Convergence of linesearch and trust-region methods using the Kurdyka-Łojasiewicz inequality

Dominikus Noll and Aude Rondepierre

Abstract We discuss backtracking linesearch and trust-region descent algorithms for unconstrained optimization and prove convergence to a critical point if the objective is of class C^1 and satisfies the Kurdyka-Łojasiewicz condition. For linesearch we investigate in which way an intelligent management memorizing the stepsize should be organized. For trust-regions we present a new curvature based acceptance test which ensures convergence under rather weak assumptions.

Key words: Nonlinear optimization, descent method, Kurdyka-Łojasiewicz inequality, linesearch, backtracking, memorized steplength, trust-region.

1 Introduction

Global convergence for linesearch descent methods traditionally only assures subsequence convergence to critical points (see e.g. [4, Proposition 1.2.1] or [13, Theorem 3.2]), while convergence of the entire sequence of iterates is not guaranteed. Similarly, subsequence convergence in trust-region methods is establish by relating the progress of trial points to the minimal progress achieved by the Cauchy point. These results are usual proved for $C^{1,1}$ or C^2-functions, see [8, Theorem 6.4.6] or [13, Theorem 4.8].

Recently Absil *et al.* [1] proved convergence of iterates of descent methods to a single limit-point for analytic objective functions, using the fact that this class satisfies the so called Łojasiewicz inequality [11, 12]. Here we prove convergence of

Dominikus Noll
Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France, e-mail: Dominikus.Noll@math.univ-toulouse.fr

Aude Rondepierre
Institut de Mathématiques de Toulouse, INSA de Toulouse, 135 avenue de Rangueil 31077 Toulouse, France, e-mail: Aude.Rondepierre@math.univ-toulouse.fr

linesearch and trust-region descent methods to a single critical point for C^1 functions satisfying the Kurdyka-Łojasiewicz (KL) inequality [10], a generalization of the Łojasiewicz inequality. This is motivated by recent convergence results based on this condition in other fields, see e.g. [2], [5], [6, 3].

For linesearch methods we prove convergence for C^1 functions, and we show that it is allowed to memorize the accepted steplength between serious steps if the objective is of class $C^{1,1}$. This option may be of interest for large scale applications, where second-order steps are not practical, and re-starting each linesearch at $t = 1$ may lead to unnecessary and costly backtracking.

For trust-region methods we discuss acceptance tests which feature conditions on the curvature of the objective along the proposed step, in tandem with the usual criteria relating the achieved progress to the minimal progress guaranteed by the Cauchy point.

The paper is organized as follows. Section 2 presents the Kurdyka-Łojasiewicz inequality. Sections 3 to 5 are devoted to the convergence of backtracking linesearch for functions satisfying the KL inequality. In section 6 convergence for trust-region methods under the KL condition is discussed and new conditions to guarantee convergence in practice are investigated.

2 The Kurdyka-Łojasiewicz condition

In 1963 S. Łojasiewicz [11, 12] proved that a real analytic function $f : \mathbb{R}^n \to \mathbb{R}$ has the following property, now called the Łojasiewicz property. Given a critical point $\bar{x} \in \mathbb{R}^n$ of f, there exists a neighborhood U of $\bar{x}$, $c > 0$ and $\frac{1}{2} \leq \theta < 1$ such that

$$|f(x) - f(\bar{x})|^\theta \leq c\|\nabla f(x)\|$$

for all $x \in U$. In 1998 K. Kurdyka presented a more general construction which applies to differentiable functions definable in an o-minimal structure [10]. The following extension to nonsmooth functions has been presented in [5]:

Definition 1. A proper lower semi-continuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has the Kurdyka-Łojasiewicz property (for short KL-property) at $\bar{x} \in \operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$ if there exist $\eta > 0$, a neighborhood U of $\bar{x}$, and a continuous concave function $\varphi : [0, \eta] \to [0, +\infty)$ such that:

1. $\varphi(0) = 0$, φ is C^1 on $(0, \eta)$, and $\varphi' > 0$ on $(0, \eta)$.
2. For every $x \in U \cap \{x \in \mathbb{R}^n : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$,

$$\varphi'(f(x) - f(\bar{x}))\operatorname{dist}(0, \partial f(x)) \geq 1. \tag{1}$$

The Łojasiewicz inequality or property is a special case of the KL-property when $\varphi(s) = s^{1-\theta}$, $\theta \in [\frac{1}{2}, 1)$. It is automatically satisfied for non-critical points, so (1) is in fact a condition on critical points. We will need the following preparatory result.

Lemma 1. *Let $K \subset \mathbb{R}^n$ be compact. Suppose f is constant on K and has the KL-property at every $\bar{x} \in K$. Then there exists $\varepsilon > 0$, $\eta > 0$ and a continuous concave function $\varphi : [0,\eta] \to [0,\infty)$, which is C^1 on $(0,\eta)$ and satisfies $\varphi(0) = 0$, $\varphi' > 0$ on $(0,\eta)$, such that*

$$\varphi'(f(x) - f(\bar{x}))\operatorname{dist}(0, \partial f(x)) \geq 1$$

for every $\bar{x} \in K$ and every x such that $\operatorname{dist}(x,K) < \varepsilon$ *and* $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$.

Proof. The proof is a slight extension of a similar result in [2] for functions having the Łojasiewicz property.

For every $\bar{x} \in K$ pick a neighborhood $B(\bar{x}, \varepsilon_{\bar{x}})$ of $\bar{x}$ and $\eta_{\bar{x}} > 0$ in tandem with a function $\varphi_{\bar{x}}$ as in definition 1. Since K is compact, there exist finitely many $\bar{x}_i \in K$, $i = 1,\ldots,N$ such that $K \subset \bigcup_{i=1}^N B(\bar{x}_i, \frac{1}{2}\varepsilon_{\bar{x}_i})$. Write for simplicity $\varepsilon_i := \varepsilon_{\bar{x}_i}$, $\eta_i := \eta_{\bar{x}_i}$, $\varphi_i := \varphi_{\bar{x}_i}$. Then put

$$\eta = \min_{i=1\ldots N} \eta_i > 0 \quad \text{and} \quad \varepsilon = \min_{i=1,\ldots,N} \tfrac{1}{2}\varepsilon_i > 0.$$

It follows immediately that: $\{x \in \mathbb{R}^n : \operatorname{dist}(x,K) < \varepsilon\} \subset \bigcup_{i=1}^N B(\bar{x}_i, \varepsilon_i)$.

