A trust region method for minimization of nonsmooth functions with linear constraints

José Mario Martínez * Antonio Carlos Moretti †

November 3, 1995

Abstract

We introduce a trust region algorithm for minimization of nonsmooth functions with linear constraints. At each iteration, the objective function is approximated by a model function that satisfies a set of assumptions stated recently by Qi and Sun in the context of unconstrained nonsmooth optimization. The trust region iteration begins with the resolution of an “easy problem”, as in recent works of Martínez and Santos and Friedlander, Martínez and Santos, for smooth constrained optimization. In practical implementations we use the infinity norm for defining the trust region, which fits well with the domain of the problem. We prove global convergence and report numerical experiments.

Key words. Trust regions, nonsmooth functions, constrained optimization.

*Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil (martinezm@ime.unicamp.br). This author was supported by FAPESP (Grant 90-3724-6), FINEP and FAEP-UNICAMP.

†Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil (moretti@ime.unicamp.br). This author was supported by FAPESP (Grant 90-3724-6), FINEP and FAEP-UNICAMP.
1 Introduction

Many real-life problems require the minimization of functions whose domain $\Omega \subset \mathbb{R}^n$ is a polyhedron:

\[
\text{Minimize } f(x) \quad \text{subject to } x \in \Omega.
\]

(1)

Trust region methods (see [15], [6], [11], [12] and references therein) are generally successful for solving (1) when $f$ is smooth and the structure of $\Omega$ is simple. A global convergence theory for trust region methods for minimizing smooth functions on arbitrary regions has been given recently in [10]. The subproblems considered in [10] have the form

\[
\text{Minimize } \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s
\]

subject to $x_k + s \in \Omega, \|s\| \leq \Delta,
\]

(2)

where $x_k \in \Omega$ is the current approximation to the solution. Since, generally, problem (2) is difficult, only an approximate solution is required. For computing the approximate solution, an “easy” subproblem is considered:

\[
\text{Minimize } \nabla f(x_k)^T s + \frac{1}{2} M_k \|s\|_2^2
\]

subject to $x_k + s \in \Omega, \|s\| \leq \Delta,
\]

(3)

where $\|B_k\|_2 \leq M_k$. The solution of this subproblem is the orthogonal projection of $-\nabla f(x_k)/M_k$ on the feasible region of (3). If $M_k$ is large, this solution is close to the null vector. If $s^Q$ is the solution of (3), the approximate solution of (2) is defined in [10] as any feasible increment that satisfies

\[
\nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \leq c_0 [\nabla f(x_k)^T s^Q + \frac{1}{2} M_k \|s^Q\|_2^2],
\]

(4)
where \( c_0 \in (0, 1] \). In particular, \( s^Q \) can be considered an approximate solution of (2).

Conn, Gould, Sartenaer and Toint [3] defined a different trust region approach, where the “easy problem” that precedes the approximate solution of (2) is a linear programming problem. The approach based on (3) has been exploited for bound constrained optimization in [7].

Some authors studied the extension of trust region methods to unconstrained nonsmooth optimization. See [20], [4], [16], among others. In this work we use the iteration function approach for linearly constrained nonsmooth problems which, in the unconstrained case, was introduced in [16]. The iteration function approach was used without connection with trust regions in [13] and [14].

This paper is organized as follows: in Section 2 we present the main model algorithm and we prove that it is well defined. In Section 3 we prove global convergence. In Section 4 we discuss the relation between critical points and Dini-stationary points. In Section 5, we give numerical experiments and in Section 6 we state some conclusions and lines for future research.

2 Main model algorithm

We assume that the feasible region \( \Omega \subset \mathbb{R}^n \) is a convex polyhedron (intersection of a finite number of half-spaces) and that \( f: \Omega \to \mathbb{R} \). Throughout this paper, \( \| \cdot \| \) will denote an arbitrary norm on \( \mathbb{R}^n \). We assume, as in [16], that \( \phi: \Omega \times \mathbb{R}^n \to \mathbb{R} \) is such that

**A1.** For all \( x \in \Omega \), \( \phi(x, 0) = 0 \) and \( \phi(x, \cdot) \) is lower semicontinuous.

**A2.** (subhomogeneity of \( \phi \) in \( s \)). For all \( x \in \Omega \), if \( x + s \in \Omega \), \( t \in [0, 1] \),

\[
\phi(x, ts) \leq t\phi(x, s).
\]
For all $x \in \Omega$, $\Delta > 0$, $M \geq 0$, we define $m(x, \Delta, M) \in \mathbb{R}$ the minimum of

$$\phi(x, s) + \frac{M}{2} \|s\|_2^2 \quad \text{subject to} \quad x + s \in \Omega, \; \|s\| \leq \Delta. \quad (4)$$

By Assumption A1, for all $\Delta > 0$ there exists $s$ such that $x + s \in \Omega$, $\|s\| \leq \Delta$ and

$$\phi(x, s) + \frac{M}{2} \|s\|_2^2 = m(x, \Delta, M).$$

We say that $x \in \Omega$ is a critical point of (1) if there exist $\Delta > 0$, $M \geq 0$ such that $m(x, \Delta, M) \geq 0$.

**Lemma 2.1.** Under Assumptions A1 and A2, $x \in \Omega$ is a critical point of (1) if, and only if, for all $\Delta > 0$, $M \geq 0$,

$$m(x, \Delta, M) \geq 0.$$  

**Proof.** Let us suppose that $\Delta_1 > 0$, $M_1 \geq 0$ and $s_1 \in \mathbb{R}^n$ are such that $x + s_1 \in \Omega$, $\|s_1\| \leq \Delta_1$ and

$$m(x, \Delta_1, M_1) = \phi(x, s_1) + \frac{M_1}{2} \|s_1\|_2^2 < 0.$$

Then $\phi(x, s_1) < 0$.

