Applied Mathematics and Computation 220 (2013) 324–330

Contents lists available at SciVerse ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Global convergence of trust-region algorithms for convex constrained minimization without derivatives

P.D. Conejo [a,*,1], E.W. Karas [b,2], L.G. Pedroso [b,3], A.A. Ribeiro [b,3], M. Sachine [b,3]

[a] State University of Paraná, Dep. of Mathematics, 85819 110 Cascavel, PR, Brazil
[b] Federal University of Paraná, Dep. of Mathematics, CP 19081, 81531 980 Curitiba, PR, Brazil

ARTICLE INFO

Keywords:
Derivative-free optimization
Convex constrained optimization
Trust region

ABSTRACT

In this work we propose a trust-region algorithm for the problem of minimizing a function within a convex closed domain. We assume that the objective function is differentiable but no derivatives are available. The algorithm has a very simple structure and allows a great deal of freedom in the choice of the models. Under reasonable assumptions for derivative-free schemes, we prove global convergence for the algorithm, that is to say, that all accumulation points of the sequence generated by the algorithm are stationary.

1. Introduction

In this work we will discuss global convergence of a general derivative-free trust-region algorithm for solving the nonlinear programming problem

$$\begin{aligned} &\text{minimize} \quad f(x), \\ &\text{subject to} \quad x \in \Omega, \end{aligned} \tag{1}$$

where $\Omega \subset \mathbb{R}^n$ is a nonempty closed convex set and $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a differentiable function. We assume that Ω is a "simple" set, in the sense that it is easy to compute the orthogonal projection of an arbitrary point onto the feasible set Ω. Although the objective function is smooth, we assume that its derivatives are not available. This situation is common in a wide range of applications [9], particularly when the objective function is provided by a simulation package or a black box. Such practical situations have been motivating research on derivative-free optimization in the recent years [9,11].

There are many derivative-free methods with global convergence results for handling the problem (1). When the feasible set is defined by linear constraints, the problem can be solved by GSS method *(Generating Set Search)* presented in [19]. This method encompasses many algorithms, including Generalized Pattern Search [20]. In [3], the authors propose an inexact-restoration scheme where the GSS algorithm is used in the optimality phase, therefore avoiding the evaluation of the gradient of the objective function. In [14,18], the authors present Augmented Lagrangian methods for generally constrained problems. In [18], the subproblems are linearly constrained and solved by GSS algorithm. In [14], any class of constraints can be considered on the subproblems, since one can provide a suitable derivative-free algorithm for handling them.

* Corresponding author.
E-mail addresses: paulo.conejo@unioeste.br (P.D. Conejo), ewkaras@ufpr.br (E.W. Karas), lucaspedroso@ufpr.br (L.G. Pedroso), ademir.ribeiro@ufpr.br (A.A. Ribeiro), mael@ufpr.br (M. Sachine).

1 Ph.D. Program in Numerical Methods in Engineering, Federal University of Paraná. Partially supported by Fundação Araucária, Brazil.
2 Partially supported by CNPq, Brazil. Grants 472313/2011-8 and 307714/2011-0.
3 Partially supported by CNPq, Brazil. Grant 472313/2011-8.

0096-3003/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.06.041

Here we consider the class of derivative-free trust-region methods, which has been pioneered by Winfield [31] and exhaustively studied for unconstrained and box-constrained problems by Powell [22–25,27], Conn and Toint [12], Conn, Scheinberg and Toint [8], Conn, Scheinberg and Vicente [11], Fasano, Morales and Nocedal [15], Gratton, Toint and Tröltzsch [17], and for linearly constrained problems by Powell [26]. Global convergence results for the unconstrained case are presented in [11,17,27]. In these works the models are based on polynomial interpolation. Essentially, the difference between them consists in the interpolation set updating and the construction of the models.

With regard to derivative based trust-region methods, for both constrained and unconstrained problems, well-established algorithms with global convergence results can be found in the literature, for example in [5–7,21,28]. When the derivatives of the objective function are not available, there are also trust-region methods with good practical performance [1,25,29,30]. However, until now, theoretical results have not been established for the constrained case. As far as we know, this work is the first one to present global convergence results for a class of derivative-free trust-region methods for constrained optimization.

