Institute of Computer Science
Academy of Sciences of the Czech Republic

Interior point methods for minimization of composite nonsmooth functions

L.Lukšan, C.Matonoha, J.Vlček

Technical report No. 987

December 2006

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

Institute of Computer Science
Academy of Sciences of the Czech Republic

Interior point methods for minimization of composite nonsmooth functions

L.Lukšan, C.Matonoha, J.Vlček [1]

Technical report No. 987

December 2006

Abstrakt:

In this report, we propose a class of primal interior point methods for minimization of composite nonsmooth functions. After a short introduction where composite nonsmooth functions are defined, we describe two algorithms based on iterative and direct determination of the minimax vector respectively. Finally, we investigate most important special cases a give results of numerical experiments, which demonstrate high efficiency of primal interior point methods for composite nonsmooth functions.

Keywords:
Numerical optimization, nonlinear programming, nonlinear approximation, algorithms, software systems.

[1] This work was supported by the Grant Agency of the Czech Academy of Sciences, project code IAA1030405, L.Lukšan is also from Technical University of Liberec, Hálkova 6, 461 17 Liberec.

1 Introduction

Functions which we need to minimize are often non-differentiable and its non-smoothness is caused by the fact that they contain absolute values or point maxima of differentiable functions. Typical examples are the norms $\|f(x)\|_1$ a $\|f(x)\|_\infty$ of a smooth mapping $f : R^n \to R^m$. Generalizations of these functions are, e.g., functions of the form $F(x) = p(f_1(x), \dots, f_m(x))$, where $p : R^m \to R$ is a non-smooth function satisfying some additional conditions (it should be convex or locally Lipschitz) and $f : R^n \to R^m$ is a smooth mapping. Such functions are usually called composite non-smooth functions and their advantage is the fact that their structure can be often advantageously used for a construction of efficient numerical algorithms for their minimization. As a special case we might introduce the Fletcher composite non-smooth functions (see [2]) of the form

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

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

In this contribution we focus our attention on a different class of composite non-smooth functions defined by the following way.

Definition 1 *We say that $F(x)$ is a composite non-smooth function if*

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

where $h : R^m \to R$ and $f_{ij} : R^n \to R$, $1 \le i \le m$, $1 \le j \le n_i$, are twice continuously differentiable functions. At the same time, $h(z) = h(z_1, \dots, z_m)$ is a convex function such that $\partial h(z)/\partial z_i \ge \underline{h}_i > 0$, $1 \le i \le m$, if $x \in R^n$ and $z \in Z(x)$, where

$$Z(x) = \{z \in R^m : z_i \ge F_i(x), \ \ 1 \le i \le m\}.$$

Conditions put on the function $h(z)$ are relatively strong, but many functions satisfy them (e.g. $h(z) = z_1 + \ldots + z_m$). It is clear that we can express in this way all the functions mentioned above. The absolute values $F_i(x) = |f_i(x)|$ can be expressed in the form $F_i(x) = \max(f_i(x), -f_i(x))$, in which case $Z(x) \subset R_+^m$, so the condition $\partial h(z)/\partial z_i \ge \underline{h}_i > 0$ if $z_i \ge 0$ suffices. To express the functions $\|f(x)\|_1$, $\|f_+(x)\|_1$ by (2) is much easier in comparison with (1), since in that case the matrix P contains 2^m columns.

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

$$h(z_1, \dots, z_m) \tag{3}$$

with constraints

$$f_{ij}(x) \le z_i, \quad 1 \le i \le m, \quad 1 \le j \le n_i \tag{4}$$

(the condition $\partial h(z)/\partial z_i \geq \underline{h}_i > 0$, $1 \leq i \leq m$, if $x \in R^n$ and $z \in Z(x)$, is sufficient for satisfying equalities $z_i = F_i(x)$, $1 \leq i \leq m$ at the minimum point). This equivalent nonlinear programming problem will be solved by the primal interior point method. For this reason we will apply the Newton minimization method to the barrier function

$$B_\mu(x,z) = h(z) - \mu \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log(z_i - f_{ij}(x)) \tag{5}$$

assuming that $\mu \to 0$. The notation

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

for $1 \leq i \leq m$, $1 \leq j \leq n_i$, is used. The primal interior point method is based on the fact that it is easy to find a vector $z \in R^m$ satisfying constraints (4). Hence, it is not necessary to introduce slack variables, add equality constraints, use a penalty function and iterate the Lagrangian multipliers. In the subsequent sections, we will describe two approaches which differ in the determination of the minimax vector $z \in R^m$. We focus on the problems whose structure allows using the sparse matrix technique.