Given any $\Delta > 0$, $M \geq 0$, we choose $t \in (0, 1)$ such that $\|ts_1\| \leq \Delta$ and

$$\frac{tM}{2} \|s_1\|_2^2 < -\frac{1}{2} \phi(x, s_1).$$

It follows from $x + ts_1 \in \Omega$ and Assumption A2 that

$$m(x, \Delta, M) \leq \phi(x, ts_1) + (M/2) \|ts_1\|_2^2 \leq t\phi(x, s_1) + (t^2M/2) \|s_1\|_2^2 \leq t\phi(x, s_1) - (t/2)\phi(x, s_1) = (t/2)\phi(x, s_1) < 0.$$
So, $m(x, \Delta_1, M_1) < 0$ if, and only if, $m(x, \Delta, M) < 0$ for all $\Delta > 0$, $M \geq 0$. This completes the proof. QED

The trust region algorithm generates a sequence of feasible points $\{x_k\}$ in a way that will be described below. In our approach, only successful iterations are numbered, so that $x_k \neq x_{k+1}$ for all $k$. At the beginning of each iteration we use a trust region radius greater than a fixed constant $\Delta > 0$. This allows us to take the risk of performing large steps when the approximation to the solution is poor. See [5], [10], [7]. Given $x_k \in \Omega$ (the $k$—th approximation to the solution of (1)), the steps for obtaining $x_{k+1}$, or for deciding to stop the algorithm, are the following (the constant 0.1 can be replaced by any other constant in $(0, 1)$):

**Algorithm 2.1.** Choose $\Delta \geq \Delta_1$. Let $B_k \in \mathbb{R}^{n \times n}$ be symmetric and $M_k \geq \|B_k\|_2$.

**Step 1. Solve the easy problem.**

Let $s^Q$ be the solution of

Minimize $Q_k(s) \equiv \phi(x_k, s) + M_k\|s\|_2^2$,

subject to $x_k + s \in \Omega$, $\|s\| \leq \Delta$. \hfill (5)

If $Q_k(s^Q) = 0$, stop.

**Step 2. Approximate solution of trust region subproblem.**

Compute $s^c \in \mathbb{R}^n$ such that

$$
\Phi_k(s^c) \leq 0.1Q_k(s^Q),
$$

$$
x_k + s^c \in \Omega, \|s^c\| \leq \Delta.
$$

where, for all $s \in \mathbb{R}^n$, $\Phi_k(s) = \phi(x_k, s) + \frac{1}{2}s^T B_k s$.

**Step 3. Accept or reject the current step.**

If

$$
f(x_k + s^c) \leq f(x_k) + 0.1\Phi_k(s^c),
$$

5
define $s_k = s^c$, $s^Q_k = s^Q$, $\Delta_k = \Delta$, $x_{k+1} = x_k + s_k$ and return.

If (7) does not hold, choose a new $\Delta$ belonging to $[0.1\Delta, 0.9\Delta]$ and return to Step 1.

**Remarks.** Problem (5) is a proximal point type (see [17], [18]) regularization of the minimization of $\phi(x_k, s)$ on its feasible region. The larger is $M_k$, the simpler and better conditioned this subproblem turns out to be. By Step 2, we see that $s^Q$ satisfies the conditions required for the current step $s^c$. Moreover, by (6), the exact solution of (5) is not necessary as well. The practical consequences of these observations are that, when the minimization of $\Phi_k$ is really hard, for example, at severely nondifferentiable points, we can take advantage of the mild conditions required for $s^c$ and $s^Q$. On the other hand, if minimizing $\Phi_k$ is not very difficult, an accurate solution of the subproblem implicit at Step 2 should improve convergence. Observe that, although $s^Q$ obviously satisfies (6), it is possible that for an increment $s^c$ satisfying (6), we have $Q_k(s^c) >> 0$. So, the convergence theory related to Algorithm 2.1 cannot be reduced to a convergence theory of an algorithm based only on Step 1.

In many well-known trust region methods, after a successful step, the new trust region radius $\Delta$ must be less than or equal to a multiple of $\Delta_k$. For example, in [16],

$$\Delta = \min\{\tilde{\Delta}, c_3\Delta_k\}, \quad (8)$$

where $c_3 > 1$ and $\tilde{\Delta}$ is related to the level set of $f(x_0)$, which is assumed to be bounded. In Algorithm 2.1, instead of an upper bound like (8) for “the new $\tilde{\Delta}$”, we impose the lower bound

$$\Delta \geq \Delta, \quad (9)$$

where $\Delta > 0$ is arbitrary. We have several practical reasons for adopting (9) instead of (8):

(a) A small trust region radius $\Delta_k$ accepted usually follows a small $\Delta'$ rejected. This means that the current model is failing to produce good trial
steps. When this happens it is useful to change the model (for example, through restarting procedures in quasi-Newton methods) instead of using the same type of model with possible poor repeated guesses. Now, when the model changes, there is no reason for linking the corresponding trust region radius to the small trust region radius that reflects failures of the previous ones. In other words, it is reasonable to give new credit to new local approximations of the objective function;

(b) Change (or “discontinuity”) of the model appears in a natural way in the nonsmooth context, as we will see in the minimax examples of Section 4. In these cases quite different iteration functions \( \phi \) can correspond to arbitrarily close points on the domain. So, the necessity of an independent new trust region radius is more apparent in the nonsmooth framework than in the differentiable one;

(c) It must be recalled that the “real” objective of minimization algorithms is to find global minimizers rather than local ones or stationary points. Large trust region radius can give many times the chance to the method of jumping over the attraction basin of a local minimizer;

(d) Allowing the method to try “full Newton steps” is generally recommendable in smooth optimization, even if previous trials were not quite successful. In the nonsmooth framework, we find it necessary to give the chance to locally accurate models to make their guesses without “historical” restrictions to its trust region;

(e) Of course, we can loose some functional evaluations trying large trust regions, when the accepted step is condemned to be very small, but we feel that this loss of efficiency is usually compensated by gains on robustness.

In this section we prove that Algorithm 2.1 is well defined. This means that, if the method does not stop at Step 1, then it finds a new point \( x_{k+1} \) satisfying (7) in finite time. We will need the following additional assumption.
A3'. If \( x + s_j \in \Omega \) for all \( j \in \mathbb{N} \equiv \{0, 1, 2, \ldots \} \) and \( s_j \to 0 \), then

\[
f(x + s_j) - f(x) \leq \phi(x, s_j) + o(\|s_j\|).
\] (10)

**Theorem 2.2.** Under Assumptions A1, A2 and A3', if \( x_k \in \Omega \), and the algorithm does not stop at Step 1, then a new point \( x_{k+1} \in \Omega \) such that \( f(x_{k+1}) < f(x_k) \) is obtained after a finite number of reductions of the trust region radius \( \Delta \).