The proposed algorithm considers quadratic models that approximate the objective function based only on zero-order information. Nevertheless, the gradient of the models must represent adequately the gradient of the objective function at the current point. This property can be achieved by many derivative-free techniques, most of them based in polynomial interpolation [4,10,11,13].

At each iteration, the algorithm considers the Euclidean projection of the gradient of the model at the current point onto the feasible set Ω. The model is minimized subject to Ω and to the trust region, in the sense that an approximate solution of this subproblem must satisfy a Cauchy-type condition. This solution will be accepted or rejected as the new iterate according to the ratio between the actual and predicted reductions, a classical trust-region procedure. In particular, the proposed algorithm can be applied for solving unconstrained problems. In this case, the considered projection is reduced to the gradient of the model. Furthermore, the classical condition on the Cauchy step can be used as an acceptance criterion for the solution of the subproblem.

The paper is organized as follows. In Section 2, we propose a derivative-free trust-region algorithm. In Section 3, we present its global convergence analysis. Conclusions are stated in Section 4. Throughout the paper, the symbol $\|\cdot\|$ denotes the Euclidean norm.

2. The algorithm

In this section we present a general trust-region algorithm for solving the problem (1) that generates a sequence of approximate minimizers of quadratic constrained subproblems. The algorithm allows a great deal of freedom in the construction and resolution of the subproblems.

At each iteration $k \in \mathbb{N}$, we consider the current iterate $x^k \in \Omega$ and the quadratic model

$$Q_k(d) = f(x^k) + (g^k)^T d + \frac{1}{2} d^T G_k d,$$

where $g^k \in \mathbb{R}^n$ and $G_k \in \mathbb{R}^{n\times n}$ is a symmetric matrix. Any quadratic model in this form can be used, as long as it provides a sufficiently accurate approximation of the objective function, in the sense that $g^k = \nabla Q_k(0)$ and G_k satisfy Hypotheses H3 and H4 discussed ahead. We assume little about the Hessian of the models, just symmetry and uniform boundedness, allowing that even linear models may be used by setting $G_k = 0$. We do not use Taylor models because we are interested in the case where derivatives are not available, although we assume they exist.

We consider the stationarity measure at x^k for the problem of minimizing Q_k over the set Ω defined by

$$\pi_k = \|P_\Omega(x^k - g^k) - x^k\|,$$

where P_Ω denotes the orthogonal projection onto Ω, which exists because it is a closed convex set. We say that a point $x^* \in \Omega$ is stationary for the original problem (1) when $\|P_\Omega(x^* - \nabla f(x^*)) - x^*\| = 0$. This is a classical definition of stationarity, since Ω is convex [2,7,28].

For proving convergence to stationary points, we assume that an approximate solution $d^k \in \mathbb{R}^n$ of the trust-region subproblem

$$\begin{array}{ll} \text{minimize} & Q_k(d), \\ \text{subject to} & x^k + d \in \Omega, \\ & \|d\| \leqslant \Delta_k \end{array} \tag{2}$$

satisfies the efficiency condition given by

$$Q_k(0) - Q_k(d^k) \ \geqslant \ c_1 \pi_k \min\left\{\frac{\pi_k}{1+\|G_k\|}, \Delta_k, 1\right\}, \tag{3}$$

where $c_1 > 0$ is a constant independent of k.

Conditions of this kind are well-known in trust-region approaches and were used by several authors in different situations. In the unconstrained case, in which $\Omega = \mathbb{R}^n$, the stationarity measure π_k is simply $\|g^k\|$ and the classical Cauchy step satisfies a similar condition, as proved in [21, Lemma 4.5] and [11, Theorem 10.1] with and without derivatives of the objective function, respectively. Conditions of this type also appear throughout [7], under different contexts. In [16], the authors prove the global convergence of a filter method for nonlinear programming by assuming that an approximate solution of the subproblems satisfies a condition analogous to (3). For bound-constrained nonlinear optimization without derivatives, Tröltzsch [30] also assumes such a condition.

Once computed an approximate solution of (2), we analyse whether or not it produces a satisfactory decrease in the model. As usual in trust-region methods, the trial step is assessed by means of the ratio

$$\gamma_k = \frac{f(x^k) - f(x^k + d^k)}{Q_k(0) - Q_k(d^k)}. \tag{4}$$

We present now a general derivative-free trust-region algorithm with no specification on the model update and on the internal solver for the subproblems. Later we will state assumptions to prove that any accumulation point of the sequence generated by the algorithm is stationary.