Suppose $f(x) = \underline{f}$ for every $x \in K$. Then (1) holds uniformly on K in the sense that given any x with $\operatorname{dist}(x,K) < \varepsilon$ and $\underline{f} < f(x) < \underline{f} + \eta$, there exists $i(x) \in \{1,\ldots,N\}$ such that

$$\varphi'_{i(x)}(f(x) - \underline{f}) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

To conclude the proof, it remains to define the function $\varphi : [0,\eta] \to [0,\infty)$. We let

$$\varphi(t) = \int_0^t \max_{i=1\ldots N} \varphi'_i(\tau)\, d\tau, \qquad t \in [0,\eta].$$

Observe that $\tau \mapsto \max_{i=1\ldots N} \varphi'_i(\tau)$ is continuous on $(0,\eta)$ and decreasing on $[0,\eta]$. Then φ is well defined and continuous on $[0,\eta]$, and of class C^1 on $(0,\eta)$. We also easily check $\varphi(0) = 0$, φ concave on $[0,\eta]$ and strictly increasing on $(0,\eta)$. Finally we have

$$\begin{aligned} \varphi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) &= \varphi'(f(x) - \underline{f}) \operatorname{dist}(0, \partial f(x)) \\ &\geq \varphi'_{i(x)}(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \geq 1 \end{aligned}$$

for all $\bar{x} \in K$ and all $x \in \mathbb{R}^n$ such that $\operatorname{dist}(x,K) < \varepsilon$ and $\underline{f} < f(x) < \underline{f} + \eta$.

Next we address convergence of linesearch methods assuming f of class C^1 and having the (KL) property. We will need the following technical lemma, whose proof can be found e.g. in [7]:

Lemma 2. *Let f be of class C^1 and $x_j \to x$, $y_j \to x$. Then*

$$\frac{f(y_j) - f(x_j) - \nabla f(x_j)^\top (y_j - x_j)}{\|y_j - x_j\|} \to 0.$$

3 Linesearch without memory

Descent methods which attempt second-order steps usually start the linesearch at the step length $t = 1$. We refer to this as *memory-free*. The challenge is to prove convergence for C^1 functions.

The algorithm discussed hereafter uses the following well-known

Definition 2. A sequence d^j of descent directions chosen by a descent algorithm at points x^j is called *gradient oriented* if there exist $0 < c < 1$ such that the angle $\phi_j := \angle\left(d^j, -\nabla f(x^j)\right)$ satisfies

$$\forall j \in \mathbb{N}, 0 < c \leq \cos\phi_j. \tag{2}$$

Algorithm 1 Linesearch descent method without memory.

Parameters: $0 < \gamma < 1$, $0 < \underline{\theta} < \overline{\theta} < 1$, $\tau > 0$, $0 < c < 1$.

1: **Initialize**. Choose initial guess x^1. Put counter $j = 1$.
2: **Stopping test**. Given iterate x^j at counter j, stop if $\nabla f(x^j) = 0$. Otherwise compute a gradient oriented descent direction d^j with $\cos\phi_j \geq c$ and goto linesearch.
3: **Initialize linesearch**. Put linesearch counter $k = 1$ and initialize steplength t_1 such that:

$$t_1 \geq \tau \frac{\|\nabla f(x^j)\|}{\|d^j\|}.$$

4: **Acceptance test.** At linesearch counter k and steplength $t_k > 0$ check whether

$$\rho_k = \frac{f(x^j) - f(x^j + t_k d^j)}{-t_k \nabla f(x^j)^\top d^j} \geq \gamma.$$

If $\rho_k \geq \gamma$, put $x^{j+1} = x^j + t_k d^j$, quit linesearch, increment counter j, and go back to step 2. On the other hand, if $\rho_k < \gamma$, reduce steplength such that $t_{k+1} \in [\underline{\theta} t_k, \overline{\theta} t_k]$, increment linesearch counter k, and continue linesearch with step 4.

Lemma 3. *Suppose f is differentiable and $\nabla f(x^j) \neq 0$ and let d^j be a descent direction at x^j. Then the linesearch described in algorithm 1 needs a finite number of backtracks to find a steplength t_k such that $x^j + t_k d^j$ passes the acceptance test $\rho_k \geq \gamma$.*

Proof. The proof is straightforward. Suppose the linesearch never ends, then $\rho_k < \gamma$ for all k and $t_k \to 0$. Since $f'(x^j, d^j) = \nabla f(x^j)^\top d^j < 0$, $\rho_k < \gamma$ transforms into

$$\frac{f(x^j + t_k d^j) - f(x^j)}{t_k} > \gamma \nabla f(x^j)^\top d^j = \gamma f'(x^j, d^j),$$

and the left hand side converges to $f'(x^j, d^j)$. This leads to $0 > f'(x^j, d^j) \geq \gamma f'(x^j, d^j)$, contradicting $0 < \gamma < 1$.

Having proved that an acceptable steplength is found in a finite number of backtracks, we now focus on convergence of the whole algorithm. The proof of Theorem 1 below first establishes stationarity of limit points, generalizing well-known results for gradient methods (see e.g. [4, Proposition 1.2.1]), and then proves the convergence of the iterates using the Kurdyka-Łojasiewicz condition.

Theorem 1. *Let $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x^1)\}$ be bounded. Suppose f is of class C^1 and satisfies the Kurdyka-Łojasiewicz condition. Then the sequence of iterates x^j generated by algorithm 1 is either finite and ends with $\nabla f(x^j) = 0$, or it converges to a critical point $\bar{x}$ of f.*

Proof. 1) We can clearly concentrate on the case of an infinite sequence x^j. Consider the following normalized sequence of descent directions $\widetilde{d^j} = (\|\nabla f(x^j)\|/\|d^j\|)d^j$. Then the directions $\widetilde{d^j}$ are also gradient oriented and $\|\widetilde{d^j}\| = \|\nabla f(x^j)\|$. A trial step $x^j + td^j$ can then also be written as $x^j + \widetilde{t}\widetilde{d^j}$, where the stepsizes t, $\widetilde{t}$ are in one-to-one correspondence via $\widetilde{t} = (\|d^j\|/\|\nabla f(x^j)\|)t$. Neither the backtracking rule in step 4 nor the acceptance test are affected if we write steps $x^j + td^j$ as $x^j + \widetilde{t}\widetilde{d^j}$. The initial condition in step 3 becomes $\widetilde{t} \geq \tau$. Switching back to the notation $x^j + td^j$, we may therefore assume $\|d^j\| = \|\nabla f(x^j)\|$ and that the linesearch is initialized at $t_1 \geq \tau$. The gradient oriented direction d^j now satisfies

$$\|\nabla f(x^j)\|^2 \geq -\nabla f(x^j)^\top d^j \geq c\|d^j\|\|\nabla f(x^j)\| = c\|\nabla f(x^j)\|^2 \tag{3}$$

2) From Lemma 3 we know that the linesearch ends after a finite number of backtracks, let us say with steplength $t_{k_j} > 0$. So $x^{j+1} = x^j + t_{k_j} d^j$. From the acceptance test $\rho_{k_j} \geq \gamma$ we know that

$$\begin{aligned} f(x^j) - f(x^{j+1}) &\geq -\gamma \nabla f(x^j)^\top (x^{j+1} - x^j), \\ &\geq -\gamma t_{k_j} \nabla f(x^j)^\top d^j \geq c\gamma t_{k_j} \|\nabla f(x^j)\|^2 \qquad \text{(according to (3))}. \end{aligned} \tag{4}$$

