})u_{kl}$. Vector $\tilde{\boldsymbol{e}}$ in the equation $\tilde{\boldsymbol{e}} - E^T\boldsymbol{u} = \boldsymbol{0}$ has unit elements, but its dimension is different from the dimension of a vector $\tilde{\boldsymbol{e}}$ in (11.109).

For solving this linear system, we cannot advantageously use the structure of a generalized minimax problem (because substituting $z_k = F_k(\boldsymbol{x}) = \max_{1\le l\le m_k} f_{kl}(\boldsymbol{x})$ 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 $\tilde{\boldsymbol{x}} = [\boldsymbol{x}^T, z^T]^T$,

$$\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}},\boldsymbol{u}) = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}) \\ \tilde{\boldsymbol{e}} - E^T\boldsymbol{u} \end{bmatrix}, \quad \tilde{G}(\tilde{\boldsymbol{x}},\boldsymbol{u}) = \begin{bmatrix} G(\boldsymbol{x},\boldsymbol{u}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \tilde{A}(\tilde{\boldsymbol{x}}) = \begin{bmatrix} A(\boldsymbol{x}) \\ -E^T \end{bmatrix}, \tag{11.111}$$

and write (11.110) in the form

$$\begin{bmatrix} \tilde{G}(\tilde{\boldsymbol{x}},\boldsymbol{u}) & \tilde{A}(\tilde{\boldsymbol{x}}) \\ \tilde{A}^T(\tilde{\boldsymbol{x}}) & -U^{-1}S \end{bmatrix} \begin{bmatrix} \Delta\tilde{\boldsymbol{x}} \\ \Delta\boldsymbol{u} \end{bmatrix} = - \begin{bmatrix} \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}},\boldsymbol{u}) \\ \boldsymbol{c}(\tilde{\boldsymbol{x}}) + \mu U^{-1}\tilde{\boldsymbol{e}} \end{bmatrix}, \tag{11.112}$$

where $\boldsymbol{c}(\tilde{\boldsymbol{x}}) = \boldsymbol{f}(\boldsymbol{x}) - E\boldsymbol{z}$. This system of equations is more advantageous against systems (11.49) and (11.93) 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 (11.112) has a dimension $n + m + \bar{m}$, while systems (11.49) and (11.93)

have dimensions n. It would be possible to eliminate the vector $\Delta \boldsymbol{u}$, so the resulting system

$$(\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) + \tilde{A}(\tilde{\boldsymbol{x}}) M^{-1} \tilde{A}^T(\tilde{\boldsymbol{x}})) \Delta \tilde{\boldsymbol{x}} = -\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) - \tilde{A}(\tilde{\boldsymbol{x}})(M^{-1} \boldsymbol{c}(\tilde{\boldsymbol{x}}) + \mu S^{-1} \tilde{\boldsymbol{e}}), \tag{11.113}$$

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 $u_{kl} \downarrow 0$ or $s_{kl} \downarrow 0$ if $\mu \downarrow 0$, so some elements of a matrix M^{-1} may tend to infinity, which increases the condition number of system (11.113). Conversely, the solution of Eq. (11.112) 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, 18]). Therefore, it is advantageous to split the constraints to active with $s_{kl} \le \tilde{\varepsilon} u_{kl}$ (we denote active quantities by $\hat{\boldsymbol{c}}(\tilde{\boldsymbol{x}})$, $\hat{A}(\tilde{\boldsymbol{x}})$, $\hat{\boldsymbol{s}}$, $\Delta \hat{\boldsymbol{s}}$, $\hat{S}$, $\hat{\boldsymbol{u}}$, $\Delta \hat{\boldsymbol{u}}$, $\hat{U}$, $\hat{M} = \hat{U}^{-1} \hat{S}$) and inactive with $s_{kl} > \tilde{\varepsilon} u_{kl}$ (we denote inactive quantities by $\check{\boldsymbol{c}}(\tilde{\boldsymbol{x}})$, $\check{A}(\tilde{\boldsymbol{x}})$, $\check{\boldsymbol{s}}$, $\Delta \check{\boldsymbol{s}}$, $\check{S}$, $\check{\boldsymbol{u}}$, $\Delta \check{\boldsymbol{u}}$, $\check{U}$, $\check{M} = \check{U}^{-1} \check{S}$). Eliminating inactive equations from (11.112) we obtain