Algorithm 1. General Algorithm

Data: $x^0 \in \Omega, \alpha > 0, \Delta_0 > 0, 0 < \tau_1 < 1 \leqslant \tau_2$, $\eta_1 \in (0,1), 0 \leqslant \eta < \eta_1 \leqslant \eta_2$.
Set $k = 0$.
Repeat
  Construct the model Q_k.
  If $\Delta_k > \alpha\pi_k$, then
    $\Delta_{k+1} = \tau_1\Delta_k, d^k = 0$ and $x^{k+1} = x^k$.
Else
  Find an approximate solution d^k of (2).
  If $\gamma_k > \eta$, then
    $x^{k+1} = x^k + d^k$.
  Else
    $x^{k+1} = x^k$.
  If $\gamma_k < \eta_1$, then
    $\Delta_{k+1} = \tau_1\Delta_k$.
  Else
    If $\gamma_k > \eta_2$ and $\|d^k\| = \Delta_k$, then
      $\Delta_{k+1} = \tau_2\Delta_k$.
    Else
      $\Delta_{k+1} = \Delta_k$.
$k = k + 1$.

Note that the model is updated whenever a new point is computed. Otherwise, the trust-region radius is reduced by the factor τ_1. We will prove in the next section that $\Delta_k \to 0$ as $k \to \infty$, what will be important in the proofs of the convergence results. This also suggests that, in light of the Hypothesis H4 stated ahead, given a tolerance $\varepsilon > 0$ and parameters $\beta_1, \beta_2 > 0$, the combination of $\Delta_k \leqslant \beta_1\varepsilon$ and $\pi_k \leqslant \beta_2\varepsilon$ could make a reasonable stopping criterion on implementations of the algorithm. When π_k is small, the iterate is probably close to a solution of the problem of minimizing the model within the feasible set Ω. On the other hand, if Δ_k is large, we cannot guarantee that the model represents properly the objective function. Hence, when $\Delta_k > \alpha\pi_k$, the trust-region radius is reduced in the attempt of finding a more accurate model. Although we could always set $\alpha = 1$, this parameter might be used to balance the magnitude of π_k and Δ_k, according to the problem.

3. Convergence

In this section we prove, under reasonable assumptions, that every accumulation point of the sequence generated by the algorithm is stationary. From now on we assume that the algorithm generates an infinite sequence $(x^k) \subset \Omega$.

For proving the convergence we consider the following hypotheses.

H1. The function f is differentiable and its gradient ∇f is Lipschitz continuous with constant $L > 0$, in Ω.
H2. The function f is bounded below in Ω.
H3. The matrices G_k are uniformly bounded, that is, there exists a constant $\beta \geqslant 1$ such that $\|G_k\| \leqslant \beta - 1$ for all $k \geqslant 0$.
H4. There exists a constant $c_2 > 0$ such that

$$\|g^k - \nabla f(x^k)\| \leqslant c_2\Delta_k$$

for all $k \in \mathbb{N}$.

Hypotheses H1 and H2 impose conditions on the objective function, whereas H3 and H4 describe properties that the interpolation models must satisfy. The first three hypotheses are usual in convergence analysis for both derivative-free and derivative-based trust-region algorithms. Hypothesis H4 states that the model has to properly represent the objective function near the current point. There are algorithms able to find models with such properties without computing $\nabla f(x^k)$, for instance [11, Chapter 6]. The proposed algorithm allows the usage of any technique to fulfill Hypothesis H4, although in literature the most usual procedure is polynomial interpolation [11,15,17,23,29].

For the purpose of our analysis, we shall consider the following sets of indices

$$\mathcal{S} = \{k \in \mathbb{N} \mid \gamma_k > \eta\} \quad \text{and} \quad \overline{\mathcal{S}} = \{k \in \mathbb{N} \mid \gamma_k \geqslant \eta_1\}. \tag{5}$$

The set $\mathcal{S}$ is referred to as the set of the *successful* iterations. Note that $\overline{\mathcal{S}} \subset \mathcal{S}$.