**Proof.** Assume that \( x_k \) is not a critical point of (1). Then, by Lemma 2.1, there exists \( \bar{\Delta} > 0 \) such that \( m(x_k, \bar{\Delta}, M_k) < 0 \). Since \( \phi \) is lower semicontinuous with respect to the variable \( s \), there exists \( \bar{s} \) such that \( \|\bar{s}\| \leq \bar{\Delta} \), \( x_k + \bar{s} \in \Omega \) and

\[
\phi(x_k, \bar{s}) + \frac{M_k}{2} \|\bar{s}\|^2 = \bar{m} \equiv m(x_k, \bar{\Delta}, M_k) < 0.
\]

Now, by Assumption A2,

\[
\phi(x_k, ts) + \frac{M_k}{2} \|ts\|^2 \leq t[\phi(x_k, \bar{s}) + \frac{M_k}{2} \|\bar{s}\|^2] = t\bar{m}
\] (11)

for all \( t \in [0, 1] \). Assume now that \( \Delta \leq \|\bar{s}\|, t = \Delta/\|\bar{s}\| \leq 1 \), \( s(\Delta) = ts \), so

\[
\|s(\Delta)\| = \Delta.
\] (12)

By (11), we have that

\[
\phi(x_k, s(\Delta)) + \frac{M_k}{2} \|s(\Delta)\|^2 \leq \Delta \frac{\bar{m}}{\|s\|} < 0.
\] (13)

Therefore, for small enough \( \Delta > 0 \), we have that

\[
m(x_k, \Delta, M_k) \leq \Delta \frac{\bar{m}}{\|s\|} < 0.
\] (14)
Let \( s^Q(\Delta) \), \( s^c(\Delta) \) be such that \( \|s^Q(\Delta)\| \leq \Delta \), \( \|s^c(\Delta)\| \leq \Delta \), \( x_k + s^Q(\Delta) \in \Omega \), \( x_k + s^c(\Delta) \in \Omega \), \( Q_k(s^Q(\Delta)) = m(x_k, \Delta, M_k) \) and

\[
\Phi_k(s^c(\Delta)) \leq 0.1Q_k(s^Q(\Delta)). \tag{15}
\]

By (14) and (15), we have that, for small enough \( \Delta \),

\[
\Phi_k(s^c(\Delta)) \leq 0.1\Delta \frac{\bar{m}}{\|s\|} < 0.
\tag{16}
\]

So, writing \( c_1 = 0.1\frac{\bar{m}}{\|s\|} < 0 \), we have that, for \( \Delta > 0 \) small enough,

\[
|\Phi_k(s^c(\Delta))| \geq |c_1|\Delta. \tag{17}
\]

Now, by Assumption A3',

\[
f(x_k + s^c(\Delta)) - f(x_k) \leq \phi(x_k, s^c(\Delta)) + o(\Delta).
\]

So, since \( \frac{1}{2}s^c(\Delta)^T B_k s^c(\Delta) = o(\Delta) \),

\[
f(x_k + s^c(\Delta)) - f(x_k) \leq \phi(x_k, s^c(\Delta)) + \frac{1}{2}s^c(\Delta)^T B_k s^c(\Delta) + o(\Delta)
= \Phi_k(s^c(\Delta)) + o(\Delta).
\]

So, for small enough \( \Delta > 0 \), we have, by (17), that

\[
f(x_k + s^c(\Delta)) - f(x_k) \leq \Phi_k(s^c(\Delta)) + |c_1|\Delta/2
\leq \Phi_k(s^c(\Delta)) + |\Phi_k(s^c(\Delta))|/2 \leq \Phi_k(s^c(\Delta))/2.
\]

This implies that, for small enough \( \Delta > 0 \), the condition (7) must be satisfied. Thus, the desired result is proved. \textbf{Q.E.D}

\textbf{Remark.} It is easy to verify that, in Theorem 2.2, only the fact that \( \Omega \) is convex was used, instead of its polyhedral characteristics. So, the algorithm is well defined even in that general case. As we will see in the
following section, for proving convergence we are going to use a property of the intersection of half-spaces that is shared by many other (but not by all) convex sets (see Lemma 3.1 below). The case where $\Omega$ is an Euclidean ball has interesting applications in the smooth context (see [10]).

3 Global convergence

For proving global convergence of Algorithm 2.1, we need three additional assumptions. Assumption A3 is a stronger version of Assumption A3’.

A3. If $x_j + s_j \in \Omega$ for all $j \in \mathbb{N}$, $\{x_j\}$ is convergent and $s_j \to 0$, then

$$f(x_j + s_j) - f(x_j) \leq \phi(x_j, s_j) + o(\|s_j\|).$$  \hfill (18)

A4. There exists $\bar{\Delta} > 0$ such that, for all $\|s\| \leq \bar{\Delta}$, $-\phi(\cdot, s)$ is lower semicontinuous.

A5. $f$ is lower semicontinuous or bounded below in $\Omega$.

By Theorem 2.2, we have that Algorithm 2.1 stops in a finite number of iterations if, and only if, it finds a critical point. If the algorithm does not stop, then it generates an infinite set of iterates $x_k \in \Omega$ such that $f(x_{k+1}) < f(x_k)$ for all $k \in \mathbb{N}$. We are going to prove that all limit points of the generated sequence are critical. First, we need to prove an auxiliary lemma.

Lemma 3.1. Assume that $\{x_j\}_{j \in \mathbb{N}}$ is a sequence in $\Omega$ that converges to $x$. Suppose that $s \in \mathbb{R}^n$ is such that $x + s \in \Omega$. Then, there exists $j_0 \in \mathbb{N}$ such
that \( x_j + s/2 \in \Omega \) for all \( j \geq j_0 \).

**Proof.** \( \Omega \) is the intersection of a finite number of half-spaces \( H_1, \ldots, H_m \). For \( i = 1, \ldots, m \), let \( a_i \in \mathbb{R}^n \), \( b_i \in \mathbb{R} \) be such that

\[
H_i = \{ x \in \mathbb{R}^n \mid a_i^T x \leq b_i \}. \tag{19}
\]

We define

\[
V_i = \{ x \in \mathbb{R}^n \mid a_i^T x = b_i \}. \tag{20}
\]

and

\[
N_i = \{ s \in \mathbb{R}^n \mid a_i^T s = 0 \}. \tag{21}
\]

Let us analyze the following cases:

(a) \( x \in V_i \);

(b) \( x \in H_i - V_i \).

If (a) holds, we have that \( a_i^T x = b_i \) and \( a_i^T (x + s) \leq b_i \). So, \( a_i^T s \leq 0 \). Since, for all \( j \in \mathbb{N} \), \( a_i^T x_j \leq b_i \), we deduce that \( a_i^T (x_j + s) \leq b_i \) for all \( j \in \mathbb{N} \). By convexity, \( x_j + s/2 \in H_i \) for all \( j \in \mathbb{N} \).