$$\Delta \check{\boldsymbol{u}} = \check{M}^{-1}(\check{\boldsymbol{c}}(\tilde{\boldsymbol{x}}) + \check{A}(\tilde{\boldsymbol{x}})^T \Delta \tilde{\boldsymbol{x}}) + \mu \check{S}^{-1} \tilde{\boldsymbol{e}} \tag{11.114}$$

and

$$\begin{bmatrix} \hat{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) & \hat{A}(\tilde{\boldsymbol{x}}) \\ \hat{A}^T(\tilde{\boldsymbol{x}}) & -\hat{M} \end{bmatrix} \begin{bmatrix} \Delta \tilde{\boldsymbol{x}} \\ \Delta \hat{\boldsymbol{u}} \end{bmatrix} = - \begin{bmatrix} \hat{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) \\ \hat{\boldsymbol{c}}(\tilde{\boldsymbol{x}}) + \mu \hat{U}^{-1} \tilde{\boldsymbol{e}} \end{bmatrix}, \tag{11.115}$$

where

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

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 (11.114) into (11.109) we can write

$$\Delta \hat{\boldsymbol{s}} = -\hat{M}(\hat{\boldsymbol{u}} + \Delta \hat{\boldsymbol{u}}) + \mu \hat{U}^{-1} \tilde{\boldsymbol{e}}, \quad \Delta \check{\boldsymbol{s}} = -(\check{\boldsymbol{c}} + \check{A}^T \Delta \tilde{\boldsymbol{x}} + \check{\boldsymbol{s}}). \tag{11.116}$$

The matrix of the linear system (11.115) is symmetric, but indefinite, so its Choleski decomposition cannot be determined. In this case, we use either dense [3] or sparse [6] Bunch–Parlett decomposition for solving this system. System (11.115) (especially if it is large and sparse) can be efficiently solved by iterative conjugate gradient method with indefinite preconditioner [20]. If the vectors $\Delta \tilde{\boldsymbol{x}}$ and $\Delta \hat{\boldsymbol{u}}$ are solutions of system (11.115), we determine vector $\Delta \check{\boldsymbol{u}}$ by (11.114) and vectors $\Delta \hat{\boldsymbol{s}}$, $\Delta \check{\boldsymbol{s}}$ by (11.116).

Having vectors $\Delta \tilde{\boldsymbol{x}}$, $\Delta \boldsymbol{s}$, $\Delta \boldsymbol{u}$, we need to determine a step length $\alpha > 0$ and set

$$\tilde{\boldsymbol{x}}_+ = \tilde{\boldsymbol{x}} + \alpha \Delta \tilde{\boldsymbol{x}}, \quad \boldsymbol{s}_+ = \boldsymbol{s}(\alpha), \quad \boldsymbol{u}_+ = \boldsymbol{u}(\alpha), \tag{11.117}$$

where $\boldsymbol{s}(\alpha)$ and $\boldsymbol{u}(\alpha)$ are vector functions such that $\boldsymbol{s}(\alpha) > 0$, $\boldsymbol{s}'(0) = \Delta\boldsymbol{s}$ and $\boldsymbol{u}(\alpha) > 0$, $\boldsymbol{u}'(0) = \Delta\boldsymbol{u}$. This step is not trivial, because we need to decrease both the value of the barrier function $\tilde{B}_\mu(\tilde{\boldsymbol{x}}, \boldsymbol{s}) = B_\mu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{s})$ and the norm of constraints $\|\boldsymbol{c}(\tilde{\boldsymbol{x}})\|$, and also to assure positivity of vectors $\boldsymbol{s}$ and $\boldsymbol{u}$. We can do this in several different ways: using either the augmented Lagrange function [20, 21] or a bi-criterial filter [9, 29] or a special algorithm [11, 16]. In this section, we confine our attention to the augmented Lagrange function which has (for the problem (11.106)) the form

$$P(\alpha) = \tilde{B}_\mu(\tilde{\boldsymbol{x}} + \alpha\Delta\tilde{\boldsymbol{x}}, \boldsymbol{s}(\alpha)) + (\boldsymbol{u} + \Delta\boldsymbol{u})^T(\boldsymbol{c}(\tilde{\boldsymbol{x}} + \alpha\Delta\tilde{\boldsymbol{x}}) + \boldsymbol{s}(\alpha)) + \frac{\sigma}{2}\|\boldsymbol{c}(\tilde{\boldsymbol{x}} + \alpha\Delta\tilde{\boldsymbol{x}}) + \boldsymbol{s}(\alpha)\|^2, \tag{11.118}$$