In the following lemma, the constants c_1, L, β and c_2 are the ones defined in (3) and in Hypotheses H1, H3, H4, respectively. The lemma establishes that if the trust-region radius is sufficiently small, then the algorithm will perform a successful iteration.

Lemma 3.1. *Suppose that Hypotheses H1, H3 and H4 hold. Consider the set*

$$\mathcal{K} = \left\{k \in \mathbb{N} \mid \Delta_k \leqslant \min\left\{\frac{\pi_k}{\beta}, \frac{(1-\eta_1)}{\pi_k}c, \alpha\pi_k, 1\right\}\right\}, \tag{6}$$

where $c = \frac{(L+c_2+\frac{\beta}{2})}{c_1}$. *If* $k \in \mathcal{K}$, *then* $k \in \overline{\mathcal{S}}$.

Proof. By the Mean Value Theorem, there exists $t_k \in (0,1)$ such that

$$f(x^k + d^k) = f(x^k) + \nabla f(x^k + t_k d^k)^T d^k.$$

Therefore, by Hypotheses H1, H3 and H4,

$$\begin{aligned}\left|f(x^k) - f(x^k+d^k) - Q_k(0) + Q_k(d^k)\right| &= \left|-\left(\nabla f(x^k+t_kd^k) - g^k\right)^T d^k + \frac{1}{2}(d^k)^T G_k d^k\right| \\ &\leqslant \left(\|\nabla f(x^k+t_kd^k) - \nabla f(x^k)\| + \|\nabla f(x^k) - g^k\|\right)\|d^k\| + \frac{1}{2}\|d^k\|^2\|G_k\| \\ &\leqslant \left(t_kL\|d^k\| + c_2\Delta_k\right)\|d^k\| + \frac{1}{2}\beta\|d^k\|^2.\end{aligned}$$

Since $\|d^k\| \leqslant \Delta_k$ and $t_k \in (0,1)$, we have that

$$\left|f(x^k) - f(x^k+d^k) - Q_k(0) + Q_k(d^k)\right| \leqslant c_0\Delta_k^2, \tag{7}$$

where $c_0 = L + c_2 + \frac{\beta}{2}$.

From (6), for every $k \in \mathcal{K}$ we have that $\Delta_k \leqslant \alpha\pi_k$ and consequently $\pi_k > 0$. Then, it follows from (3) that $Q_k(0) - Q_k(d^k) \neq 0$. Thus, from expressions (4), (7) and (3), for all $k \in \mathcal{K}$

$$|\gamma_k - 1| = \left|\frac{f(x^k) - f(x^k+d^k) - Q_k(0) + Q_k(d^k)}{Q_k(0) - Q_k(d^k)}\right| \leqslant \frac{c_0\Delta_k^2}{c_1\pi_k\min\left\{\frac{\pi_k}{\beta}, \Delta_k, 1\right\}} = \frac{c\Delta_k^2}{\pi_k\min\left\{\frac{\pi_k}{\beta}, \Delta_k, 1\right\}}.$$

By (6),

$$\Delta_k = \min\left\{\frac{\pi_k}{\beta}, \Delta_k, 1\right\} \quad \text{and} \quad \frac{c\Delta_k}{\pi_k} \leqslant 1 - \eta_1.$$

Therefore $|\gamma_k - 1| \leqslant 1 - \eta_1$ and hence $\gamma_k \geqslant \eta_1$. Consequently $k \in \overline{\mathcal{S}}$. □

Hypothesis H4 says that the smaller Δ_k, the better the models represent the objective function. Therefore, it is reasonable to expect that the trust-region radius converges to zero. In the following lemma, we show that the proposed algorithm has this property.

