A recursive trust-region method in infinity norm for bound-constrained nonlinear optimization

by S. Gratton[1], M. Mouffe[1],
Ph. L. Toint[2] and M. Weber-Mendonça[2]

Report 07/01 23 April 2007

[1] C.E.R.F.A.C.S.
Avenue Gustave Coriolis, Toulouse, France
Email: gratton@cerfacs.fr, melodie.mouffe@cerfacs.fr

[2] Department of Mathematics,
University of Namur,
61, rue de Bruxelles, B-5000 Namur, Belgium,
Email: philippe.toint@fundp.ac.be, mwm@math.fundp.ac.be

A recursive trust-region method in infinity norm for bound-constrained nonlinear optimization

Serge Gratton Mélodie Mouffe Philippe L. Toint
Melissa Weber-Mendonça

23 April 2007

Abstract

A recursive trust-region method is introduced for the solution of bound-constrained nonlinear nonconvex optimization problems for which a hierarchy of descriptions exists. Typical cases are infinite-dimensional problems for which the levels of the hierarchy correspond to discretization levels, from coarse to fine. The new method uses the infinity norm to define the shape of the trust region, which is well adapted to the handling of bounds and also to the use of successive coordinate minimization as a smoothing technique. Some numerical tests are presented to motivate a theoretical analysis showing convergence to first-order critical points irrespective of the given starting point.

Keywords: Recursive methods, multilevel problems, nonlinear optimization, convergence theory.

1 Introduction

Trust-region methods are among the most popular and efficient methods for nonlinear optimization, and they are supported by an extensive theory (see Conn, Gould and Toint, 2000 for a more complete coverage of this subject). Such methods proceed iteratively by minimizing a model of the objective function in a region where the model can be trusted and which is defined in a specific norm. When the problem at hand can be decomposed into hierarchical levels, advantage can be taken of this structure. Several authors have proposed methods that take multilevel hierarchies into account, such as Fisher (1998), Nash (2000), Lewis and Nash (2002, 2005, and Oh, Milstein, Bouman and Webb, 2005). Gratton, Sartenaer and Toint (2004) have proposed a recursive trust-region algorithm that is tailored to the multilevel case and uses the Euclidean norm to define the trust region. The initial numerical experiments with this algorithm are extremely promising (see Gratton, Sartenaer and Toint, 2006*a*) and motivate further analysis of methods of this type.

While theoretically satisfying and practically reasonable, the choice of the Euclidean norm for the trust region definition is not without drawbacks. In particular, it implies the definition of computationally expensive preconditioners that specify compatible norms at the different levels. The technique used for updating the radius of the trust region also has the consequence that steps may, in some cases, be unduly restricted at lower levels, thereby potentially limiting the efficiency of the minimization. Moreover, the combination of Gauss-Seidel-like smoothing iterations (a very common feature of multilevel algorithms) with the Euclidean trust region is unnatural and complicates the practical implementations. Finally, and crucially for our concern in this paper, Euclidean trust regions do not mix naturally with bound-constrained problems.

In order to avoid these difficulties, an alternative multilevel algorithm for bound-constrained optimization can be defined using the infinity- (or max-) norm for the trust

region definition. The first purpose of this paper is to describe this algorithm, which does not require any imposed preconditioner and is much less restrictive for the lower-levels steps than its Euclidean relative for the unconstrained case. Moreover, smoothing iterations, which explore directions aligned with the coordinate vectors, are well integrated.

Unfortunately, the convergence theory presented in Gratton et al. (2004, 2006*b*) cannot be applied to this case without significant modifications, not only because of the possible presence of bounds, but also because the algorithm analyzed in these references is itself very dependent on the choice of the Euclidean norm. Our second purpose is thus to prove global convergence of the new algorithm to first-order critical points, that is convergence from arbitrary starting points to limit points satisfying the first-order optimality conditions.

Remarkably, the algorithm and theory presented here also apply, with minimal adaptations, to the problem of solving sets of nonlinear equations. Indeed, one of the most common techniques in this area is to consider the minimization of some (smooth) norm of the residuals, which can then be viewed as an unconstrained minimization problem, whose solution yields the desired roots if the residuals converge to zero. As a consequence, the multilevel algorithm discussed in the paper also applies to the multilevel solution of nonlinear equations, as does the associated global convergence proof.

The paper is organized as follows. Section 2 introduces the considered problem more formally and describes the alternative algorithm. Section 3 presents some numerical motivation and Section 4 analyzes its convergence properties. Some conclusions and perspectives are finally discussed in Section 5.

2 The problem and algorithm

In what follows, we wish to solve the bound-constrained minimization problem

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x), \\ \text{such that} & l \leq x \leq u \end{array} \tag{2.1}$$

where the $\leq$ signs (here and below) are understood componentwise, where l and u are vectors in $\mathbb{R}^n$ such that $l \leq u$, and where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function which is bounded below on the feasible set $\{x \in \mathbb{R}^n \mid l \leq x \leq u\}$.

This problem is viewed as an accurate representation of a more general underlying problem (such as, for instance, a contact problem in infinite dimensions). Since our interest is in the multilevel case, we also suppose that we know a set of functions $\{f_i\}_{i=0}^r$ which give alternative and potentially less accurate descriptions of the same underlying problem. Each of these f_i is assumed to be itself a twice continuously differentiable function from $\mathbb{R}^{n_i}$ to $\mathbb{R}$ (with $n_i \geq n_{i-1}$), $n_r = n$ and $f_r(x) = f(x)$ for all $x \in \mathbb{R}^n$. Each of these descriptions is said to define a *level*, which we index by i. As the existence of more than one level would not be very useful without ways to relate them among themselves, we also assume that, for each $i = 1, \ldots, r$, there exists an operator $R_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{i-1}}$ (the *restriction*) and another operator $P_i : \mathbb{R}^{n_{i-1}} \to \mathbb{R}^{n_i}$ (the *prolongation*) such that

$$R_i^T = \sigma_i P_i \tag{2.2}$$

for some known constant $\sigma_i > 0$. The prolongations and restrictions therefore define a hierarchy of levels, from lowest ($i = 0$) to highest ($i = r$). These assumptions are common to a number of multilevel approaches in optimization (Fisher, 1998, Nash, 2000, Gratton et al., 2004)) or in the solution of nonlinear systems of equations (see Briggs, Henson and McCormick, 2000 and the references therein). For simplicity of notations, and because this is often the case in practice, we assume, without loss of generality, that $\|R_i\|_\infty = 1$ for all i.

As indicated above, trust-region methods compute a step by minimizing a model of the objective function in the trust region, but we also choose to keep our iterates feasible throughout the process, which implies that the model minimization must take place in

the intersection of the feasible set of (2.1) with the trust region. In the classical (single level) case, the step from the iterate x_k (at iteration k) is thus obtained from the (possibly approximate) solution of the subproblem

$$\begin{aligned} \min_{s \in \mathcal{B}_k} \quad & m_k(x_k + s), \\ \text{such that} \quad & l \leq x_k + s \leq u \end{aligned} \tag{2.3}$$

where $m_k(x_k + s)$ is the objective function's model around x_k and the trust region $\mathcal{B}_k$ is defined as a ball in a suitable norm, whose radius is iteratively adapted by the method. In what follows we will use the infinity norm in this definition, as motivated in the introduction. In particular, if

$$\mathcal{F} \stackrel{\rm def}{=} \{x \in \mathbb{R}^n \mid l \leq x \leq u\}, \tag{2.4}$$

then the set $\mathcal{B}_k \cap \mathcal{F}$ is also a n-dimensional box. Classical trust-region methods are based on the Taylor's quadratic model for f given by

$$m_k(x_k + s) = f(x_k) + \langle \nabla f(x_k), s \rangle + \tfrac{1}{2} \langle s, H_k s \rangle,$$

where H_k is a symmetric approximation of $\nabla_{xx} f(x_k)$.

When the problem has two levels (r and $r-1$), the main idea is to use f_{r-1} to construct a model h_{r-1} for $f_r = f$ in the neighbourhood of the current iterate, which is cheaper than Taylor's quadratic model at level r, and to use this "lower-level" model whenever possible to define a step in a "feasible trust region" at level $r-1$. As this sentence suggests, we really need to handle two types of constraints for this subproblem, which we consider successively.

The first set of constraints are the "hard" bound constraints specified in the problem statement (2.1) which define the feasible region (2.4) at the top level ($i = r$). Because we wish to *maintain all our iterates feasible with respect to these bounds* and because the concatenation $P_r R_r$ of the restriction and prolongation operators may not be contractive, the "feasible domain" at level $r-1$ must be chosen with care. We choose here to follow the proposal by Gelman and Mandel (1990) and to define this set by

$$\mathcal{F}_{r-1} \stackrel{\rm def}{=} \{x_{r-1} \mid l_{r-1} \leq x_{r-1} \leq u_{r-1}\} \tag{2.5}$$

where $\mathcal{F}_r = \mathcal{F}$ and where the vectors l_{r-1} and u_{r-1} are defined componentwise by

$$[l_{r-1}]_j \stackrel{\rm def}{=} [R_r x_{r,k}]_j + \frac{1}{\|P_r\|_\infty} \max_{t=1,\ldots,n_r} \left\{ \begin{array}{lll} [l - x_{r,k}]_t & \text{when} & [P_r]_{tj} > 0 \\ [x_{r,k} - u]_t & \text{when} & [P_r]_{tj} < 0 \end{array} \right\} \tag{2.6}$$

and

$$[u_{r-1}]_j \stackrel{\rm def}{=} [R_r x_{r,k}]_j + \frac{1}{\|P_r\|_\infty} \min_{t=1,\ldots,n_r} \left\{ \begin{array}{lll} [u - x_{r,k}]_t & \text{when} & [P_r]_{tj} > 0 \\ [x_{r,k} - l]_t & \text{when} & [P_r]_{tj} < 0 \end{array} \right\} \tag{2.7}$$

for $j = 1, \ldots, n_{r-1}$. Gelman and Mandel (1990) show, in a less general case, that this definition enforces the inclusion

$$x_{r,k} + P_r(x - R_r x_{r,k}) \in \mathcal{F}_r \qquad \text{for all } \; x \in \mathcal{F}_{r-1}.$$

We verify this property in our more general setting in Lemma 4.3 below.

The second set of constraints at level $r-1$ are (infinity-norm) trust-region constraints inherited from iterate $x_{r,k}$ at level r (for consistency, where we define $\mathcal{A}_r = \mathbb{R}^{n_r}$), which we represent componentwise ($t = 1, \ldots, n_{r-1}$) by

$$\mathcal{A}_{r-1} \stackrel{\rm def}{=} \{x_{r-1} \mid v_{r-1} \leq x_{r-1} \leq w_{r-1}\},$$

where

$$[v_{r-1}]_t = \sum_{u=1,[R_r]_{tu}>0}^{n_r} [R_r]_{tu}[\max(v_r, x_{r,k} - \Delta_{r,k}e)]_u + \sum_{u=1,[R_r]_{tu}<0}^{n_r} [R_r]_{tu}[\min(w_r, x_{r,k} + \Delta_{r,k}e)]_u$$

and

$$[w_{r-1}]_t = \sum_{u=1,[R_r]_{tu}>0}^{n_r} [R_r]_{tu}[\min(w_r, x_{r,k} + \Delta_{r,k}e)]_u + \sum_{u=1,[R_r]_{tu}<0}^{n_r} [R_r]_{tu}[\max(v_r, x_{r,k} - \Delta_{r,k}e)]_u$$

where $\Delta_{r,k}$ is the radius of the trust region

$$\mathcal{B}_{r,k} \stackrel{\rm def}{=} \{x_{r,k} + s \in \mathbb{R}^{n_r} \mid \|s\|_\infty \leq \Delta_{r,k}\}$$

around the iterate $x_{r,k}$ at level r, and where e is the vector of all ones of appropriate dimension and the min and max operators are understood componentwise when applied on vectors. These constraints are essentially "softer" than the hard bound constraints in the sense that it is not vital that they are satisfied exactly by all iterates: as will be shown below, a bounded violation of the trust-region constraints is acceptable in our algorithm.

We then propose to take both sets of constraints into account at level $r-1$, and to *minimize the (potentially nonquadratic) model h_{r-1} on the intersection of these two domains*, that is on the box

$$\mathcal{L}_{r-1} \stackrel{\rm def}{=} \mathcal{F}_{r-1} \cap \mathcal{A}_{r-1},$$

starting from the initial point $x_{r-1,0} = R_r x_{r,k}$. This minimization is again computed *using a trust-region algorithm.* Thus we perform one or more steps of a trust-region method starting from $x_{r-1,0}$. This introduces a "level $r-1$" trust region

$$\mathcal{B}_{r-1,\ell} \stackrel{\rm def}{=} \{x_{r-1,\ell} + s \in \mathbb{R}^{n_{r-1}} \mid \|s\|_\infty \leq \Delta_{r-1,\ell}\}$$

for each step $s_{r-1,p}$ ($p \geq 0$) in this lower minimization process. The effective domain (working set) in which the $p+1$-th iterate at level $r-1$ is computed is thus, finally,

$$\mathcal{W}_{r-1,\ell} \stackrel{\rm def}{=} \mathcal{F}_{r-1} \cap \mathcal{A}_{r-1} \cap \mathcal{B}_{r-1,\ell}. \tag{2.8}$$

The minimization of h_{r-1} is then carried out using our trust-region algorithm until an approximate constrained minimizer $x_{r-1,*}$ is found in $\mathcal{L}_{r-1}$, and the resulting step $x_{r-1,*} - x_{r-1,0}$ is then prolongated to level r by computing $s_{r,k} = P_r(x_{r-1,*} - x_{r-1,0})$.