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

Theorem 11.10 *Let $\boldsymbol{s} > 0$, $\boldsymbol{u} > 0$ and let vectors $\Delta\tilde{\boldsymbol{x}}$, $\Delta\hat{\boldsymbol{u}}$ be solutions of the linear system*

$$\begin{bmatrix} \hat{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) & \hat{A}(\tilde{\boldsymbol{x}}) \\ \hat{A}^T(\tilde{\boldsymbol{x}}) & -\hat{M} \end{bmatrix} \begin{bmatrix} \Delta\tilde{\boldsymbol{x}} \\ \Delta\hat{\boldsymbol{u}} \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) \\ \hat{\boldsymbol{c}}(\tilde{\boldsymbol{x}}) + \mu\hat{U}^{-1}\tilde{\boldsymbol{e}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r} \\ \hat{\boldsymbol{r}} \end{bmatrix}, \tag{11.119}$$

where $\boldsymbol{r}$ and $\hat{\boldsymbol{r}}$ are residual vectors, and let vectors $\Delta\check{\boldsymbol{u}}$ and $\Delta\boldsymbol{s}$ be determined by (11.114) and (11.116). Then

$$P'(0) = -(\Delta\tilde{\boldsymbol{x}})^T\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})\Delta\tilde{\boldsymbol{x}} - (\Delta\boldsymbol{s})^T M^{-1}\Delta\boldsymbol{s} - \sigma\|\boldsymbol{c}(\tilde{\boldsymbol{x}}) + \boldsymbol{s}\|^2 + (\Delta\tilde{\boldsymbol{x}})^T\boldsymbol{r} + \sigma(\hat{\boldsymbol{c}}(\tilde{\boldsymbol{x}}) + \hat{\boldsymbol{s}})^T\hat{\boldsymbol{r}}. \tag{11.120}$$

If

$$\sigma > -\frac{(\Delta\tilde{\boldsymbol{x}})^T\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})\Delta\tilde{\boldsymbol{x}} + (\Delta\boldsymbol{s})^T M^{-1}\Delta\boldsymbol{s}}{\|\boldsymbol{c}(\tilde{\boldsymbol{x}}) + \boldsymbol{s}\|^2} \tag{11.121}$$

and if system (11.115) is solved in such a way that

$$(\Delta\tilde{\boldsymbol{x}})^T\boldsymbol{r} + \sigma(\hat{\boldsymbol{c}}(\tilde{\boldsymbol{x}}) + \hat{\boldsymbol{s}})^T\hat{\boldsymbol{r}} < (\Delta\tilde{\boldsymbol{x}})^T\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})\Delta\tilde{\boldsymbol{x}} + (\Delta\boldsymbol{s})^T M^{-1}\Delta\boldsymbol{s} + \sigma(\|\boldsymbol{c}(\tilde{\boldsymbol{x}}) + \boldsymbol{s}\|^2), \tag{11.122}$$

then $P'(0) < 0$.

Inequality (11.122) is significant only if linear system (11.115) is solved iteratively and residual vectors $\boldsymbol{r}$ and $\hat{\boldsymbol{r}}$ are nonzero. If these vectors are zero, then (11.122) follows immediately from (11.121). Inequality (11.121) serves for determination of a penalty parameter, which should be as small as possible. If the matrix $\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})$ is positive semidefinite, then the right-hand side of (11.121) is negative and we can choose $\sigma = 0$.

11.4.2 Implementation

The algorithm of the primal-dual interior point method consists of four basic parts: determination of the matrix $G(\boldsymbol{x}, \boldsymbol{u})$ or its approximation, solving linear system (11.115), a step length selection, and an update of the barrier parameter μ. The matrix $G(\boldsymbol{x}, \boldsymbol{u})$ has form (11.33), so its approximation can be determined in the one of the ways introduced in Remark 11.13.

The linear system (11.115), 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

$$C = \begin{bmatrix} \hat{D} & \hat{A}(\tilde{\boldsymbol{x}}) \\ \hat{A}^T(\tilde{\boldsymbol{x}}) & -\hat{M} \end{bmatrix}, \tag{11.123}$$

where $\hat{D}$ is a positive definite diagonal matrix that approximates matrix $\hat{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})$. An iterative process is terminated if residual vectors satisfy conditions (11.122) and