By construction we have: $t_{k_j} = \|x^{j+1} - x^j\|/\|d^j\| = \|x^{j+1} - x^j\|/\|\nabla f(x^j)\|$, so that:

$$f(x^j) - f(x^{j+1}) \geq c\gamma \|\nabla f(x^j)\|\|x^{j+1} - x^j\|, \tag{5}$$

in which we recognize the so-called strong descent condition in [1]. Summing (5) from $j = 1$ to $j = m-1$ gives

$$\sum_{j=1}^{m-1} \|\nabla f(x^j)\|\|x^{j+1} - x^j\| \leq (c\gamma)^{-1} \sum_{j=1}^{m-1} f(x^j) - f(x^{j+1}) = (c\gamma)^{-1} \left(f(x^1) - f(x^m) \right).$$

Since the algorithm is of descent type, the right hand side is bounded above, so the series on the left is summable. In particular, $\|\nabla f(x^j)\|\|x^{j+1} - x^j\| \to 0$, or equivalently $t_{k_j}\|\nabla f(x^j)\|^2 \to 0$.

3) Fix an accumulation point $\bar{x}$ of x^j and select a subsequence $j \in J$ such that $x^j \to \bar{x}$, $j \in J$. To show that $\bar{x}$ is critical, it suffices to find a subsequence $j' \in J'$ such that $\nabla f(x^{j'}) \to 0$.

Suppose on the contrary that no such subsequence exists, so that $\|\nabla f(x^j)\| \geq \mu > 0$ for some $\mu > 0$ and all $j \in J$. To obtain a contradiction, we will focus on the last step before acceptance.

3.1) First note that we must have $t_{k_j} \to 0$, $j \in J$. Indeed using $\|\nabla f(x^j)\|\|x^{j+1} - x^j\| \geq \mu\|x^{j+1} - x^j\|$, $j \in J$ in tandem with the results from part 2), we see that $\|x^{j+1} - x^j\| \to 0$, $j \in J$. Then, knowing that

$$t_{k_j} = \|x^{j+1} - x^j\|/\|\nabla f(x^j)\| \leq \mu^{-1}\|x^{j+1} - x^j\|,$$

we deduce $t_{k_j} \to 0$ and by boundedness of the x^j also $t_{k_j}\|\nabla f(x^j)\| \to 0$, $j \in J$.

3.2) We now claim that there exists an infinite subsequence J' of J such that (i) $\|\nabla f(x^j)\| \geq \mu > 0$, $j \in J'$, (ii) $t_{k_j} \to 0$, $j \in J'$, and (iii) $k_j \geq 2$ for $j \in J'$ i.e. for $j \in J'$, there was at least one backtrack during the j^{th} linesearch. Item (iii) is a consequence of the initial condition $t_1 \geq \tau$ in step 3 of the algorithm. Namely, in tandem with $t_{k_j} \to 0$, $j \in J$, this condition says that the set $J' = \{j \in J : k_j \geq 2\} = \{j \in J : t_{k_j} < t_1\}$ cannot be finite.

This sequence $j \in J'$ satisfies $\rho_{k_j} \geq \gamma$, $\rho_{k_j-1} < \gamma$, $t_{k_j} \to 0$, $\|\nabla f(x^j)\| \geq \mu > 0$. Because of the backtracking rule, we then also have $t_{k_j-1} \to 0$. Putting $y^{k_j-1} = x^j + t_{k_j-1}d^j$, given that $x^j \to \bar{x}$, $t_{k_j}\|\nabla f(x^j)\| \to 0$, $j \in J'$, and $t_{k_j-1}\|d^j\| = t_{k_j-1}\|\nabla f(x^j)\| \leq \underline{\theta}^{-1}t_{k_j}\|\nabla f(x^j)\|$, we have $y^{k_j-1} \to \bar{x}$, $j \in J'$.

Note that d^j is gradient oriented so that $y^{k_j-1} - x^j$ is also gradient oriented and

$$-\nabla f(x^j)^\top(y^{k_j-1} - x^j) \geq c\|\nabla f(x^j)\|\|y^{k_j-1} - x^j\| \geq c\mu\|y^{k_j-1} - x^j\|. \tag{6}$$

3.3) Now we expand

$$\begin{aligned}\rho_{k_j-1} &= \frac{f(x^j) - f(y^{k_j-1})}{-\nabla f(x^j)^\top(y^{k_j-1} - x^j)} = 1 - \frac{f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)^\top(y^{k_j-1} - x^j)}{-\nabla f(x^j)^\top(y^{k_j-1} - x^j)}\\ &=: 1 - R_j.\end{aligned}$$

Using (6) gives

$$\begin{aligned}|R_j| &= \frac{\left|f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)^\top(y^{k_j-1} - x^j)\right|}{-\nabla f(x^j)^\top(y^{k_j-1} - x^j)}\\ &\leq \frac{\left|f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)^\top(y^{k_j-1} - x^j)\right|}{c\mu\|y^{k_j-1} - x^j\|}.\end{aligned}$$

Since f is of class C^1, and since $x^j \to \bar{x}$, $y^{k_j-1} \to \bar{x}$, Lemma 2 guarantees the existence of a sequence $\varepsilon_j \to 0$ such that

$$\left|f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)^\top(y^{k_j-1} - x^j)\right| \leq \varepsilon_j\|y^{k_j-1} - x^j\|.$$

We deduce $|R_j| \leq \varepsilon_j/(c\mu) \to 0$, hence $\rho_{k_j-1} \to 1$ contradicting $\rho_{k_j-1} < \gamma$. This proves that $\|\nabla f(x^j)\| \geq \mu > 0$ for all $j \in J$ was impossible. Therefore $\bar{x}$ is critical, and so are all the accumulation points of x^j.

4) By boundedness of the sequence x^j the set K of its accumulation points $\bar{x}$ is bounded and consists of critical points of f. It is also closed, as can be shown by a diagonal argument. Hence K is compact. Since the algorithm is of descent type, f has constant value on K.