If (b) holds, we have that \( a_i^T x < b_i \) and \( a_i^T (x + s) \leq b_i \). Therefore, \( a_i^T (x + s/2) < b_i \). Since \( x_j + s/2 \to x + s/2 \), this implies that there exists \( j_i \in \mathbb{N} \) such that

\[
a_i^T (x_j + s/2) < b_i \text{ for all } j \geq j_i.
\]

We proved that, for all \( i = 1, \ldots, m \), there exists \( j_i \in \mathbb{N} \) such that \( x_j + s/2 \in H_i \) for all \( j \geq j_i \). Defining

\[
\text{j}_0 = \max \{ j_1, \ldots, j_m \},
\]

we obtain the desired result. **QED**

**Remark.** Lemma 3.1 is not true for an arbitrary closed and convex \( \Omega \), as the following example, due to R. H. Bielschowsky, shows. Consider

\[
C = \{ z \in \mathbb{R}^3 \text{ such that } z_3 = 0 \text{ and } z_1^2 + z_2^2 \leq 1 \}
\]

11
and let $\Omega$ be the convex hull of $C \cup \{(0,1,1)^T\}$. Define $x = (0,1,0)^T, s = (0,0,1)^T$. We consider the sequence $x_j = (\cos \theta_j, \sin \theta_j, 0)$, where $\theta_j = \pi/2 - 1/j$ for all $j \in \mathbb{N}$. Then, $x$ and $x + s$ belong to $\Omega$ but $x_j + ts \notin \Omega$ for all $t > 0$.

**Theorem 3.2.** Under Assumptions A1–A5, if $x_*$ is a limit point of the infinite sequence $\{x_k\}_{k \in \mathbb{N}}$ generated by Algorithm 2.1, and $\{M_k\}_{k \in \mathbb{N}}$ is bounded, then $x_*$ is a critical point.

**Proof.** Let $K_1$ be an infinite subset of $\mathbb{N}$ such that

$$\lim_{k \in K_1} x_k = x_*.$$  \hfill (22)

Since $\Omega$ is closed, $f(x_k)$ is decreasing and $f$ is lower semicontinuous or bounded below, we have that there exists $L \in \mathbb{R}$ such that

$$\lim_{k \in K_1} f(x_k) \geq L,$$  \hfill (23)

where $L = f(x_*)$ or $L$ is a lower bound of $f$ in $\Omega$. Since $f(x_{k+1}) \leq f(x_k)$ for all $k \in \mathbb{N}$ this implies that

$$\lim_{k \to \infty} f(x_k) - f(x_{k+1}) = 0.$$  \hfill (24)

By (5), (6), (7) and (24) we have that

$$\lim_{k \to \infty} m(x_k, \Delta_k, M_k) = \lim_{k \to \infty} Q_k(\delta_k) = 0.$$  \hfill (25)

Suppose that $x_*$ is not a critical point. So, by Lemma 2.1,

$$\beta \equiv m(x_*, \bar{\Delta}, 0) < 0.$$  \hfill (26)

where $\bar{\Delta}$ is defined in Assumption A4. Let $\bar{s}$ be such that $x_* + \bar{s} \in \Omega, \|\bar{s}\| \leq \bar{\Delta}$ and

$$\phi(x_*, \bar{s}) = m(x_*, \bar{\Delta}, 0) = \beta < 0.$$
Therefore, 
\[ m(x_*, \|\bar{s}\|, 0) = \phi(x_*, \bar{s}) = \beta < 0. \]

So, by Assumption A2, 
\[ \phi(x_*, \bar{s}/2) \leq \beta/2 < 0. \]

Now, by Lemma 3.1, there exists \( k_0 \in \mathbb{N} \) such that \( x_k + \bar{s}/2 \in \Omega \) for all \( k \in K_1, k \geq k_0 \). Since \( x_k + \bar{s}/2 \) converges to \( x_* + \bar{s}/2 \), Assumption A4 implies that there exists \( k_1 \geq k_0 \) such that for all \( k \in K_1, k \geq k_1 \), \( \Delta \leq \|\bar{s}\|/2 \),

\[
m(x_k, \Delta, 0) \leq \phi(x_k, \Delta, \|\bar{s}/2\|/2) \leq \frac{\Delta}{\|\bar{s}/2\|} \phi(x_k, \bar{s}/2) \leq \frac{\Delta}{\|\bar{s}/2\|} \beta = \frac{\beta}{4} = \bar{\varepsilon} \Delta,
\]
where
\[
\bar{\varepsilon} = \frac{\beta}{2 \|\bar{s}\|} < 0.
\]

But, clearly, if \( M_k \leq M \) for all \( k \in K_1, k \geq k_1 \), we have:

\[
m(x_k, \Delta, M_k) \leq m(x_k, \Delta, 0) + \frac{M_k}{2} \Delta^2 \\
\leq m(x_k, \Delta, 0) + \frac{M}{2} \Delta^2 \leq \Delta (\bar{\varepsilon} + M \Delta/2).
\]

So, if \( \Delta \leq \bar{\Delta} = \min\{\|\bar{s}\|/2, |\bar{\varepsilon}|/M\} \), we have that

\[
m(x_k, \Delta, M_k) \leq \Delta \bar{\varepsilon}/2 < 0, \tag{27}
\]
while, if \( \Delta > \bar{\Delta} \),

\[
m(x_k, \Delta, M_k) \leq \bar{\Delta} \bar{\varepsilon}/2 < 0. \tag{28}
\]
By (25), (27) and (28), we have that
\[
\lim_{k \in K_1} \Delta_k = 0. \tag{29}
\]

Therefore, there exists \( k_2 \in \mathbb{N} \) such that \( \Delta_k \leq \Delta \) for all \( k \in K_1, k \geq k_2 \). This implies that for all \( k \in K_1, k \geq k_2 \), \( \Delta_k \) cannot be the first trust region radius tried at iteration \( k \). That is, \( \Delta_k \) was necessarily preceded by a trust region radius \( \bar{\Delta}_k \) such that \( \bar{\Delta}_k \leq 10\Delta_k \) and that the test (7) was not satisfied when \( \Delta = \bar{\Delta}_k \). Let \( k_3 \geq k_2 \) such that \( \bar{\Delta}_k \leq \Delta \) for all \( k \geq k_3 \), \( k \in K_1 \). Then, by (27),
\[
m(x_k, \bar{\Delta}_k, M_k) \leq \bar{\Delta}_k \bar{c}/2 < 0,
\tag{30}
\]
for all \( k \in K_1, k \geq k_3 \). Therefore, calling \( s^c(\bar{\Delta}_k) \) the step \( s^c \) computed at Step 2 of Algorithm 2.1 when \( \Delta = \bar{\Delta}_k \), and \( c' = 0.05\bar{c} \), we have that \( \Phi_k(s^c(\bar{\Delta}_k)) \leq c' \bar{\Delta}_k < 0 \) and, thus,
\[
|\Phi_k(s^c(\bar{\Delta}_k))| \geq |c'| \bar{\Delta}_k > 0 \tag{31}
\]
for all \( k \in K_1, k \geq k_3 \).