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

where $0 < \omega < 1$ is a prescribed precision. The directional derivative $P'(0)$ given by (11.118) should be negative. There are two possibilities how this requirement can be achieved. We either determine the value $\sigma \ge 0$ satisfying inequality (11.121), which implies $P'(0) < 0$ if (11.122) holds (Theorem 11.10), or set $\sigma = 0$ and ignore inequality (11.122). If $P'(0) \ge 0$, we determine a diagonal matrix $\tilde{D}$ with elements

$$\begin{cases} \tilde{D}_{jj} = \underline{\Gamma}, & \text{if} \quad \frac{\|\tilde{\boldsymbol{g}}\|}{10}|\tilde{G}_{jj}| < \underline{\Gamma}, \\ \tilde{D}_{jj} = \frac{\|\tilde{\boldsymbol{g}}\|}{10}|\tilde{G}_{jj}|, & \text{if} \quad \underline{\Gamma} \le \frac{\|\tilde{\boldsymbol{g}}\|}{10}|\tilde{G}_{jj}| \le \overline{\Gamma}, \\ \tilde{D}_{jj} = \overline{\Gamma}, & \text{if} \quad \overline{\Gamma} < \frac{\|\tilde{\boldsymbol{g}}\|}{10}|\tilde{G}_{jj}|, \end{cases} \tag{11.124}$$

for $1 \le j \le n + m$, where $\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u})$ and $0 < \underline{\Gamma} < \overline{\Gamma}$, set $\tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) = \tilde{D}$ and restart the iterative process by solving new linear system (11.115).

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

$$\begin{cases} \alpha_{s_{kl}} = \alpha, & \text{if} \quad \Delta s_{kl} \ge 0, \\ \alpha_{skl} = \min\{\alpha, -\gamma \frac{s_{kl}}{\Delta s_{kl}}\}, & \text{if} \quad \Delta s_{kl} < 0, \\ \alpha_{u_{kl}} = \alpha, & \text{if} \quad \Delta u_{kl} \ge 0, \\ \alpha_{u_{kl}} = \min\{\alpha, -\gamma \frac{u_{kl}}{\Delta u_{kl}}\}, & \text{if} \quad \Delta u_{kl} < 0 \end{cases}$$

when choosing a step length using the augmented Lagrange function. A parameter $0 < \gamma < 1$ (usually $\gamma = 0.99$) assures the positivity of vectors $\boldsymbol{s}_+$ and $\boldsymbol{u}_+$ in (11.117). 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 [28] seems to be very efficient.

Procedure C

Compute the centrality measure

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

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

The considerations up to now are summarized in Algorithm 11.3 introduced in the Appendix.

11.5 Numerical Experiments

The methods studied in this contribution were tested by using two collections of test problems TEST14 and TEST15 described in [19], which are the parts of the UFO system [24] and can be downloaded from the web-page www.cs.cas.cz/luksan/test.html. Both these collections contain 22 problems with functions $f_k(\boldsymbol{x})$, $1 \le k \le m$, $\boldsymbol{x} \in \mathbb{R}^n$, where n is an input parameter and $m \ge n$ depends on n (we have used the values $n = 100$ and $n = 1000$ for numerical experiments). Functions $f_k(\boldsymbol{x})$, $1 \le k \le m$, have a sparse structure (the Jacobian matrix of a mapping $\boldsymbol{f}(\boldsymbol{x})$ is sparse), so sparse matrix decompositions can be used for solving linear equation systems.

The tested methods, whose results are reported in Tables 11.1, 11.2, 11.3, 11.4, and 11.5 introduced in the Appendix, are denoted by five letters. The first pair of letters gives the problem type: either a classical minimax MX (when a function $F(\boldsymbol{x})$ has form (11.10) or $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_\infty$ holds) or a sum of absolute values SA (when $F(\boldsymbol{x}) = \|f(\boldsymbol{x})\|_1$ holds). Further two letters specify the method used:

`PI` –the primal interior point method (Sect. 11.2),
`SM` –the smoothing method (Sect. 11.3),
`DI` –the primal-dual interior point method (Sect. 11.4).

The last letter denotes the procedure for updating a barrier parameter μ (procedures A and B are described in Sect. 11.2.4 and procedure C is described in Sect. 11.4.2).