If more than two levels are available ($r > 1$), this can be done recursively, the approximation process stopping at level 0, where Taylor's model is always used. Let us reconsider the details of this process in this more general situation. Consider iteration k at level i, and assume that $x_{i,k}$ is an iterate in the minimization of h_i initiated from iterate $x_{i+1,q}$. At level i, we know

$$\mathcal{F}_i \stackrel{\rm def}{=} \{x \mid l_i \leq x \leq u_i\} \tag{2.9}$$

the "restricted" feasible domain at level i, where

$$[l_i]_j \stackrel{\rm def}{=} [x_{i,0}]_j + \frac{1}{\|P_{i+1}\|_\infty} \max_{t=1,\ldots,n_{i+1}} \left\{ \begin{array}{lll} [l_{i+1} - x_{i+1,q}]_t & \text{when} & [P_{i+1}]_{tj} > 0 \\ [x_{i+1,q} - u_{i+1}]_t & \text{when} & [P_{i+1}]_{tj} < 0 \end{array} \right\} \tag{2.10}$$

and

$$[u_i]_j \stackrel{\rm def}{=} [x_{i,0}]_j + \frac{1}{\|P_{i+1}\|_\infty} \min_{t=1,\ldots,n_{i+1}} \left\{ \begin{array}{lll} [u_{i+1} - x_{i+1,q}]_t & \text{when} & [P_{i+1}]_{tj} > 0 \\ [x_{i+1,q} - l_{i+1}]_t & \text{when} & [P_{i+1}]_{tj} < 0 \end{array} \right\} \tag{2.11}$$

for $j = 1, \ldots, n_i$. We also know

$$\mathcal{A}_i = \{x \mid v_i \leq x \leq w_i\}, \tag{2.12}$$

the restriction of the trust-region constraints inherited from levels r to $i+1$ through $x_{i+1,q}$, where the t-th components of v_i and w_i are defined by

$$\begin{aligned}[v_i]_t &= \sum_{u=1,[R_{i+1}]_{tu}>0}^{n_{i+1}} [R_{i+1}]_{tu}[\max(v_{i+1}, x_{i+1,q} - \Delta_{i+1,q}e)]_u \\ &\quad + \sum_{u=1,[R_{i+1}]_{tu}<0}^{n_{i+1}} [R_{i+1}]_{tu}[\min(w_{i+1}, x_{i+1,q} + \Delta_{i+1,q}e)]_u\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}[w_i]_t &= \sum_{u=1,[R_{i+1}]_{tu}>0}^{n_{i+1}} [R_{i+1}]_{tu}[\min(w_{i+1}, x_{i+1,q} + \Delta_{i+1,q}e)]_u \\ &\quad + \sum_{u=1,[R_{i+1}]_{tu}<0}^{n_{i+1}} [R_{i+1}]_{tu}[\max(v_{i+1}, x_{i+1,q} - \Delta_{i+1,q}e)]_u\end{aligned} \tag{2.14}$$

(as indicated above, we define $v_r = -\infty$ and $w_r = +\infty$ for consistency). We finally know

$$\mathcal{B}_{i,k} = \{x_{i,k} + s \in \mathbb{R}^{n_i} \mid \|s\|_\infty \leq \Delta_{i,k}\},$$

the current trust region at level i associated with $x_{i,k}$. We then have to find a step $s_{i,k}$ which sufficiently reduces a model of h_i in the region

$$\mathcal{W}_{i,k} = \mathcal{F}_i \cap \mathcal{A}_i \cap \mathcal{B}_{i,k}. \tag{2.15}$$

Observe that the set $\mathcal{W}_{i,k}$ can either be viewed both as

$$\mathcal{W}_{i,k} = \mathcal{L}_i \cap \mathcal{B}_{i,k}, \tag{2.16}$$

the intersection of a level dependent domain $\mathcal{L}_i \stackrel{\text{def}}{=} \mathcal{F}_i \cap \mathcal{A}_i$ with an iteration dependent trust-region $\mathcal{B}_{i,k}$, or as

$$\mathcal{W}_{i,k} = \mathcal{F}_i \cap \mathcal{S}_{i,k},$$

the intersection of $\mathcal{F}_i$, the feasible set for hard constraints, with $\mathcal{S}_{i,k} = \mathcal{A}_i \cap \mathcal{B}_{i,k}$, the feasible set for soft ones. Figure 2.1 illustrates this decomposition. Note that all the involved sets are boxes, which makes their representation and intersection computationally easy.

Once $\mathcal{W}_{i,k}$ is known, we then choose a model for h_i between

$$m_{i,k}(x_{i,k} + s_i) = h_i(x_{i,k}) + \langle g_{i,k}, s_i\rangle + \tfrac{1}{2}\langle s_i, H_{i,k}s_i\rangle, \tag{2.17}$$

the usual truncated Taylor series for h_i (with $g_{i,k} = \nabla_x h_i(x_{i,k})$ and $H_{i,k}$ being a symmetric approximation of $\nabla_{xx}h_i(x_{i,k})$), or

$$h_{i-1}(x_{i-1,0} + s_{i-1}) \stackrel{\text{def}}{=} f_{i-1}(x_{i-1,0} + s_{i-1}) + \langle R_i g_{i,k} - \nabla_x f_{i-1}(x_{i-1,0}), s_{i-1}\rangle. \tag{2.18}$$

Observe that this last definition implies that $g_{i-1,0} = R_i g_{i,k}$.

If one chooses the model (2.18) (which is only possible if $i > 0$), the determination of the step then consists in (possibly approximately) solving the lower-level bound-constrained problem

$$\min_{x_{i-1,0}+s_{i-1}\in\mathcal{L}_{i-1}} h_{i-1}(x_{i-1,0} + s_{i-1}). \tag{2.19}$$

This minimization produces a step s_{i-1} such that $h_{i-1}(x_{i-1,0}+s_{i-1}) < h_{i-1}(x_{i-1,0})$ which must be then brought back to level i by the prolongation P_i, i.e. $s_i = P_i s_{i-1}$. We then obtain that

$$\langle g_{i,k}, s_i\rangle = \langle g_{i,k}, P_i s_{i-1}\rangle = \frac{1}{\sigma_i}\langle R_i g_{i,k}, s_{i-1}\rangle. \tag{2.20}$$

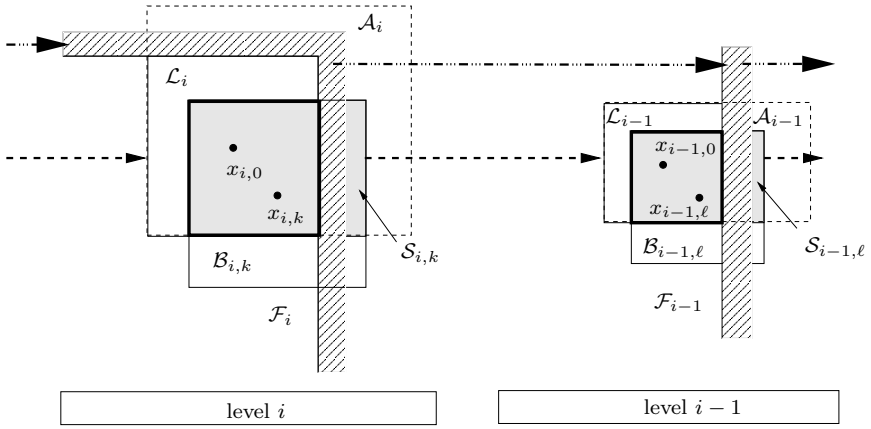

Figure 2.1: The various sets at levels i and $i-1$. The region infeasible with respect to the bounds defining $\mathcal{F}_i$ and $\mathcal{F}_{i-1}$ are shaded and the effect of the Gelman-Mandel restriction operator between levels is shown by the arrows on top of the figure. The sets $\mathcal{W}_{i,k}$ and $\mathcal{W}_{i-1,\ell}$ are delineated in bold and the sets $\mathcal{A}_i$ and $\mathcal{A}_{i-1}$ in dashed. The sets $\mathcal{S}_{i,k}$ and $\mathcal{S}_{i-1,\ell}$ are lightly shaded. The effect of the restriction operator R_i between $\mathcal{S}_{i,k}$ and $\mathcal{A}_{i-1}$ is illustrated by the (lower set of) dashed arrows.

But does it always make sense to use the lower level model (2.18)? The answer obviously depends on the benefit expected from the solution of (2.19). In Gratton et al. (2004), it sufficed to test if $\|g_{i-1,0}\|_2 = \|R_i g_{i,k}\|_2$ was large enough compared to $\|g_{i,k}\|_2$. However, this criticality measure is inadequate in our context because (2.19) is now a bound-constrained problem. In the sequel of this paper, we choose to follow Conn et al. (2000) and use, for each $x_{i,k} \in \mathcal{L}_i$, the criticality measure[(1)] defined by

$$\chi_{i,k} \stackrel{\rm def}{=} \chi(x_{i,k}) = |\min_{\substack{x_{i,k}+d\in\mathcal{L}_i \\ \|d\|_\infty\leq 1}} \langle g_{i,k}, d\rangle| \stackrel{\rm def}{=} |\langle g_{i,k}, d_{i,k}\rangle|. \tag{2.21}$$

If the restriction of the problem from the non-critical iterate $x_{i,k}$ at level i to level $i-1$ is not already first-order critical, that is if

$$\chi_{i-1,0} \geq \kappa_\chi \chi_{i,k}, \tag{2.22}$$

for some constant $\kappa_\chi \in (0,1)$, then we may proceed at this lower level. Otherwise, the recursion is useless and we should use (2.17) instead.

Once we have decided to approximately solve (2.19), we must also decide what we mean by "approximately". In the spirit of (2.22), we choose to terminate the minimization at level r if $\chi_{r,k} \leq \epsilon_r$ for some $\epsilon_r > 0$ and to terminate the lower level minimization at iterate $(i-1,\ell)$ as soon as the inequality

$$\chi_{i-1,\ell} < \epsilon_{i-1} \stackrel{\rm def}{=} \kappa_\chi \chi_{i,k}, \tag{2.23}$$

holds. We then define $x_{i-1,*} = x_{i-1,\ell}$, $s_{i-1} = x_{i-1,*} - x_{i-1,0}$ and $s_{i,k} = P_i s_{i-1}$.

If, on the other hand, we decide at iteration (i,k) to use Taylor's model $m_{i,k}$ given by (2.17), a step $s_{i,k}$ is then computed that produces a sufficient decrease in the value of this model in its usual meaning for trust-region methods with convex constraints (defined here by the set $\mathcal{L}_i$), that is, $s_{i,k}$ is such that it satisfies

$$m_{i,k}(x_{i,k}) - m_{i,k}(x_{i,k}+s_{i,k}) \geq \kappa_{\rm red}\chi_{i,k} \min\left[1, \frac{\chi_{i,k}}{\beta_{i,k}}, \Delta_{i,k}\right], \tag{2.24}$$

[(1)] Other criticality measures are possible, such as $\mu_{i,k} = \|{\rm Proj}_{i,k}(x_{i,k} - g_{i,k}) - x_{i,k}\|_2$ where ${\rm Proj}_{i,k}$ is the orthogonal projection onto the box $\mathcal{L}_i$ but we will not investigate this alternative here.

for some constant $\kappa_{\rm red} \in (0, \frac{1}{2})$ and $\beta_{i,k} \stackrel{\rm def}{=} 1 + \|H_{i,k}\|_2$. Despite its apparently technical character, this requirement, known as the modified Cauchy condition, is not overly restrictive and can be guaranteed in practical algorithms, as described for instance in Section 12.2.1 of Conn et al. (2000).

We now specify our algorithm formally, as Algorithm RMTR$_\infty$. As most trust-region methods, it uses the constants $0 < \eta_1 \leq \eta_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1$.

Algorithm 2.1: RMTR$_\infty(i, x_{i,0}, g_{i,0}, \chi_{i,0}, \mathcal{F}_i, \mathcal{A}_i, \epsilon_i)$

Step 0: Initialization. Compute $h_i(x_{i,0})$ and $g_{i,0} - \nabla_x f_i(x_{i,0})$. Set

$$\mathcal{L}_i = \mathcal{F}_i \cap \mathcal{A}_i \quad \text{and} \quad \mathcal{W}_{i,0} = \mathcal{L}_i \cap \{x \,|\, \|x - x_{i,0}\|_\infty \leq \Delta_{i,0}\},$$

with $\Delta_{i,0} = \Delta_i^{\rm s}$ and $k = 0$.