Lemma 3.2. *Suppose that Hypotheses H2 and H3 hold. Then the sequence* (Δ_k) *converges to zero.*

Proof. If $\overline{\mathcal{S}}$ is finite, then there exists $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0, \Delta_{k+1} \leqslant \tau_1\Delta_k$. Thus, (Δ_k) converges to zero. We assume henceforth $\overline{\mathcal{S}}$ is infinite. For any $k \in \overline{\mathcal{S}}$, using (3) and Hypothesis H3 we have

$$f(x^k) - f(x^{k+1}) \geqslant \eta_1\left(Q_k(0) - Q_k(d^k)\right) \geqslant \eta_1 c_1\pi_k\min\left\{\frac{\pi_k}{\beta}, \Delta_k, 1\right\}.$$

As $k \in \overline{\mathcal{S}}$, we have that $\Delta_k \leqslant \alpha\pi_k$ and hence

$$f(x^k) - f(x^{k+1}) \geqslant \eta_1 c_1 \frac{\Delta_k}{\alpha} \min\left\{\frac{\Delta_k}{\alpha\beta}, \Delta_k, 1\right\}.$$

Since $(f(x^k))$ is a nonincreasing sequence and bounded below by Hypothesis H2, the left-hand side of the above expression converges to zero. Then,

$$\lim_{k\in\overline{\mathcal{S}}} \Delta_k = 0. \tag{8}$$

Consider the set

$$\mathcal{U} = \left\{k \in \mathbb{N} \mid k \notin \overline{\mathcal{S}}\right\}.$$

If $\mathcal{U}$ is finite, then by (8) we have that $\lim_{k\to\infty} \Delta_k = 0$. Now suppose that $\mathcal{U}$ is infinite. Consider $k \in \mathcal{U}$ and define ℓ_k the last index in $\overline{\mathcal{S}}$ before k. Then ℓ_k is well-defined for all large k and $\Delta_k \leqslant \tau_2\Delta_{\ell_k}$, which implies that

$$\lim_{k\in\mathcal{U}} \Delta_k \leqslant \tau_2 \lim_{k\in\mathcal{U}} \Delta_{\ell_k} = \tau_2 \lim_{\ell_k\in\overline{\mathcal{S}}} \Delta_{\ell_k}.$$

By (8) it follows that $\lim_{k\in\mathcal{U}} \Delta_k = 0$ which completes the proof. □

The next lemma provides a weak convergence result for the problem of minimizing the model within the feasible set Ω. We prove that the sequence (π_k) has a subsequence converging to zero.

Lemma 3.3. *Suppose that Hypotheses H1 to H4 hold. Then*

$$\liminf_{k\to\infty} \pi_k = 0.$$

Proof. The proof is by contradiction. Suppose that there exist a constant $\varepsilon > 0$ and an integer $K > 0$ such that $\pi_k \geqslant \varepsilon$ for each $k \geqslant K$. Take $\tilde{\Delta} = \min\left\{\frac{\varepsilon}{\beta}, \frac{(1-\eta_1)\varepsilon}{c}, \alpha\varepsilon, 1\right\}$ where β is the constant of Hypothesis H3, c is defined in Lemma 3.1, η_1 and $\alpha > 0$ are given in Algorithm 1.
Consider $k \geqslant K$. If $\Delta_k \leqslant \tilde{\Delta}$, then $k \in \mathcal{K}$, with $\mathcal{K}$ given in (6). By Lemma 3.1, $k \in \overline{\mathcal{S}}$ and thus $\Delta_{k+1} \geqslant \Delta_k$. It follows that the radius of the trust region can only decrease if $\Delta_k > \tilde{\Delta}$, and in this case, $\Delta_{k+1} = \tau_1\Delta_k > \tau_1\tilde{\Delta}$.

Therefore, one can see that for all $k \geqslant K$

$$\Delta_k \geqslant \min\left\{\tau_1\tilde{\Delta}, \Delta_K\right\}, \tag{9}$$

which contradicts Lemma 3.2 and concludes the proof. □

Assuming a sufficient decrease in the objective function, that is, setting $\eta > 0$ in the algorithm, we can prove that not only there exists a subsequence of π_k converging to zero, as stated in previous lemma, but also that the convergence prevails for the whole sequence.

Lemma 3.4. *Suppose that Hypotheses H1 to H4 hold and $\eta > 0$. Then*

$$\lim_{k\to\infty} \pi_k = 0.$$

Proof. Suppose by contradiction that for some $\varepsilon > 0$ the set

$$\mathbb{N}' = \{k \in \mathbb{N} \mid \pi_k \geqslant \varepsilon\} \tag{10}$$

is infinite. By Lemma 3.2, the sequence (Δ_k) converges to zero. Then, there exists $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0$,

$$\Delta_k \leqslant \min\left\{\frac{\varepsilon}{\beta}, \frac{(1-\eta_1)\varepsilon}{c}, \alpha\varepsilon, 1\right\}, \tag{11}$$

where the constants β and c are given in Lemma 3.1, and $\alpha > 0$ and η_1 are defined in Algorithm 1.
By (10), for all $k \in \mathbb{N}'$ with $k \geqslant k_0$,

$$\Delta_k \leqslant \min\left\{\frac{\pi_k}{\beta}, \frac{(1-\eta_1)\pi_k}{c}, \alpha\pi_k, 1\right\}, \tag{12}$$

consequently by Lemma 3.1, $k \in \overline{\mathcal{S}} \subset \mathcal{S}$.