Table 11.1 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
MXPI-A	2232	7265	11575	0.74	4	-	2849	5078	2821	0.32	2	-
MXPI-B	2184	5262	9570	0.60	1	-	1567	2899	1589	0.24	1	-
MXSM-A	3454	11682	21398	1.29	5	-	4444	12505	4465	1.03	-	-
MXSM-B	10241	36891	56399	4.15	3	-	8861	32056	8881	2.21	1	1
MXDI-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
MXPI-A	1386	3735	7488	5.58	4	-	3237	12929	3258	5.91	6	-
MXPI-B	3153	6885	12989	9.03	4	-	1522	3287	1544	2.68	5	-
MXSM-A	10284	30783	82334	54.38	7	-	4221	9519	4242	8.00	8	-
MXSM-B	18279	61180	142767	87.76	6	-	13618	54655	13639	45.10	9	1
MXDI-C	3796	6677	48204	49.95	6	-	2371	5548	2371	18.89	3	-

Table 11.2 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
MXPI-A	2194	5789	10553	0.67	3	-	2890	5049	2912	0.48	1	-
MXPI-B	6767	17901	39544	3.79	4	-	1764	3914	1786	0.37	2	-
MXSM-A	3500	9926	23568	1.79	7	-	8455	23644	8476	4.69	4	-
MXSM-B	15858	48313	92486	8.33	5	-	9546	34376	9566	2.59	9	1
MXDI-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
MXPI-A	4110	14633	20299	18.89	4	-	1549	2636	1571	2.51	3	-
MXPI-B	6711	31618	29939	30.73	7	-	1992	6403	2013	4.96	4	-
MXSM-A	9994	24333	88481	67.45	11	-	6164	15545	6185	29.37	8	-
MXSM-B	23948	84127	182604	149.63	8	-	24027	92477	24048	132.08	8	1
MXDI-C	3528	9084	26206	49.78	12	-	1932	2845	1932	18.73	5	-

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 [12] (Remark 11.13).

Table 11.3 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
MXPI-A	15525	20272	55506	4.41	1	-	6497	8204	6518	1.37	3	-
MXPI-B	7483	17999	27934	3.27	5	-	1764	7598	2488	0.74	2	-
MXSM-A	17574	29780	105531	11.09	4	-	9879	15305	9900	5.95	-	-
MXSM-B	13446	29249	81938	6.80	9	1	9546	34376	9566	2.59	3	-
MXDI-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
MXPI-A	10325	15139	32422	39.30	6	-	6484	9904	6502	13.77	2	-
MXPI-B	14836	30724	46864	68.70	10	-	7388	15875	7409	19.98	3	-
MXSM-A	11722	24882	69643	61.65	10	-	6659	12824	6681	41.55	8	-
MXSM-B	13994	31404	86335	78.45	9	1	15125	25984	15147	61.62	10	-
MXDI-C	1408	2406	10121	15.63	6	-	2228	3505	2228	35.13	10	-

Table 11.4 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
SAPI-A	1647	5545	8795	0.63	5	-	12265	23579	12287	1.37	2	1
SAPI-B	1957	7779	10121	0.67	6	-	4695	6217	10608	0.67	3	-
SASM-A	1677	4505	16079	0.74	3	-	20025	27369	20047	2.83	4	-
SASM-B	2389	8085	23366	1.18	4	-	5656	11637	5678	1.02	2	-
SADI-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
SAPI-A	7592	19621	46100	15.39	4	-	22277	36610	22298	19.09	7	1
SAPI-B	9067	35463	56292	19.14	6	-	16650	35262	16672	14.47	6	1
SASM-A	5696	13534	41347	15.28	4	-	20020	30736	20042	23.05	5	1
SASM-B	8517	30736	57878	23.60	6	-	18664	28886	18686	18.65	5	1
SADI-C	6758	11011	47960	94.78	11	1	13123	14610	13124	295.46	8	2

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 step length $\overline{\Delta}$ is used for three reasons. First, too large steps may lead to overflows if arguments of functions (roots,

Table 11.5 TEST15 (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
SAPI-A	15760	21846	58082	4.24	8	-	39469	58157	39486	6.28	4	1
SAPI-B	4592	17050	17778	1.46	5	-	5932	25035	5952	1.48	6	1
SASM-A	10098	14801	610511	3.54	5	-	9162	28421	9184	3.65	6	1
SASM-B	4528	14477	290379	2.94	8	-	3507	9036	3528	1.27	6	-
SADI-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
SAPI-A	18519	39319	70951	61.04	5	-	27308	44808	27327	36.64	4	1
SAPI-B	12405	57969	43189	55.06	7	-	12712	32179	12731	21.48	8	1
SASM-A	19317	32966	113121	62.65	8	-	22264	42908	22284	62.46	7	1
SASM-B	14331	33572	86739	57.56	6	-	12898	42479	12919	47.05	7	1
SADI-C	2093	3681	12616	20.01	3	1	23957	28000	23960	186.92	5	3

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 step length 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.

- 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 (11.115) is indefinite, so we cannot use the Choleski (or the Gill–Murray [10]) decomposition. If the matrix of linear system (11.115) is large and sparse, we can use the Bunch–Parlett decomposition [6]. 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 [18], 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).