Step 1: Model choice. If $i = 0$, go to Step 3. Else, compute $\mathcal{L}_{i-1}$ and $\chi_{i-1,0}$. If (2.23) holds, go to Step 3. Otherwise, choose to go to Step 2 or to Step 3.

Step 2: Recursive step computation. Call Algorithm

$$\textbf{RMTR}_\infty(i-1, R_i x_{i,k}, R_i g_{i,k}, \chi_{i-1,0}, \mathcal{F}_{i-1}, \mathcal{A}_{i-1}, \kappa_\chi \chi_{i,k}),$$

yielding an approximate solution $x_{i-1,*}$ of (2.19). Then define $s_{i,k} = P_i(x_{i-1,*} - R_i x_{i,k})$, set $\delta_{i,k} = \frac{1}{\sigma_i}\big[h_{i-1}(R_i x_{i,k}) - h_{i-1}(x_{i-1,*})\big]$ and go to Step 4.

Step 3: Taylor step computation. Choose $H_{i,k}$ and compute a step $s_{i,k} \in \mathbb{R}^{n_i}$ that sufficiently reduces the model $m_{i,k}$ given by (2.17) in the sense of (2.24) and such that $x_{i,k} + s_{i,k} \in \mathcal{W}_{i,k}$. Set $\delta_{i,k} = m_{i,k}(x_{i,k}) - m_{i,k}(x_{i,k} + s_{i,k})$.

Step 4: Acceptance of the trial point. Compute $h_i(x_{i,k} + s_{i,k})$ and

$$\rho_{i,k} = \big[h_i(x_{i,k}) - h_i(x_{i,k} + s_{i,k})\big]/\delta_{i,k}. \tag{2.25}$$

If $\rho_{i,k} \geq \eta_1$, then define $x_{i,k+1} = x_{i,k} + s_{i,k}$; otherwise, define $x_{i,k+1} = x_{i,k}$.

Step 5: Termination. Compute $g_{i,k+1}$ and $\chi_{i,k+1}$. If $\chi_{i,k+1} \leq \epsilon_i$ or $x_{i,k+1} \notin \mathcal{L}_i$, then return with the approximate solution $x_{i,*} = x_{i,k+1}$.

Step 6: Trust-Region Update. Set

$$\Delta_{i,k+1} \in \begin{cases} [\Delta_{i,k}, +\infty) & \text{if } \rho_{i,k} \geq \eta_2, \\ [\gamma_2 \Delta_{i,k}, \Delta_{i,k}] & \text{if } \rho_{i,k} \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_{i,k}, \gamma_2 \Delta_{i,k}] & \text{if } \rho_{i,k} < \eta_1, \end{cases} \tag{2.26}$$

and

$$\mathcal{W}_{i,k+1} = \mathcal{L}_i \cap \{x \,|\, \|x - x_{i,k+1}\|_\infty \leq \Delta_{i,k+1}\}. \tag{2.27}$$

Increment k by one and go to Step 1.

Some comments are now necessary for a full understanding of this algorithm.

1. As usual in trust-region algorithms, iterations at which $\rho_{i,k} \geq \eta_1$ are called *successful*. At such iterations, the trial point $x_{i,k} + s_{i,k}$ is accepted as the new iterate and the radius of the correponding trust region is possibly enlarged. If the iteration is unsuccessful, the trial point is rejected and the radius is reduced.

2. The test for the value of i at the begining of Step 1 is designed to identify the lowest level, at which no further recursion is possible. In this case, a Taylor's iteration is the only choice left.

3. As we have observed above, we have that

$$\|x_{i,k+1} - x_{i,k}\|_\infty \leq \|P_i\|_\infty \|x_{i-1,*} - x_{i-1,0}\|_\infty,$$

which does not imply that $x_{i,k+1}$ belongs to $\mathcal{A}_i$. Even if this does not prevent convergence, there is no point continuing the recursive iteration started from $x_{i,k}$ since we have already taken a step which is too large. This is why we have included the second termination condition in Step 5. Note that we always have by construction that $x_{i,k+1} \in \mathcal{F}_i$, while the inclusion $x_{i,k+1} \in \mathcal{A}_i$ may fail at a recursive iteration.

4. The difference between the "restriction formulae" (2.9)-(2.11) for the hard bounds and (2.12)-(2.14) for the soft ones makes it necessary to pass both $\mathcal{A}_i$ and $\mathcal{F}_i$ to the algorithm at level i, as it is necessary to compute $\mathcal{L}_i$ at each level independently.

5. The choice left in Step 1 between recursive or Taylor's iterations is important in practice as it allows the exploitation of efficient multilevel techniques such as the smoothing iterations mentioned above (see Briggs et al., 2000 for an excellent introduction). In Gratton et al. (2006*a*), the same freedom is exploited (in the context of an Euclidean-norm method) to introduce periodic Gauss-Seidel-type smoothing iterations in the spirit of classical multigrid algorithms. The choice of the infinity norm instead of the Euclidean one is in fact more consistent. Indeed this smoothing process, which successively explores the coordinate directions, needs not be stopped or modified when the boundary of the trust region is met, since the process can then be continued in one of the faces of this boundary.

6. The original problem (2.1) is solved by calling RMTR$_\infty$ from a virtual $(r+1)$-rst level at which we assume the trust region to be infinite.

3 Numerical motivation

We now present some motivation and preliminary results showing the potential of this new method as a bound constrained minimizer and also as an alternative to the original Euclidean-norm RMTR algorithm by Gratton et al. (2004) for unconstrained problems. This last comparison is possible because RMTR$_\infty$ may be applied with arbitrarily large bounds to solve unconstrained problems.

Our purpose here is not to discuss an implementation of RMTR$_\infty$ or RMTR, but rather to provide some numerical intuition. We therefore refer the reader to Gratton et al. (2006*a*) for details beyond the brief description which follows (see also Herty and Thömmes, 2006). A crucial ingredient of our implementation, inspired by multilevel/multigrid algorithms, is the presence of smoothing iterations which reduce the oscillatory components of the gradient. As in Gratton et al. (2006*a*), we use here the Sequential Coordinate Minimization, which we abbreviate as SCM, since it is the exact equivalent of the Gauss-Seidel process in our optimization setting. The SCM smoother consists in the minimization of the model along each coordinate direction successively. Note that one SCM smoothing may consist of one or more cycles, a cycle being defined as a succession of n_i (at level i) one-dimensional minimizations of the model. We use V-type recursions, analogous to those defined for traditional multigrid (we discuss these in the context of RMTR at the end of Section 4). However, recursive iterations are still subject to satisfying (2.22), and may thus not always be possible. As alternative Taylor iterations, we then use the Projected Truncated Conjugate Gradient (PTCG) algorithm (see, for instance, Calamai and Moré, 1987, Conn, Gould and Toint, 1988, Lin and Moré, 1999 or Conn, Gould and Toint, 1992). Gratton et al. (2004) solved the coarsest level Euclidean-norm subproblem accurately by the Hebden-Moré algorithm (Moré, 1978). However, this efficient technique is not applicable when the infinity norm is used to define the trust-region, and we therefore chose to also use the PTCG algorithm to obtain an approximate solution at this level.

The test problems presented here are three of the problems previously discussed in Gratton et al. (2006*a*) and are defined on the two-dimensional unit square S_2.

Surf: A minimum surface problem. We want to find the solution to

$$\min_{v\in\mathcal{K}} \int_0^1 \int_0^1 (1+(\partial_x v)^2+(\partial_y v)^2)^{\frac{1}{2}} dx\, dy,$$

where $\mathcal{K} = \{v \in H^1(S_2) \mid v(x,y) = v_0(x,y) \text{ on } \partial S_2\}$. The boundary condition v_0 is chosen as

$$v_0(x,y) = \begin{cases} f(x), & \text{for} \quad y = 0, \quad 0 \le x \le 1, \\ 0, & \text{for} \quad x = 0, \quad 0 \le y \le 1, \\ f(x), & \text{for} \quad y = 1, \quad 0 \le x \le 1, \\ 0, & \text{for} \quad x = 1, \quad 0 \le y \le 1. \end{cases}$$

where $f(x) = x(1-x)$. To do this, we discretize the problem using a finite element basis, defined by a uniform triangulation of S_2, with same grid spacing h along the 2 coordinate directions. We use the classical P1 functions which are linear on each triangle and take value 0 or 1 at each vertex as basis functions.

Q2: A simple quadratic problem, where we consider the two-dimensional boundary value problem defined by

$$\begin{aligned} -\Delta u(x,y) &= f \quad \text{in } S_2 \\ u(x,y) &= 0 \quad \text{on } \partial S_2, \end{aligned}$$

where f is such that the analytical solution to this problem is

$$u(x,y) = \sin(2\pi x(1-x))\sin(2\pi y(1-y)).$$

This problem is discretized using a 5-points finite-difference scheme, giving linear systems $A_i x = b_i$ at level i where each A_i is a symmetric positive-definite matrix. This problem is the typical *model problem* for multigrid solvers. Here, we apply Algorithm RMTR and RMTR$_\infty$ on its variational formulation

$$\min_{x\in\mathbb{R}^{n_r}} \frac{1}{2} x^T A_r x - x^T b_r.$$

NC: A nonconvex optimal control problem, defined as the nonlinear least squares problem

$$\min_{u,\gamma} \mathcal{I}(u,\gamma) = \int_{S_2} (u-u_0)^2 + \int_{S_2} (\gamma-\gamma_0)^2 + \int_{S_2} f^2,$$

where

$$-\Delta u + \gamma u - f_0 = f \quad \text{in } S_2, \tag{3.28}$$

$$u = 0 \quad \text{on } \partial S_2. \tag{3.29}$$

The unknown functions u and γ are defined in the unit square S_2, and $\gamma_0(x,y)$ and $u_0(x,y)$ are also defined on S_2 by $\gamma_0(x,y) = u_0(x,y) = \sin(x(1-x))\sin(y(1-y))$. The function f_0 is such that $-\Delta u_0 + \gamma_0 u_0 = f_0$ on S_2. This problem corresponds to a penalized version of a constrained optimal control problem, and is discretized using finite differences.

The results obtained are shown in Tables 3.1 and 3.2. We first consider unconstrained optimization problems. The first two problems were tested using 7 levels of discretization, with 3^2 variables in the coarsest level and 255^2 variables in the finest level, and for problem NC we used 7 levels of discretization, with 18 variables in the coarsest level and 130050 variables in the finest level. Since in the highest level the cost is dominated by the number

	RMTR		RMTR_∞	
	f	s/c	f	s/c
Surf	21	10/10	15	10/10
Q2	73	9/12	10	3/6
NC	18	9/33	14	5/15

Table 3.1: Computational work at the finest level for the ∞- and Euclidian-norm versions of the RMTR algorithm

of evaluation of the cost function f and the number of smoothing iterations s and smoothing cycles c, we report these numbers in the table for both algorithms using respectively, f and s/c. All the tests reported were obtained using an experimental MATLAB code.

We see that RMTR requires more work (function evaluations and smoothing) at the finest level, that is where these calculations are more costly. RMTR_∞, on the other hand, tends to exploit lower levels more, where computation is cheap. Although this is not visible in the table, our experience indicates that the comparative advantage of RMTR_∞ becomes even more pronounced when the number of variables at the finest level grows.

We now compare RMTR_∞ with mesh refinement optimization, a technique for multilevel optimization problems, where the discretized problems are solved in turn from the coarsest level to the finest one (see, for example, Frese, Bouman and Sauer, 1999, Lemarchand, Pironneau and Polak, 2001, or Borzi and Kunisch, 2006). This technique is known to be considerably more efficient that the solution of the problem using its fine level representation only. The successive solutions are obtained, in our test, by a classical single-level trust-region algorithm in which PTCG is used to solve the trust-region subproblem. The starting point at level $i+1$ is obtained by prolongating (using P_{i+1}) the solution obtained at level i. The following results are obtained on the minimal surface problem *Surf* with an obstacle defined by the lower bound

$$l = \begin{cases} 2 & \text{if} \quad \frac{4}{9} \leq x \leq \frac{5}{9} \text{ and } \frac{4}{9} \leq y \leq \frac{5}{9} \\ 0 & \text{else} \end{cases} .$$

As previously, the problem was tested using 7 levels of discretization, with 3^2 variables in the coarsest level and 255^2 variables in the finest level. We present the number of function evaluation f on the finest level and we also compare the number of SCM cycles for RMTR_∞ with the number of PTCG iterations for the mesh refinement on the finest level (note that the work for these procedures is comparable, since they both require a number of operations proportional to the number of nonzeros in H_r).