Since f satisfies the Kurdyka-Łojasiewicz condition at every $\bar{x} \in K$, Lemma 1 gives us $\varepsilon > 0$, $\eta > 0$, and a continuous concave function $\varphi : [0,\eta] \to [0,\infty)$ with $\varphi(0) = 0$ and $\varphi' > 0$ on $(0,\eta)$ such that for every $\bar{x} \in K$ and every x with $\mathrm{dist}(x,K) < \varepsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$ we have

$$\varphi'\left(f(x) - f(\bar{x})\right) \|\nabla f(x)\| \geq 1. \tag{7}$$

5) Assume without loss of generality that $f(\bar{x}) = 0$ on K. Then $f(x^j) > 0$ for all j, because our algorithm is of descent type. Concavity of φ implies

$$\varphi\left(f(x^j)\right) - \varphi\left(f(x^{j+1})\right) \geq \varphi'\left(f(x^j)\right)\left(f(x^j) - f(x^{j+1})\right). \tag{8}$$

Using $f(\bar{x}) = 0$, the Kurdyka-Łojasiewicz estimate (7) gives

$$\varphi'\left(f(x^j)\right) = \varphi'\left(f(x^j) - f(\bar{x})\right) \geq \|\nabla f(x^j)\|^{-1}. \tag{9}$$

Hence by (8)

$$\begin{aligned}
\varphi\left(f(x^j)\right) - \varphi\left(f(x^{j+1})\right) &\geq \|\nabla f(x^j)\|^{-1}\left(f(x^j) - f(x^{j+1})\right) \\
&\geq c\gamma\|x^{j+1} - x^j\| \qquad \text{(using (5))}.
\end{aligned}$$

Summing from $j = 1$ to $j = m-1$ gives

$$c\gamma \sum_{j=1}^{m-1} \|x^j - x^{j+1}\| \leq \varphi\left(f(x^1)\right) - \varphi\left(f(x^m)\right),$$

and since the term on the right hand side is bounded, the series on the left converges. This shows that x^j is a Cauchy sequence, which converges therefore to some $\bar{x} \in K$, proving that $K = \{\bar{x}\}$ is singleton.

4 Memorizing the steplength

In Newton type descent schemes it is standard to start the linesearch at steplength $t = 1$. However, if a first-order method is used, a different strategy may be more promising. To avoid unnecessary backtracking, we may decide to start the $(j+1)^{\text{st}}$ linesearch where the j^{th} ended. Such a concept may be justified theoretically if f is of class $C^{1,1}$.

Standard proofs for backtracking linesearch algorithms use indeed $C^{1,1}$ functions. The Lipschitz constant of ∇f on Ω allows a precise estimation of the Armijo stepsize

$$t_\gamma = \sup\{t > 0 : f(x+td) - f(x) < \gamma t \nabla f(x)^\top d\}.$$

As long as the linesearch starts with large steps, $t > t_\gamma$, backtracking $t_{k+1} \in [\underline{\theta} t_k, \overline{\theta} t_k]$ will lead to an acceptable steplength t^* such that $\underline{\theta} t_\gamma \leq t^* \leq t_\gamma$. This mechanism guarantees that the accepted steplength is not too small and replaces the usual conditions against small stepsizes. However, what we plan to do in this section is memorize the last accepted steplength. So the above argument will not work, because our linesearch may already start small, and we will have no guarantee to end up in the interval $[\underline{\theta} t_\gamma, t_\gamma]$. In that situation the safeguard against too small steps is more subtle to assure. We propose the following

Algorithm 2 Descent method with memorized steplength.

Parameters: $0 < \gamma < \Gamma < 1$, $0 < c < 1$, $0 < \underline{\theta} < \overline{\theta} < 1$, $\Theta > 1$.

1: **Initialize**. Choose initial guess x^1. Fix memory steplength $\tau_1 = 1$. Put counter $j = 1$.

2: **Stopping test**. Given iterate x^j at counter j, stop if $\nabla f(x^j) = 0$. Otherwise compute descent direction d^j with $\|d^j\| = \|\nabla f(x^j)\|$ and $\cos\phi_j \geq c$ and goto linesearch.

3: **Initialize linesearch**. Put linesearch counter $k = 1$ and use memory steplength τ_j to initialize linesearch at steplength $t_1 = \tau_j$.

4: **Acceptance test.** At linesearch counter k and steplength $t_k > 0$ check whether

$$\rho_k = \frac{f(x^j) - f(x^j + t_k d^j)}{-t_k \nabla f(x^j)^\top d^j} \geq \gamma.$$

If $\rho_k \geq \gamma$ put $x^{j+1} = x^j + t_k d^j$, quit linesearch and goto step 5. On the other hand, if $\rho_k < \gamma$ backtrack by reducing steplength to $t_{k+1} \in [\underline{\theta} t_k, \overline{\theta} t_k]$ and continue linesearch with step 4.

5: **Update memory steplength**. Define the new memory steplength τ_{j+1} as

$$\tau_{j+1} = \begin{cases} t_k & \text{if } \gamma \leq \rho_k < \Gamma \\ \Theta t_k & \text{if } \rho_k \geq \Gamma \end{cases},$$

where t_k is the accepted steplength in step 4. Increment counter j and go back to step 2.

Theorem 2. *Let $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x^1)\}$ be bounded, and suppose f satisfies the Kurdyka-Łojasiewicz condition and is of class $C^{1,1}(\Omega)$. Let x^j be the sequence of steps generated by the descent algorithm 2. Then either $\nabla f(x^j) = 0$ for some j, or x^j converges to a critical point of f.*

Proof. 1) As in the proof of Theorem 1 we concentrate on the case where the sequence x^j is infinite. As required by algorithm 2, the sequence d^j is already normalized to $\|d^j\| = \|\nabla f(x^j)\|$. We now follow the proof of Theorem 1 until the end of part 2), where $t_{k_j} \|\nabla f(x^j)\|^2 \to 0$ is proved.

2) We wish to prove $\nabla f(x^j) \to 0$, $j \in \mathbb{N}$. Assume on the contrary that there exists an infinite set $J \subset \mathbb{N}$ such that $\|\nabla f(x^j)\| \geq \mu > 0$ for all $j \in J$. Then we must have $t_{k_j} \to 0$, $j \in J$. This is shown precisely as in part 3.1) of the proof of Theorem 1.

3) Using the sequence $j \in J$ which satisfies $\|\nabla f(x^j)\| \geq \mu$ and $t_{k_j} \to 0$, $j \in J$, we now have the first substantial modification. We construct another infinite sequence $J' \subset \mathbb{N}$ such that $t_{k_j} \to 0$, $j \in J'$, and such that in addition for every $j \in J'$ the j^{th} linesearch did at least one backtrack. In other words, $k_j \geq 2$ for every $j \in J'$. In contrast with Theorem 1 we do not claim that J' is a subsequence of J. Neither do we have any information as to whether $\|\nabla f(x^j)\| \geq \mu$ for $j \in J'$, and we therefore cannot use such an estimate, as we did in the proof of Theorem 1.

Now J' can be constructed as follows. Put

$$j'(j) = \min\{j' \in \mathbb{N} : j' \geq j, k_{j'} \geq 2\}, \quad \text{and} \quad J' = \{j'(j) : j \in J\}.$$

We claim that $j'(j) < \infty$ for every $j \in J$. For suppose there exists $j \in J$ such that $k_{j'} = 1$ for all $j' \geq j$. Then no backtracking is done in any of the linesearches j' following j. Since the stepsize t is not decreased between linesearches, it is not decreased at all, so it cannot become arbitrarily small any more. This contradicts $t_{k_j} \to 0$, $j \in J$. This argument shows $j \leq j'(j) < \infty$ for all $j \in J$, so J' is an infinite set.