Appendix

Algorithm 11.1: 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 step length selection $\varepsilon_1 > 0$. A maximum step length $\overline{\Delta} > 0$.

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

Step 1. *(Initiation)* Choose $\mu \leq \overline{\mu}$. Determine a sparse structure of the matrix $W = W(\boldsymbol{x}; \mu)$ from the sparse structure of the matrix $A(\boldsymbol{x})$ and perform a symbolic decomposition of the matrix W (described in [2, Section 1.7.4]). Compute values $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, values $F_k(\boldsymbol{x}) = \max_{1 \leq l \leq m_k} f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, and the value of objective function (11.4). Set $r = 0$ (restart indicator).

Step 2. *(Termination)* Solve nonlinear equations (11.44) with precision $\underline{\delta}$ to obtain a minimax vector $\boldsymbol{z}(\boldsymbol{x}; \mu)$ and a vector of Lagrange multipliers $\boldsymbol{u}(\boldsymbol{x}; \mu)$. Determine a matrix $A = A(\boldsymbol{x})$ and a vector $\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{x}; \mu) = A(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}; \mu)$. If $\mu \leq \underline{\mu}$ and $\|\boldsymbol{g}\| \leq \underline{\varepsilon}$, terminate the computation.

Step 3. *(Hessian matrix approximation)* Set $G = G(\boldsymbol{x}; \mu)$ or compute an approximation G of the Hessian matrix $G(\boldsymbol{x}; \mu)$ using gradient differences or using quasi-Newton updates (Remark 11.13).

Step 4. *(Direction determination)* Determine a matrix $\nabla^2 \hat{B}(\boldsymbol{x}; \mu)$ by (11.48) and a vector $\Delta\boldsymbol{x}$ by solving Eq. (11.49) with the right-hand side defined by (11.47).

Step 5. *(Restart)* If $r = 0$ and (11.54) does not hold, set $G = I$, $r = 1$ and go to Step 4. If $r = 1$ and (11.54) does not hold, set $\Delta\boldsymbol{x} = -\boldsymbol{g}$. Set $r = 0$.

Step 6. *(Step length selection)* Determine a step length $\alpha > 0$ satisfying inequalities (11.55) (for a barrier function $\hat{B}(\boldsymbol{x}; \mu)$ defined by (11.46)) and $\alpha \leq \overline{\Delta}/\|\Delta\boldsymbol{x}\|$. Note that nonlinear equations (11.44) are solved at the point $\boldsymbol{x} + \alpha\Delta\boldsymbol{x}$. Set $\boldsymbol{x} := \boldsymbol{x} + \alpha\Delta\boldsymbol{x}$. Compute values $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, values $F_k(\boldsymbol{x}) = \max_{1 \leq l \leq m_k} f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, and the value of objective function (11.4).

Step 7. *(Barrier parameter update)* Determine a new value of a barrier parameter $\mu \geq \underline{\mu}$ 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.

Algorithm 11.2: 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 step length selection $\varepsilon_1 > 0$. A maximum step length $\overline{\Delta} > 0$.

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

Step 1. *(Initiation)* Choose $\mu \leq \overline{\mu}$. Determine a sparse structure of the matrix $W = W(\boldsymbol{x}; \mu)$ from the sparse structure of the matrix $A(\boldsymbol{x})$ and perform a symbolic decomposition of the matrix W (described in [2, Section 1.7.4]). Compute values $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, values $F_k(\boldsymbol{x}) = \max_{1 \leq l \leq m_k} f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, and the value of objective function (11.4). Set $r = 0$ (restart indicator).

Step 2. *(Termination)* Determine a vector of smoothing multipliers $\boldsymbol{u}(\boldsymbol{x}; \mu)$ by (11.87). Determine a matrix $A = A(\boldsymbol{x})$ and a vector $\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{x}; \mu) = A(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}; \mu)$. If $\mu \leq \underline{\mu}$ and $\|\boldsymbol{g}\| \leq \underline{\varepsilon}$, terminate the computation.

Step 3. *(Hessian matrix approximation)* Set $G = G(\boldsymbol{x}; \mu)$ or compute an approximation G of the Hessian matrix $G(\boldsymbol{x}; \mu)$ using gradient differences or using quasi-Newton updates (Remark 11.13).

Step 4. *(Direction determination)* Determine the matrix W by (11.94) and the vector $\Delta\boldsymbol{x}$ by (11.93) using the Gill–Murray decomposition of the matrix W.