Now, by Assumption A3,
\[
f(x_k + s^c(\bar{\Delta}_k)) - f(x_k) \leq \phi(x_k, s^c(\bar{\Delta}_k)) + o(\bar{\Delta}_k).
\]
But, by the boundedness of \( \|B_k\|_2, \frac{1}{2}s^c(\bar{\Delta}_k)^T B_k s^c(\bar{\Delta}_k) = o(\bar{\Delta}_k) \), so
\[
f(x_k + s^c(\bar{\Delta}_k)) - f(x_k) \leq \phi(x_k, s^c(\bar{\Delta}_k)) + \frac{1}{2}s^c(\bar{\Delta}_k)^T B_k s^c(\bar{\Delta}_k) + o(\bar{\Delta}_k)
\]
\[
= \Phi_k(s^c(\bar{\Delta}_k)) + o(\bar{\Delta}_k).
\]
By (31), there exists \( k_4 \in \mathbb{N} \) such that for all \( k \in K_1, k \geq k_4 \),
\[
f(x_k + s^c(\bar{\Delta}_k)) - f(x_k) \leq \Phi_k(s^c(\bar{\Delta}_k)) + |c'| \bar{\Delta}_k / 2
\]
\[
\leq \Phi_k(s^c(\bar{\Delta}_k)) + |\Phi_k(s^c(\bar{\Delta}_k))|/2 \leq \Phi_k(s^c(\bar{\Delta}_k))/2
\]
So, for all \( k \in K_1 \) such that \( k \geq k_4 \), the trust region radius \( \bar{\Delta}_k \) is accepted, contradicting our previous assumption. This completes the proof. QED
4 Relations with optimality

Let us show that any local minimizer of (1) is a critical point.

**Theorem 4.1.** Under Assumptions A2 and A3', every local minimizer of (1) is a critical point.

**Proof.** Assume that $x \in \Omega$ is not a critical point of (1). Then, there exists $s \in \mathbb{R}^n$, such that $x + s \in \Omega$ and $\phi(x, s) < 0$. Then, by Assumptions A2 and A3', we have, for all $1 > t \downarrow 0$:

$$f(x + ts) - f(x) \leq \phi(x, ts) + o(ts)$$

$$\leq t\phi(x, s) + o(ts).$$

So,

$$\frac{f(x + ts) - f(x)}{t} \leq \phi(x, s) + \frac{o(ts)}{t}$$

This implies that for all $t$ small enough,

$$\frac{f(x + ts) - f(x)}{t} \leq \frac{\phi(x, s)}{2} < 0.$$

Therefore, $x$ is not a local minimizer. **QED**

A simple example shows that, without further assumptions on $\phi$, the set of critical points of (1) can be much larger than the set of local minimizers. In fact assume that $f(x) = x^2/2$, $\Omega = [0, 1]$, and $\phi(x, s) = |s|$ for all $x, s$. It is easy to verify that Assumptions A1 – A5 hold in this case. However, $x$ is a critical point of (1) for all $x \in \Omega$. The next assumption, following the lines of [16], restricts the set of critical points in a way that it resembles better the set of local minimizers of the problem.
**A0.** Let us assume that, for all \( x \in \Omega, s \in \mathbb{R}^n \) such that \( x + s \in \Omega \), the upper Dini directional derivative of \( f \) at \( x \) in the direction \( s \), given by

\[
 f^+(x; s) = \limsup_{t \downarrow 0} \frac{f(x + ts) - f(x)}{t},
\]

exists and that there exists \( \theta > 0 \) such that

\[
 \liminf_{t \downarrow 0} \frac{\phi(x, ts)}{t} \leq \theta f^+(x; s). \tag{32}
\]

A point \( x \in \Omega \) is a *Dini stationary point* of \((1)\) if for all \( x \in \Omega, x + s \in \Omega \),

\[
 f^+(x; s) \geq 0.
\]

In smooth optimization Dini stationary points coincide with first-order stationary points. The following theorem shows that, with Assumption A0, critical points of \((1)\) are Dini stationary points.

**Theorem 4.2.** *Under Assumption A0, every critical point of \((1)\) is a Dini stationary point.*

**Proof.** Let \( x \in \Omega \) be a critical point of \((1)\) and let \( s \in \mathbb{R}^n \) such that \( x + s \in \Omega \). Since \( x \) is a critical point, we have that \( m(x, \Delta, 0) \geq 0 \) for all \( \Delta > 0 \). This implies that, for all \( t > 0 \),

\[
 \phi(x, ts) \geq 0.
\]

Therefore,

\[
 \liminf_{t \downarrow 0} \frac{\phi(x, ts)}{t} \geq 0.
\]

So, by (32), \( f^+(x; s) \geq 0 \). This completes the proof. **Q.E.D.**
5 Numerical experiments

The transmission $T(\lambda, s, d, \bar{n}, \alpha)$ of a thin absorbing film on a transparent substrate (see [19]) is given by the following formulae:

$$T(\lambda, s, d, \bar{n}, \alpha) = \frac{A\bar{x}}{B - C\bar{x} + D\bar{x}^2} \quad (33)$$

where

$$A = 16\bar{s}(\bar{n}^2 + \bar{k}^2) \quad (34)$$

$$B = [(\bar{n} + 1)^2 + \bar{k}^2][(\bar{n} + 1)(\bar{s}^2 + \bar{k}^2) + \bar{k}^2] \quad (35)$$

$$C = [(\bar{n}^2 - 1 + \bar{k}^2)(\bar{n}^2 - \bar{s}^2 + \bar{k}^2) - 2\bar{k}^2(\bar{s}^2 + 1)]2 \cos \varphi$$

$$- \bar{k}[2(\bar{n}^2 - \bar{s}^2 + \bar{k}^2) + (\bar{s}^2 + 1)(\bar{n}^2 - 1 + \bar{k}^2)]2 \sin \varphi \quad (36)$$

$$D = [(\bar{n} - 1)^2 + \bar{k}^2][(\bar{n} - 1)(\bar{s}^2 + \bar{k}^2) + \bar{k}^2] \quad (37)$$

$$\varphi = 4\pi\bar{n}d/\lambda, \quad \bar{x} = \exp(-\alpha d), \quad \alpha = 4\pi\bar{k}/\lambda. \quad (38)$$

In formulae (34)–(38) the following notation is used:

(a) $\lambda$ is the wavelength;

(b) $\bar{s}$ is the refractive index of the substrate, assumed to be constant for all $\lambda$;

(c) $\bar{n} = \bar{n}(\lambda)$ is the refractive index of the film;

(d) $\alpha = \alpha(\lambda)$ is the absorption index of the film;

(e) $d$ is the thickness of the film.

