A globally convergent trust-region algorithm for unconstrained derivative-free optimization

P. S. Ferreira* E. W. Karas† M. Sachine‡

March 10, 2014

Abstract

In this work we explicit a derivative-free trust-region algorithm for unconstrained optimization based on the paper (Computational Optimization and Applications 53: 527–555, 2012) proposed by Powell. The objective function is approximated by quadratic models obtained by polynomial interpolation. The number of points of the interpolation set is fixed. In each iteration only one interpolation point can be replaced and consequently the objective function is evaluated only once per iteration. The method involves two types of steps: trust-region and alternative steps. In a trust-region step the model is minimized in the hope that the reduction achieved by the model is inherited by the objective function. An alternative step aims to keep the good position of the interpolation points. We present the method in an algorithmic scheme and we discuss in details the results related to the global convergence of the algorithm, proving that, under reasonable hypotheses, any accumulation point of the sequence generated by the algorithm is stationary.

Keywords: Unconstrained minimization, derivative-free optimization, trust-region methods, polynomial interpolation and global convergence

1 Introduction

In this paper we address unconstrained nonlinear programming problems in which the derivatives of the objective function are unavailable or unreliable and the function evaluations are computationally costly and possibly noisy. There are many applications of derivative-free optimization in diverse areas such as engineering design, circuit design, dynamic pricing, molecular geometry, groundwater community and medical image registration as mentioned in [8]. Such situations motivated researchers to pursuit techniques for derivative-free optimization.

In particular, M. J. D. Powell has contributed largely in this area. In 1964 he started his exhaustive study in this context with the paper [14]. For unconstrained optimization problems, Powell proposed UOBYQA [15] and NEWUOA [16] trust-region algorithms with good performance on practical deals. Later he suggested developments of NEWUOA in [17]. Based on [16], in [1] and [11], the authors presented an algorithm for solving bound and linearly constrained problems, respectively. In [18] Powell proposed the BOBYQA algorithm for box constrained problems. Despite the practical goals, the algorithms [15, 16, 18] do not have global convergence results. In order to ensure theoretical results, he presented in [19] a framework for unconstrained problems.

Several authors extended classical techniques for solving derivative-free optimization problems. In [6, 7] the authors proposed globally convergent trust-region algorithms for unconstrained optimization and in [4] for closed domains constrained problems. The use of multivariate

*E-mail: pri.savulski@hotmail.com
† ewkaras@ufpr.br, ewkaras@gmail.com
‡ mael@ufpr.br
Department of Mathematics, Federal University of Paraná, CP 19081, 81531-980, Curitiba, PR, Brazil.
This work was supported by CNPq under grants 133154/2011-4, 472313/2011-8, 477611/2013-3, 307714/2011-0.

interpolation techniques are explored in [9] for derivative-free optimization problems. In [12], the authors proposed proximal point method for unconstrained optimization. Augmented Lagrangian methods for constrained optimization are proposed in [10]. The inexact restoration algorithm can be too adapted to constrained derivative-free problems as presented in [3]. A review of derivative-free algorithms can be found in [8, 20].

Due to the importance of the work of Powell, we present an explicit trust-region algorithm, based in [19], for solving unconstrained optimization problems without using derivatives. As usual in trust-region algorithms [2, 5, 13], this approach employs a quadratic approximation of the objective function. In general the quadratic model is obtained by a Taylor approximation, using thereby derivatives. Since the approach discussed in this paper does not utilize derivatives, the models are built through a quadratic polynomial interpolation as in [6]. We also present a detailed theoretical study of the global convergence results of the algorithm.

This paper is organized as follows. In Section 2 we describe the proposed algorithm for derivative-free unconstrained minimization. Some properties of the algorithm are analyzed in Section 3. The global convergence of the algorithm is discussed in Section 4. Finally, Section 5 is devoted to conclusions and lines for future research.

2 The algorithm

The purpose of this work is to discuss the global convergence of a trust-region algorithm for minimizing a twice differentiable objective function $F : \mathbb{R}^n \to \mathbb{R}$ without using its derivatives. The algorithm is based on the method described by Powell [19]. In each iteration k, a quadratic model Q^k of F is constructed by polynomial interpolation with $n+1$ affine linear points, which defines the model uniquely at the linear case. The algorithm can perform two types of attempts: trust-region and alternative attempts. The trust-region ones aim to minimize the model in hope that most of the reduction obtained by the model is inhered by the objective function. The alternative ones aim to improve the geometry of the interpolation set. These alternative attempts are divided in two variations, alpha and beta attempts. The objective function is evaluated at just one point on each iteration. So at some iteration the algorithm performs the three types of attempts until one of these compute F at some point, then this iteration receives the name of this accepted attempt and the index of the iteration is increased.

The interpolation sets are represented by $P^k = \{y_0,\ y_1,\ \ldots,\ y_n\}$ where

$$F(y_0) \ \leq \ F(y_i) \tag{1}$$

for all $i \in \{1, 2, \ldots, n\}$.

We start with $x^1 = y_0$ for $y_0 \in P^1$. At iteration k, once computed the step d^k by some type of attempt, we set

$$x^{k+1} = \begin{cases} x^k + d^k, & \text{if } \ F(x^k + d^k) < F(x^k) \\ x^k, & \text{otherwise.} \end{cases} \tag{2}$$

Moreover since (1) holds at each iteration k it is reasonable to take $y_0 \in P^k$ as $y_0 = x^k$. Thus the sequence $\left(F(x^k)\right)_{\mathbb{R}^n}$ is monotonically nonincreasing. Consider $y_0 \in P^k$ and $y_0^+ \in P^{k+1}$, then

$$\|y_0^+ - y_0\| \leq \left\|d^k\right\|, \tag{3}$$

where $\| \,.\, \|$ denotes the Euclidean norm on $\mathbb{R}^n$.

The quadratic model Q^k that approximates F around the point x^k is of the form

$$Q^k(x) \ = \ F(x^k) \ + \ (x - x^k)^T g^k \ + \ \frac{1}{2}(x - x^k)^T G^k (x - x^k), \tag{4}$$

where $g^k \in \mathbb{R}^n$ and $G^k \in \mathbb{R}^{n\times n}$. The Hessians G^k are taken arbitrarily provided that they are bounded symmetric matrices, they may even be zero, unless an exception discussed later. The gradients g^k are obtained uniquely with the following $n+1$ interpolation conditions

$$Q^k(y_i) = F(y_i), \tag{5}$$

for all $y_i \in P^k$ and $i \in \{0, 1, \ldots, n\}$.

2.1 Trust-region attempt

At each iteration k we consider a trust region, centered at $y_0 \in P^k$, where we believe that the model represents the objective function adequately. The radius of this region is denoted by ρ which can be reduced or not. Moreover at iteration k, ν receives the index of the last iteration in which the trust-region radius was reduced. Note that $\nu \le k$.

In a trust-region attempt the step d^k is computed as a solution of the problem

$$\begin{array}{ll} \text{minimize} & Q^k(x^k + d) \\ \text{subject to} & \|d\| \le \rho. \end{array} \tag{6}$$

Considering the Cauchy step d_c^k as the multiple of the gradient g^k that minimizes the model Q_k in the trust region, the step d^k is defined as any vector satisfying

$$Q^k(x^k + d^k) \le Q^k(x^k + d_c^k) \quad \text{and} \quad \left\|d^k\right\| \le \rho, \tag{7}$$

except if $k \ge \nu + 5$ and $\eta^k = 0$, where

$$\eta^k = \begin{cases} 0, & \text{if } k = \nu \\ \max\left\{\left|Q^j(x^j + d^j) - F(x^j + d^j)\right| : j = \nu, \ldots, k-1\right\}, & \text{if } k > \nu. \end{cases} \tag{8}$$

In this case d^k must be an exact solution of (6).

In order to avoid unnecessary function evaluations, we just compute $F(x^k + d^k)$ whenever the following conditions are satisfied

$$Q^k(x^k) \;-\; Q^k(x^k + d^k) \;>\; \gamma\eta^k \qquad \text{and} \qquad \left\|d^k\right\| \;\ge\; \frac{1}{2}\rho, \tag{9}$$

for a given $\gamma > 0$. If one of these conditions does not hold, this trust-region attempt is called unsuccessful and another type of attempt is tried. Otherwise, the objective function is evaluated at $x^k + d^k$ and we compare the actual and the predicted reductions. If

$$F(x^k) \;-\; F(x^k + d^k) \;\ge\; 0.1\left\{Q^k(x^k) \;-\; Q^k(x^k + d^k)\right\}, \tag{10}$$

this trust-region attempt is called successful. If it does not hold this attempt is unsuccessful too, but since the function F was evaluated, the index of the iteration is incremented.