2 Iterative determination of the minimax vector

The necessary conditions for (x,z) to be a minimum of function (5) have the form

$$\nabla_x B_\mu(x,z) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} A_{ij}(x) \frac{\mu}{z_i - f_{ij}(x)} = 0 \tag{6}$$

and

$$\frac{\partial B_\mu(x,z)}{\partial z_i} = h_i(z) - \sum_{j=1}^{n_i} \frac{\mu}{z_i - f_{ij}(x)} = 0, \quad 1 \leq i \leq m, \tag{7}$$

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

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

$$= - \begin{bmatrix} \sum_{i=1}^m A_i(x)u_i(x,z) \\ h_1(z) - e_1^T u_1(x,z) \\ \dots \\ h_m(z) - e_m^T u_m(x,z) \end{bmatrix}, \tag{8}$$

where

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

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

for $1 \le i \le m$, $1 \le j \le n_i$, and where $A_i(x)=[A_{i1}(x),\dots,A_{in_i}(x)]$,

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

This formula is verified from the differentiation of (6) and (7) by x and z. Setting

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

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

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

we can rewrite equation (8) in the form

$$\begin{bmatrix} W(x,z) & -C(x,z) \\ -C^T(x,z) & H(z)+V(x,z) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = -\begin{bmatrix} g(x,z) \\ c(x,z) \end{bmatrix}. \tag{9}$$

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

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

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

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

The solution is determined from the formulas

$$\Delta z=(C^TW^{-1}C-H-V)^{-1}(C^TW^{-1}g+c), \tag{10}$$

$$\Delta x=W^{-1}(C\Delta z-g). \tag{11}$$

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

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

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

is sparse. Then we can use the fact that

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

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

The solution is determined from the formulas

$$\Delta x = -(W - CD^{-1}C^T)^{-1}(g + CD^{-1}c), \tag{12}$$

$$\Delta z = D^{-1}(C^T \Delta x - c). \tag{13}$$

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

In every step of the primal interior point method with the iterative determination of the minimax vector we know the value of the parameter μ and the vectors $x \in R^n$, $z \in R^m$ such that $z_i > F_i(x)$, $1 \le i \le m$. Using (10)–(11) or (12)–(13) we determine direction vectors Δx, Δz and select a step-size α in such a way that

$$B_\mu(x + \alpha\Delta x, z + \alpha\Delta z) < B_\mu(x, z) \tag{14}$$

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

Inequality (14) is satisfied for sufficiently small values of the step-size α, if the matrix of system (9) is positive definite.

Theorem 1 *Let the matrix $G = \sum_{i=1}^m \sum_{j=1}^{n_i} G_{ij} u_{ij}$ be positive definite. Then the matrix of system (9) is positive definite.*

Proof. The matrix of equation (9) is positive definite if and only if the matrix $D = H + V$ as well as its Schur complement $W - CD^{-1}C^T$ are both positive definite. The matrix $D = H + V$ is positive definite since both matrices H and V are positive definite. Now we use the fact that the matrix $V^{-1} - D^{-1}$ is positive definite, since the matrix $H = D - V$ is positive definite (see [8]). Thus $v^T(W - CD^{-1}C^T)v \ge v^T(W - CV^{-1}C^T)v$ $\forall v \in R^n$ so it suffices to prove that the matrix $W - CV^{-1}C^T$ is positive definite. But