We assume that $d$ and $\bar{s}$ are known, a set of experimental data $(\lambda_i, T_i), i = 1, \ldots, N, \ (\lambda_i < \lambda_{i+1}$ for all $i = 1, \ldots, N), \ are \ given \ and \ we \ wish \ to \ estimate, \ \bar{n}(\lambda) \ and \ \alpha(\lambda)$.

We established explicitly the physical constraints that must be satisfied by $\bar{n}$ and $\alpha$. Writing, from now on, $\bar{n}_i = \bar{n}(\lambda_i), \alpha_i = \alpha(\lambda_i)$ in order to simplify the notation, the constraints are:

(a) $\alpha_i \geq 0, \ \bar{n}_i \geq 1$ for all $i = 1, \ldots, N$.

(b) The functions $\alpha(\lambda)$ and $\bar{n}(\lambda)$ must be nonincreasing.
(c) The functions $\alpha(\lambda)$ and $\bar{n}(\lambda)$ must be convex ($\alpha''(\lambda) \geq 0$ and $\bar{n}''(\lambda) \geq 0$ for all $\lambda$).

By the considerations above, we decided to model the problem of determining $\bar{n}$ and $\alpha$ by the following optimization problem:

\[
\text{Minimize } \max_{i=1, \ldots, N} |T(\lambda_i, s, d, \bar{n}_i, \alpha_i) - T_i| \quad (39)
\]

subject to

\[
\alpha_i \geq 0, \bar{n}_i \geq 0 \text{ for all } i = 1, \ldots, N; \quad (40)
\]

\[
\alpha_{i+1} \leq \alpha_i \text{ and } \bar{n}_{i+1} \leq \bar{n}_i \text{ for all } i = 1, \ldots, N - 1; \quad (41)
\]

\[
\bar{n}_i \leq \bar{n}_{i-1} + \frac{\bar{n}_{i+1} - \bar{n}_{i-1}}{\lambda_{i+1} - \lambda_{i-1}}(\lambda_i - \lambda_{i-1}), \text{ and } \alpha_i \leq \alpha_{i-1} + \frac{\alpha_{i+1} - \alpha_{i-1}}{\lambda_{i+1} - \lambda_{i-1}}(\lambda_i - \lambda_{i-1}) \quad (42)
\]

for all $i = 2, \ldots, N - 1$. In our experiments we used $N = 30$, so, the optimization problem has 60 variables. With a different objective function, this was called the \textit{pointwise constrained optimization} approach for the resolution of the estimation problem in [1].

For a given value of the thickness $d$, the constraints (40)-(42) define a polyhedron $\Omega$ with nonempty interior in $\mathbb{R}^{2N}$.

The objective function of (39)-(42) is of minimax type, that is:

\[
f(x) = \max\{f_i(x), i = 1, \ldots, 2N\}, \quad (43)
\]

where, for all $j = 1, \ldots, N$,

\[
f_{2j-1}(x) = T(\lambda_j, s, d, \bar{n}_j, \alpha_j) - T_j
\]

and

\[
f_{2j}(x) = -(T(\lambda_j, s, d, \bar{n}_j, \alpha_j) - T_j).
\]

Following [16], we define, for this problem,
\[
\phi(x, s) = \max_{i \in I(x)} \left\{ f_i(x) + \nabla f_i(x)^T s \right\} - f(x),
\]

where
\[
I(x) = \{ i \in \{1, \ldots, 2N\} \mid f(x) - f_i(x) < p \}. \tag{45}
\]

Therefore, problem (5) at Step 1 of Algorithm 2.1 takes the form

Minimize \( z + \frac{1}{2} M_k \| s \|_2^2 \)

subject to \( x_k + s \in \Omega \),

\[
z \geq f_i(x_k) + \nabla f_i(x_k)^T s, \text{ for all } i \in I(x_k),
\]

\[
\| s \|_\infty \leq \Delta. \tag{46}
\]

The quadratic programming problem (46) will not be very difficult because of the regularization term \( (1/2) M_k \| s \|_2^2 \).

At Step 2 of Algorithm 2.1 we must solve, approximately,

Minimize \( z + \frac{1}{2} s^T B_k s \)

subject to

\[
x_k + s \in \Omega,
\]

\[
z \geq f_i(x_k) + \nabla f_i(x_k)^T s \text{ for all } i \in I(x_k),
\]

\[
\| s \|_\infty \leq \Delta. \tag{47}
\]

However, the criterion (7) for acceptance of an approximate solution of (47) is very mild since even the initial approximation \( s^Q \) (solution of (46)) satisfies it.

In our numerical experiments, we used the subroutine EKKQSLV of [9] for solving the quadratic programs (46) and (47). We use the solution \( s^Q \) of (46) as initial estimate for the resolution of (47). In these experiments,
we used $B_k \equiv 0$, $\Delta = 1$, and $M_k = 1000$ for all $k$. The initial $\Delta$ at each iteration was $\Delta$. The "new $\Delta$" at Step 3 of Algorithm 2.1 was $\Delta/8$ if $\Delta = 1$ and $\Delta/2$ otherwise. Since we only need an approximate solution of (47) satisfying very mild conditions, we stopped the execution of OSL for solving (47) when the default stopping criterion (see [9]) is satisfied or when the number of iterations exceeded 10. We stopped the execution of Algorithm 2.1 when

$$f(x_k) - f(x_{k+1}) \leq 0.01f(x_k).$$

We are interested in comparing the behavior of the algorithm for different choices of $p$ in (45). The size of $p$ determines the number of elements of $I(x)$, and, so, the number of constraints of the quadratic programs (46) and (47). So, the steps of the main algorithm are more expensive when $p$ is large. The data $T_i$ and the computed values of $T(\lambda_i, s, d, \bar{n}_i, \alpha_i)$ are adimensional quantities between 0 and 1. Therefore, for all feasible $x$ and for all $i$ we have that $f_i(x) \leq 1$. So, if $p \geq 1$ we necessarily have that $I(x) = \{1, \ldots, 2N\}$. On the other hand, if we defined $I(x)$ as the set of indices such that $f_i(x) = f(x)$ (which corresponds to $p \to 0$), the algorithm can converge to a nondifferentiable and noncritical point. It can be conjectured that, for $p$ very small, difficulties can be encountered by the method to get away from nondifferentiable points. Therefore, there exists a conflict between global convergence and expensiveness of the iteration and we want to study experimentally if this contradiction appears in practice.