	Mesh refinement		RMTR_∞	
	f	PTCG	f	s/c
Surf	145	5553	246	182/182

Table 3.2: Compared results for the mesh refinement and RMTR_∞

We see that RMTR_∞ requires considerably less iterations than the mesh refinement technique. Conversely, RMTR_∞ needs more objective function evaluations. As the cost of an evaluation is the same order as a smoothing cycle, we conclude that the total amount of work required by RMTR_∞ to solve the problem remains significantly smaller.

As can be seen from these (admittedly limited) examples, the new algorithm is both more efficient than its Euclidean-norm predecessor, and even more advantageous compared to mesh-refinements techniques.

4 Convergence theory

Having motivated our interest in the new method, both as an efficient solver for bound-constrained problems and as a significant improvement on the existing RMTR algorithm for the unconstrained case, we are now interested in obtaining a theoretical guarantee that RMTR$_\infty$ converges to a first-order critical point of the problem from any starting point. The theory proposed in this section differs significantly from the proof for the RMTR algorithm in Gratton et al. (2004), mostly because of the form of the new criticality measure (imposed by the bounds and the choice of the infinity norm) and because the new algorithm allows for potentially very asymmetric trust regions.

We start by making our assumptions more formal. First, we assume that the Hessians of each h_i and their approximations are bounded above by the constant $\kappa_{\rm H} \geq 1$, i.e., more formally, that for $i = 0, \ldots, r$,

$$1 + \|\nabla_{xx} h_i(x_i)\|_2 \leq \kappa_{\rm H} \tag{4.30}$$

for all $x_i \in \mathcal{F}_i$. and

$$\beta_{i,k} \leq \kappa_{\rm H} \tag{4.31}$$

for all k. We also assume that all gradients at all levels remain uniformly bounded, which is to say that there exists $\kappa_{\rm g} \geq 1$ such that

$$\|\nabla_x h_j(x_i)\|_2 \leq \kappa_{\rm g} \quad \text{for all} \quad j = 0, \ldots, r, \quad \text{and all} \quad x_i \in \mathcal{F}_i. \tag{4.32}$$

This assumption is not overly restrictive and, for instance, automatically holds by continuity if all iterates $x_{j,\ell}$ remain in a bounded domain, which is the case if both l and u are finite in (2.1).

We also define some additional notation and concepts. We first choose the constant $\kappa_{\rm P} \geq 1$ such that

$$\|P_i\|_\infty \leq \kappa_{\rm P} \qquad \text{for all } \; i = 1, \ldots, r. \tag{4.33}$$

If we choose to go to Step 2 (i.e. we choose to use the model (2.18)) at iteration (i, k) , we say that this iteration initiates a *minimization sequence* at level $i-1$, which consists of all successive iterations *at this level* (starting from the point $x_{i-1,0} = R_i x_{i,k}$) until a return is made to level i within iteration (i, k). In this case, we say that iteration (i, k) is the *predecessor* of the minimization sequence at level $i-1$. If $(i-1, \ell)$ belongs to this minimization sequence, this is written as $(i, k) = \pi(i-1, \ell)$. We also denote by p_{i-1} the index of the penultimate iterate in the minimization sequence $\{x_{i-1,0}, \ldots, x_{i-1,p_{i-1}}, x_{i-1,*}\}$. Note that (2.15) implies that $\mathcal{W}_{i,k} \subseteq \mathcal{B}_{i,k}$.
We first prove a useful property of the criticality measure $\chi(\cdot)$ in our context.

Lemma 4.1 *Consider the optimization problem (2.1) and define the function $\chi(x)$ by*

$$\chi(x) = |\min_{\substack{x+d \in \mathcal{F} \\ \|d\|_\infty \leq 1}} \langle \nabla_x f(x), d \rangle| \tag{4.34}$$

(as in (2.21)). Then, for all $x, y \in \mathcal{F}$, we have that

$$|\chi(x) - \chi(y)| \leq (n\kappa_H + \kappa_g)\sqrt{n}\|x - y\|_\infty.$$

Proof. Denote by $d(x)$ the argument of the minimum in (4.34) and consider the continuous feasible path from x to y given by the segment $[x, y]$. At any point z on this path, the vector $d(z)$ can be chosen as an extreme point of the polytope $\mathcal{P}(z) = \{z + d \in \mathcal{F} \mid \|d\|_\infty \leq 1\}$ which is determined by the set of bounds from $\mathcal{F}$ or from the inequality $\|d\|_\infty \leq 1$ which are active at $z + d(z)$. It is always possible to decompose the segment $[x, y]$ into a (possibly infinite) set of subsegments such that the set of active constraints

(for (4.34)) is constant along each subsegment. Consider now z_1 and z_2 the endpoints of one such subsegment. Then

$$\begin{aligned} |\chi(z_1) - \chi(z_2)| &= |\langle \nabla_x f(z_2), d(z_2)\rangle - \langle \nabla_x f(z_1), d(z_1)\rangle| \\ &= |\langle \nabla_x f(z_2) - \nabla_x f(z_1), d(z_2)\rangle - \langle \nabla_x f(z_1), d(z_1) - d(z_2)\rangle| \\ &\leq \|\nabla_x f(z_2) - \nabla_x f(z_1)\|_2 \|d(z_2)\|_2 + \|\nabla_x f(z_1)\|_2 \|d(z_1) - d(z_2)\|_2 \end{aligned} \tag{4.35}$$

From the mean-value theorem, we know that $\nabla_x f(z_1) = \nabla_x f(z_2) + G_{[z1,z2]}(z_1 - z_2)$, where, from (4.30)

$$\|G_{[z1,z2]}\|_2 = \| \int_0^1 \nabla_{xx} f(z_1 + t(z_2 - z_1))\, dt\|_2 \leq \max_{z \in [z_1, z_2]} \|\nabla_{xx} f(z)\|_2 \leq \kappa_{\rm H} \sqrt{n}$$

Hence, since $\|d(z_2)\|_2 \leq \sqrt{n} \|d(z_2)\|_\infty \leq \sqrt{n}$,

$$|\chi(z_1) - \chi(z_2)| \leq n\kappa_{\rm H} \|z_1 - z_2\|_2 + \kappa_{\rm g} \|d(z_1) - d(z_2)\|_2.$$

Now, remembering that the set of active constraints is the same at z_1 and z_2, we consider the j-th component of z_1 and distinguish two cases. The first is when a bound from $\mathcal{F}$, $[u]_j$ say, is active at both $z_1 + d(z_1)$ and $z_2 + d(z_2)$. In this case,

$$[z_1 + d(z_1)]_j = [z_2 + d(z_2)]_j = [u]_j$$

and we deduce that

$$[d(z_1) - d(z_2)]_j = -[z_1 - z_2]_j.$$

The second case is when a unit bound from the inequality $\|d(z_1)\|_\infty \leq 1$ is active. Assuming without loss of generality that the upper bound is active, we then have that

$$[d(z_1)]_j = [d(z_2)]_j = 1$$

yielding that $[d(z_1) - d(z_2)]_j = 0$. Combining the two cases, we conclude that

$$\|d(z_1) - d(z_2)\|_2 \leq \|z_1 - z_2\|_2,$$

and hence, using (4.35) that, for z_i in the path from x to y,

$$\begin{aligned} |\chi(x) - \chi(y)| &= \big|\sum_i \chi(z_i) - \chi(z_{i+1})\big| \\ &\leq \sum_i |\chi(z_i) - \chi(z_{i+1})| \\ &\leq n\kappa_{\rm H} \sum_i \|z_i - z_{i+1}\|_2 + \kappa_{\rm g} \sum_i \|d(z_i) - d(z_{i+1})\|_2 \\ &\leq n\kappa_{\rm H} \sum_i \|z_i - z_{i+1}\|_2 + \kappa_{\rm g} \sum_i \|z_i - z_{i+1}\|_2 \\ &= (n\kappa_{\rm H} + \kappa_{\rm g}) \|x - y\|_2 \\ &\leq (n\kappa_{\rm H} + \kappa_{\rm g}) \sqrt{n} \|x - y\|_\infty, \end{aligned}$$

which gives the desired conclusion. □

The algorithm also ensures the following technical lemma.

Lemma 4.2 *There exists an $\epsilon_{\min} \in (0, 1]$ such that, for each iteration $(i, k) \neq (i, *)$ (i.e., for all iterates at level i but the last one),*

$$\chi_{i,k} \geq \epsilon_{\min}. \tag{4.36}$$

Proof. The inequality (2.23), which is the stopping criteria for minimization at level j, in Step 5 of the algorithm, implies that for all (i,k) and all $(j,\ell) \in \mathcal{R}(i,k)$,

$$\chi_{j,\ell} \geq \epsilon_j = \kappa_\chi \chi_{\pi(j,\ell)} \geq \kappa_\chi \epsilon_{j+1} = \kappa_\chi^2 \chi_{\pi^2(j,\ell)} \geq \cdots \geq \kappa_\chi^{i-j} \chi_{i,k} \geq \cdots \geq \kappa_\chi^r \epsilon_r.$$

This proves (4.36) with $\epsilon_{\min} = \min[1, \kappa_\chi^r \epsilon_r]$. □

We now prove the general version of the Gelman and Mandel's result stating that "bound constraints are preserved" by the prolongation operator.

Lemma 4.3 *The definitions (2.10)–(2.11) enforce the inclusion*

$$x_{i,k} + P_i(x_{i-1} - x_{i-1,0}) \in \mathcal{F}_i \quad \text{for all } x_{i-1} \in \mathcal{F}_{i-1} \tag{4.37}$$

for $i = 1, \ldots, r$. As a consequence $x_{i,k} \in \mathcal{F}_i$ for all $i = 0, \ldots, r$ and all $k \geq 0$.

Proof. For $t = 1, \ldots, n_i$, define $\phi_{i,t} = \sum_{j=1}^{n_{i-1}} |[P_i]_{t,j}|$ and observe that $\phi_{i,t} \leq \|P_i\|_\infty$ for all t. Consider now any $x_{i-1} \in \mathcal{F}_{i-1}$ and the corresponding lower level step $s_{i-1} = x_{i-1} - x_{i-1,0}$. Then (2.10) and (2.11) imply that

$$\begin{aligned}
&[x_{i,k}]_t + \sum_{j=1}^{n_{i-1}} [P_i]_{tj}[s_{i-1}]_j \\
&\quad = [x_{i,k}]_t + \sum_{j=1,[P_i]_{tj}<0}^{n_{i-1}} |[P_i]_{tj}|(-[s_{i-1}]_j) + \sum_{j=1,[P_i]_{tj}>0}^{n_{i-1}} |[P_i]_{tj}|[s_{i-1}]_j \\
&\quad \geq [x_{i,k}]_t + \sum_{j=1,[P_i]_{tj}<0}^{n_{i-1}} |[P_i]_{tj}| \frac{(-\min_t[x_{i,k} - l_i]_t)}{\|P_i\|_\infty} + \sum_{j=1,[P_i]_{tj}>0}^{n_{i-1}} |[P_i]_{tj}| \frac{\max_t[l_i - x_{i,k}]_t}{\|P_i\|_\infty} \\
&\quad \geq [x_{i,k}]_t + \sum_{j=1,[P_i]_{tj}<0}^{n_{i-1}} |[P_i]_{tj}| \frac{[l_i - x_{i,k}]_t}{\|P_i\|_\infty} + \sum_{j=1,[P_i]_{tj}>0}^{n_{i-1}} |[P_i]_{tj}| \frac{[l_i - x_{i,k}]_t}{\|P_i\|_\infty} \\
&\quad \geq [x_{i,k}]_t + \phi_{i,t} \frac{[l_i - x_{i,k}]_t}{\|P_i\|_\infty} \\
&\quad = \frac{\phi_{i,t}}{\|P_i\|_\infty}[l_i]_t + \left(1 - \frac{\phi_{i,t}}{\|P_i\|_\infty}\right)[x_{i,k}]_t \\
&\quad \geq [l_i]_t
\end{aligned}$$

where the last inequality results from the fact that $[x_{i,k}]_t \geq [l_i]_t$. A similar reasoning gives that

$$[x_{i,k}]_t + \sum_{j=1}^{n_{i-1}} [P_i]_{t,j}[s_{i-1}]_j \leq [u_i]_t$$

for all t, thereby concluding the proof of (4.37). The feasibility of every iterate with respect to the level-dependent bound constraints then results from the fact that all trial points at level i belong to $\mathcal{L}_i \subseteq \mathcal{F}_i$ by construction. □

We next show that the distance from all iterates in a single minimization sequence at level i to the starting point of that sequence is bounded above by a multiple of the trust-region radius at the predecessor's level.

Lemma 4.4 *The definitions (2.13)-(2.14) imply that, for $0 \leq j < r$,*

$$\|x - x_{j,0}\|_\infty \leq 2\Delta_{\pi(j,0)} \tag{4.38}$$

for all $x \in \mathcal{L}_j$.