Given $k \in \mathbb{N}'$ with $k \geqslant k_0$, consider ℓ_k the first index such that $\ell_k > k$ and $\pi_{\ell_k} \leqslant \varepsilon/2$. The existence of ℓ_k is ensured by Lemma 3.3. So, $\pi_k - \pi_{\ell_k} \geqslant \varepsilon/2$. Using the definition of π_k, the triangle inequality and the contraction property of projections, we have that

$$\begin{aligned}
\frac{\varepsilon}{2} &\leqslant \|P_\Omega(x^k - g^k) - x^k\| - \|P_\Omega(x^{\ell_k} - g^{\ell_k}) - x^{\ell_k}\| \\
&\leqslant \|P_\Omega(x^k - g^k) - x^k - P_\Omega(x^{\ell_k} - g^{\ell_k}) + x^{\ell_k}\| \leqslant 2\|x^k - x^{\ell_k}\| + \|g^k - g^{\ell_k}\| \\
&= 2\|x^k - x^{\ell_k}\| + \|g^k - \nabla f(x^k) + \nabla f(x^k) - \nabla f(x^{\ell_k}) + \nabla f(x^{\ell_k}) - g^{\ell_k}\| \\
&\leqslant 2\|x^k - x^{\ell_k}\| + \|g^k - \nabla f(x^k)\| + \|\nabla f(x^k) - \nabla f(x^{\ell_k})\| + \|\nabla f(x^{\ell_k}) - g^{\ell_k}\|.
\end{aligned}$$

So, using Hypotheses H1 and H4,

$$\frac{\varepsilon}{2} \leqslant (2 + L)\|x^k - x^{\ell_k}\| + c_2(\Delta_k + \Delta_{\ell_k}). \tag{13}$$

Let us consider $C_k = \{i \in \mathcal{S} \mid k \leqslant i < \ell_k\}$. Note that, by (12), $k \in \mathcal{S}$, so $C_k \neq \emptyset$. For each $i \in C_k$, using the fact that $i \in \mathcal{S}$, condition (3) and Hypothesis H3, we conclude that

$$f(x^i) - f(x^{i+1}) \geqslant \eta\Big(Q_i(0) - Q_i(d^i)\Big) \geqslant \eta c_1 \pi_i \min\left\{\frac{\pi_i}{\beta}, \Delta_i, 1\right\}.$$

By the definition of ℓ_k, we have that $\pi_i > \varepsilon/2$ for all $i \in C_k$. As $i \geqslant k$, by (11) $\Delta_i \leqslant \varepsilon/\beta$ and $\Delta_i \leqslant 1$. Therefore

$$\frac{\Delta_i}{2} \leqslant \frac{\varepsilon}{2\beta} \leqslant \frac{\pi_i}{\beta}.$$

It follows that

$$f(x^i) - f(x^{i+1}) > \frac{\eta c_1 \varepsilon \Delta_i}{4}$$

and hence

$$\Delta_i < \frac{4}{\eta c_1 \varepsilon}(f(x^i) - f(x^{i+1})). \tag{14}$$

On the other hand,

$$\|x^k - x^{\ell_k}\| \leqslant \sum_{i \in C_k} \|x^i - x^{i+1}\| \leqslant \sum_{i \in C_k} \Delta_i,$$

which combined with (14) provides

$$\|x^k - x^{\ell_k}\| < \frac{4}{\eta c_1 \varepsilon}(f(x^k) - f(x^{\ell_k})).$$

By Hypothesis H2, the sequence $(f(x^k))$ is bounded below, and since it is nonincreasing, $f(x^k) - f(x^{\ell_k}) \to 0$. Therefore the subsequence $(\|x^k - x^{\ell_k}\|)_{k \in \mathbb{N}'}$ converges to zero, which together with Lemma 3.2 contradicts (13), completing the proof. □

Now we have all the ingredients for proving global convergence to first-order stationary points. In the following theorem we establish a relation between the measure of stationarity for the original problem and the measure of stationarity given in Lemmas 3.3 and 3.4, which provides the global convergence result.