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

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

3 Direct determination of the minimax vector

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

$$z(x) = \arg \min_{z \in Z(x)} B_\mu(x, z), \tag{15}$$

$$x = \arg \min_{x \in R^n} B_\mu(x), \quad B_\mu(x) \triangleq B_\mu(x, z(x)), \tag{16}$$

where $Z(x)$ is the set used in Definition 1. Equation (15) serves for a determination of an optimal vector $z(x) \in R^m$ corresponding to a given vector $x \in R^n$. The function $B_\mu(x, z)$ is a convex function of a vector z for a given vector x, since it is a sum of convex functions $h(z)$ and $-\log(z_i - f_{ij}(x))$, $1 \le i \le m$, $1 \le j \le n_i$. As a stationary point, its minimum is determined by the set of equations (7).

Theorem 2 *The system of equations*

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

for an arbitrary vector $x \in R^n$ and for $z \in Z(x)$, has the unique solution $z(x)$ such that

$$F_i(x) < z_i(x) \le F_i(x) + n_i\mu/\underline{h}_i, \quad 1 \le i \le m.$$

Proof. Denoting $\overline{z}_i = F_i(x) + n_i\mu/\underline{h}_i$, $1 \le i \le m$, functions $\partial h(z)/\partial z_i$, $1 \le i \le m$, attain their maximum values $\overline{h}_i \ge \underline{h}_i > 0$, $1 \le i \le m$, on the compact set determined by inequalities $F_i(x) \le z_i \le \overline{z}_i$, $1 \le i \le m$, since they are continuous there. Denoting $\underline{z}_i = F_i(x) + \mu/\overline{h}_i$, $1 \le i \le m$, and choosing an arbitrary (sufficiently small) number $\varepsilon > 0$, the function $B_\mu(x, z)$ attains its minimum on the compact set $Z_\varepsilon(x) \subset \text{int } Z(x)$ determined by equations $\underline{z}_i - \varepsilon\mu/\overline{h}_i \le z_i \le \overline{z}_i + \varepsilon n_i\mu/\underline{h}_i$, $1 \le i \le m$, since it is continuous on int $Z(x)$. Now we will show that this minimum cannot lie on the boundary of $Z_\varepsilon(x)$. It is clear that for every point of this boundary there is at least one index $1 \le i \le m$ such that either $z_i = \underline{z}_i - \varepsilon\mu/\overline{h}_i$ or $z_i = \overline{z}_i + \varepsilon n_i\mu/\underline{h}_i$ holds. If $z_i = \underline{z}_i - \varepsilon\mu/\overline{h}_i$, then

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

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

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

so a small decrease of the variable z_i can decrease the function value of $B_\mu(x,z)$. The above considerations imply that the minimum of the function $B_\mu(x,z)$ is an interior point of the set $Z_\varepsilon(x)$ and since $B_\mu(x,z)$ is continuously differentiable on $Z_\varepsilon(x)$, necessary conditions (7) have to be satisfied. Since the number $\varepsilon > 0$ can be chosen arbitrarily, the solution satisfies inequalities $F_i(x) < \underline{z}_i \le z_i(x) \le \overline{z}_i$, $1 \le i \le m$. □

System of equations (7) can be solved by the Newton method started, e.g., from the point z such that $z_i = \overline{z}_i$, $1 \le i \le m$. If the Hessian matrix of the function $h(z)$ is diagonal, then system (7) is decomposed on m scalar equations, which can be efficiently solved by methods described in [4], [5] (see [9]).

If we are able to find a solution of system (7) for an arbitrary vector $x \in R^n$, we can restrict our attention to the unconstrained minimization of the function $B_\mu(x) = B_\mu(x, z(x))$, which has n variables. It is suitable to know the gradient and the Hessian matrix of the function $B_\mu(x)$.