Proof. Consider an $x \in \mathcal{L}_j \subseteq \mathcal{A}_j$. If we now denote the bounds defining the set $\mathcal{S}_{\pi(j,0)}$ by

$$\overline{v}_{j+1} \stackrel{\rm def}{=} \max\left[v_{j+1}, x_{\pi(j,0)} - \Delta_{\pi(j,0)}e\right] \quad \text{and} \quad \overline{w}_{j+1} \stackrel{\rm def}{=} \min\left[w_{j+1}, x_{\pi(j,0)} + \Delta_{\pi(j,0)}e\right],$$

we then verify that

$$\begin{aligned}
[w_j - v_j]_t &= \sum_{q=1,[R_{j+1}]_{tq}>0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{w}_{j+1}]_u + \sum_{q=1,[R_{j+1}]_{tq}<0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{v}_{j+1}]_u \\
&\qquad - \sum_{q=1,[R_{j+1}]_{tq}<0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{v}_{j+1}]_u - \sum_{q=1,[R_{j+1}]_{tq}<0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{w}_{j+1}]_u \\
&= \sum_{q=1,[R_{j+1}]_{tq}>0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{w}_{j+1} - \overline{v}_{j+1}]_u + \sum_{q=1,[R_{j+1}]_{tq}<0}^{n_{j+1}} [R_{j+1}]_{tu}[\overline{v}_{j+1} - \overline{w}_{j+1}]_u \\
&\stackrel{\rm def}{=} [R_{j+1}z(t)]_t,
\end{aligned}$$

where we have used (2.13) and (2.14), and where, for $t = 1, \ldots, n_{j+1}$,

$$[z(t)]_u = \mathrm{sign}([R_{j+1}]_{tu})[\overline{w}_{j+1} - \overline{v}_{j+1}]_u.$$

This last definition implies that $\|z(t)\|_\infty = \|\overline{w}_{j+1} - \overline{v}_{j+1}\|_\infty$ for $t = 1, \ldots, n_{j+1}$. Taking norms and using the identity $\|R_{j+1}\|_\infty = 1$, we therefore obtain that

$$\begin{aligned}
\|w_j - v_j\|_\infty &= \max_t |[R_{j+1}z(t)]_t| \\
&\leq \max_t \|R_{j+1}z(t)\|_\infty \\
&\leq \max_t \|z(t)\|_\infty \\
&= \|\overline{w}_{j+1} - \overline{v}_{j+1}\|_\infty.
\end{aligned} \tag{4.39}$$

Remembering now the definition of $\overline{w}_{j+1}$ and $\overline{v}_{j+1}$, we see that

$$\begin{aligned}
\|\overline{w}_{j+1} - \overline{v}_{j+1}\|_\infty &= \|\min\left[w_{j+1}, x_{\pi(j,0)} + \Delta_{\pi(j,0)}e\right] - \max\left[v_{j+1}, x_{\pi(j,0)} - \Delta_{\pi(j,0)}e\right]\|_\infty \\
&\leq \|\min\left[w_{j+1}, x_{\pi(j,0)} + \Delta_{\pi(j,0)}e\right] - x_{\pi(j,0)}\|_\infty \\
&\quad + \|x_{\pi(j,0)} - \max\left[v_{j+1}, x_{\pi(j,0)} - \Delta_{\pi(j,0)}e\right]\|_\infty \\
&\leq 2\Delta_{\pi(j,0)}.
\end{aligned}$$

Combining now this bound with (4.39) and our assumption that $x \in \mathcal{A}_j$, we obtain that

$$\|x - x_{j,0}\|_\infty \leq \|w_j - v_j\|_\infty \leq 2\Delta_{\pi(j,0)}.$$

□

To each iteration (i,k) at level i, we now associate the set

$$\mathcal{R}(i,k) \stackrel{\rm def}{=} \{(j,\ell) \mid \text{iteration } (j,\ell) \text{ occurs within iteration } (i,k)\}.$$

This set always contains the pair (i,k) and only contains that pair if a Taylor step is used at iteration (i,k). If we choose a recursive step, then it also contains the pairs of level and iteration number of all iterations that occur in the potential recursion started in Step 2 and terminating on return within iteration (i,k), but it does not contain the pairs of indices corresponding to the terminating iterates $(j,*)$ of its internal minimization sequences. It is easy to verify that $j \leq i$ for every j such that $(j,\ell) \in \mathcal{R}(i,k)$ for some non-negative k and ℓ. Note also that $\mathcal{R}(i,k)$ contains at most one minimization sequence at level $i-1$, but may contain more than one at level $i-2$ and below, since each iteration at level $i-1$ may generate its own. Associated with $\mathcal{R}(i,k)$, we also define

$$\mathcal{T}(i,k) \stackrel{\rm def}{=} \{(j,\ell) \in \mathcal{R}(i,k) \mid (j,\ell) \text{ is a Taylor iteration}\}.$$

Our next proposition indicates that, if $\Delta_{i,k}$ becomes too small, then the method reduces, at level i, to the standard trust-region method using Taylor's iterations only.

Lemma 4.5 *Assume that, for some iteration (i,k),*

$$\Delta_{i,k} \leq \frac{1}{2}\min\left[1, \frac{\epsilon_{\min}}{2\kappa_g\sqrt{n_r}}, \Delta^s_{\min}\right] \stackrel{\rm def}{=} \kappa_2 \in (0,1), \tag{4.40}$$

where $\Delta^s_{\min} \stackrel{\rm def}{=} \min_{i=0,\ldots,r}\Delta^s_i$ Then no recursion occurs in iteration (i,k) and $\mathcal{R}(i,k) = \mathcal{T}(i,k) = \{(i,k)\}$.

Proof. Assume that iteration (i,k) is recursive and that iteration $(i-1,0)$ exists. Since (4.40) implies that $2\Delta_{i,k} < 1$, we deduce from (4.38) that $\mathcal{L}_{i-1} \subset \{x_{i-1,0}+d \mid \|d\|_\infty \leq 1\}$ and thus that

$$\chi_{i-1,0} = |\min_{x_{i-1,0}+d\in\mathcal{L}_{i-1}} \langle g_{i-1,0}, d\rangle| = |\langle g_{i-1,0}, d_{i-1,0}\rangle| \tag{4.41}$$

with

$$\|d_{i-1,0}\|_\infty \leq 2\Delta_{i,k}. \tag{4.42}$$

Using (4.40), (4.36), the Cauchy-Schwarz inequality, (4.41), (4.42) and (4.32) successively, we conclude that

$$\Delta_{i,k} \leq \frac{\epsilon_{\min}}{4\kappa_{\rm g}\sqrt{n_r}} \leq \frac{\chi_{i-1,0}}{4\kappa_{\rm g}\sqrt{n_r}} = \frac{|\langle g_{i-1,0}, d_{i-1,0}\rangle|}{4\kappa_{\rm g}\sqrt{n_r}} \leq \frac{\kappa_{\rm g}\sqrt{n_{i-1}}\,\|d_{i-1,0}\|_\infty}{4\kappa_{\rm g}\sqrt{n_r}} \leq \tfrac{1}{2}\Delta_{i,k}$$

which is impossible. Hence our initial assumption that iteration (i,k) is recursive cannot hold and the proof is complete. □

This lemma essentially states that when the trust-region becomes too small compared to the current criticality level, then too little can be gained from lower level iterations to allow recursion. This has the following important consequence.

Lemma 4.6 *Consider an iteration (i,k) for which $\chi_{i,k} > 0$ and*

$$\Delta_{i,k} \leq \min\left[\kappa_2, \kappa_3\chi_{i,k}\right], \tag{4.43}$$

where κ_2 is defined in (4.40) and $\kappa_3 \in (0,1)$ is given by

$$\kappa_3 = \min\left[1, \frac{\kappa_{red}(1-\eta_2)}{n_r\kappa_H}\right].$$

Then iteration (i,k) is very successful and $\Delta_{i,k+1} \geq \Delta_{i,k}$.

Proof. Because of (4.40) and Lemma 4.5, we know that iteration (i,k) is a Taylor iteration. Thus, using (2.24),

$$\delta_{i,k} \geq \kappa_{\rm red}\chi_{i,k}\min\left[1, \frac{\chi_{i,k}}{\beta_{i,k}}, \Delta_{i,k}\right].$$

But, because $\kappa_{\rm red}(1-\eta_2)/n_r \leq 1$ and (4.31), (4.43) implies that $\Delta_{i,k} \leq \min\left[1, \frac{\chi_{i,k}}{\beta_{i,k}}\right]$ and hence that

$$\delta_{i,k} \geq \kappa_{\rm red}\chi_{i,k}\Delta_{i,k}. \tag{4.44}$$

We now observe that the mean-value theorem, (2.17) and the definition of $g_{i,k}$ ensure that

$$h_i(x_{i,k}+s_{i,k}) - m_{i,k}(x_{i,k}+s_{i,k}) = \tfrac{1}{2}\langle s_{i,k}, [\nabla_{xx}h_i(\xi_{i,k}) - H_{i,k}]s_{i,k}\rangle$$

for some $\xi_{i,k} \in [x_{i,k}, x_{i,k}+s_{i,k}]$, and thus, using (4.30), (4.31), the Cauchy-Schwartz inequality and the bound $\|s_{i,k}\|_2 \leq \sqrt{n_i}\|s_{i,k}\|_\infty \leq \sqrt{n_i}\Delta_{i,k}$, that

$$|h_i(x_{i,k}+s_{i,k}) - m_{i,k}(x_{i,k}+s_{i,k})| \leq \tfrac{1}{2}\left[\,\|\nabla_{xx}h_i(\xi_{i,k})\|_2 + \|H_{i,k}\|_2\,\right]\|s_{i,k}\|_2^2 \leq n_i\kappa_{\rm H}\Delta^2_{i,k}.$$

Combining now (4.43), (4.44) and this last inequality, we verify that

$$|\rho_{i,k} - 1| \leq \left| \frac{h_i(x_{i,k} + s_{i,k}) - m_{i,k}(x_{i,k} + s_{i,k})}{\delta_{i,k}} \right| \leq \frac{n_i \kappa_{\rm H}}{\kappa_{\rm red} \chi_{i,k}} \Delta_{i,k} \ \leq 1 - \eta_2.$$

Thus iteration (i,k) must be very successful and, because of (2.26), the trust-region radius cannot decrease. □

This last result implies the following useful consequence.

Lemma 4.7 *Each minimization sequence contains at least one successful iteration.*

Proof. This follows from the fact that unsuccessful iterations cause the trust-region radius to decrease, until (4.43) is eventually satisfied and a (very) successful iteration occurs because of Lemma 4.6. □

The attentive reader will have noticed that the term in $\Delta^{\rm s}_{\rm min}$ in the minimum defining κ_2 in (4.40) has not been used in Lemma 4.5. This term is however crucial in the following further consequence of (4.40).

Lemma 4.8 *For every iteration (j,ℓ), with $j = 0, \ldots, r$ and $\ell > 0$, we have that*

$$\Delta_{j,\ell} \geq \Delta_{min} \stackrel{\rm def}{=} \gamma_1 \min[\kappa_2, \kappa_3 \epsilon_j]. \tag{4.45}$$

Proof. Suppose that (j,ℓ) is the first iteration such that

$$\Delta_{j,\ell} < \gamma_1 \min[\kappa_2, \kappa_3 \epsilon_j]. \tag{4.46}$$

Since $\gamma_1 < 1$ and $\kappa_2 \leq \Delta^{\rm s}_{\rm min}$, we then obtain that

$$\Delta_{j,0} = \Delta^{\rm s}_j \geq \Delta^{\rm s}_{\rm min} \geq \gamma_1 \Delta^{\rm s}_{\rm min} \geq \gamma_1 \min[\kappa_2, \kappa_3 \epsilon_j],$$

and, because of (4.46), we have that $\ell > 0$. This in turn implies that $\Delta_{j,\ell}$ is computed using Step 6 of the algorithm. But, the mechanism of the algorithm imposes that $\Delta_{j,\ell} \geq \gamma_1 \Delta_{j,\ell-1}$ an thus (4.46) also yields that

$$\Delta_{j,\ell-1} < \min[\kappa_2, \kappa_3 \epsilon_j] \leq \min[\kappa_2, \kappa_3 \chi_{j,\ell-1}],$$

where we have used the mechanism of the algorithm to derive the last inequality. Hence, we may apply Lemma 4.6 to conclude that iteration $(j,\ell-1)$ is very successful and that $\Delta_{j,\ell} \geq \Delta_{j,\ell-1}$. Thus, iteration (j,ℓ) cannot be the first such that (4.46) holds. This implies that (4.46) is impossible, which completes the proof. □

We next show the crucial result that the algorithm is well defined, and that all the recursions are finite.

Theorem 4.9 *The number of iterations in each level is finite. Moreover, there exists $\kappa_h \in (0,1)$ such that, for every minimization sequence at level $i = 0, \ldots, r$ and every $t \geq 0$,*

$$h_i(x_{i,0}) - h_i(x_{i,t+1}) \geq \tau_{i,t} \mu^{i+1} \kappa_h,$$

where $\tau_{i,t}$ is the total number of successful Taylor iterations in $\displaystyle\bigcup_{\ell=0}^{t} \mathcal{R}(i,\ell)$ and $\mu = \eta_1/\sigma_{\max}$ with $\sigma_{\max} = \max_{i=1,\ldots,r} \sigma_i$.