Step 5. *(Restart)* If $r = 0$ and (11.54) does not hold, set $G = I$, $r = 1$ and go to Step 4. If $r = 1$ and (11.54) does not hold, set $\Delta\boldsymbol{x} = -\boldsymbol{g}$. Set $r = 0$.

Step 6. *(Step length selection)* Determine a step length $\alpha > 0$ satisfying inequalities (11.55) (for a smoothing function $S(\boldsymbol{x}; \mu)$) and $\alpha = \overline{\Delta}/\|\Delta\boldsymbol{x}\|$. Set $\boldsymbol{x} := \boldsymbol{x} + \alpha\Delta\boldsymbol{x}$. Compute values $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, values $F_k(\boldsymbol{x}) = \max_{1 \leq l \leq m_k} f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, and the value of the objective function (11.4).

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

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.

Algorithm 11.3: Primal-dual interior point method

Data: A tolerance for the gradient norm $\underline{\varepsilon} > 0$. A parameter for determination of active constraints $\tilde{\varepsilon} > 0$. A parameter for initiation of slack variables and Lagrange multipliers $\delta > 0$. An initial value of the barrier parameter $\overline{\mu} > 0$. A precision for the direction determination $0 \leq \omega < 1$. A parameter for the step length selection $0 < \gamma < 1$. A tolerance for the step length selection $\varepsilon_1 > 0$. Maximum step length $\overline{\Delta} > 0$.

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

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

Step 2. *(Termination)* Determine a matrix $\tilde{A}(\tilde{\boldsymbol{x}})$ and a vector $\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) = \tilde{A}(\tilde{\boldsymbol{x}})\boldsymbol{u}$ by (11.111). If the KKT conditions $\|\tilde{\boldsymbol{g}}(\tilde{\boldsymbol{x}}, \boldsymbol{u})\| \leq \underline{\varepsilon}$, $\|\boldsymbol{c}(\tilde{\boldsymbol{x}}) + \boldsymbol{s}\| \leq \underline{\varepsilon}$, and $\boldsymbol{s}^T\boldsymbol{u} \leq \underline{\varepsilon}$ are satisfied, terminate the computation.

Step 3. *(Hessian matrix approximation)* Set $G = G(\boldsymbol{x}, \boldsymbol{u})$ or compute an approximation G of the Hessian matrix $G(\boldsymbol{x}, \boldsymbol{u})$ using gradient differences or utilizing quasi-Newton updates (Remark 11.13). Determine a parameter $\sigma \geq 0$ by (11.121) or set $\sigma = 0$. Split the constraints into active if $\hat{s}_{kl} \leq \tilde{\varepsilon}\hat{u}_{kl}$ and inactive if $\check{s}_{kl} > \tilde{\varepsilon}\check{u}_{kl}$.

Step 4. *(Direction determination)* Determine the matrix $\tilde{G} = \tilde{G}(\tilde{\boldsymbol{x}}, \boldsymbol{u})$ by (11.111) (where the Hessian matrix $G(\boldsymbol{x}, \boldsymbol{u})$ is replaced with its approximation G). Determine vectors $\Delta\tilde{\boldsymbol{x}}$ and $\Delta\hat{\boldsymbol{u}}$ by solving linear system (11.115), a vector $\Delta\check{\boldsymbol{u}}$ by (11.114), and a vector $\Delta\boldsymbol{s}$ by (11.116). Linear system (11.115) 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 (11.123). Compute the derivative of the augmented Lagrange function by formula (11.120).

Step 5. *(Restart)* If $P'(0) \geq 0$, determine a diagonal matrix $\tilde{D}$ by (11.124), set $\tilde{G} = \tilde{D}$, $\sigma = 0$, and go to Step 4.

Step 6. *(Step length selection)* Determine a step length parameter $\alpha > 0$ satisfying inequalities $P(\alpha) - P(0) \leq \varepsilon_1 \alpha P'(0)$ and $\alpha \leq \overline{\Delta}/\|\Delta\boldsymbol{x}\|$. Determine new vectors $\tilde{\boldsymbol{x}} := \tilde{\boldsymbol{x}} + \alpha\Delta\tilde{\boldsymbol{x}}$, $\boldsymbol{s} := \boldsymbol{s}(\alpha)$, $\boldsymbol{u} := \boldsymbol{u}(\alpha)$ by (11.117). Compute values $f_{kl}(\boldsymbol{x})$, $1 \leq k \leq m$, $1 \leq l \leq m_k$, and set $c_{kl}(\tilde{\boldsymbol{x}}) = f_{kl}(\boldsymbol{x}) - z_k$, $1 \leq k \leq m$, $1 \leq l \leq m_k$. Compute the value of the barrier function $\tilde{B}_\mu(\tilde{\boldsymbol{x}}, \boldsymbol{s})$.