Theorem 3 *One has*

$$\nabla B_\mu(x) = \sum_{i=1}^{m} A_i(x) u_i(x) \tag{17}$$

and

$$\nabla^2 B_\mu(x) = W(x,z(x)) - C(x,z(x))\left(H(z(x)) + V(x,z(x))\right)^{-1} C^T(x,z(x)), \tag{18}$$

where $W(x,z(x))$, $C(x,z(x))$, $H(z(x))$, $V(x,z(x))$ are the matrices used in the previous section. If the matrix $H(z(x))$ is diagonal, we can express (18) in the form

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

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

Proof. Differentiating function (5), where $z = z(x)$, we obtain

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

Formula (18) can be derived by an additional differentiation of relations (7) and (17). Simpler way is based on the use of formula (12). Since (7) implies $c(x,z(x)) = 0$, after substitution $c = 0$ into (12) we will obtain the relation

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

which confirms a validity of formula (18). □

To determine the Hessian matrix inverse, we can use relations (10)–(11) which, after substitution $c(x, z(x)) = 0$, give

$$\begin{aligned}(\nabla^2 B_\mu(x))^{-1} &= W(x, z(x))^{-1} - W(x, z(x))^{-1} C(x, z(x)) \\ &\quad \Big(C^T(x, z(x)) W^{-1}(x, z(x)) C(x, z(x)) - H(z(x)) - V(x, z(x))\Big)^{-1} \\ &\quad C^T(x, z(x)) W(x, z(x))^{-1}. \end{aligned} \tag{19}$$

If system (7) is not solved with a sufficient precision, we use (12)–(13) rather than (18) and (10)–(11) rather than (19) (where the actual vector $c(x, z(x)) \neq 0$ is substituted).

In every step of the primal interior point method with the direct determination of the minimax vector we know the value of the parameter μ and the vector $x \in R^n$. Solving system (7) we determine the vector $z(x)$, using Hessian matrix (18) or its inverse (19) we determine a direction vector Δx and select a step-size α in such a way that

$$B_\mu(x + \alpha\Delta x, z(x + \alpha\Delta x)) < B_\mu(x, z(x)) \tag{20}$$

(the vector $z(x + \alpha\Delta x)$ is obtained as a solution of system (7), in which x is replaced by $x + \alpha\Delta x$). Finally, we set $x^+ = x + \alpha\Delta x$ and determine a new value $\mu^+ < \mu$. Conditions for the direction vector Δx to be descent are the same as in Theorem 1. It suffices when the matrix $G(x, z(x))$ is positive definite.

4 Special cases and numerical experiments

The simplest function of form (2) is the sum

$$F(x) = \sum_{i=1}^{m} F_i(x) = \sum_{i=1}^{m} \max_{1 \le j \le n_i} f_{ij}(x). \tag{21}$$

In this case, $\partial h(z)/\partial z_i = 1$, $1 \le i \le m$, for an arbitrary vector z and the matrix $H(z)$ is diagonal. System of equations (7) decomposes on m scalar equations

$$1 - \sum_{j=1}^{n_i} \frac{\mu}{z_i - f_{ij}(x)} = 0, \qquad 1 \le i \le m, \tag{22}$$

whose solutions lie in the intervals

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

as follows from the proof of Theorem 2 substituting $\overline{h}_i = \underline{h}_i = 1$. For $m = 1$ we obtain the classic minimax problem and the primal interior point method for it is described in [9]. Table 1, taken from [9], contains a comparison of the primal interior point method PI described in that contribution with the smoothing method SM described in [13], the primal-dual interior point method DI described in [6] and the non-smooth equation method NE described in [7]. All these methods were realized as the line-search methods with two modifications: NM denotes the discrete Newton method with the Hessian matrix computed using the differences by the way described in [1] and VM denotes the variable metric method with the partitioned updates described in [3]. The tests were carried out using a collection of 22 test problems introduced in [11] (the source texts can be downloaded from the web page `www.cs.cas.cz/~luksan/test.html` as Test 14). In Table 1, `NIT` denotes the total number of iterations, `NFV` denotes the total number of function evaluations, `NFG` denotes the total number of gradient evaluations, `NR` denotes the total number of restarts, `NL` denotes the number of problems for which the lowest known local minimum was not found, `NF` denotes the number of failures, `NT` denotes the number of problems for which some parameters of the method had to be tuned, and `Time` denotes the total computational time in seconds.