Proof. We will show this by induction on the levels, starting from level 0. First, let us define $\omega_{i,t}$ as the number of successful Taylor iterations in $\mathcal{R}(i,t)$. Thus,

$$\tau_{i,t} = \sum_{\ell=0}^{t} \omega_{i,\ell}.$$

Note that, if iteration (i,ℓ) is successful, then $\omega_{i,\ell} \geq 1$.

Consider first a minimization sequence started at level 0, and assume without loss of generality, that it belongs to $\mathcal{R}(r,k)$ for some $k \geq 0$. Every iteration in this minimization sequence has to be a Taylor iteration, which implies the sufficient decrease condition (2.24) is satisfied, and in particular, for all successful iterations,

$$\begin{aligned} h_0(x_{0,\ell}) - h_0(x_{0,\ell+1}) &\geq \eta_1 \delta_{0,\ell} \geq \eta_1 \kappa_{\rm red} \chi_{0,\ell} \min\left[1, \frac{\chi_{0,\ell}}{\beta_{0,\ell}}, \Delta_{0,\ell}\right] \\ &\geq \omega_{0,\ell} \eta_1 \kappa_{\rm red} \epsilon_{\min} \min\left[1, \frac{\epsilon_{\min}}{\kappa_{\rm H}}, \Delta_{\min}\right] \end{aligned} \tag{4.47}$$

where we used Lemma 4.8, (4.31), (4.36) and the fact that $\omega_{0,\ell} = 1$ for every successful iteration $(0,\ell)$, since $\mathcal{R}(0,\ell) = \{(0,\ell)\}$. Since we know from Lemma 4.7 that every minimization sequence has at least one successful iteration, we can sum up the reductions obtained at level 0, which gives us

$$h_0(x_{0,0}) - h_0(x_{0,t+1}) = \sum_{\ell=0}^{t}{}^{(\rm S)} [h_0(x_{0,\ell}) - h_0(x_{0,\ell+1})] \geq \tau_{0,t} \eta_1 \kappa_{\rm h} \geq \tau_{0,t} \mu \kappa_{\rm h} \tag{4.48}$$

where the superscript (S) indicates that the sum is restricted to successful iterations and where

$$\kappa_{\rm h} \stackrel{\rm def}{=} \kappa_{\rm red} \epsilon_{\min} \min\left[1, \frac{\epsilon_{\min}}{\kappa_{\rm H}}, \Delta_{\min}\right] = \kappa_{\rm red} \epsilon_{\min} \min\left[\frac{\epsilon_{\min}}{\kappa_{\rm H}}, \Delta_{\min}\right], \tag{4.49}$$

where the last equality results from the inequalities $\epsilon_{\min} \leq 1$ and $\kappa_{\rm H} \geq 1$. If $r = 0$, since $h_0 = f$ is bounded below by assumption, then (4.48) implies that $\tau_{0,t}$ is finite. If $r > 0$, f_0 is continuous implies that h_0 is also continuous, and thus it is bounded below on the set $\{x \in \mathbb{R}^{n_0} | \|x - x_{0,0}\|_\infty \leq 2\Delta_{r,k}\}$, and again, $\tau_{0,t}$ has to be finite. Since $\tau_{0,t}$ accounts for all successful iterations in the minimization sequence, we obtain that there must be a last finite successful iteration $(0,p_0)$. If, on the contrary, the sequence is infinite, then all iterations $(0,\ell)$ would be unsuccessful for $\ell > p_0$, causing $\Delta_{0,\ell}$ to converge to zero, which is impossible in view of Lemma 4.8. Hence, the minimization sequence is finite. The same reasoning may be applied to every such sequence at level 0.

Now, consider an arbitrary minimization sequence at level i within $\mathcal{R}(r,k)$ for some $k > 0$, and assume that each minimization sequence at level $i-1$ is finite and also that each successful iteration $(i-1,u)$ in every minimization sequence at this lower level satisfies

$$h_{i-1}(x_{i-1,u}) - h_{i-1}(x_{i-1,u+1}) \geq \omega_{i-1,u} \mu^i \kappa_{\rm h}. \tag{4.50}$$

Consider a successful iteration (i,ℓ), whose existence is ensured by Lemma 4.7. If it is a Taylor iteration, we obtain that

$$h_i(x_{i,\ell}) - h_i(x_{i,\ell+1}) \geq \eta_1 \kappa_{\rm h} \geq \mu^{i+1} \kappa_{\rm h} = \omega_{i,\ell} \mu^{i+1} \kappa_{\rm h}, \tag{4.51}$$

since $\eta_1 \in (0,1)$, $\sigma_{\max} > 1$ and $\omega_{i,\ell} = 1$ for every successful Taylor iteration (i,ℓ). If, on the other hand, iteration (i,ℓ) uses Step 2, then we obtain that

$$\begin{aligned} h_i(x_{i,\ell}) - h_i(x_{i,\ell+1}) &\geq \frac{\eta_1}{\sigma_i} [h_{i-1}(x_{i-1,0}) - h_{i-1}(x_{i-1,*})] \\ &\geq \mu \sum_{u=0}^{p_{i-1}}{}^{(\rm S)} [h_{i-1}(x_{i-1,u}) - h_{i-1}(x_{i-1,u+1})]. \end{aligned}$$

Since $\omega_{i,\ell} = \tau_{i-1,p_{i-1}}$, the definition of $\tau_{i-1,t}$ and (4.50) give that

$$h_i(x_{i,\ell}) - h_i(x_{i,\ell+1}) \geq \mu^{i+1} \kappa_{\rm h} \sum_{u=0}^{p_{i-1}} \omega_{i-1,u} = \tau_{i-1,p_{i-1}} \mu^{i+1} \kappa_{\rm h} = \omega_{i,\ell} \mu^{i+1} \kappa_{\rm h}. \tag{4.52}$$

Combining (4.51) and (4.52), we see that (4.50) again holds at level i instead of $i-1$. Moreover, as above,

$$h_i(x_{i,0}) - h_i(x_{i,t+1}) = \sum_{\ell=0}^{t} {}^{(\mathrm{S})} [h_i(x_{i,\ell}) - h_i(x_{i,\ell+1})] \geq \tau_{i,t}\mu^{i+1}\kappa_{\mathrm{h}}, \tag{4.53}$$

for the minimization sequence including iteration (i,ℓ). If $i=r$, $h_i = f$ is bounded below by assumption and (4.53) imposes that the number of successful iterations in this sequence must again be finite. The same conclusion holds if $i<r$, since h_i is continuous and hence bounded below on the set $\{x \in \mathbb{R}^{n_i} | \|x - x_{i,0}\|_\infty \leq 2\Delta_{r,k}\}$ which contains $x_{i,t+1}$ because of Lemma 4.4. As for level 0, we may then conclude that the number of iterations (both successful and unsuccessful) in the minimization sequence is finite. Moreover, the same reasoning holds for every minimization sequence at level i, and the induction is complete. □

Corollary 4.10 *Assume that one knows a constant f_{low} such that $h_r(x_r) = f(x) \geq f_{\mathrm{low}}$ for every $x \in \mathbb{R}^n$. Then Algorithm RMTR$_\infty$ needs at most*

$$\left\lceil \frac{f(x_{r,0}) - f_{\mathrm{low}}}{\theta(\epsilon_{\mathrm{min}})} \right\rceil \tag{4.54}$$

successful Taylor iterations at any level to obtain an iterate $x_{r,k}$ such that $\chi_{r,k} < \epsilon_r$, where

$$\theta(\epsilon) = \mu^{r+1}\kappa_{red}\epsilon \min\left[\frac{\epsilon}{\kappa_H}, \gamma_1 \min\left[\kappa_2, \kappa_3\epsilon\right]\right]. \tag{4.55}$$

Proof. The desired bound directly follows from Theorem 4.9, (4.49), (4.45) and the definition of ϵ_{min}. □

This complexity result for general nonconvex problems is similar to Corollary 3.8 in Gratton et al. (2004), and may also be very pessimistic. It is of the same order as the corresponding bound for the pure gradient method (see (Nesterov 2004), page 29). This is not surprising given that it is based on the Cauchy condition, which itself results from a step in the steepest-descent direction. Note that the bound is in terms of iteration numbers, and only implicitly accounts for the cost of computing a Taylor step satisfying (2.24). As was the case for the Euclidean norm, this suggests several comments.

1. The bound (4.54) is expressed in terms of the number of successful Taylor iterations, that is successful iterations where the trial step is computed without resorting to further recursion. This provides an adequate measure of the linear algebra effort for all successful iterations, since successful iterations using the recursion of Step 2 cost little beyond the evaluation of the level-dependent objective function and its gradient. Moreover, the number of such iterations is, by construction, at most equal to r times that of Taylor iterations (in the worst case where each iteration at level r includes a full recursion to level 0 with a single successful iteration at each level $j>0$). Hence the result shows that the number of necessary successful iterations, all levels included, is of order $1/\epsilon^2$ for small values of ϵ. This order is not qualitatively altered by the inclusion of unsuccessful iterations either, provided we replace the very successful trust-region radius update (top case in (2.26)) by

$$\Delta_{i,k}^{+} \in [\Delta_{i,k}, \gamma_3\Delta_{i,k}] \quad \text{if} \quad \rho_{i,k} \geq \eta_2,$$

for some $\gamma_3 > 1$. Indeed, Lemma 4.8 imposes that the decrease in radius caused by unsuccessful iterations must asymptotically be compensated by an increase at successful ones. This is to say that, if α is the average number of unsuccessful iterations per sucessful one at any level, then one must have that $\gamma_3\gamma_2^{\alpha} \geq 1$, and

therefore that $\alpha \leq -\log(\gamma_3)/\log(\gamma_2)$. Thus the complexity bound in $1/\epsilon^2$ for small ϵ is only modified by a constant factor if all iterations (successful and unsucessful) are considered. This therefore also gives a worst case upper bound on the number of function and gradient evaluations.

2. Moreover, (4.54) involves the number of successful Taylor iterations *summed up on all levels* (as a result of Theorem 4.9). Thus such successful iterations at cheap low levels decrease the number of necessary expensive ones at higher levels, and the multilevel algorithm requires (at least in the theoretical worst case) fewer Taylor iterations at the upper level than the single-level variant. This provides theoretical backing for the practical observation that the structure of multilevel bound-constrained optimization problems can be used to advantage.

3. The definition of $\theta(\epsilon)$ in (4.55) is interesting in that it does not depend on the problem dimension, but rather on the properties of the problem or of the algorithm itself. Thus, if we consider the case where different levels correspond to different discretization meshes and make the mild assumption that r and $\kappa_{\rm H}$ are uniformly bounded above, we deduce that our complexity bound is mesh-independent.

A second important consequence of Theorem 4.9 is that the algorithm is globally convergent, in the sense that, if ϵ_r is "driven to zero", it generates a subsequence of iterates that are asymptotically first-order critical. More specifically, we examine the sequence of iterates $\{x_{r,k}\}$ generated as follows. We consider, at level r, a sequence of tolerances $\{\epsilon_{r,j}\} \in (0,1)$ monotonically converging to zero, start the algorithm with $\epsilon_r = \epsilon_{r,0}$ and alter slightly the mechanism of Step 5 (at level r only) to reduce ϵ_r from $\epsilon_{r,j}$ to $\epsilon_{r,j+1}$ as soon as $\chi_{r,k+1} \leq \epsilon_{r,j}$. The calculation is then continued with this more stringent threshold until it is also attained, $\epsilon_r^{\rm g}$ is then again reduced and so on.

Theorem 4.11 *Assume that ϵ_r is "driven to zero" in Algorithm RMTR$_\infty$. Then*

$$\liminf_{k\to\infty} \chi_{r,k} = 0. \tag{4.56}$$

Proof. Since $\Delta_{r+1,0} = \infty$ ensures that $\mathcal{L}_r = \mathcal{F}_r$, Lemma 4.3 implies that each successive minimization at level r can only stop at iteration k if

$$\chi_{r,k+1} \leq \epsilon_{r,j}. \tag{4.57}$$

Theorem 4.9 then implies that there are only finitely many successful iterations between two reductions of ϵ_r. We therefore obtain that for each $\epsilon_{r,j}$ there is an arbitrarily large k such that (4.57) holds. The desired result then follows immediately from our assumption that $\{\epsilon_{r,j}\}$ converges to zero. □

Of course, the interest of this result is mostly theoretical, since most practical applications of Algorithm RMTR$_\infty$ consider a nonzero gradient tolerance ϵ_r.

Observe that our definition of ϵ_i in (2.23) implies that, if ϵ_r is driven to zero, then so is $\epsilon_i = \kappa_\chi^{r-i}\epsilon_r$. As for the Euclidean case, and assuming the trust region becomes asymptotically inactive at every level (as is most often the case in practice), each minimization sequence in the algorithm becomes infinite (as if it were initiated with a zero gradient threshold and an infinite initial radius). Recursion to lower levels then remains possible for arbitrarily small gradients, and may therefore occur arbitrarily far in the sequence of iterates. Moreover, we may still apply Theorem 4.11 at each level and deduce that, if the trust region becomes asymptotically inactive,

$$\liminf_{k\to\infty} \chi_{i,k} = 0 \tag{4.58}$$

for all $i = 0, \ldots, r$.