Thus, when (9) and (10) hold, the trust-region iteration is successful. In this case, the point $x^k + d^k$ is accepted and the interpolation set is updated as $P^{k+1} = P^k \cup \left\{x^k + d^k\right\} \setminus \{y_t\}$, where $t = \text{argmax}\left\{|\theta_i| : i = 1, \ldots, n\right\}$ and θ_i, for $i = 0, 1, \ldots, n$ are such that

$$x^k + d^k = \sum_{i=0}^{n} \theta_i y_i \qquad \text{and} \qquad \sum_{i=0}^{n} \theta_i = 1. \tag{11}$$

Since $x^k = y_0$, (11) can be rewritten as

$$d^k = \sum_{i=1}^{n} \theta_i \left(y_i - y_0\right). \tag{12}$$

Note that at least one of θ_i, $i = 1, \ldots, n$, is not null. Otherwise $d^k = 0$ which is impossible in successful trust-region iterations.

2.2 Alternative attempt

The alternative attempts aim to improve the geometry of the interpolation set. In these attempts first is computed a candidate to leave the interpolation set and then if necessary an alternative step d^k is computed. The choice of the candidate to leave the interpolation set is performed differently by an alpha or a beta attempt. But in both cases the step d^k is computed the same way.

- **Alpha attempt** An alpha attempt aims to maximize the volume of the convex hull provided by the interpolation set. Fixed $i \in \{1, \dots, n\}$, consider $\mathcal{H}_i \subset \mathbb{R}^n$ the hyperplane that contains the interpolation points y_j, for $j \in \{0, 1, \dots, n\} \setminus \{i\}$. Let σ_i be the distance from y_i to $\mathcal{H}_i$. Then the point y_t candidate to leave the interpolation set is the one that provides the shortest distance σ_i, i.e., $\sigma_t \leq \sigma_i$ for all $i = 1, \dots, n$. If this distance is not sufficiently small, then we do not change the interpolation set in order to avoid unneeded evaluations of F. Otherwise, if
$$\sigma_t < \alpha\rho, \tag{13}$$
for $\alpha \in (0, 1)$, this attempt is successful and the interpolation point y_t will not belong to P^{k+1}.

- **Beta attempt** The purpose of this beta attempt is to prevent the points from getting too far from the trust-region center. For this, the distances between each interpolation point y_i to the center x^k are computed. In this attempt an auxiliary set $\mathcal{B} \subset \{1, 2, \dots, n\}$ is used. The point y_t candidate to not belong to the new interpolation set is the one that provides
$$\left\| y_t - x^k \right\| \geq \left\| y_i - x^k \right\|, \tag{14}$$
for all $i \in \mathcal{B}$. In order to avoid unneeded evaluations at F, the point y_t that satisfies (14) will be accepted if, and only if,
$$\mathcal{B} \neq \emptyset \quad \text{and} \quad \left\| y_t - x^k \right\| > \beta\rho, \tag{15}$$
for $\beta > 1$. If these conditions hold then this beta attempt is successful. Otherwise, this attempt is unsuccessful and other type of attempt is tried. Moreover if this attempt is unsuccessful and the last trust-region attempt was also unsuccessful, then the radius is updated $\left(\rho = \dfrac{\rho}{10}\right)$, the iteration is not incremented and ν receives the index of this iteration.

Therefore, if (13) or (15) holds in a successful alpha or beta iteration, respectively, the point $x^k + d^k$ is taken as one of the points at the interpolation set that maximize the volume of the convex hull of the set $\{y_0, \dots, y_n\} \setminus \{y_t\} \cup \left\{x^k + d^k\right\}$. This set will be the interpolation set for the next iteration. It is easy to see that the step d^k is the orthogonal vector to the hyperplane $\mathcal{H}_t$, with norm ρ. Among the two possible direction, the alternative step d^k is chosen as the one that provides the smallest value of the model.

The set $\mathcal{B}$ is full whenever a trust-region attempt is successful or the radius decreases. Moreover it loses an element t, unless $t \notin \mathcal{B}$, whenever the interpolation point $y_t \in P^k$ will not belong to the new interpolation set, i. e., whenever an alpha or beta attempt is successful.

Note that the interpolation points in a successful iteration k satisfy
$$y_0^+ = x^k \qquad \text{and} \qquad y_t^+ = x^k + d^k \tag{16}$$
or
$$y_0^+ = x^k + d^k \qquad \text{and} \qquad y_t^+ = x^k \tag{17}$$
in an alternative one or just (17) in a trust-region one, where $y_0^+ = x^{k+1}$ and $y_t^+ \in P^{k+1}$. The other interpolation points keep the same.

Some general requirements are needed, the first and second attempts performed by the method are alpha and trust-region one, respectively. This pattern is always repeated after each reduction of the radius. Moreover, between two trust-region attempts with the same radius the algorithm can perform at most an alpha and a beta attempt. In order to guarantee a good geometry of the interpolation set, we perform at most τ_α (τ_β) trust-region attempts between two consecutive alpha (beta) attempts with the same radius, for a given $\tau_\alpha > 0$ ($\tau_\beta > 0$). The Hessian of the model is not arbitrary whenever $\eta^{k+1} = 0$ and $k \geq \nu + 5$. In this case set

$$G^{k+1} = G^k,$$

and consequently $Q^{k+1} \equiv Q^k$ since the freedom in the model lies in the choice of the Hessian.

Finally, we express an algorithm as a finite list of well-defined instructions for the method proposed by Powell [19]. The algorithm has some auxiliary variables: c_α, utr, aux_ν, c_{τ_α} and c_{τ_β}. The three first ones are binary and the two last variables are integers. The variable c_α assumes the value 1 whenever an alpha attempt is performed. When a trust-region attempt or iteration is unsuccessful the variable utr assumes the value 1. The variable aux_ν becomes 1 immediately after a decrease at the trust-region radius. At last, c_{τ_α} (c_{τ_β}) counts the quantity of performed trust-region attempts with the same radius between two consecutive alpha (beta) attempts.

The indices k of the interpolation points and the trust-region radius are omitted to not overload the notation.

Algorithm 1

Data: $y_0, y_1, \ldots, y_n \in \mathbb{R}^n,\ G^1 \in \mathbb{R}^{n\times n},\ \rho > 0$
Parameters: $\alpha \in (0,1),\ \beta > 1,\ \gamma > 0,\ \tau_\alpha \in \mathbb{N},\ \tau_\beta \in \mathbb{N}$
Initialization: $c_\alpha = 0,\ c_{\tau_\alpha} = 0,\ c_{\tau_\beta} = 0,\ utr = 0,\ \nu = 1,\ aux_\nu = 1,\ \eta^1 = 0$
$k = 1,\ x^1 = y_0,\ \mathcal{B} = \{1, 2, \ldots, n\}$, determine the model Q^1

REPEAT

- If $c_{\tau_\alpha} \neq \tau_\alpha, c_{\tau_\beta} \neq \tau_\beta, aux_\nu \neq 1$ and $utr = 0$, then (trust-region attempt)
  - $c_\alpha = 0,\ c_{\tau_\alpha} = c_{\tau_\alpha} + 1$ and $c_{\tau_\beta} = c_{\tau_\beta} + 1$
  - If $k \geq \nu + 5$ and $\eta^k = 0$, then
    - Compute the exact solution d^k of the problem (6)
  - Else compute an approximate solution d^k of the problem (6)
  - If $Q^k(x^k) - Q^k(x^k + d^k) > \gamma\eta^k$ and $\|d^k\| \geq \frac{1}{2}\rho$, then
    - Evaluate $F(x^k + d^k)$ and compute η^{k+1}
    - If $F(x^k) - F(x^k + d^k) \geq 0.1\left[Q^k(x^k) - Q^k(x^k + d^k)\right]$, then (★)
      - $t = \operatorname{argmax}\{|\theta_i| : i = 1, 2, \ldots, n\}$, with θ_i satisfying (12)
      - $x^{k+1} = x^k + d^k,\ y_0 = x^{k+1},\ y_t = x^k$
      - Determine the model Q^{k+1} and set $\mathcal{B} = \{1, 2, \ldots, n\}$
    - Else $utr = 1,\ x^{k+1} = x^k$ and $Q^{k+1} = Q^k$ (X)
    - $k = k + 1$
  - Else $utr = 1$ (X)
- Else If $c_\alpha = 0$, then (alpha attempt)
  - ERRO faltou $c_\alpha = 1;\ c_{\tau_\alpha} = 0;\ aux_\nu = 0$
  - $t = \operatorname{argmin}\{\sigma_i : i = 1, 2, \ldots, n\}$
  - If $\sigma_t < \alpha\rho$, then (★)
    - Compute the alternative step d^k and evaluate $F(x^k + d^k)$
    - If $F(x^k + d^k) < F(x^k)$, then $x^{k+1} = x^k + d^k, y_0 = x^{k+1}$ and $y_t = x^k$
    - Else $x^{k+1} = x^k$ and $y_t = x^k + d^k$
    - Compute η^{k+1}, determine the model Q^{k+1} and set $\mathcal{B} = \mathcal{B} \setminus \{t\}$
    - $k = k + 1$
- Else (beta attempt)
  - $c_{\tau_\beta} = 0$
  - If $\mathcal{B} \neq \emptyset$, then $t = \operatorname{argmax}\left\{\|y_i - x^k\| : i \in \mathcal{B}\right\}$
  - Else $t = 0$
  - If $\|y_t - x^k\| > \beta\rho$, then (★)
    - Compute the alternative step d^k and evaluate $F(x^k + d^k)$
    - If $F(x^k + d^k) < F(x^k)$, then $x^{k+1} = x^k + d^k, y_0 = x^{k+1}$ and $y_t = x^k$
    - Else $x^{k+1} = x^k$ and $y_t = x^k + d^k$
    - Compute η^{k+1}, determine the model Q^{k+1} and set $\mathcal{B} = \mathcal{B} \setminus \{t\}$
    - $k = k + 1$
  - Else (X)
    - If $utr = 1$, then
      - $\rho = 0.1\rho,\ aux_\nu = 1,\ \mathcal{B} = \{1, 2, \ldots, n\}$
      - $c_\alpha = 0,\ \nu = k$ and $\eta^k = 0$
  - $utr = 0$