Method	NIT	NFV	NFG	NR	NL	NF	NT	Time
PI-NM	1682	3771	11173	325	-	-	4	1.75
SM-NM	4213	12632	32451	823	1	-	8	7.78
DI-NM	1718	3561	16989	74	1	-	10	6.11
NE-NM	5159	22195	42161	2363	2	-	14	32.86
PI-VM	1632	2266	1654	23	-	-	2	1.00
SM-VM	7192	20710	7214	22	1	-	8	6.42
DI-VM	2172	5283	2172	27	1	-	8	6.97
NE-VM	2756	5667	2756	49	1	-	9	5.31

Table 1. Test 14: minimax with 200 variables

If $n_i = 2$, $1 \le i \le m$, equations (22) are quadratic and their solution has the form

$$z_i(x) = \mu + \frac{f_{i1}(x) + f_{i2}(x)}{2} + \sqrt{\mu^2 + \left(\frac{f_{i1}(x) - f_{i2}(x)}{2}\right)^2}, \quad 1 \le i \le m. \tag{23}$$

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

$$z_i(x) = \mu + \sqrt{\mu^2 + f_i^2(x)}, \quad 1 \le i \le m. \tag{24}$$

The primal interior point method for the sums of absolute values is described in [10]. Table 2 contains a comparison of two realizations of the primal interior point method (the trust region realization PT and the line-search realization PL) with the primal-dual interior point method DI described in [6] and the bundle variable metric method BM described in [12]. These methods were realized in two modifications: NM denotes the discrete Newton method with the Hessian matrix computed using the differences and VM denotes the variable metric method with the partitioned updates (BM is principally the variable metric method, so it could not be realized as NM). The tests were again carried out using a collection of 22 test problems introduced in [11]. The meaning of the columns is the same as in Table 1.

Method	NIT	NFV	NFG	NR	NL	NF	NT	Time
PT-NM	2784	3329	23741	1	2	-	4	3.72
PL-NM	5093	12659	30350	1	1	-	6	4.49
DI-NM	4565	6301	37310	212	2	-	12	30.63
PT-VM	5390	5578	5414	22	1	1	1	2.31
PL-VM	4145	8669	4167	23	1	1	2	2.75
DI-VM	6903	14259	14259	29	3	-	9	89.37
BM-VM	34079	34111	34111	22	1	1	11	25.72

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

Tables 1 and 2 indicate that the primal interior point methods are very suitable for minimization of composite nonsmooth functions. They are more efficient than special bundle methods and also than general primal-dual interior point methods applied to problem (3)–(4). This is especially caused by the fact that the primal-dual interior point methods require the introduction of an additional slack vector $s \in R^m$ so that the resulting optimization problem contains $n+2m$ variables x, z, s, which considerably increases the computational time.

Reference

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

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

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

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

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

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

[7] L.Lukšan, C.Matonoha, J.Vlček: Nonsmooth equation method for nonlinear nonconvex optimization. In: Conjugate Gradient Algorithms and Finite Element

Methods (M.Křížek, P.Neittaanmäki, R.Glovinski, S.Korotov eds.). Springer Verlag, Berlin 2004.

[8] Lukšan L., Matonoha C., Vlček J.: A shifted Steihaug-Toint method for computing trust-region step. Technical Report V-914. Prague, ICS AS CR, 2004.

[9] Lukšan L., Matonoha C., Vlček J.: Primal interior-point method for large sparse minimax optimization. Technical Report V-941, Prague, ICS AS CR, 2005. Submitted to Computational Optimization and Applications.

[10] Lukšan L., Matonoha C., Vlček J.: Trust-region interior point method for large sparse l_1 optimization. Technical Report V-942, Prague, ICS AS CR, 2005. Submitted to Optimization Methods and Software.

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

[12] Lukšan L., Vlček J.: Variable metric method for minimization of partially separable nonsmooth functions. Pacific Journal on Optimization 2 (2006) 59-70.

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