As is the case for single-level trust-region algorithms, we now would like to prove that the limit inferior in (4.56) and (4.58) can be replaced by a true limit. This requires the

notion of *recursively successful iteration.* We say that iteration $(j,\ell) \in \mathcal{R}(i,k)$ is *recursively successful for* (i,k) whenever iterations $(j,\ell), \pi(j,0), \pi^2(j,0), \ldots, \pi^{i-j}(j,0) = (i,k)$ are all successful. This is to say that the decrease in objective function obtained at iteration (j,ℓ) effectively contributes to the reduction obtained at iteration (i,k). We start by stating a result on the relative sizes on the objective function decreases in the course of a recursive iteration.

Lemma 4.12 *Assume that some iteration $(j,\ell) \in \mathcal{R}(i,k)$ is recursively successful for (i,k). Then*

$$h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) \leq h_j(x_{j,0}) - h_j(x_{j,*}) \leq \mu^{j-i} \left[h_i(x_{i,k}) - h_i(x_{i,k+1}) \right]. \tag{4.59}$$

Proof. The first inequality immediately results from the monotonicity of the sequence of objective function values in a minimization sequence. To prove the second inequality, consider iteration $(j+1,q) = \pi(j,0)$. Then

$$h_j(x_{j,0}) - h_j(x_{j,*}) = \sigma_{j+1}\delta_{j+1,q} \leq \eta_1^{-1}\sigma_{\max} \left[h_{j+1}(x_{j+1,q}) - h_{j+1}(x_{j+1,q+1}) \right]$$

where we used the definition of $\delta_{j+1,q}$, the definition of $\sigma_{\max}$ and the fact that iteration $(j+1,q)$ must be successful since (j,ℓ) is recursively successful for (i,k). But this argument may now be repeated at level $j+2,\ldots,i$, yielding the desired bound, given that $\mu = \eta_1/\sigma_{\max} < 1$. □

This lemma then allows us to express a simple relation between the size of Taylor steps at recursively successful iterations and the associated objective decrease.

Lemma 4.13 *Assume that the Taylor iteration $(j,\ell) \in \mathcal{R}(i,k)$ is recursively successful for (i,k) and that, for some $\epsilon \in (0,1)$,*

$$\chi_{j,\ell} \geq \epsilon \tag{4.60}$$

and

$$h_i(x_{i,k}) - h_i(x_{i,k+1}) < \frac{\mu^r \eta_1 \kappa_{red} \epsilon^2}{\kappa_H}. \tag{4.61}$$

Then

$$\|x_{j,\ell} - x_{j,\ell+1}\|_\infty \leq \frac{1}{\kappa_{red}\eta_1\epsilon} \left[h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) \right]. \tag{4.62}$$

Proof. We know from (2.24), (4.31), (4.60) and the successful nature of iteration (j,ℓ) that

$$\begin{aligned} h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) &\geq \eta_1\kappa_{\rm red}\chi_{j,\ell} \min\left[\tfrac{\chi_{j,\ell}}{\kappa_{\rm H}}, \Delta_{j,\ell}, 1\right] \\ &\geq \eta_1\kappa_{\rm red}\epsilon \min\left[\tfrac{\epsilon}{\kappa_{\rm H}}, \Delta_{j,\ell}, 1\right] \\ &= \eta_1\kappa_{\rm red}\epsilon \min\left[\tfrac{\epsilon}{\kappa_{\rm H}}, \Delta_{j,\ell}\right] \end{aligned} \tag{4.63}$$

where we used (4.31) and the inequality $\epsilon < 1$ to deduce the last equality. But Lemma 4.12 gives that

$$\begin{aligned} h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) &\leq \mu^{j-i} \left[h_i(x_{i,k}) - h_i(x_{i,k+1}) \right] \\ &\leq \mu^{-r} \left[h_i(x_{i,k}) - h_i(x_{i,k+1}) \right] \\ &\leq \tfrac{\eta_1\kappa_{\rm red}\epsilon^2}{\kappa_{\rm H}}, \end{aligned}$$

where we used (4.61) to deduce the last inequality. Hence we see that only the second term in the last minimum of (4.63) can be active, which gives that

$$h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) \geq \eta_1\kappa_{\rm red}\epsilon\Delta_{j,\ell}.$$

We then obtain (4.62) from the observation that $x_{j,\ell+1} = x_{j,\ell} + s_{j,\ell} \in \mathcal{W}_{j,\ell} \subseteq \mathcal{B}_{j,\ell}$. □

We next prove the following useful technical lemma.

Lemma 4.14 *Assume that a minimization sequence at level j $(0 \leq j \leq r)$ is such that*

$$\chi_{j,0} \geq \epsilon_{\clubsuit} \tag{4.64}$$

for some $\epsilon_{\clubsuit} \in (0,1)$, but also that

$$\|s_{j,\ell}\|_{\infty} \leq \kappa_{\clubsuit} [\, h_j(x_{j,\ell}) - h_j(x_{j,\ell+1}) \,] \tag{4.65}$$

for some $\kappa_{\clubsuit} > 0$ as long as iteration (j,ℓ) is successful and $\chi_{j,\ell} \geq \frac{1}{2}\epsilon_{\clubsuit}$. Assume finally that

$$h_j(x_{j,0}) - h_j(x_{j,*}) \leq \frac{\epsilon_{\clubsuit}}{2\kappa_{\clubsuit}(n_r\kappa_H + \kappa_g)\sqrt{n_r}}. \tag{4.66}$$

Then $\chi_{j,\ell} \geq \frac{1}{2}\epsilon_{\clubsuit}$ and (4.65) hold for all $\ell \geq 0$.

Proof. Assume that there exists a (first) successful iteration (j,s) such that

$$\chi_{j,s} < \tfrac{1}{2}\epsilon_{\clubsuit}, \tag{4.67}$$

which implies that $\chi_{j,\ell} \geq \frac{1}{2}\epsilon_{\clubsuit}$ for all $0 \leq \ell < s$. We now use (4.65) and the triangle inequality, and sum on all successful iterations (at level j) from 0 to $s-1$, yieding

$$\|x_{j,0} - x_{j,s}\|_{\infty} \leq \sum_{\ell=0}^{s-1}{}^{(S)} \|x_{j,\ell} - x_{j,\ell+1}\|_{\infty} \leq \kappa_{\clubsuit} [\, h_j(x_{j,0}) - h_j(x_{j,s}) \,]. \tag{4.68}$$

Applying now Lemma 4.1, the monotonicity of h_j within the minimization sequence, the bound $n_j \leq n_r$ and (4.66), we obtain from (4.68) that

$$\begin{aligned} |\chi_{j,0} - \chi_{j,s}| &\leq \kappa_{\clubsuit}(n_j\kappa_{\rm H} + \kappa_{\rm g})\sqrt{n_j}\,[\, h_j(x_{j,0}) - h_j(x_{j,s}) \,] \\ &\leq \kappa_{\clubsuit}(n_r\kappa_{\rm H} + \kappa_{\rm g})\sqrt{n_r}\,[\, h_j(x_{j,0}) - h_j(x_{j,*}) \,] \\ &\leq \tfrac{1}{2}\epsilon_{\clubsuit}. \end{aligned}$$

But this last inequality is impossible since we know from (4.64) and (4.67) that $\chi_{j,0} - \chi_{j,s} > \frac{1}{2}\epsilon_{\clubsuit}$. Hence our assumption (4.67) is itself impossible and we obtain that, for all $\ell \geq 0$, $\chi_{j,\ell} \geq \frac{1}{2}\epsilon_{\clubsuit}$. This and the lemma's assumptions then ensure that (4.65) also holds for all $j \geq 0$. □

We now consider the case of recursive iterations.

Lemma 4.15 *Assume that, for some recursive successful iteration (i,k),*

$$\chi_{i,k} \geq \epsilon_{\diamondsuit} \tag{4.69}$$

and

$$h_i(x_{i,k}) - h_i(x_{i,k+1}) \leq \frac{\kappa_{\chi}\epsilon_{\diamondsuit}}{2\kappa_{\diamondsuit}(n_r\kappa_H + \kappa_g)\sqrt{n_r}} \tag{4.70}$$

for some $\epsilon_{\diamondsuit} \in (0,1)$ and some $\kappa_{\diamondsuit} > 0$. Assume also that

$$\|s_{i-1,\ell}\|_{\infty} \leq \kappa_{\diamondsuit} [\, h_{i-1}(x_{i-1,\ell}) - h_{i-1}(x_{i-1,\ell+1}) \,] \tag{4.71}$$

for all (recursively) successful iterations in the minimization sequence initiated at level $i-1$ by iteration (i,k) as long as

$$\chi_{i-1,\ell} \geq \tfrac{1}{2}\kappa_{\chi}\epsilon_{\diamondsuit}. \tag{4.72}$$

Then

$$\|s_{i,k}\|_{\infty} \leq \mu^{-1}\kappa_P\kappa_{\diamondsuit} [\, h_i(x_{i,k}) - h_i(x_{i,k+1}) \,]. \tag{4.73}$$

Proof. Consider the minimization sequence initiated at level $i-1$ by iteration (i,k). Because of (2.22) and (4.69), we have that $\chi_{i-1,0} \geq \kappa_\chi \epsilon_\diamondsuit$. We may now apply Lemma 4.14 with $\epsilon_\clubsuit = \kappa_\chi \epsilon_\diamondsuit$ and $\kappa_\clubsuit = \kappa_\diamondsuit$, given that (4.70) ensures (4.66). As a result, we know that $\chi_{i-1,\ell} \geq \frac{1}{2}\kappa_\chi \epsilon_\diamondsuit$ and (4.71) hold for all successful iterations $(i-1,\ell)$ $(\ell \geq 0)$. Using the triangle inequality and summing on all successful iterations at level $i-1$, we find that

$$\|x_{i-1,0} - x_{i-1,*}\|_\infty \leq \sum_{\ell=0}^{p_{i-1}} {}^{(S)}\|x_{i-1,\ell} - x_{i-1,\ell+1}\|_\infty \leq \kappa_\diamondsuit \, [\, h_{i-1}(x_{i-1,0}) - h_{i-1}(x_{i-1,*}) \,].$$

This inequality, the definition of $s_{i,k}$, (4.33) and Lemma 4.12 in turn imply that

$$\begin{aligned} \|s_{i,k}\|_\infty &\leq \|P_i\|_\infty \|x_{i-1,0} - x_{i-1,*}\|_\infty \\ &\leq \kappa_{\rm P}\kappa_\diamondsuit \, [\, h_{i-1}(x_{i-1,0}) - h_{i-1}(x_{i-1,*}) \,] \\ &\leq \mu^{-1}\kappa_{\rm P}\kappa_\diamondsuit \, [\, h_i(x_{i,k}) - h_i(x_{i,k+1}) \,]. \end{aligned}$$

□

Our next step is to consider the cumulative effect of all the complete recursion for an iteration at the finest level.