Theorem 3.5. *Suppose that Hypotheses H1 to H4 hold. Then*

(i) *If* $\eta = 0$, *then* $\liminf_{k \to \infty} \|P_\Omega(x^k - \nabla f(x^k)) - x^k\| = 0$.
(ii) *If* $\eta > 0$, *then* $\lim_{k \to \infty} \|P_\Omega(x^k - \nabla f(x^k)) - x^k\| = 0$.

Proof. By the triangle inequality, the contraction property of projections and Hypothesis H4, we have that

$$\begin{aligned}
\|P_\Omega(x^k - \nabla f(x^k)) - x^k\| &= \|P_\Omega(x^k - \nabla f(x^k)) - P_\Omega(x^k - g^k) + P_\Omega(x^k - g^k) - x^k\| \\
&\leqslant \|P_\Omega(x^k - \nabla f(x^k)) - P_\Omega(x^k - g^k)\| + \|P_\Omega(x^k - g^k) - x^k\| \\
&\leqslant \|\nabla f(x^k) - g^k\| + \|P_\Omega(x^k - g^k) - x^k\| \\
&\leqslant c_2\Delta_k + \pi_k.
\end{aligned}$$

Using Lemmas 3.2, 3.3 and 3.4, we complete the proof. □

We emphasize that liminf-type convergence result is still guaranteed if the new iterate is moved to a point with lower objective function value, since Lemmas 3.1, 3.2 and 3.3 remain valid. This can be an interesting strategy if a better point is found during the construction of the model, which can occur, for example, when the model is obtained by interpolation.

4. Conclusions

In this work we proposed a general derivative-free trust-region algorithm for minimizing a smooth objective function in a closed convex set. The algorithm has a very simple structure and it allows a certain degree of freedom on the choice of the models, as long as they approximate sufficiently well the objective function, in the sense of Hypothesis H4. Furthermore, any internal algorithm can be utilized for solving the subproblems, provided that it generates a sequence of points satisfying the efficiency condition (3). Under these hypotheses and other standard assumptions, we have established global convergence of the algorithm in a neat way. Further research is necessary in order to extend the analysis to derivative-free problems in general domains.

Acknowledgement

The authors are grateful to José Mario Martínez who encouraged them to pursue this investigation and to Celso Penteado Serra (*in memorian*) and Luis Mauricio Graña Drummond for their valuable comments and suggestions. We also thank the anonymous referees whose suggestions led to great improvements in the paper.