3 Properties of the algorithm

The algorithm above provides the following remarks which guarantee that the descriptions and the exigencies of the method are satisfied.

- At each iteration the algorithm tries at least one of the three attempts identified by their respective names: trust-region, alpha and beta attempts. Each attempt is considered successful if the conditions at the line marked by star (★) are satisfied and then, this iteration receives the name of this accepted attempt and k is increased. Otherwise, this attempt is unsuccessful, the commands after the line identified by (X) are performed and k does not change. The only situation in which k is increased at an unsuccessful attempt is in a trust-region iteration where (9) holds, because in this case $F(x^k + d^k)$ was computed.
- Whenever the objective function is computed, the iteration counter is increased.
- Many consecutive trust-region attempts are performed until c_{τ_α} becomes τ_α or c_{τ_β} reaches τ_β or one of these attempts is unsuccessful.
- After an unsuccessful trust-region attempt, we perform an alpha and a beta attempt, in this order, before the next trust-region attempt.
- If $\mathcal{B}$ is empty and we perform a beta attempt, then this attempt is unsuccessful and the radius is decreased, because the last trust-region one was unsuccessful, unless $n = 1$. Indeed, if the last trust-region iteration was successful, then $\mathcal{B} \neq \emptyset$ in this beta attempt.

The next result establishes the maximum number of iterations performed between two beta attempts.

Lemma 3.1 *The algorithm performs at most* $2\tau_\beta + 1$ *iterations between two consecutive beta attempts.*

Proof. Note that between two consecutive beta attempts, the trust-region radius is the same one, since a change of the radius occurs in a beta attempt. Moreover, the algorithm performs at most τ_β trust-region attempts between two consecutive beta attempts. Consequently, we have at most τ_β trust-region iterations between these beta attempts when all the trust-region attempts compute the objective function in the point obtained. Besides, between each pair of trust-region iterations there is at most an alpha iteration. The proof is completed as it is included an alpha iteration after the first and before the last beta attempts. □

A consequence of the result above is that all considered trust-region iterations are successful except possibly the last one, since whenever a trust-region attempt is unsuccessful we must perform a beta attempt before the next trust-region one.

Next lemma ensures that at least one trust-region attempt will be successful in $(3n+3)$ iterations with the same radius.

Lemma 3.2 *[19, page 538] For every run of* $(3n+3)$ *consecutive iterations with the same radius, at least one is a successful trust-region iteration.*

Proof. Let p and q be iterations ($p < q$) with the same radius such that between them inclusive, all trust-region attempts are unsuccessful. Thus, we must proof that $q-p+1 < 3n+3$ or equivalently $q - p \leq 3n + 1$. We consider just the case $q \geq p + 4$ because, otherwise, $q - p < 4 \leq 3n + 1$, and there is nothing to prove.

From algorithm requirements the possible sequences of three attempts from p to q inclusive are: trust-region/alpha/beta; alpha/beta/trust-region and beta/trust-region/alpha. Moreover, if p was an alpha attempt, other possibility is the sequence alpha/trust-region/alpha, because the last trust-region attempt before p was not necessarily unsuccessful. Thus, every three consecutive iterations between p and q inclusive there exists at least a trust-region attempt. Besides from $p+3$ to q inclusive, every three consecutive iterations includes also at least a beta attempt. Note that all these beta attempts are successful, because otherwise the radius should decrease, since the last trust-region attempt was unsuccessful.

The cardinality of the set $\mathcal{B}$ ($|\mathcal{B}|$) is monotonically decreasing between the iterations p and q inclusive. Besides $|\mathcal{B}| \leq n-1$ at the beginning of the iteration $p+3$, because at least an alpha or a beta iteration is performed until then. As at beta iterations the set $\mathcal{B}$ loses one element, between $p+3$ and q we have at most $n-1$ beta iterations. Moreover, since in this interval every three consecutive iterations includes at least a beta iteration, at most $2(n-1)+2 = 2n$ trust-region and alpha iterations are performed between $p+3$ and q. Therefore, the total number of iterations between them inclusive is at most $3n-1$ which yields the desired conclusion. □

The next result guarantees that the interpolation points obtained at a successful attempt are not getting too far from the center of the new trust region.

Lemma 3.3 *[19, (4.2)] Consider a successful iteration k with radius $\rho > 0$. Then, for all $i \in \{0, 1, \ldots, n\}$,*

$$\left\|y_i^+ - y_0^+\right\| \leq \|y_i - y_0\| + \rho, \tag{18}$$

where $y_i \in P^k$ and $y_i^+ \in P^{k+1}$.

Proof. Since k is a successful iteration, the sets P^k and P^{k+1} differ in one element when (16) or (17) holds. In both cases, we have $y_i^+ = y_i$ for $i \in \{1, 2, \ldots, n\} \setminus \{t\}$ and $y_0 = x^k$. Thus, for $i \in \{1, 2, \ldots, n\} \setminus \{t\}$,

$$\left\|y_i^+ - y_0^+\right\| \leq \left\|y_i^+ - y_i\right\| + \|y_i - y_0\| + \left\|y_0 - y_0^+\right\| \leq \|y_i - y_0\| + \rho,$$

because (3) holds and $\left\|d^k\right\| \leq \rho$. For $i = t$,

$$\left\|y_t^+ - y_0^+\right\| = \left\|d^k\right\| \leq \rho \leq \|y_t - y_0\| + \rho$$

and the proof is complete. □

Consider two matrices $Y^k \in \mathbb{R}^{n\times n}$ and $D^k \in \mathbb{R}^{(n+1)\times(n+1)}$ defined by

$$Y^k = (y_1 - y_0,\ y_2 - y_0,\ \ldots,\ y_n - y_0)$$

and

$$D^k = \begin{pmatrix} y_0 & y_1 & \cdots & y_n \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{19}$$

Note that by subtracting the first column of D^k from the other ones it follows that

$$\det D^k = (-1)^{(n+1)+1} \begin{vmatrix} y_1 - y_0 & \cdots & y_n - y_0 \end{vmatrix} = (-1)^n \det(Y^k),$$

consequently

$$\left|\det D^k\right| = \left|\det Y^k\right|. \tag{20}$$

Moreover, at a successful iteration the interpolation point y_t is replaced by $y_0 + d^k$, then

$$\frac{\left|\det D^{k+1}\right|}{\left|\det D^k\right|} = \frac{\sigma_t^+}{\sigma_t} \tag{21}$$

where σ_t and σ_t^+ are the distances from y_t and $y_0 + d^k$, respectively, to the hyperplane $\mathcal{H}$ defined previously.

The following result provides a relation between D^k and D^{k+1}, when k is a successful trust-region iteration.

Lemma 3.4 *[19, (3.9)] Consider k a successful trust-region iteration. Then*

$$\left|\det D^{k+1}\right| = |\theta_t| \, \left|\det D^k\right|.$$

Proof. Since k is a successful trust-region iteration, (17) holds and $y_i^+ = y_i$, for $i \in \{1, 2, \dots, n\} \setminus \{t\}$. Using (11), $x^k + d^k = \sum_{i=0}^{n} \theta_i y_i$, and subtracting from the first column a linear combination of the other ones with coefficients θ_i, we obtain that

$$\det D^{k+1} = \left| \begin{matrix} \theta_t y_t & y_1 & \cdots & y_{t-1} & y_0 & y_{t+1} & \cdots & y_n \\ \left(1 - \sum\limits_{i=0, i\neq t}^{n} \theta_i\right) & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{matrix} \right|.$$

By (11), we have $1 - \sum\limits_{i=0,\ i\neq t}^{n} \theta_i = \theta_t$. Using properties of determinant we conclude the proof.