Our experiments were generated as follows: for $\lambda_i = 866 + 10(i - 1)$, $i = 1, \ldots, 30$ we generated transmission data using the formulae (33)-(38) with $d = 1150$, $\bar{s} = 1.51$,

$$n_i^* = n^*(\lambda_i) = 1.5\exp\left(-\frac{\lambda_i - 866}{200}\right) + 3.5$$

(48)

and

$$\alpha_i^* = \alpha^*(\lambda_i) = 4.0\exp\left(-\frac{\lambda_i - 866}{200}\right)$$

(49)
Using the generated transmission curve with noisy errors randomly generated between $qT_i$ and $-qT_i$, we solved (39)-(42). We use these artificial perturbations for simulating different types of experimental errors. In practice, it is important to obtain robustness of the algorithms with respect to these perturbations.

In all the tests, we used the following initial guesses for $\bar{n}$ and $\alpha$:

$$\bar{n}_0(\lambda_i) = -0.0009\lambda_i + 5.8$$  \hspace{1cm} (50)

and

$$\alpha_0(\lambda_i) = -0.008\lambda_i + 12$$  \hspace{1cm} (51)

All the experiments were run using a Pentium 66 MHz computer, double precision and the Fortran 77 compiler. The results are given in Table 1 below. For each value of $p$ and $q$ we report (a) error, the final value of the objective function of (39)-(42), (b) kon, the number of iterations performed, (c) nfe, the number of functional evaluations and (d) time, the computer time, in seconds. The overall performance of the method for this problem is very good, when compared with previous attempts for solving this type of problems. In particular, the use of the objective function (39) is better than the use of the sum of squares, which is more usual. The reasons for this effectiveness can be explained in terms of the structure of the problem. In fact, for $p$ small, the algorithm works actively reducing one or two terms of the form $|T(\lambda_i, s, d, \bar{n}_i, \alpha_i) - T_i|$. Only two variables are involved on each term of this type. However, the decrease and convexity constraints (which imply “continuity”) automatically cause an improvement of the remaining variables. Since these variables are not controlled directly, the iteration is cheap but the error on practically all the components is reduced. We performed additional experiments “jumping” Step 2 of the algorithm, with completely different practical results. In fact, in all the cases where Step 2 was omitted convergence did not occur after 1000 iterations and, in fact, the progress on the objective function value was negligible.
\[ f_i(x) = \frac{1}{2}(x - 1)^T A_i(x - 1), \quad (52) \]

Table 1: Numerical experiments: optical problem

We performed a second set of experiments using problems of type (43). The functions \( f_i \) were defined as
with \( x \in \mathbb{R}^{10} \), \( i = 1, \ldots, 10 \), \( f_{11}(x) \equiv 0 \) and \( \mathbf{1} \equiv (1, \ldots, 1)^T \in \mathbb{R}^{10} \). The feasible set \( \Omega \) is the box \([1, 11]^{10}\). Therefore the practical problem is to find a feasible point of

\[
f_i(x) \leq 0, \quad i = 1, \ldots, 10, \quad x \in \Omega
\]

and

\[
f(x) = \max \{ f_i(x), i = 1, \ldots, 11 \}
\]

for all \( x \in \Omega \). The matrices \( A_i \) where defined as \( A_i = I + A'_i \) where each entry of \( A'_i \) is a random number uniformly generated between \(-5\) and \(5\). These matrices are, almost always, indefinite. Since \( \mathbf{1} \) satisfies (53), we know that \( f(x) = 0 \) at global minimizers of \( f \) on \( \Omega \).

The objective of these easily generated experiments was to test different definitions for the matrix \( B_k \), used at Step 2 of Algorithm 2.1. The following alternatives were tested:

(a) \( B_k = 1000I \); in this case, defining \( M_k = 1000 \), Step 2 of the algorithm is not executed at all;

(b) \( B_k = 0 \);

(c) \( B_k = A_{j_k} \) where

\[
j_k = \operatorname{Argmax} \{ f_i(x_k), i = 1, \ldots, 10 \};
\]

(d) \( B_0 = I \) and \( B_{k+1} = B_k + \Delta B_k \), where \( \Delta B_k \) is a rank-one quasi-Newton correction of \( B_k \).

For computing \( B_{k+1} \) in (d) we used the classical symmetric rank-one correction formula of smooth optimization (see, for example [6]). For all \( k = 0, 1, 2, \ldots \), the following vectors were defined

\[
s_k = x_{k+1} - x_k, \quad y_k = \nabla f_{j_k+1}(x_{k+1}) - \nabla f_{j_k}(x_k), \quad v_k = y_k - B_k s_k.
\]

If \( |v_k^T s_k| \geq \max \{0.001\|v_k\|_2\|s_k\|_2, 10^{-6}\} \), we define \( \sigma_k = v_k^T s_k \). Otherwise, we take \( \sigma_k = \operatorname{sgn}(v_k^T s_k) \max \{0.001\|v_k\|_2\|s_k\|_2, 10^{-6}\} \). Finally, the new up-
dated matrix is computed as

\[ B_{k+1} = B_k + \frac{v_k v_k^T}{\sigma_k}. \]

Therefore, the “secant equation” \( B_{k+1} s_k = y_k \) is satisfied, provided that 
\( |v_k^T s_k| \) is not very small.

In all our tests we chose \( x_0 = (6, \ldots, 6)^T \). This initial point was never a solution of the problem under consideration. Convergence to a global solution was declared when \( f(x_k) = 0 \). The executions were also interrupted when 100 iterations or 300 functional evaluations were performed.

The numerical results are surveyed in Table 2. For each experiment such that convergence to a global solution was detected, we report the pair \( kon, nfe \), where \( kon \) is the number of iterations performed and \( nfe \) is the number of functional evaluations. In the cases where convergence was not achieved, we also report, between parentheses, the final value of the objective function.

Let us briefly describe the main implementation features of Algorithm 2.1 for this family of problems.

(i) The “easy problem” at Step 1 is a convex quadratic program with linear constraints. It was reduced to a bound constrained minimization problem using the transformation described in [8] and solved using the method introduced in [7] (called BOX from here on).