For the indices $j \in J'$ we have $k_j \geq 2$ and

$$t_{k_j} \text{ accepted}, \qquad t_{k_j-1} \text{ rejected}, \qquad \underline{\theta} t_{k_j-1} \leq t_{k_j} \leq \overline{\theta} t_{k_j-1}.$$

In particular, $\rho_{k_j-1} < \gamma$, $\rho_{k_j} \geq \gamma$. Moreover, $t_{k_j-1} \to 0$, $j \in J'$. Writing $y^{k_j-1} = x^j + t_{k_j-1} d^j$, we see that $x^j - y^{k_j-1} \to 0$, $j \in J'$. Now we expand

$$\begin{aligned}\rho_{k_j-1} &= \frac{f(x^j) - f(y^{k_j-1})}{-t_{k_j-1} \nabla f(x^j)^\top d^j} = 1 - \frac{f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)^\top (y^{k_j-1} - x^j)}{-t_{k_j-1} \nabla f(x^j)^\top d^j} \\ &=: 1 + R_j.\end{aligned}$$

Since f is of class $C^{1,1}$, and since the sequences x^j and y^{k_j-1} are bounded and $x^j - y^{k_j-1} \to 0$, there exists a constant $L > 0$ (the Lipschitz constant of ∇f on Ω) such that

$$\left| f(y^{k_j-1}) - f(x^j) - \nabla f(x^j)(y^{k_j-1} - x^j) \right| \leq \tfrac{L}{2} \|y^{k_j-1} - x^j\|^2 = \tfrac{L}{2} t_{k_j-1}^2 \|d^j\|^2$$

for all $j \in J'$. Gradient orientedness of d^j implies $|\nabla f(x^j)^\top d^j| \geq c\|d^j\|^2$, so the residual term R_j may be estimated as

$$|R_j| \leq \frac{\frac{L}{2} t_{k_j-1}^2 \|d^j\|^2}{c\, t_{k_j-1} \|d^j\|^2} = (L/2c)\, t_{k_j-1} \to 0 \quad (j \in J').$$

That shows $\rho_{k_j-1} \to 1$, $(j \in J')$, contradicting $\rho_{k_j-1} < \gamma$. This argument proves $\nabla f(x^j) \to 0$, $j \to \infty$. In consequence, every accumulation point $\bar{x}$ of the sequence x^j is a critical point.

4) The remainder of the proof is now identical with 4) - 5) in the proof of Theorem 1, and the conclusion is the same.

5 A practical method

In algorithm 2 we cannot a priori exclude the possibility that τ_j becomes arbitrarily small, even though it has in principle the possibility to recover if good steps are made (see step 5 of Algorithm 2). Let us see what happens if $d^j = -P_j^{-1}\nabla f(x^j)$, where P_j is the Hessian of f or a quasi-Newton substitute of the Hessian. The crucial question is, will this method eventually produce good steps $\rho_k \geq \Gamma$, so that the memorized steplength increases to reach $\tau_j = 1$, from whereon the full Newton step is tried first?

Theorem 3. *Let $0 < \gamma < \Gamma < \frac{1}{2}$. Suppose the Newton steps $d^j = -\nabla^2 f(x^j)^{-1}\nabla f(x^j)$ at x^j form a sequence of gradient oriented descent directions. Let $\bar{x}$ be a local minimum of f satisfying the second order sufficient optimality condition.*

Then there exists a neighborhood V of $\bar{x}$ such that as soon as $x^j \in V$, the iterates stay in V, the first trial step $x^{j+1} = x^j + t_1 d^j$ is accepted with $\rho_1 \geq \Gamma$, so that the memory steplength is increased from τ_j to $\tau_{j+1} = \min\{\Theta\tau_j, 1\}$, until it reaches 1 after a finite number of steps. From that moment on the full Newton step is tried and accepted, and the method converges quadratically to $\bar{x}$.

Proof. This theorem is similar to theorem 6.4 from [9] with the following differences: the step t_k does not necessarily satisfy the second Wolfe condition, and the sequence x^j of iterates is not assumed to converge towards $\bar{x}$. Instead we have to use the hypothesis of gradient-orientedness, and the backtracking process of the line-search to prove the same result.

Since the local minimum $\bar{x}$ satisfies the second order sufficient optimality condition, the Hessian of f at $\bar{x}$ is positive definite, and we have $\mu := \lambda_{min}(\nabla^2 f(\bar{x})) > 0$. Using a well-known result on Newton's method (see e.g. [9, theorem 2.1]), there exists an open neighborhood U of the local minimum $\bar{x}$, where the Newton iterates are well-defined, remain in U, converge to $\bar{x}$ and

$$\lambda_{min}(\nabla^2 f(x)) \geq \frac{\mu}{2} \quad \text{and} \quad \lambda_{max}(\nabla^2 f(x)) \leq K < \infty \tag{10}$$

for every $x \in U$.

Assume now that the iterates x^j reach U. We first prove that the Newton step is acceptable in the sense that $f(x^j + d^j) - f(x^j) < \gamma\nabla f(x^j)^\top d^j$ because of $\gamma < \frac{1}{2}$. Indeed, as in the proof of Theorem 6.4 in [9], the combined use of the mean value theorem, gradient-orientedness and hypothesis (10) imply that for all j with $x^j \in U$, the Newton iterate $x^j + d^j$ is accepted by any Armijo parameter $< \frac{1}{2}$, so that it even passes the acceptance test with the larger constant Γ instead of γ due to $0 < \gamma < \Gamma < \frac{1}{2}$. Note that the same is then true for every damped Newton step, namely as soon as $t = 1$ passes the acceptance test, so does any $t < 1$.

The last point is to prove that if the iterates x^j enter U with $\tau_j < 1$, then our algorithm starts to increase τ until the Newton step is actually made. Indeed, even though at the beginning a smaller step $x^j + td^j$ with $t < 1$ is made, according to what was previously shown, this step is accepted at once with $\rho_1 > \Gamma$ and remains in U.

We then update $\tau_{j+1} = \Theta \tau_j$ (with a fixed $\Theta > 1$), meaning that τ_j is increased until it hits 1 after a finite number of iterations j. From that moment onward the Newton step is tried first, accepted at once, and quadratic convergence prevails.

Remark 1. This result indicates that Γ should be only slightly larger than γ, at least near the second order minimum.

Remark 2. The following modification of the update rule of τ seems interesting. Fix $1 < \Theta < \Xi$ and put

$$\tau_{j+1} = \begin{cases} t_{k_j} & \text{if } \gamma \leq \rho_{k_j} < \Gamma \\ \Theta t_{k_j} & \text{if } \rho_{k_j} \geq \Gamma \text{ and } k_j \geq 2 \\ \Xi t_{k_j} & \text{if } \rho_{k_j} \geq \Gamma \text{ and } k_j = 1 \end{cases} .$$

This accelerates the increase of τ if acceptance is immediate and helps to get back to $\tau = 1$ faster if the neighborhood of attraction of Newton's method is reached. Our convergence analysis covers this case as well.