□

Since $|\theta_t| > 0$, the previous lemma implies that a successful trust-region iteration keeps the nonsingularity of the interpolation matrix, consequently it keeps the affine linearity of the interpolation set. The following result ensures that this is also kept in alternative iterations.

Lemma 3.5 *[19, (3.3) and (3.4)] For $i = 1, \dots, n$, consider σ_i the distance from $y_i \in P^k$ to the hyperplane $\mathcal{H}_i$ that contains the interpolation points y_j, for $j \in \{0, 1, \dots, n\} \setminus \{i\}$. Then*

$$\prod_{i=1}^{n} \sigma_i \ \leq \ \left|\det D^k\right| \ \leq \ \prod_{i=1}^{n} \|y_i - y_0\|. \tag{22}$$

Proof. For $i = 1, \dots, n$, let s_i be the vector defined by

$$s_i = \begin{cases} (y_i - y_0), & \text{for } i = 1, \\ (y_i - y_0) - \sum\limits_{j=1}^{i-1} \phi_{ij} (y_j - y_0), & \text{for } i = 2, \dots, n, \end{cases}$$

where the coefficients ϕ_{ij} are the values that minimize $\|s_i\|$. Consider S^k the matrix whose columns are the vectors s_i. Denoting by $Y_i = y_i - y_0$ the i-th column of Y^k, the matrix S^k can be written as

$$S^k = \left(Y_1 \quad Y_2 - \phi_{21} Y_1 \quad \cdots \quad Y_n - \sum_{j=1}^{n-1} \phi_{nj} Y_j \right).$$

Using properties of determinant, we conclude that $\det S^k = \det Y^k$. By this and (20), we have

$$\left|\det D^k\right| = \left|\det S^k\right| = \left[\left(\det S^k\right)^2\right]^{1/2} = \left[\det\left(\left(S^k\right)^T S^k\right)\right]^{1/2}.$$

Note that by definition, the vectors s_i are the result of the Gram-Schmidt orthogonalization process of the columns of Y^k. Thus, $\left(S^k\right)^T S^k = \operatorname{diag}\left(\|s_i\|^2\right)$. By the definition of ϕ_{ij}

$$\|s_i\| = \left\| y_i - y_0 - \sum_{j=1}^{i-1} \phi_{ij} (y_j - y_0) \right\| \leq \|y_i - y_0\|,$$

whose inequality holds for all $i \in \{1, 2, \ldots, n\}$. Therefore,

$$\left|\det D^k\right| = \prod_{i=1}^{n} \|s_i\| \leq \prod_{i=1}^{n} \|y_i - y_0\| \tag{23}$$

and the upper bound of (22) is verified.

Now, for $i = 1, 2, \ldots, n$, let $\widehat{s}_i$ be the vector defined by

$$\widehat{s}_i = y_i - \left(y_0 + \sum_{j=1,\ j\neq i}^{n} \psi_{ij} \left(y_j - y_0\right) \right),$$

where ψ_{ij} are the coefficients that minimize $\|\widehat{s}_i\|$. Thus

$$\|\widehat{s}_i\| \leq \left\| y_i - \left(y_0 + \sum_{j=1}^{i-1} \phi_{ij} \left(y_j - y_0\right) \right) \right\|.$$

Therefore, for all $i \in \{1, 2, \ldots, n\}$, we have that

$$\|\widehat{s}_i\| \leq \|s_i\|. \tag{24}$$

Note also that $\widehat{s}_i$ can be rewritten by

$$\widehat{s}_i = y_i - \left[\left(1 - \sum_{j=1,\ j\neq i}^{n} \psi_{ij} \right) y_0 + \sum_{j=1,\ j\neq i}^{n} \psi_{ij} y_j \right] = y_i - \bar{y}_i,$$

where $\bar{y}_i$ is an affine combination of the points $\{y_0,\ y_1,\ \ldots,\ y_n\} \setminus \{y_i\}$, in other words, a vector in the hyperplane $\mathcal{H}_i$. By the definition of ψ_{ij}, it follows that $\|\widehat{s}_i\| = \sigma_i$. Combining this with (23) and (24) we have that

$$\left|\det D^k\right| \geq \prod_{i=1}^{n} \sigma_i,$$

and the proof is completed. □

Note that since the interpolation sets are affine linear, we have the distances σ_i are not null, for all $i \in \{1, 2, \ldots, n\}$. Then the previous result guarantees that the alternative iterations keep the interpolation set affine linear.

The next result provides that the distances from y_0 to each y_i, for all $i \notin \mathcal{B}$, are bounded at each iteration.

Lemma 3.6 *Consider the set $\mathcal{B}$ at the beginning of the iteration k with trust-region radius ρ. If $i \notin \mathcal{B}$, then*

$$\|y_i - y_0\| \leq (3n + 2)\rho.$$

Proof. Consider $\ell < k$ the last iteration before k in which the index i left $\mathcal{B}$. The iteration ℓ is an alternative one and $\mathcal{B} \neq \{1, 2, \ldots, n\}$ from $\ell + 1$ to k. Thus, in these iterations the radius is the same and the trust-region attempts are unsuccessful. By Lemma 3.2, $k - \ell < 3n + 3$ and consequently $\ell \geq k - 3n - 2$. Consider $\bar{\ell}$, with $\ell \leq \bar{\ell} < k$, the iteration such that

$$y_i^{\bar{\ell}} \neq y_i^{\bar{\ell}+1} \qquad \text{and} \qquad y_i^{j+\bar{\ell}} = y_i^{j+\bar{\ell}+1},$$

for $j = 1, 2, \ldots, k - \bar{\ell} - 1$. By the definition of the index $\bar{\ell}$, this iteration is an alternative one and

$$\bar{\ell} \geq k - 3n - 2.$$

Now we prove by induction that, for $j = 1, \ldots, k - \bar{\ell}$,

$$\left\| y_i^{j+\bar{\ell}} - y_0^{j+\bar{\ell}} \right\| \leq j\rho. \tag{25}$$

For $j = 1$, we have that (16) or (17) hold, then $\left\| y_i^{1+\bar{\ell}} - y_0^{1+\bar{\ell}} \right\| = \left\| d^{\bar{\ell}} \right\| = \rho$. Using the triangle inequality, the induction hypothesis and (3) we have that

$$\left\| y_i^{j+\bar{\ell}+1} - y_0^{j+\bar{\ell}+1} \right\| \leq \left\| y_i^{j+\bar{\ell}+1} - y_i^{j+\bar{\ell}} \right\| + \left\| y_i^{j+\bar{\ell}} - y_0^{j+\bar{\ell}} \right\| + \left\| y_0^{j+\bar{\ell}} - y_0^{j+\bar{\ell}+1} \right\| \leq (j+1)\rho.$$

Thus, (25) holds, in particular for $j = k - \bar{\ell} \leq 3n + 2$, completing the proof.

□

The next lemma provides that immediately after a change of the trust-region radius, the interpolation points are bounded by a constant that depends on this radius. Consequently, as the radius is reduced, the interpolation points get closer.

Lemma 3.7 *[19, (4.4)] Consider $\bar{\rho} > 0$ the trust-region radius at the end of an iteration k that reduces the radius. Then, for all $y_i \in P^k$, $i = 1, 2, \ldots, n$,*

$$\|y_i - y_0\| \leq 10 \max\{(3n + 2),\ \beta\}\, \bar{\rho},$$

where $\beta > 1$ is given in the algorithm.

Proof. Consider an arbitrary $i \in \{1, 2, \ldots, n\}$. If $i \in \mathcal{B}$, then $\mathcal{B} \neq \emptyset$. Since the radius was reduced to $\bar{\rho}$, then the last beta attempt with radius $10\bar{\rho}$ was unsuccessful. Thus by (14), (15) and the mechanism of the algorithm we have

$$\|y_i - y_0\| \leq \|y_t - y_0\| \leq 10\beta\bar{\rho}. \tag{26}$$

On the other hand, if $i \notin \mathcal{B}$, the trust-region radius at the beginning of this iteration is $10\bar{\rho}$. By Lemma 3.6

$$\|y_i - y_0\| \leq 10(3n + 2)\bar{\rho},$$

which combined with (26) completes the proof. □

For next results, consider

$$\Gamma_k = \sum_{i=1}^{n} \|y_i - y_0\|, \tag{27}$$

where $y_i \in P^k$, for $i = 1, 2, \ldots, n$. Note that, if the iteration k is successful we obtain from (18) that

$$\Gamma_{k+1} \leq \Gamma_k + n\rho, \tag{28}$$

where ρ is the trust-region radius at iteration k. At an unsuccessful iteration, $\Gamma_{k+1} = \Gamma_k$ and (28) also holds.