Lemma 4.16 *Assume that, for some successful iteration (r,k) $(k \geq 0)$,*

$$\chi_{r,k} \geq \epsilon \tag{4.74}$$

and

$$f(x_{r,k}) - f(x_{r,k+1}) < \frac{\eta_1 \kappa_{red} (\frac{1}{2}\kappa_\chi)^{2r} \epsilon^2}{2(n_r \kappa_H + \kappa_g)\sqrt{n_r}} \tag{4.75}$$

for some $\epsilon \in (0,1)$. Then

$$\|s_{r,k}\|_\infty \leq \kappa_\heartsuit \, [\, f(x_{r,k}) - f(x_{r,k+1}) \,], \tag{4.76}$$

where

$$\kappa_\heartsuit \stackrel{\rm def}{=} \left(\frac{\kappa_P}{\mu}\right)^r \frac{1}{\kappa_{red}\eta_1(\frac{1}{2}\kappa_\chi)^r \epsilon}. \tag{4.77}$$

Proof. Assume that (4.74) and (4.75) hold at the successful iteration (r,k) and consider the subset of iterations given by $\mathcal{R}(r,k)$. If (r,k) is a Taylor iteration, then $\mathcal{R}(r,k) = \{(r,k)\}$ and the desired result follows from Lemma 4.13 and the inequality

$$\frac{1}{\kappa_{\rm red}\eta_1\epsilon} \leq \kappa_\heartsuit. \tag{4.78}$$

If iteration (r,k) is recursive, consider a minimization sequence containing a recursively successful iteration for (r,k) at the deepest possible level in $\mathcal{R}(r,k)$. Let the index of this deepest level be d and note that every successful iteration in this minimization sequence must be recursively successful for (r,k). Let now $(d+1,q) = \pi(d,0)$ and assume that

$$\chi_{d+1,q} \geq (\tfrac{1}{2}\kappa_\chi)^{r-d-1}\epsilon, \tag{4.79}$$

which gives, in view of (2.22), that $\chi_{d,0} \geq (\frac{1}{2})^{r-d-1}\kappa_\chi^{r-d}\epsilon$. Each (recursively) successful iteration of our deepest minimization sequence must thus be a Taylor iteration. Because of Lemma 4.13, we then obtain that, as long as $\chi_{d,\ell} \geq (\frac{1}{2}\kappa_\chi)^{r-d}\epsilon$ and iteration (d,ℓ) is successful, we have that

$$\|s_{d,\ell}\|_\infty = \|x_{d,\ell} - x_{d,\ell+1}\|_\infty \leq \frac{1}{\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d}\epsilon} \, [\, h_d(x_{d,\ell}) - h_d(x_{d,\ell+1}) \,], \tag{4.80}$$

We could then apply Lemma 4.15 for iteration $(d+1,q) = \pi(d,0)$ with

$$\epsilon_\diamondsuit = (\tfrac{1}{2}\kappa_\chi)^{r-d-1}\epsilon \quad \text{and} \quad \kappa_\diamondsuit = \frac{1}{\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d}\epsilon},$$

if (4.70) holds. But note that Lemma 4.12 implies that

$$h_{d+1}(x_{d+1,q}) - h_{d+1}(x_{d+1,q+1}) \leq \mu^{d+1-r}\,[\,f(x_{r,k}) - f(x_{r,k+1})\,]$$

which in turn gives (4.70) in view of (4.75), as desired. As a result of Lemma 4.15, we then deduce that

$$\begin{array}{rcl} \|s_{d+1,q}\|_\infty & \leq & \mu^{-1}\kappa_{\rm P}\kappa_\diamondsuit\,[\,h_{d+1}(x_{d+1,q}) - h_{d+1}(x_{d+1,q+1})\,] \\ & = & \left(\frac{\kappa_{\rm P}}{\mu}\right)\frac{1}{\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d}\epsilon}\,[\,h_{d+1}(x_{d+1,q}) - h_{d+1}(x_{d+1,q+1})\,]. \end{array} \tag{4.81}$$

Consider now a minimization sequence at level $d+1$ and assume that $d+1 < r$. Then define $(d+2,t) = \pi(d+1,0)$ and assume, in line with (4.79), that $\chi_{d+2,t} \geq (\frac{1}{2}\kappa_\chi)^{r-d-2}\epsilon$ which yields in particular that $\chi_{d+1,0} \geq (\frac{1}{2})^{r-d-2}\kappa_\chi^{r-d-1}\epsilon$. The reasoning for level $d+1$ is entirely similar to that for level d, except that we now have to consider a possible mixture of Taylor and recursive steps. As long as

$$\chi_{d+1,\ell} \geq (\tfrac{1}{2}\kappa_\chi)^{r-d-1}\epsilon \tag{4.82}$$

and iteration $(d+1,\ell)$ is (recursively) successful, we have that, for Taylor steps,

$$\|s_{d+1,\ell}\|_\infty = \|x_{d+1,\ell} - x_{d+1,\ell+1}\|_\infty \leq \frac{h_{d+1}(x_{d+1,\ell}) - h_{d+1}(x_{d+1,\ell+1})}{\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d-1}\epsilon} \tag{4.83}$$

because of Lemma 4.13, while we know that (4.81) holds for recursive steps. But since $\kappa_{\rm P} > \mu(\frac{1}{2}\kappa_\chi)$, we may deduce that (4.81) thus holds for successful iterations $(d+1,q)$ as long as (4.82) is satisfied. As for level $d+1$, we may now apply Lemma 4.15 for iteration $(d+2,t)$, with

$$\epsilon_\diamondsuit = (\tfrac{1}{2}\kappa_\chi)^{r-d-2}\epsilon \quad \text{and} \quad \kappa_\diamondsuit = \frac{\kappa_{\rm P}}{\mu\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d}\epsilon},$$

since, as above, (4.70) results from Lemma 4.12 and (4.75). This new application of Lemma 4.15 then ensures that

$$\begin{array}{rcl} \|s_{d+2,t}\|_\infty & \leq & \mu^{-1}\kappa_{\rm P}\kappa_\diamondsuit\,[\,h_{d+2}(x_{d+2,t}) - h_{d+2}(x_{d+2,t+1})\,] \\ & = & \left(\frac{\kappa_{\rm P}}{\mu}\right)^2\frac{1}{\kappa_{\rm red}\eta_1(\frac{1}{2}\kappa_\chi)^{r-d}\epsilon}\,[\,h_{d+1}(x_{d+1,q}) - h_{d+1}(x_{d+1,q+1})\,]. \end{array} \tag{4.84}$$

The proof is then completed by applying the argument repeatedly up to level r. □

We finally prove the main result.

Theorem 4.17 *Assume that ϵ_r is "driven to zero" in Algorithm RMTR$_\infty$. Then*

$$\lim_{k\to\infty}\chi_{r,k} = 0. \tag{4.85}$$

Proof. As in Theorem 4.11, we identify our sequence of iterates with that generated by considering a sequence of tolerances $\{\epsilon_{r,j}\} \in (0,1)$ monotonically converging to zero. We start our proof by observing that the monotonic nature of the sequence $\{f(x_{r,\ell})\}_{\ell\geq 0}$ and the fact that $f(x)$ is bounded below impose that

$$f(x_{r,k}) - f(x_{r,k+1}) \to 0 \tag{4.86}$$

for all successful iterations (r,k). Assume now, for the purpose of deriving a contradiction, that

$$\limsup_{k\to\infty}\chi_{r,k} \geq 3\epsilon > 0 \tag{4.87}$$

for some $\epsilon \in (0,1)$ and consider a $k_0 > 0$ such that $\chi_{r,k_0} \geq 2\epsilon$ and such that both (4.75) and

$$f(x_{r,k}) - f(x_{r,k+1}) \leq \frac{\epsilon}{\kappa_\heartsuit(n_r\kappa_{\rm H} + \kappa_{\rm g})\sqrt{n_r}}$$

hold for all $k \geq k_0$. Without loss of generality, we may assume that the minimization sequence at level r starts at iteration k_0. But Lemma 4.16 ensures that (4.76) holds for each successful iteration (r,k) $(k \geq k_0)$ as long as (4.74) holds. We may therefore apply Lemma 4.14 with

$$\epsilon_\clubsuit = 2\epsilon \quad \text{and} \quad \kappa_\clubsuit = \kappa_\heartsuit$$

to the (truncated) minimization sequence at level r and deduce that (4.74) holds for all $k \geq k_0$, which is impossible in view of Theorem 4.11. Hence (4.87) is impossible and our proof complete. □

Theorem 4.17 implies, in particular, that any limit point of the infinite sequence $\{x_{r,k}\}$ is first-order critical for problem (2.1). But we may draw stronger conclusions: if we additionally assume that the trust region becomes asymptotically inactive at all levels, then, as explained above, each minimization sequence in the algorithm becomes infinite, and we may apply Theorem 4.17 to each of them, concluding that

$$\lim_{k\to\infty} \chi_{i,k} = 0$$

for every level $i = 0, \ldots, r$. The behaviour of Algorithm RMTR$_\infty$ is therefore truly coherent with its multilevel formulation, since the same convergence results hold for each level.

The convergence results at the upper level are unaffected if minimization sequences at lower levels are "prematurely" terminated, provided each such sequence contains at least one successful iteration. Indeed, none of the proofs depends on the actual stopping criterion used. Thus, one might think of stopping a minimization sequence after a preset number of successful iterations: in combination with the freedom left at Step 1 to choose the model whenever (2.22) holds, this strategy allows a straightforward implementation of fixed lower-iterations patterns, like the V or W cycles in multigrid methods. This is what we have done in Section 3.

Our theory also remains essentially unchanged if we merely insist on first-order coherence (i.e., definition (2.18)) to hold only for small enough trust-region radii $\Delta_{i,k}$, or only up to a perturbation of the order of $\Delta_{i,k}$ or $\|g_{i,k}\|\Delta_{i,k}$. Other generalizations may be possible. Similarly, although we have assumed for motivation purposes that each f_i is "more costly" to minimize that f_{i-1}, we have not used this feature in the theory presented above, nor have we used the form of the lower levels objective functions. In particular, the choice of Section 3 to define f_i as identically zero for $i = 0, \ldots, r-1$ satisfies all our assumptions. Nonconstant prolongation and restriction operators of the form $P_i(x_{i,k})$ and $R_i(x_{i,k})$ may also be considered, provided the singular values of these operators remain uniformly bounded. We also refer the reader to (Gratton et al. 2006*b*) for a discussion of convergence properties of multilevel trust-region methods to second-order critical points.

5 Conclusion and perspectives

We have presented a variant of the recursive multilevel RMTR algorithm for unconstrained nonlinear optimization that clearly has advantages over the original method in terms of computational costs and flexibility. The use of the infinity norm (as opposed to the Euclidean norm used in the original algorithm) removes the need for costly preconditioning of the trust-region and adapts very naturally to bound constrained problems. However, and despite the conceptual similarity between RMTR and the new algorithm, their convergence theory differ significantly. Fortunately, the same strong global convergence results can be proved (with somewhat simpler arguments) for the new algorithm, which makes it very attractive for practical use.

Acknowledgements

The work of Melissa Weber-Mendonça is made possible by a grant of CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) from Brazil.

References

A. Borzi and K. Kunisch. A globalisation strategy for the multigrid solution of elliptic optimal control problems. *Optimization Methods and Software*, **21**(3), 445–459, 2006.

W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, USA, 2nd edn, 2000.

P. H. Calamai and J. J. Moré. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, **39**, 93–116, 1987.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Testing a class of methods for solving minimization problems with simple bounds on the variables. *Mathematics of Computation*, **50**, 399–430, 1988.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A)*. Number 17 *in* 'Springer Series in Computational Mathematics'. Springer Verlag, Heidelberg, Berlin, New York, 1992.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. Number 01 *in* 'MPS-SIAM Series on Optimization'. SIAM, Philadelphia, USA, 2000.

M. Fisher. Minimization algorithms for variational data assimilation. *in* 'Recent Developments in Numerical Methods for Atmospheric Modelling', pp. 364–385. ECMWF, 1998.

T. Frese, Ch. Bouman, and K. Sauer. Multiscale bayesian methods for discrete tomography. *in* G. Herman and A. Kuba, eds, 'Discrete Tomography: Foundations, Algorithms and Applications', pp. 237–261. Birkhauser, Boston, 1999.

E. Gelman and J. Mandel. On multilevel iterative methods for optimization problems. *Mathematical Programming*, **48**(1), 1–17, 1990.

S. Gratton, A. Sartenaer, and Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. Technical Report 04/06, Department of Mathematics, University of Namur, Namur, Belgium, 2004.

S. Gratton, A. Sartenaer, and Ph. L. Toint. Numerical experience with a recursive trust-region method for multilevel nonlinear optimization. Technical Report 06/01, Department of Mathematics, University of Namur, Namur, Belgium, 2006*a*.

S. Gratton, A. Sartenaer, and Ph. L. Toint. Second-order convergence properties of trust-region methods using incomplete curvature information, with an application to multigrid optimization. *Journal of Computational and Applied Mathematics*, **24**(6), 676–692, 2006*b*.

M. Herty and G. Thömmes. A two-level trust-region method for optimal control with radiative transfer. *Advanced Modelling and Optimization*, **8**(2), 187–207, 2006.

G. Lemarchand, O. Pironneau, and E. Polak. A mesh refinement method for optimization with DDM. *in* N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes and Y. Kuznetsov, eds, 'Thirteenth International Conference on Domain Decomposition Methods', pp. 419–325, 2001.

M. Lewis and S. G. Nash. Practical aspects of multiscale optimization methods for VLSICAD. *in* J. Cong and J. R. Shinnerl, eds, 'Multiscale Optimization and VLSI/CAD', pp. 265–291, Dordrecht, The Netherlands, 2002. Kluwer Academic Publishers.

M. Lewis and S. G. Nash. Model problems for the multigrid optimization of systems governed by differential equations. *SIAM Journal on Scientific Computing*, **26**(6), 1811–1837, 2005.

C. Lin and J. J. Moré. Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, **9**(4), 1100–1127, 1999.

J. J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. *in* G. A. Watson, ed., 'Numerical Analysis, Dundee 1977', number 630 *in* 'Lecture Notes in Mathematics', pp. 105–116, Heidelberg, Berlin, New York, 1978. Springer Verlag.

S. G. Nash. A multigrid approach to discretized optimization problems. *Optimization Methods and Software*, **14**, 99–116, 2000.

Y. Nesterov. *Introductory Lectures on Convex Optimization*. Applied Optimization. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

S. Oh, A. Milstein, Ch. Bouman, and K. Webb. A general framework for nonlinear multigrid inversion. *IEEE Transactions on Image Processing*, **14**(1), 125–140, 2005.