6 Convergence of trust-region methods for functions of class C^1

The idea of memorizing the step length in a linesearch method is paralleled by the trust-region strategy. The basic trust-region algorithm uses a quadratic model

$$m(y, x^j) = f(x^j) + \nabla f(x^j)^\top (y - x^j) + \frac{1}{2}(y - x^j)^\top B_j (y - x^j)$$

to approximate the objective function f within the trust-region $\{x \in \mathbb{R}^n : \|y - x^j\| \leq \Delta_k\}$ around the current iterate x^j, where $\Delta_k > 0$ is the trust-region radius, and B_j an approximation of the Hessian at x^j. One then computes an approximate solution y^{k+1} of the tangent program

$$\min\{m(y, x^j) : \|y - x^j\| \leq \Delta_k, y \in \mathbb{R}^n\}. \tag{11}$$

Instead of minimizing the trust-region model, the step y^{k+1} is only supposed to achieve a decrease of $m(\cdot, x^j)$, which is at least a given percentage of the reduction obtained by the Cauchy point x_C^{j+1}. This means, y^{k+1} satisfies

$$f(x^j) - m(y^{k+1}, x^j) \geq c \left[f(x^j) - m(x_C^{j+1}, x^j) \right] \tag{12}$$

where $0 < c < 1$ is fixed once and for all and where the Cauchy point x_C^{j+1} is defined as the solution of the one-dimensional problem:

$$\min \left\{ m\left(x^j - t \frac{\nabla f(x^j)}{\|\nabla f(x^j)\|}, x^j \right) : t \in \mathbb{R}, 0 \leq t \leq \Delta_k \right\}. \tag{13}$$

Here we follow the line of Conn *et al.* [8], who determine a step y^{k+1} satisfying the weaker condition

$$f(x^j) - m(y^{k+1}, x^j) \geq c\|\nabla f(x^j)\| \min\left(\Delta_k, \frac{\|\nabla f(x^j)\|}{1+\|B_j\|}\right). \tag{14}$$

It can be shown that (12) implies (14), and that the exact solution of (11) satisfies (14). With these preparations we can now state our algorithm.

Algorithm 3 Trust-region method.

Parameters: $0 < \gamma < \Gamma < 1$, $0 < \underline{\theta} < \overline{\theta} < 1$, $\tau > 0$.

1: **Initialize**. Choose initial guess x^1 and initial trust-region radius $\Delta_1^\sharp > 0$. Put counter $j = 1$.

2: **Stopping test**. Given iterate x^j at counter j, stop if $\nabla f(x^j) = 0$. Otherwise goto step 3.

3: **Model definition.** Define a model $m(\cdot, x^j)$ of f in $\{x \in \mathbb{R}^n : \|x - x^j\| \leq \Delta_j^\sharp\}$:

$$m(y, x^j) = f(x^j) + \nabla f(x^j)^\top (y - x^j) + \tfrac{1}{2}(y - x^j)^\top B_j (y - x^j).$$

4: **Initialize inner loop.** Put counter $k = 1$ and $\Delta_1 = \Delta_j^\sharp$.

5: **Tangent program.** At inner loop counter k let y^{k+1} be an approximate solution of

$$\min\{m(y, x^j) : \|y - x^j\| \leq \Delta_k, y \in \mathbb{R}^n\}$$

in the sense of (12).

6: **Acceptance test.** At counter k, check whether

$$\rho_k = \frac{f(x^j) - f(y^{k+1})}{f(x^j) - m(y^{k+1}, x^j)} \geq \gamma. \tag{15}$$

- If $\rho_k \geq \gamma$ put $x^{j+1} = y^{k+1}$, and update:

$$\Delta_{j+1}^\sharp \in \begin{cases} [\Delta_k, +\infty[& \text{if } \rho_k > \Gamma \text{ and } \|y^{k+1} - x^j\| = \Delta_k \\ [\bar{\theta}\Delta_k, \Delta_k] & \text{otherwise.} \end{cases}$$

  Increment outer counter j, and go back to step 2.
- If $\rho_k < \gamma$, then: $\Delta_{k+1} \in [\underline{\theta}\Delta_k, \overline{\theta}\Delta_k]$. Increment inner counter k and go to step 5.

The trial point y^{k+1} computed in step 5 of the algorithm is called a *serious step* if accepted as a new iterate x^{j+1}, and a *null step* if rejected. To decide whether a step y^{k+1} is accepted, we compute the ratio

$$\rho_k = \frac{f(x^j) - f(y^{k+1})}{f(x^j) - m(y^{k+1}, x^j)},$$

which reflects the agreement between f and its model at y^{k+1}. If the model $m(\cdot, x^j)$ is a good approximation of f at y^{k+1}, we expect $\rho_k \approx 1$, so here y^{k+1} is a good point and should be accepted. If $\rho_k \ll 1$, y^{k+1} is bad and we reject it. Step 6 of the algorithm formalizes this decision.

The proof of the global convergence of the trust-region algorithm for functions of class C^1 in the sense of subsequences can be found in e.g. [13, theorem 4.8]. One first proves finiteness of the inner loop and then global convergence of algorithm 3.

Our issue here is to prove convergence of the sequence, which requires the Kurdyka-Łojasiewicz condition and the so-called strong descent condition in [1]:

Theorem 4. *Let $\Omega = \{x \in \mathbb{R}^n : f(x) \le f(x^1)\}$ be bounded. Suppose f is of class C^1 and satisfies the Kurdyka-Łojasiewicz condition. Let the Hessian matrices B_j be uniformly bounded. If the sequence x^j, $j \in \mathbb{N}$, of iterates of algorithm 3 satisfies the strong descent condition*

$$f(x^j) - f(x^{j+1}) \ge \sigma \|\nabla f(x^j)\| \|x^{j+1} - x^j\|, \tag{16}$$

then it is either finite and ends with $\nabla f(x^j) = 0$, or it converges to a critical point $\bar{x}$ of f.

Proof. Let K be the set of the accumulation points of the sequence x^j, $j \in \mathbb{N}$. As in the proof of theorem 1 we prove compactness of K and show that f is constant on K. Then the Kurdyka-Łojasiewicz condition gives

$$\begin{aligned} \varphi(f(x^j)) - \varphi(f(x^{j+1})) &\ge \varphi'(f(x^j)) \left(f(x^j) - f(x^{j+1}) \right) \\ &\ge \|\nabla f(x^j)\|^{-1} \left(f(x^j) - f(x^{j+1}) \right). \end{aligned}$$

Assuming the strong descent condition $f(x^j) - f(x^{j+1}) \ge \sigma \|\nabla f(x^j)\| \|x^{j+1} - x^j\|$ as in [1] now yields

$$\varphi(f(x^j)) - \varphi(f(x^{j+1})) \ge \sigma \|x^{j+1} - x^j\|.$$

Using the series argument from theorem 1 proves convergence of the sequence of iterates x^j to some $\bar{x} \in K$, and then $K = \{\bar{x}\}$.