Lemma 3.8 *[19, (4.8) and (4.9)] Consider ℓ and m ($\ell < m$) two consecutive beta attempts with the same radius ρ and let $C_1 > 0$ be the constant*

$$C_1 = \max\{(3n + 2),\ \beta,\ (2\tau_\beta + 2)n\}\, n.$$

If $\Gamma_\ell > C_1\rho$, then

$$\Gamma_k < \Gamma_\ell, \tag{29}$$

otherwise,

$$\Gamma_k \leq [C_1 + (2\tau_\beta + 2)\, n]\, \rho,$$

for $\ell < k \leq m$.

Proof. Consider first the case that $\Gamma_\ell > C_1\rho$. By (27)

$$C_1\rho < \Gamma_\ell \leq \sum_{i=1}^{n} \max_{1\leq j\leq n} \|y_j - y_0\| = n\,\|y_{\bar{t}} - y_0\|\,, \tag{30}$$

where $\bar{t} = \text{argmax}\left\{\|y_i - y_0\|\,,\ \text{for } y_i \in P^\ell \text{ and } i = 1,2,\ldots,n\right\}$. From the definition of C_1, $\|y_{\bar{t}} - y_0\| > (3n+2)\rho$ and by Lemma 3.6, $\bar{t} \in \mathcal{B}$. Moreover, from (30), we have that $\|y_{\bar{t}} - y_0\| > \beta\rho$. Therefore, (15) holds and consequently ℓ is a beta iteration with $t = \bar{t}$. Thus, from (16), (17) and the mechanism of an alternative step, we obtain that $\left\|y_t^+ - y_0^+\right\| = \rho$ and

$$\Gamma_{\ell+1} = \sum_{i=1, i\neq t}^{n} \left\|y_i^+ - y_0^+\right\| + \rho.$$

Using Lemma 3.3, (30) and the fact that $t = \bar{t}$ we conclude that

$$\Gamma_{\ell+1} \leq \sum_{i=1, i\neq t}^{n} \|y_i - y_0\| + n\rho = \Gamma_\ell - \|y_t - y_0\| + n\rho \leq \Gamma_\ell - \frac{\Gamma_\ell}{n} + n\rho.$$

Moreover, as $\Gamma_\ell > C_1\rho \geq (2\tau_\beta + 2)\,n^2\rho$, we have that

$$\Gamma_{\ell+1} < \Gamma_\ell - (2\tau_\beta + 1)\,n\rho. \tag{31}$$

Since m is the next beta attempt after the beta iteration ℓ, by Lemma 3.1 between ℓ and m there are at most $2\tau_\beta + 1$ iterations. Once ℓ is a beta iteration, we have that

$$k - (\ell+1) \leq 2\tau_\beta + 1, \tag{32}$$

for all k, such that $\ell < k \leq m$. On the other hand, by (28) it follows that $\Gamma_k \leq \Gamma_{k-1} + n\rho$. Using this recursively, (31) and (32), we have that

$$\Gamma_k \leq \Gamma_{\ell+1} + [k - (\ell+1)]\,n\rho k \leq \Gamma_{\ell+1} + (2\tau_\beta + 1)n\rho < \tau_\beta,$$

for $\ell < k \leq m$, which proves (29).

Consider now the case that $\Gamma_\ell \leq C_1\rho$. Using (28) recursively and Lemma 3.1, we have that

$$\Gamma_k \leq [\Gamma_\ell + (k-\ell)n]\,\rho \leq [C_1 + (2\tau_\beta + 2)n]\,\rho,$$

for $k < \ell \leq m$ and the proof is completed. □

The previous result states that, during all iterations between two consecutive beta attempts, the sum of the distances from the current point to each interpolation point is bounded. Next lemma establishes an analogous result for all iterations between the first one with some radius and the next iteration that performs a beta attempt with this radius.

Lemma 3.9 *[19, (4.10)] Consider ν the first iteration with some trust-region radius ρ and ω the first one after ν that performs a beta attempt. Define $C_2 > 0$ as the constant*

$$C_2 = 10\max\{(3n+2)\,,\ \beta\}\,n. \tag{33}$$

Then

$$\Gamma_k \leq [C_2 + (2\tau_\beta + 2)n]\,\rho,$$

for $\nu \leq k \leq \omega$.

Proof. Since the radius was reduced to ρ at the iteration ν, by (27), Lemma 3.7 and (33)

$$\Gamma_\nu \leq \sum_{i=1}^{n} 10 \max\{(3n+2), \beta\}\, \rho = C_2 \rho.$$

Using (28) recursively and this we obtain that

$$\Gamma_k \leq \Gamma_\nu + (k-\nu) n\rho \leq [C_2 + (\omega - \nu) n]\, \rho, \tag{34}$$

for $\nu \leq k \leq \omega$. Moreover, the iteration ν performs a beta attempt, once the radius was reduced. By Lemma 3.1, we have that $\omega - \nu \leq 2\tau_\beta + 2$ which combined with (34) completes the proof. □

The next result establishes that at all iterations the interpolation points are in a bounded region that depends on the current trust-region radius.

Lemma 3.10 *[19, Lemma 2] Consider k an iteration with trust-region radius ρ. Then, there exists a constant $C_3 > 1$ such that for all $y_i \in P^k$ and $i = 0, 1, \ldots, n$,*

$$\|y_i - y_0\| \leq C_3 \rho.$$

Proof. Consider ν the first iteration with trust-region radius ρ and ω the first one after ν that performs a beta attempt. Since the radius can only decrease in beta attempts, the last attempt computed with this radius is also a beta one. Thus, the iteration k is either between ν and ω or between two consecutive beta attempts. If $\nu \leq k \leq \omega$, by (27) and Lemma 3.9 it follows that, for all $i \in \{1, 2, \ldots, n\}$,

$$\|y_i - y_0\| \leq \Gamma_k \leq [C_2 + (2\tau_\beta + 2) n]\, \rho. \tag{35}$$

Suppose now that $k > \omega$. Then there exist ℓ and m two consecutive iterations that perform a beta attempt such that $\omega \leq \ell \leq k \leq m$. Note that

$$\Gamma_k \leq \max_{\ell \leq j \leq m} \{\Gamma_j\} = \widehat{\Gamma}.$$

As $\widehat{\Gamma} \geq \Gamma_\ell$, by Lemma 3.8,

$$\Gamma_k \leq \widehat{\Gamma} \leq [C_1 + (2\tau_\beta + 2)\, n]\, \rho. \tag{36}$$

Define the constant C_3 by

$$C_3 = \max\{C_1,\ C_2\} + (2\tau_\beta + 2)\, n.$$

Since $C_1, C_2, \tau_\beta > 0$, we have $C_3 > 2n > 1$. Combining (35) and (36), we get the result. □

The iterations that occur between two consecutive alpha attempts also satisfy important properties. First note that, the first attempt performed after each decrease at the trust-region radius is an alpha attempt. Then the attempts performed between two consecutive alpha ones consider the same radius.

Lemma 3.11 *[19, (5.6)] Let $C_3 > 1$ be the constant defined in the previous lemma. Consider ℓ and m two consecutive iterations that performs an alpha attempt. Then, for $\ell < k < m$,*

$$\left|\det D^{k+1}\right| \geq \frac{\left|\det D^k\right|}{2nC_3}, \tag{37}$$

where D^k is given by (19). In particular, if the iteration ℓ performs an unsuccessful alpha attempt, then (37) holds for $k = \ell$.

Proof. Consider ρ the current trust-region radius from $\ell+1$ to m. Suppose first that k is a trust-region iteration. At this case the obtained step d^k can satisfy or not the condition (10). If it does not, then $P^{k+1}=P^k$, consequently $D^{k+1}=D^k$ and (37) is achieved. Otherwise, the values θ_i are computed, for $i=0,1,\ldots,n$, and d^k can be rewritten as (12). Using the triangle inequality, the fact that $t=\operatorname{argmax}\{|\theta_i| : i=1,\ldots,n\}$ and Lemma 3.10,

$$\|d^k\| \le \sum_{i=1}^{n} |\theta_t| \, \|y_i - y_0\| \le |\theta_t| \, nC_3\rho.$$

From this and Lemma 3.4,

$$\left|\det D^{k+1}\right| \ge \frac{\|d^k\|}{nC_3\rho} \left|\det D^k\right|.$$

Moreover, since k is a successful trust-region iteration, the conditions (9) are satisfied. Therefore, (37) holds for all trust-region iteration from ℓ to m.