(ii) The computation of \( s^c \) at Step 2 of the algorithm is also associated to a (generally nonconvex) quadratic minimization problem. We computed a very rough estimate of the minimizer of this subproblem, defining an associated bound constrained minimization problem as in [8] and applying BOX for finding an approximate solution. Since this process does not guarantee convergence to an appropriate solution (the quadratic is not convex) we stopped the executions of BOX when 10 iterations were completed or when the progress of one iteration was less than 0.01\( \times \) the maximum progress obtained at previous iterations. If the output of BOX, obtained in this way, satisfies (6), it is accepted as “trial increment” \( s^c \). Otherwise we define
$s^e = s^Q$ (this never happens in the experiments). Of course, this procedure is far from being optimal, but it will be useful to illustrate the potentiality of using a very rough approximate minimizer of $\Phi$ at Step 2 of the algorithm.

(iii) At Step 1, we always used $M_k = 1000$. This bound was adequate even for the choice (d) of $B_k$ in all cases. Other meaningful parameters were: $\Delta = 0.01$, $p = 0.1$. We began iteration 0 with $\Delta = 1$. The new $\Delta$ for $k = 1, 2, \ldots$ after a successful step was chosen as $\max\{4\Delta_k, \Delta\}$.

These experiments seem to confirm that, although Step 1 alone is sufficient to guarantee convergence of the method, executions that “jump” Step 2 are not efficient. In fact, in nearly all cases, (a) was the worst choice of $B_k$. The choice (c) (“true Hessian”) was better than (b) in most cases and, surprisingly, the symmetric rank-one correction of $B_k$ defined the best algorithm, in terms of iterations and functional evaluations.

<table>
<thead>
<tr>
<th>Problem</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56, 167</td>
<td>15, 51</td>
<td>13, 43</td>
<td>7, 23</td>
</tr>
<tr>
<td>2</td>
<td>80, 305 (343)</td>
<td>12, 40</td>
<td>9, 31</td>
<td>9, 27</td>
</tr>
<tr>
<td>3</td>
<td>95, 303</td>
<td>13, 44</td>
<td>21, 69</td>
<td>9, 24</td>
</tr>
<tr>
<td>4</td>
<td>31, 97</td>
<td>81, 301 (363)</td>
<td>15, 51</td>
<td>12, 38</td>
</tr>
<tr>
<td>5</td>
<td>87, 304</td>
<td>7, 23</td>
<td>6, 22</td>
<td>6, 19</td>
</tr>
<tr>
<td>6</td>
<td>44, 140</td>
<td>100, 276 (166)</td>
<td>16, 55</td>
<td>6, 20</td>
</tr>
<tr>
<td>7</td>
<td>68, 301 (306)</td>
<td>9, 30</td>
<td>7, 24</td>
<td>4, 10</td>
</tr>
<tr>
<td>8</td>
<td>67, 200</td>
<td>62, 315 (74)</td>
<td>8, 28</td>
<td>9, 24</td>
</tr>
<tr>
<td>9</td>
<td>75, 300 (110)</td>
<td>75, 300 (136)</td>
<td>15, 52</td>
<td>12, 38</td>
</tr>
<tr>
<td>10</td>
<td>35, 116</td>
<td>11, 36</td>
<td>8, 27</td>
<td>6, 18</td>
</tr>
</tbody>
</table>

Table 2: Numerical experiments: feasibility problems

25
6 Final remarks

One main concern in this paper was to define trust region subproblems for constrained nonsmooth problems with a similar complexity to the corresponding smooth situations. In the linearly constrained case, this objective is achieved, at least in the very important family of minimax problems. In fact, the subproblems associated to the use of the infinity norm have only a few constraints more than those associated to a differentiable objective function, when $p$ is small. An obvious consequence of this fact is that, if we have a practical linearly constrained problem where the natural objective function $f$ is nonsmooth, it is worthwhile to try to find an iteration function $\phi(x, s)$ satisfying the Assumptions, instead of replacing $f$ by a less natural smooth one. In [16], it is proved that, for many interesting problems, suitable iteration functions can be defined.

The subproblem defined in [16] for the unconstrained case does not require and exact solution, but, on the other hand, does not provide a criterion for deciding if a sufficiently good approximate solution has been achieved. Our suggestion in this paper (following [10] and [7]) seems to be satisfactory from the computational point of view. We solve first an easy problem which turns out to be a regularization of the minimization of $\phi(x, s)$. (In fact, the easy problem also needs only an approximate solution in the sense of ([16]).) Then, we use the solution of the easy problem to define the stopping criterion for the quadratic programming algorithm that solves the main subproblem. The solution of the easy problem plays the role of the approximate Cauchy point of other trust region algorithms for large scale smooth problems (see, for example [2]). However, the differentiability of the objective function is essential for defining the Cauchy point, which belongs to the projected half-line determined by the negative gradient. On the other hand, the generalization of the easy problem to the nonsmooth case is quite natural.

We performed numerical experiments using a real-life physical problem.
For this problem, the difference between performances of the algorithm with different $p$ was not impressive. On one hand, small values of $p$ produced, in general, more economic executions in terms of iterations and evaluations but, on the other hand, smaller values of $f$ were obtained with $p \approx 1$.

The flexibility of our approach allows us to solve problems with different strategies for defining the bounds $M_k$ and with different stopping criteria for the subproblems. However, we are aware that practitioners want to be provided with good default parameteres that should produce the best results, in most cases. This will be the subject of most experimental research in the near future. It is also interesting to investigate the best possible choices of $B_k$ with the aim of incorporating second order information in an efficient way. The striking results obtained with the rank-one correction formula on the second set of experiments suggest that the role of quasi-Newton approximations in nonsmooth optimization can be quite different from the role played by these updates in the differentiable case.

From a strictly theoretical point of view, other questions remain to be investigated. Our convergence proof does not use upper bounds for the quotient between consecutive trust region radius (see the remark on Section 2) but uses a lower bound $\Delta > 0$. It should be interesting to verify if every limit point is critical when neither the upper bound nor the lower bound hypotheses are assumed. Moreover, under the "upper bound assumption" and other suitable hypotheses, it can be proved for many trust region methods that there exists a critical limit point admitting a not very fast growth of $\|B_k\|$. See [15], [20], [16]. It remains to be investigated if, with our approach, a limited growth of $M_k$ can also be admitted.

**Acknowledgements**

We are indebted to two anonymous referees for useful comments. In particular, the proof of Lemma 2.1 was considerably simplified following a referee’s suggestion. We also acknowledge Silvia Jamesch for correcting several mistakes and R. H. Bielschowsky for the conical counterexample on
Section 3.

References