Now we have to give practical criteria which imply the strong descent condition (16). Several easily verified conditions for the iterates of the trust-region algorithm are given in [1]. Here we focus on conditions involving the curvature of the model along the search direction. Let $\omega(y, x^j)$ denote the curvature of the model $m(\cdot, x^j)$ between x^j and y^{k+1}, namely:

$$\omega(y^{k+1}, x^j) = \frac{(y^{k+1} - x^j)^\top B_j (y^{k+1} - x^j)}{\|y^{k+1} - x^j\|^2}.$$

Note that the curvature along the Cauchy point direction is:

$$\omega(x_C^{j+1}, x^j) = \frac{\nabla f(x^j)^\top B_j \nabla f(x^j)}{\|\nabla f(x^j)\|^2}.$$

We propose the following modified tangent program in algorithm 3:

Fix $0 < \mu < 1$.

5': **Tangent program.** Compute an approximate solution y^{k+1} of

$$\min\{m(y,x^j) : \|y - x^j\| \leq \Delta_k, y \in \mathbb{R}^n\}$$

in the sense of (12), such that in addition

$$\omega(y^{k+1},x^j) \geq \mu\,\omega(x_C^{j+1},x^j) \geq 0 \tag{17}$$

as soon as the Cauchy point lies in the interior of the trust-region, i.e., if $\|\nabla f(x^j)\| \leq \Delta_k\omega(x_C^{j+1},x^j)$.

This modified step (5') in the algorithm has a solution y^{k+1}, because the Cauchy point satisfies the two conditions (12) and (17). We have to prove the convergence of the modified trust-region algorithm, which we will achieve by proving the strong descent condition. We will need the following preparatory

Lemma 4. *When y^{k+1} is a descent step of the model $m(\cdot,x^j)$ away from x^j, then it satisfies*

$$\|\nabla f(x^j)\| \geq \frac{1}{2}\omega(y^{k+1},x^j)\|y^{k+1} - x^j\|.$$

Each serious step x^{j+1} generated by algorithm 3 satisfies:

$$\|\nabla f(x^j)\| \geq \tfrac{1}{2}\omega(x^{j+1},x^j)\|x^{j+1} - x^j\|.$$

Proof. By definition every descent step y^{k+1} of the model $m(\cdot,x^j)$ at the current iterate x^j, has to verify $-\nabla f(x^j)^\top(y^{k+1} - x^j) > 0$ and $f(x^j) - m(y^{k+1},x^j) \geq 0$, so that

$$-\nabla f(x^j)^\top(y^{k+1} - x^j) \geq \tfrac{1}{2}(y^{k+1} - x^j)^\top B_j(y^{k+1} - x^j).$$

Using the Cauchy-Schwarz inequality $\|\nabla f(x^j)\|\|y^{k+1} - x^j\| \geq -\nabla f(x^j)^\top(y^{k+1} - x^j)$, we obtain

$$\|\nabla f(x^j)\| \geq \frac{1}{2}\frac{(y^{k+1} - x^j)^\top B_j(y^{k+1} - x^j)}{\|y^{k+1} - x^j\|} = \frac{1}{2}\omega(y^{k+1},x^j)\|y^{k+1} - x^j\|.$$

According to the acceptance test, any serious step is also a descent step of the model at the current iterate, which proves the second part of the lemma.

Note that the previous result is only useful when the curvature is positive.

Proposition 1. *The iterates x^j generated by the algorithm (3) with step 5' replacing the original step 5 satisfy the strong descent condition (16).*

Proof. The idea here is to show that the Cauchy step is bounded below by a fraction of the step i.e. there exists $\eta \in (0,1)$ such that

$$\|x_C^{j+1} - x^j\| \geq \eta\|x^{j+1} - x^j\|. \tag{18}$$

Indeed, the sufficient decrease condition (12) together with (18) gives strong descent (see theorem 4.4 from [1]). By the definition of the Cauchy point we have

$$\|x_C^{j+1}-x^j\| = \begin{cases} \dfrac{\|\nabla f(x^j)\|}{\omega(x_C^{j+1},x^j)} & \text{if } \|\nabla f(x^j)\| \le \Delta_{k_j}\omega(x_C^{j+1},x^j) \\ \Delta_{k_j} & \text{otherwise.} \end{cases}$$

In the first case, that is, when $\|\nabla f(x^j)\| \le \Delta_{k_j}\omega(x_C^{j+1},x^j)$, the curvature condition (17) gives

$$\|x_C^{j+1}-x^j\| = \frac{\|\nabla f(x^j)\|}{\omega(x_C^{j+1},x^j)} \ge \mu\frac{\|\nabla f(x^j)\|}{\omega(x^{j+1},x^j)} \ge \frac{\mu}{2}\|x^{j+1}-x^j\|$$

according to Lemma 4. In the second case we have $\|x_C^{j+1}-x^j\| = \Delta_{k_j} \ge \|x^{j+1}-x^j\|$, since x^{j+1} has to belong to the trust-region. Thus (18) is satisfied in both case with $\eta = \frac{\mu}{2}$.

In the last part of the paper we present yet another version (5") of the tangent program based on condition (14) from Conn *et al.* [8], which allows to prove convergence, and yet is weaker than the sufficient decrease condition. Note that this condition is at least satisfied by the Cauchy point and the exact solution of the tangent program.

5": **Tangent program.** Compute an approximate solution y^{k+1} of

$$\min\{m(y,x^j) : \|y-x^j\| \le \Delta_k, y \in \mathbb{R}^n\}$$

in the sense of (14), i.e., $f(x^j) - m(y^{k+1},x^j) \ge c\|\nabla f(x^j)\| \min\left(\Delta_k, \dfrac{\|\nabla f(x^j)\|}{1+\|B_j\|}\right)$.

Now with 5" each serious step satisfies

$$\begin{aligned} f(x^j) - m(x^{j+1},x^j) &\ge c\|\nabla f(x^j)\| \min\left(\Delta_{k_j}, \frac{\|\nabla f(x^j)\|}{\|B_j\|}\right) \\ &\ge c\|\nabla f(x^j)\| \min\left(\|x^{j+1}-x^j\|, \frac{\|\nabla f(x^j)\|}{\|B_j\|}\right) \\ &\ge c \min\left(1, \frac{\|\nabla f(x^j)\|}{\|B_j\|\|x^{j+1}-x^j\|}\right) \|\nabla f(x^j)\|\|x^{j+1}-x^j\|. \end{aligned} \tag{19}$$

To infer the strong descent condition (16), the question is how to guarantee that $\frac{\|\nabla f(x^j)\|}{\|B_j\|\|x^{j+1}-x^j\|}$ remains bounded away from 0? Let us first consider the simpler case when the matrix B_j is positive.