Suppose now that k is a beta iteration. Consider $t \in \{1,2,\ldots,n\}$ the index that satisfies (14) and $\sigma_t = \min\{\|y_t - x\| : x \in \mathcal{H}_t\}$ the distance from $y_t \in P^k$ to the hyperplane $\mathcal{H}_t$. Using the fact that $y_0 \in \mathcal{H}_t$ and Lemma 3.10, it follows that

$$\sigma_t \le \|y_t - y_0\| \le C_3\rho.$$

By (21) and this, we have that

$$\frac{\left|\det D^{k+1}\right|}{\left|\det D^k\right|} = \frac{\sigma_t^+}{\sigma_t} = \frac{\rho}{\sigma_t} \ge \frac{1}{C_3} \ge \frac{1}{2nC_3},$$

where $\sigma_t^+ = \min\left\{\left\|x^k + d^k - x\right\| : x \in \mathcal{H}_t\right\} = \rho$ since an alternative step was computed.

In particular, if the iteration ℓ performs an unsuccessful alpha attempt, ℓ is a beta or a trust-region iteration and (37) holds for $k=\ell$. □

The Lemmas 3.10 and 3.11 are fundamental to prove the global convergence of the algorithm since they ensure that the affine linearity of the interpolation points is kept along the iterations even when the points became closer and closer.

By Lemma 3.2, in $3n+3$ consecutive iterations with the same radius at least one is a successful trust-region iteration. Moreover, at most τ_α trust-region attempts are performed between two consecutive alpha attempts with the same radius. So, there exists $\widehat{\tau}_\alpha > 0$ such that, every run of $\widehat{\tau}_\alpha$ consecutive iterations with the same radius, at least an alpha attempt is performed.

Lemma 3.12 *[19, (5.7) and (5.10)] Consider $C_3 > 1$ given by Lemma 3.10, $\widehat{\tau}_\alpha > 0$ and $\ell < m$ two consecutive alpha attempts. Let ρ be the trust-region radius from ℓ to $m-1$ and $C_4 > 0$ be the constant defined by*

$$C_4 = \left[\min\left\{\alpha,\ (2nC_3)^{-\widehat{\tau}_\alpha}\right\}\right]^n. \tag{38}$$

If $\left|\det D^\ell\right| < C_4\rho^n$, then ℓ is an alpha iteration and

$$\left|\det D^{k+1}\right| > \left|\det D^\ell\right|, \tag{39}$$

otherwise,

$$\left|\det D^{k+1}\right| \ \ge\ \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}, \tag{40}$$

for all $k=\ell,\ldots,m-1$.

Proof. Suppose first that $\left|\det D^\ell\right| < C_4\rho^n$. By Lemma 3.5 and the definition of C_4,

$$\sigma_t^n \;\le\; \prod_{i=1}^{n}\sigma_i \;\le\; \left|\det D^\ell\right| \;<\; (\alpha\rho)^n\,, \tag{41}$$

where $\sigma_t = \min_{1\le i\le n}\{\sigma_i\}$. Thus, the condition (13) holds, ℓ is an alpha iteration and $P^\ell \neq P^{\ell+1}$. Using (21), (41), the hypothesis, (38) and the fact that $C_3 > 1$, we have that

$$\left|\det D^{\ell+1}\right| \ge \rho\frac{\left|\det D^\ell\right|}{\left|\det D^\ell\right|^{\frac{1}{n}}} \;>\; (2nC_3)^{\widehat{\tau}_\alpha}\left|\det D^\ell\right| \;>\; \left|\det D^\ell\right|, \tag{42}$$

which proves (39) for $k=\ell$. Using Lemma 3.11 for $k=\ell+1$ recursively, we have

$$\left|\det D^{\ell+1+i}\right| \;\ge\; \frac{\left|\det D^{\ell+1}\right|}{(2nC_3)^i},$$

for $i = 1, 2, \ldots, m-(\ell+1)$. By (42),

$$\left|\det D^{\ell+1+i}\right| \;>\; \left|\det D^\ell\right|(2nC_3)^{\widehat{\tau}_\alpha - i} \ge \left|\det D^\ell\right|,$$

since, $i \le m-(\ell+1) < \widehat{\tau}_\alpha$, which proves the result in this case.

Suppose now that $\left|\det D^\ell\right| \ge C_4\rho^n$. If ℓ is an alpha iteration, then the volume of the convex hull increases, i.e., $\left|\det D^{\ell+1}\right| > \left|\det D^\ell\right|$. On the other hand, if the alpha attempt is unsuccessful, (37) holds for $k=\ell$. Since $(2nC_3) > 1$, in both cases,

$$\left|\det D^{\ell+1}\right| \ge \frac{\left|\det D^\ell\right|}{2nC_3}. \tag{43}$$

As $\widehat{\tau}_\alpha > 1$, (40) holds for $k=\ell$. Using again Lemma 3.11 recursively and (43)

$$\left|\det D^{k+1}\right| \ge \frac{\left|\det D^{\ell+1}\right|}{(2nC_3)^{k-\ell}} \ge \frac{\left|\det D^\ell\right|}{(2nC_3)^{k-\ell+1}}$$

for $\ell < k < m$. As $k-\ell+1 \le \widehat{\tau}_\alpha$ and $(2nC_3) > 1$, we conclude the proof. □

Corollary 3.13 *[19, (5.11)] Let $C_3 > 1$ and $C_4 > 0$ be the constants given by Lemmas 3.10 and 3.12, respectively. Consider $\ell < m$ two consecutive iterations that perform alpha attempts and ρ the trust-region radius from ℓ to $m-1$. Then, for all $\ell \le k < m$,*

$$\left|\det D^{k+1}\right| \ge \min\left\{\left|\det D^\ell\right|,\; \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\}. \tag{44}$$

Proof. Consequence of Lemma 3.12. □

The inequality (44) holds for iterations between two consecutive ones that performs an alpha attempt. The next lemma generalizes this result for all iterations with the same radius.

Lemma 3.14 *[19, (5.12)] Let $C_3 > 1$ and $C_4 > 0$ be the constants given by Lemmas 3.10 and 3.12, respectively. Consider ν the iteration in which the trust-region radius decreases to ρ, then*

$$\left|\det D^{k+1}\right| \ge \min\left\{\left|\det D^\nu\right|,\; \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\}, \tag{45}$$

for all $k \ge \nu$ that consider the radius ρ.

Proof. Note that by hypothesis, the iteration ν performs an alpha attempt. Consider $\nu^+ > \nu$ the first iteration that consider the radius 0.1ρ and $\widehat{\ell}$ the first iteration that performs an alpha attempt with $\nu < \widehat{\ell} \leq \nu^+$. By Corollary 3.13, the inequality (45) holds for all $\nu \leq k < \widehat{\ell}$, which concludes the proof for $\widehat{\ell} = \nu^+$.

Consider then the case $\widehat{\ell} \leq k < \nu^+$. Assume by contradiction that (45) does not hold. Consider without loss of generality that $\bar{k}$ is the smallest index in $[\widehat{\ell}, \nu^+)$ such that

$$\left|\det D^{\bar{k}+1}\right| < \min\left\{|\det D^{\nu}|\,,\ \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\}. \tag{46}$$

Thus, for all $i = \nu, \nu+1, \ldots, \bar{k}-1$,

$$\left|\det D^{i+1}\right| \geq \min\left\{|\det D^{\nu}|\,,\ \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\} > \left|\det D^{\bar{k}+1}\right|. \tag{47}$$

Consider $\ell \geq \widehat{\ell}$ and m the first indices before and after $\bar{k}$, respectively, that perform alpha attempts. By (47), $\left|\det D^{\ell}\right| > \left|\det D^{\bar{k}+1}\right|$. Using this and Corollary 3.13, we have that $\left|\det D^{k+1}\right| \geq \dfrac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}$, which contradicts (46) and completes the proof. □

The next result establishes a lower bound for the determinant of the interpolation matrix, which depends on the trust-region radius.