Step 7. *(Barrier parameter update)* Determine a new value of the barrier parameter $\mu \geq \underline{\mu}$ 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.

References

1. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numer. **14**, 1–137 (2005)
2. Björck, A.: Numerical Methods in Matrix Computations. Springer, New York (2015)
3. Bunch, J.R., Parlett, B.N.: Direct methods for solving symmetric indefinite systems of linear equations. SIAM J. Numer. Anal. **8**, 639–655 (1971)
4. Coleman, T.F., Moré, J.J.: Estimation of sparse Hessian matrices and graph coloring problems. Math. Program. **28**, 243–270 (1984)
5. Di Pillo, G., Grippo, L., Lucidi, S.: Smooth Transformation of the generalized minimax problem. J. Optim. Theory Appl. **95**, 1–24 (1997)
6. Duff, I.S., Gould, N.I.M., Reid, J.K., Turner, K.: The factorization of sparse symmetric indefinite metrices. IMA J. Numer. Anal. **11**, 181–204 (1991)
7. Fiedler, M.: Special Matrices and Their Applications in Numerical Mathematics. Dover, New York (2008)
8. Fletcher, R.: Practical Methods of Optimization. Wiley, New York (1987)
9. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Program. **91**, 239–269 (2002)
10. Gill, P.E., Murray, W.: Newton type methods for unconstrained and linearly constrained optimization. Math. Program. **7**, 311–350 (1974)
11. Gould, N.I.M., Toint, P.L.: Nonlinear programming without a penalty function or a filter. Math. Program. **122**, 155–196 (2010)
12. Griewank, A., Toint, P.L.: Partitioned variable metric updates for large-scale structured optimization problems. Numer. Math. **39**, 119–137 (1982)
13. Griewank, A., Walther, A.: Evaluating Derivatives. SIAM, Philadelphia (2008)
14. Le, D.: Three new rapidly convergent algorithms for finding a zero of a function. SIAM J. Sci. Stat. Comput. **6**, 193–208 (1985)
15. Le, D.: An efficient derivative-free method for solving nonlinear equations. ACM Trans. Math. Softw. **11**, 250–262 (1985)
16. Liu, X., Yuan, Y.: A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization. SIAM J. Optim. **21**, 545–571 (2011)
17. Lukšan, L., Spedicato, E.: Variable metric methods for unconstrained optimization and nonlinear least squares. J. Comput. Appl. Math. **124**, 61–93 (2000)
18. Lukšan, L., Vlček, J.: Indefinitely preconditioned inexact Newton method for large sparse equality constrained nonlinear programming problems. Numer. Linear Algebra Appl. **5**, 219–247 (1998)
19. 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)
20. Lukšan, L., Matonoha, C., Vlček, J.: Interior-point method for non-linear non-convex optimization. Numer. Linear Algebra Appl. **11**, 431–453 (2004)
21. Lukšan, L., Matonoha, C., Vlček, J.: Interior-point method for large-scale nonlinear programming. Optim. Methods Softw. **20**, 569–582 (2005)
22. Lukšan, L., Matonoha, C., Vlček, J.: Primal interior-point method for large sparse minimax optimization. Kybernetika **45**, 841–864 (2009)
23. Lukšan, L., Matonoha, C., Vlček, J.: Primal interior-point method for minimization of generalized minimax functions. Kybernetika **46**, 697–721 (2010)
24. Lukšan, L., Tůma, M., Matonoha, C., Vlček, J., Ramešová, N., Šiška, M., Hartman, J.: UFO 2017. Interactive System for Universal Functional Optimization. Technical Report V-1252, Prague, ICS AS CR (2017)
25. Mäkelä, M.M., Neittaanmäki, P.: Nonsmooth Optimization. World Scientific, London (1992)
26. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)

27. Tůma, M.: A note on direct methods for approximations of sparse Hessian matrices. Appl. Math. **33**, 171–176 (1988)
28. Vanderbei, J., Shanno, D.F.: An interior point algorithm for nonconvex nonlinear programming. Comput. Optim. Appl. **13**, 231–252 (1999)
29. Wächter, A., Biegler, L.: Line search filter methods for nonlinear programming. Motivation and global convergence. SIAM J. Comput. **16**, 1–31 (2005)