References

[1] B.M. Arouxét, N. Echebest, A. Pilotta, Active-set strategy in Powell's method for optimization without derivatives, Computational and Applied Mathematics 30 (1) (2011) 171–196.
[2] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, USA, 1995.
[3] L.F. Bueno, A. Friedlander, J.M. Martínez, F.N.C. Sobral, Inexact restoration method for derivative-free optimization with smooth constraints. SIAM Journal on Optimization 23 (2) (2013) 1189–1213.
[4] P.G. Ciarlet, P.A. Raviart, General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite elements methods, Archive for Rational Mechanics and Analysis 46 (1972) 177–199.
[5] A.R. Conn, N.I.M. Gould, A. Sartenaer, Ph.L. Toint, Convergence properties of minimization algorithms for convex constraints using a structured trust region, SIAM Journal on Optimization 6 (4) (1996) 1059–1086.
[6] A.R. Conn, N.I.M. Gould, Ph.L. Toint, Global convergence of a class of trust region algorithms for optimization with simple bounds, SIAM Journal on Numerical Analysis 25 (2) (1988) 433–460.
[7] A.R. Conn, N.I.M. Gould, Ph.L. Toint, Trust-Region Methods, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.
[8] A.R. Conn, K. Scheinberg, Ph.L. Toint, On the convergence of derivative-free methods for unconstrained optimization, in: M.D. Buhmann, A. Iserles (Eds.), Approximation Theory and Optimization: Tributes to M.J.D. Powell, Cambridge University Press, 1997, pp. 83–108.
[9] A.R. Conn, K. Scheinberg, Ph.L. Toint. A derivative free optimization algorithm in practice, in: Proceedings of the AIAA Conference, St Louis, 1998.
[10] A.R. Conn, K. Scheinberg, L.N. Vicente, Geometry of interpolation sets in derivative free optimization, Mathematical Programming 111 (2008) 141–172.
[11] A.R. Conn, K. Scheinberg, L.N. Vicente, Introduction to Derivative-Free Optimization, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.
[12] A.R. Conn, Ph. L. Toint, An algorithm using quadratic interpolation for unconstrained derivative free optimization, in: G. Di Pillo, F. Gianessi (Eds.), Nonlinear Optimization and Applications, Plenum, 1996, pp. 27–47.
[13] P.J. Davis, Interpolation and Approximation, Blaisdell, New York, 1963.
[14] M.A. Diniz-Ehrhardt, J.M. Martínez, L.G. Pedroso, Derivative-free methods for nonlinear programming with general lower-level constraints, Computational and Applied Mathematics 30 (2011) 19–52.
[15] G. Fasano, J.L. Morales, J. Nocedal, On the geometry phase in model-based algorithms for derivative-free optimization, Optimization Methods and Software 24 (2009) 145–154.
[16] C.C. Gonzaga, E.W. Karas, M. Vanti, A globally convergent filter method for nonlinear programming, SIAM Journal on Optimization 14 (3) (2003) 646–669.
[17] S. Gratton, Ph.L. Toint, A. Tröltzsch, An active set trust-region method for derivative-free nonlinear bound-constrained optimization, Optimization Methods and Software 26 (2011) 873–894.
[18] T.G. Kolda, R.M. Lewis, V. Torczon, A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints, Technical Report SAND2006-5315, Sandia National Laboratories, 2006.
[19] T.G. Kolda, R.M. Lewis, V. Torczon, Stationarity results for generating set search for linearly constrained optimization, SIAM Journal on Optimization 17 (2006) 943–968.
[20] R.M. Lewis, V. Torczon, Pattern search algorithms for linearly constrained minimization, SIAM Journal on Optimization 10 (2000) 917–941.
[21] J. Nocedal, S.J. Wright, Numerical Optimization, Springer Series in Operations Research, Springer-Verlag, 1999.
[22] M.J.D. Powell, A Direct Search Optimization Method that Models the Objective and Constraint Functions by Linear Interpolation, in: S. Gomez, J.P. Hennart (Eds.), Advances in Optimization and Numerical Analysis, Kluwer Academic, Dordrecht, 1994, pp. 51–67.
[23] M.J.D. Powell, The NEWUOA software for unconstrained optimization without derivatives, in: G. Di Pillo, M. Roma (Eds.), Large-Scale Nonlinear Optimization, Springer, New York, 2006, pp. 255–297.
[24] M.J.D. Powell, Developments of NEWUOA for minimization without derivatives, IMA Journal of Numerical Analysis 28 (2008) 649–664.
[25] M.J.D. Powell, The BOBYQA algorithm for bound constrained optimization without derivatives, Technical Report DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, Cambridge, England, August 2009.
[26] M.J.D. Powell, On derivative-free optimization with linear constraints, in: 21st ISMP, Berlin, Germany, 2012.
[27] M.J.D. Powell, On the convergence of trust region algorithms for unconstrained minimization without derivatives, Computational Optimization and Applications, 2012, 1–29. (Online First).
[28] A.A. Ribeiro, E.W. Karas. Otimização Contínua: aspectos teóricos e computacionais, Cengage Learning, São Paulo, Brazil, in press (in Portuguese).
[29] K. Scheinberg, Ph.L. Toint, Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization, SIAM Journal on Optimization 20 (6) (2010) 3512–3532.
[30] A. Tröltzsch. An active-set trust-region method for bound-constrained nonlinear optimization without derivatives applied to noisy aerodynamic design problems, Ph.D. thesis, Université de Toulouse, 2011.
[31] D. Winfield, Function minimization by interpolation in a data table, IMA Journal of Applied Mathematics 12 (3) (1973) 339–347.