Lemma 3.15 *[19, Lemma 3] Let $C_4 > 0$ be the constant given by Lemma 3.12. Consider ρ_1 the first trust-region radius, ρ the trust-region radius at some iteration, $\widehat{\tau}_\alpha \in \mathbb{N}$ and $C_5 > 0$ the constant given by*

$$C_5 = \min\left\{\frac{\left|\det D^1\right|}{{\rho_1}^n},\ \frac{C_4}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\}.$$

Then, for all iteration k with the radius ρ,

$$\left|\det D^{k}\right| \geq C_5\rho^n. \tag{48}$$

Proof. The proof is by induction in k considering all iterations with each trust-region radius. By definition of C_5 we have that $\left|\det D^1\right| \geq C_5{\rho_1}^n$. By Lemma 3.14, for all iterations $k \geq 1$ with the radius $\rho = \rho_1$,

$$\left|\det D^{k+1}\right| \geq \min\left\{\left|\det D^{1}\right|,\ \frac{C_4\ {\rho_1}^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\} \ = \ C_5\ {\rho_1}^n.$$

Consider ν the first iteration with radius ρ and ν_+ the iteration in which the radius decreases to $\rho^+ = 0.1\rho$. Assume by induction that (48) holds for all $k \in [\nu, \nu_+)$. By Lemma 3.14, the induction hypothesis and the definition of C_5, we have that, for all $\nu \leq k < \nu^+$,

$$\left|\det D^{k+1}\right| \geq \min\left\{|\det D^{\nu}|\,, \frac{C_4\rho^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\} \geq C_5\rho^n.$$

In particular, for the iteration $\nu^+ - 1$ we have that

$$\left|\det D^{\nu^+}\right| \geq C_5\rho^n > C_5\left(\rho^+\right)^n. \tag{49}$$

Moreover, for all iteration $k \geq \nu^+$ with radius ρ^+, by Lemma 3.14 and (49) we have that

$$\left|\det D^{k+1}\right| \geq \min\left\{C_5\left(\rho^+\right)^n,\ \frac{C_4\left(\rho^+\right)^n}{(2nC_3)^{\widehat{\tau}_\alpha}}\right\} \ = \ C_5\left(\rho^+\right)^n,$$

which completes the proof. □

4 Global Convergence

In this section the global convergence of the algorithm is established. From now on we assume that the algorithm generates an infinite sequence (x^k).

For proving the global convergence we consider the following hypotheses.

H1 *The objective function F is bounded below.*

H2 *The Hessians of F are bounded, i.e., there exists a constant $N > 0$ such that $\left\|\nabla^2 F(x)\right\| \leq N$, for all $x \in \mathbb{R}^n$.*

H3 *The Hessians G^k of the models Q^k are uniformly bounded, i.e., there exists a constant $M > 0$ such that $\left\|G^k\right\| = \left\|\nabla^2 Q^k(x)\right\| \leq M$, for all $k \in \mathbb{N}$ and $x \in \mathbb{R}^n$.*

4.1 Weak convergence.

This subsection establishes that there exists an point of the sequence (x^k) that is stationary.

Consider the Lagrange functions Λ_j, $j = 0, 1, \ldots, n$, of the interpolation equations (5), as the linear polynomial $\Lambda_j(x)$, with $x \in \mathbb{R}^n$, which satisfies

$$\Lambda_j(y_i) = \delta_{ij}, \tag{50}$$

for $y_i \in P^k$, $i = 0, 1, \ldots, n$, where δ_{ij} is the Kronecker delta. The coefficients of each Λ_j are defined uniquely due to the nonsingularity of the matrix D^k given by (19). Although these coefficients depend on the interpolation set P^k, the index k will be omitted.

Lemma 4.1 *[19, (3.2)] Consider ρ the trust-region radius at the iteration k. If $\left\|x - x^k\right\| \leq \rho$, then there exists a constant $C_6 > 0$ such that*

$$\sum_{j=0}^{n} |\Lambda_j(x)| \leq C_6.$$

Proof. For each $j = 0, 1, \ldots, n$, consider $\Delta_j(x) \in \mathbb{R}^{(n+1)\times(n+1)}$ a variation of the matrix D^k, where the column j is replaced by the vector $(x, 1) \in \mathbb{R}^{n+1}$. Note that

$$\det \Delta_j(y_i) = \begin{cases} 0, & i \neq j \\ \det D^k, & i = j \end{cases}. \tag{51}$$

Given $x \in \mathbb{R}^n$, the solution in $\mathbb{R}^{n+1}$ of the linear system $D^k\, z = (x, 1)$ is

$$z_j(x) = \frac{\det \Delta_j(x)}{\det D^k}. \tag{52}$$

From (50), (51) and (52), we obtain for $i = 0, 1, \ldots, n$

$$z_j(y_i) = \delta_{ij} = \Lambda_j(y_i),$$

where $y_i \in P^k$. Comparing this with (52), by the uniqueness of the Lagrange polynomial, we have, for all $x \in \mathbb{R}^n$ and $j = 0, 1, \ldots, n$, that

$$\Lambda_j(x) = \frac{\det \Delta_j(x)}{\det D^k}.$$

By Lemma 3.15, there exists a constant $C_5 > 0$, such that $\left|\det D^k\right| \geq C_5\rho^n$ and consequently,

$$\sum_{j=1}^{n} |\Lambda_j(x)| = \sum_{j=1}^{n} \left|\frac{\det \Delta_j(x)}{\det D^k}\right| \leq \sum_{j=1}^{n} \frac{|\det \Delta_j(x)|}{C_5\rho^n}. \tag{53}$$

By the same argument used to prove Lemma 3.5, with y_j replaced by x, we have, for $j = 0, 1, \ldots, n$,

$$|\det \Delta_j(x)| \leq \|x - y_0\| \prod_{i=1, i\neq j}^{n} \|y_i - y_0\|.$$

Using this and Lemma 3.10, there exists a constant $C_3 > 1$ such that

$$|\det \Delta_j(x)| \leq \rho \prod_{i=1, i\neq j}^{n} C_3\rho = {C_3}^{n-1}\rho^n, \tag{54}$$

because $x^k = y_0$ and $\left\|x - x^k\right\| \leq \rho$ by hypothesis. Thus, by (53) and (54) we have that

$$\sum_{j=1}^{n} |\Lambda_j(x)| \leq \sum_{j=1}^{n} \frac{{C_3}^{n-1}\rho^n}{C_5\rho^n} = n\frac{{C_3}^{n-1}}{C_5}. \tag{55}$$

Since $\{\Lambda_j\}_{j=0,1,\ldots,n}$ is a base for the space $\mathcal{P}_n^1$ of the polynomials of degree less than or equal to 1 defined in $\mathbb{R}^n$, then any polynomial $\ell \in \mathcal{P}_n^1$ can be written as

$$\ell(x) = \sum_{j=0}^{n} \ell(y_j)\Lambda_j(x).$$

In particular, for the constant polynomial $\ell(x) \equiv 1$, we have

$$1 \;=\; \sum_{j=0}^{n} \Lambda_j(x) \;=\; \Lambda_0(x) + \sum_{j=1}^{n} \Lambda_j(x).$$

Thus, using this and (55) we have that

$$|\Lambda_0(x)| \;=\; \left|1 - \sum_{j=1}^{n} \Lambda_j(x)\right| \;\leq\; 1 + n\frac{{C_3}^{n-1}}{C_5}.$$

Therefore

$$\sum_{j=0}^{n} |\Lambda_j(x)| \;=\; |\Lambda_0(x)| + \sum_{j=1}^{n} |\Lambda_j(x)| \leq 1 + 2n\frac{{C_3}^{n-1}}{C_5}.$$

Taking $C_6 = 1 + 2n{C_3}^{n-1}/C_5 > 0$, we conclude the proof. □

Lemma 4.2 *[19, Lemma 1] Consider Q^k the model defined by (4) and ρ the trust-region radius at the iteration k. Then there exists a constant $C_7 > 0$ such that*

$$\left\|\nabla Q^k(x^k) - \nabla F(x^k)\right\| \leq C_7\rho.$$

Proof. For each $j \in \{1, 2, \ldots, n\}$, denote $d_j = y_j - y_0$. Since the functions F and Q^k are twice differentiable, by the Taylor theorem there exist t_1 and $t_2 \in (0, 1)$ such that

$$F(y_j) = F(y_0) + \nabla F(y_0)^T d_j + \frac{1}{2} d_j^T \nabla^2 F\left(y_0 + t_1 d_j\right) d_j$$

and

$$Q^k(y_j) = Q^k(y_0) + \nabla Q^k(y_0)^T d_j + \frac{1}{2} d_j^T \nabla^2 Q^k\left(y_0 + t_2 d_j\right) d_j.$$

Subtracting these expressions and using the interpolation conditions (5), we have

$$\left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T d_j = \frac{1}{2} d_j^T \left(\nabla^2 F\left(y_0 + t_1 d_j\right) - \nabla^2 Q^k\left(y_0 + t_2 d_j\right)\right) d_j.$$

Multiplying both sides by the Lagrange function $\Lambda_j(x)$, summing in j, using the Cauchy-Schwarz and triangular inequalities, it follows that

$$\begin{aligned}
&\left|\sum_{j=0}^{n} \left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T d_j \Lambda_j(x)\right| \\
&= \left|\sum_{j=0}^{n} \frac{1}{2} d_j^T \left(\nabla^2 F(y_0 + t_1 d_j) - \nabla^2 Q^k(y_0 + t_2 d_j)\right) d_j \Lambda_j(x)\right| \\
&\le \sum_{j=0}^{n} \frac{1}{2} \|d_j\|^2 \left(\left\|\nabla^2 F\left(y_0 + t_1 d_j\right)\right\| + \left\|\nabla^2 Q^k\left(y_0 + t_2 d_j\right)\right\|\right) |\Lambda_j(x)|.
\end{aligned}$$