Proposition 2. *Consider the following conditions:*

(H_1) *B_j is positive definite and there exists a $\kappa \geq 1$ such that:*

$$\text{cond}(B_j) := \|B_j\| \|B_j^{-1}\| \leq \kappa \quad \textit{(using the matrix 2-norm).}$$

(H_2) *There exists $\bar{\sigma} > 0$ and $\underline{\sigma} > 0$ such that $\bar{\sigma} I \succ B_j \succeq \underline{\sigma} I \succ 0$.*

Then $(H_2) \Rightarrow (H_1)$. Moreover condition (H_1) in tandem with the acceptance condition (14) used within the modified step 5" of algorithm 3 guarantees strong descent.

Proof. Clearly (H_2) implies (H_1). Now for the second part assume that the matrix B_j is positive definite. Then the curvature of the model $m(\cdot, x^j)$ is also positive and by (19) and Lemma 4:

$$\begin{aligned} f(x^j) - m(x^{j+1}, x^j) &\geq c \min\left(1, \frac{\|\nabla f(x^j)\|}{\|B_j\| \|x^{j+1} - x^j\|}\right) \|\nabla f(x^j)\| \|x^{j+1} - x^j\|, \\ &\geq c \min\left(1, \frac{1}{2} \frac{\omega(x^{j+1}, x^j)}{\|B_j\|}\right) \|\nabla f(x^j)\| \|x^{j+1} - x^j\|. \end{aligned}$$

Note that $\dfrac{\omega(x^{j+1}, x^j)}{\|B_j\|} \leq 1$, therefore

$$f(x^j) - m(x^{j+1}, x^j) \geq \frac{c}{2} \frac{\omega(x^{j+1}, x^j)}{\|B_j\|} \|\nabla f(x^j)\| \|x^{j+1} - x^j\|.$$

Condition (H_1) clearly guarantees that $\omega(x^{j+1}, x^j)/\|B_j\|$ stays bounded away from 0, hence we have strong descent (16).

In order to cover also those cases where B_j is not positive, we propose to replace the acceptance test (15) by the following. Fix $0 < \mu < 1$. The trial step y^{k+1} is accepted to become x^{j+1} if it satisfies

$$\rho_k = \frac{f(x^j) - f(y^{k+1})}{f(x^j) - m(y^{k+1}, x^j)} \geq \gamma \quad \text{and} \quad \|\nabla f(x^j)\| \geq \mu \|B_j\| \|x^{j+1} - x^j\|. \tag{20}$$

The following result shows that condition (20) is eventually satisfied by the trial steps y^{k+1} according to 5". Convergence of the trust-region algorithm with the tangent program 5" follows then with the same method of proof.

Proposition 3. *Let $x \in \mathbb{R}^n$ be the current iterate. Suppose f differentiable and $\nabla f(x) \neq 0$. Then the inner loop of the trust-region algorithm with condition (14) and acceptance test (20) finds a serious iterate after a finite number of trial steps.*

Proof. Suppose on the contrary that the inner loop turns forever. Then $\Delta_k \to 0$ and $y^{k+1} \to x$ $(k \to \infty)$. Now we expand

$$\rho_k = \frac{f(x) - f(y^{k+1})}{f(x) - m(y^{k+1}, x)} = 1 - \frac{f(y^{k+1}) - m(y^{k+1}, x)}{f(x) - m(y^{k+1}, x)}.$$

By condition (14) at each inner iteration k we have

$$
\begin{aligned}
f(x) - m(y^{k+1},x) &\geq c\|\nabla f(x)\| \min\left(\frac{\|\nabla f(x)\|}{1+\|B\|}, \Delta_k\right) \\
&\geq c\|\nabla f(x)\|\Delta_k && \text{for sufficiently large } k. \\
&\geq c\|\nabla f(x)\|\,\|y^{k+1}-x\| && \text{for sufficiently large } k.
\end{aligned}
$$

On the other hand we also have

$$
\begin{aligned}
|f(y^{k+1}) - m(y^{k+1},x)| &\leq |f(y^{k+1}) - f(x) - \nabla f(x)^\top (y^{k+1}-x)| \\
&\qquad + \tfrac{1}{2}|(y^{k+1}-x)^\top B(y^{k+1}-x)| \\
&\leq \|y^{k+1}-x\|\varepsilon_k + \tfrac{1}{2}\|B\|\|y^{k+1}-x\|^2,
\end{aligned}
$$

where the existence of $\varepsilon_k \to 0$ follows from Lemma 2. Combining the previous inequalities, we obtain

$$
\begin{aligned}
\left|\frac{f(y^{k+1}) - m(y^{k+1},x)}{f(x) - m(y^{k+1},x)}\right| &\leq \frac{\|y^{k+1}-x\|\varepsilon_k + \frac{1}{2}\|B\|\|y^{k+1}-x\|^2}{c\|\nabla f(x)\|.\|y^{k+1}-x\|} \\
&\leq \frac{\varepsilon_k + \frac{1}{2}\|B\|\|y^{k+1}-x\|}{c\|\nabla f(x)\|} \to 0 \quad (k \to \infty),
\end{aligned}
$$

which implies $\rho_k \to 1$ $(k \to \infty)$. By our working hypothesis the acceptance test (20) fails. Since it requires two conditions, and since the first of these two conditions, $\rho_k \geq \gamma$, is satisfied for large k, the second condition must eventually fail, i.e; there must exists $K \in \mathbb{N}$ such that

$$\|\nabla f(x)\| < \mu\|B\|\|y^{k+1}-x\|$$

for all $k \geq K$. But from $y^{k+1} \to x$ $(k \to \infty)$ we deduce $\nabla f(x) = 0$, a contradiction.

References

1. P.A. Absil, R. Mahony, B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005.
2. H. Attouch, J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1-2, Ser. B):5–16, 2009.
3. H. Attouch, J. Bolte, P. Redont, A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Journal Mathematics of Operations Research*, 35(2), 2010.
4. D.P. Bertsekas. *Nonlinear programming*. Optimization and Neural computation series. Athena Scientific, 2nd edition edition, 1999.
5. J. Bolte, A. Daniilidis, A. Lewis, M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572 (electronic), 2007.
6. J. Bolte, A. Daniilidis, O. Ley, L. Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.

7. H. Cartan. *Calcul différentiel*. Hermann, Paris, 1967.
8. A.R. Conn, N.I.M. Gould, Ph.L. Toint. *Trust-region methods*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000.
9. J.E. Dennis, J.J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19(1), 1977.
10. K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)*, 48(3):769–783, 1998.
11. S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles (Paris, 1962)*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.
12. S. Łojasiewicz. Sur les ensembles semi-analytiques. In *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2*, pages 237–241. Gauthier-Villars, Paris, 1971.
13. J. Nocedal, S.J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.