Defining $K = \max\{M, N\}$ where M and N are given in Hypotheses 2 and 3, and using Lemmas 3.10 and 4.1, we have, for all $x \in \mathbb{R}^n$ with $\|x - y_0\| \le \rho$, that

$$\left|\sum_{j=0}^{n} \left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T d_j \Lambda_j(x)\right| \le {C_3}^2 \rho^2 K \sum_{j=0}^{n} |\Lambda_j(x)| \le {C_3}^2 \rho^2 K C_6. \tag{56}$$

On the other hand, consider the linear polynomial given by

$$\ell(x) = \left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T (x - y_0), \tag{57}$$

which can be written uniquely as a linear combination of the Lagrange polynomials $\Lambda_j(x)$, i.e.,

$$\ell(x) = \sum_{j=0}^{n} \ell(y_j) \Lambda_j(x) = \sum_{j=0}^{n} \left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T d_j\, \Lambda_j(x). \tag{58}$$

Thus, in particular for

$$\bar{x} = y_0 + \rho \frac{\nabla Q^k(y_0) - \nabla F(y_0)}{\|\nabla Q^k(y_0) - \nabla F(y_0)\|},$$

using (56)-(58) and the fact that $\|\bar{x} - y_0\| = \rho$, we have that

$$\ell(\bar{x}) = \rho \left\|\nabla Q^k(y_0) - \nabla F(y_0)\right\| = \sum_{j=0}^{n} \left(\nabla Q^k(y_0) - \nabla F(y_0)\right)^T d_j \Lambda_j(\bar{x}) \le {C_3}^2 \rho^2 K C_6.$$

Defining $C_7 = K {C_3}^2 C_6 > 0$, we complete the proof. □

The next lemma gives an upper bound to the decrease in the model obtained by the Cauchy step.

Lemma 4.3 *[19, (6.3)] Let $C_7 > 0$ be the constant given in Lemma 4.2. Consider Q^k the model defined by* (4). *If d_c^k is the Cauchy step, then*

$$Q^k(x^k + d_c^k) - Q^k(x^k) \le \frac{\rho}{2}\left(-\left\|g^k\right\| - \left\|\nabla F(x^k)\right\| + (C_7 + M)\,\rho\right).$$

Proof. Consider $\bar{d} = -\rho \frac{g^k}{\|g^k\|}$. By the definition of the Cauchy step,

$$Q^k(x^k + d_c^k) \leq Q^k(x^k + \bar{d}) = Q^k(x^k) - \rho \left\| g^k \right\| + \frac{\rho^2}{2\left\| g^k \right\|^2} {g^k}^T G^k g^k.$$

By the Cauchy-Schwarz inequality and the Hypothesis H3, we have that

$$Q^k(x^k + d_c^k) - Q^k(x^k) \leq -\frac{\rho}{2} \left\| g^k \right\| - \frac{\rho}{2} \left\| g^k \right\| + \frac{1}{2}\rho^2 M.$$

Using the triangle inequality and Lemma 4.2

$$-\left\| g^k \right\| \leq -\left\| \nabla F(x^k) \right\| + \left\| g^k - \nabla F(x^k) \right\| \leq -\left\| \nabla F(x^k) \right\| + C_7 \rho,$$

which completes the proof. □

The next result gives a bound to the difference between the function and the model values at iteration.

Lemma 4.4 *[19, (6.6)] Let $C_7 > 0$ be the constant given in Lemma 4.2. Then, for all k,*

$$\left| Q^k(x^k + d^k) - F(x^k + d^k) \right| \leq \left(C_7 + \frac{1}{2}(M + N) \right) \rho^2.$$

Proof. By the Taylor theorem, there exist $t_1, t_2 \in (0, 1)$ such that

$$Q^k(x^k + d^k) = Q^k(x^k) + \nabla Q^k(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 Q^k(x^k + t_1 d^k) d^k,$$
$$F(x^k + d^k) = F(x^k) + \nabla F(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 F(x^k + t_2 d^k) d^k.$$

Subtracting these expressions, using the interpolation condition (5), triangle inequality, Cauchy-Schwarz inequality and Hypothesis H2 and H3, we have

$$\left| Q^k(x^k + d^k) - F(x^k + d^k) \right| \leq \left\| \nabla Q^k k(x^k) - \nabla F(x^k) \right\| \left\| d^k \right\| + \frac{1}{2}(M + N) \left\| d^k \right\|^2.$$

By the fact that $\left\| d^k \right\| \leq \rho$ and Lemma 4.2, we complete the proof. □

Consider the constant $C_8 > 0$ defined by

$$C_8 \;=\; (3C_7 \;+\; 2M \;+\; 2N) \max\{1\;;\;\gamma\}, \tag{59}$$

where γ, C_7, N, M are the positive constants given in the algorithm, Lemma 4.2, Hypotheses H2 and H3, respectively.

The next lemma ensures that under certain hypotheses, a trust-region attempt is successful. Remember that it occurs when the trust-region step d^k satisfies conditions (9) and (10).

Lemma 4.5 *[19, (6.2)] Let $C_8 > 0$ be the constant given by* (59). *Consider k an iteration that performs a trust-region attempt with radius ρ. If*

$$\left\| \nabla F(x^k) \right\| > C_8 \rho, \tag{60}$$

then the trust-region attempt is successful.

Proof. By (7), Lemma 4.3, (60) and (59) we have

$$Q^k(x^k + d^k) \leq Q^k(x^k + d_c^k) < Q^k(x^k) + \frac{\rho}{2}\left(-\left\|g^k\right\| + (-C_8 + C_7 + M)\,\rho\right)$$
$$< Q^k(x^k) - \frac{\rho}{2}\left\|g^k\right\| - \frac{1}{8}M\rho^2. \tag{61}$$

Suppose by contradiction that $\left\|d^k\right\| < \dfrac{1}{2}\rho$. By the definition of the model Q^k, the Cauchy-Schwarz inequality and Hypothesis H3, it follows that

$$Q^k(x^k + d^k) > Q^k(x^k) - \frac{\rho}{2}\left\|g^k\right\| - \frac{1}{8}M\rho^2,$$

which contradicts (61). Thus, the second inequality of condition (9) holds.

Using the definition of η^k in (8) and Lemma 4.4, we conclude that

$$\eta^k \leq \left(C_7 + \frac{1}{2}(M + N)\right)\rho^2. \tag{62}$$

On the other hand, using the definition of d_c^k, the Cauchy-Schwarz inequality and the Hypothesis H3, we have that

$$Q^k(x^k + d^k) \leq Q^k(x^k + d_c^k) \leq Q^k\left(x^k - \rho\frac{g^k}{\|g^k\|}\right) \leq Q^k(x^k) - \rho\left\|g^k\right\| + \frac{1}{2}\rho^2 M. \tag{63}$$

Using triangle inequality, (60) and Lemma 4.2

$$\left\|g^k\right\| \geq \left\|\nabla F(x^k)\right\| - \left\|g^k - \nabla F(x^k)\right\| > (C_8 - C_7)\rho.$$

Replacing this in (63), we have

$$Q^k(x^k) - Q^k(x^k + d^k) > \left(C_8 - C_7 - \frac{M}{2}\right)\rho^2. \tag{64}$$

However, by (59),

$$C_8 - C_7 - \frac{1}{2}M > C_8 - (1+\gamma)C_7 - \frac{1}{2}(1+\gamma)M - \frac{1}{2}\gamma N > \gamma\left(C_7 + \frac{1}{2}(M + N)\right).$$

Using this in (64) and by (62), it follows that $Q^k(x^k) - Q^k(x^k + d^k) > \gamma\,\eta^k$, which proves the first inequality in (9).

Now we will prove that (10) holds. From Lemma 4.4, (59) and (64), we have that

$$F(x^k + d^k) - Q^k(x^k + d^k) < \frac{9}{10}\left(C_8 - C_7 - \frac{1}{2}M\right)\rho^2 < \frac{9}{10}\left(Q^k(x^k) - Q^k(x^k + d^k)\right).$$

Adding $Q^k(x^k + d^k) - Q^k(x^k)$ in both sides and using (5), we obtain

$$F(x^k) - F(x^k + d^k) > 0.1\left(Q^k(x^k) - Q^k(x^k + d^k)\right),$$

which proves (10) and completes the proof. □

Our task now is to prove that the number of iterations with the same radius is finite.

Lemma 4.6 *[19, Lemma 4] The number of iterations with each trust-region radius is finite.*

Proof. By (2), the sequence $\left(F(x^k)\right)$ is monotonically decreasing, then by H1

$$\lim_{k\to\infty}\left(F(x^k) - F(x^{k+1})\right) = 0. \tag{65}$$

Consider $\rho > 0$ a trust-region radius. Suppose by contradiction that the number of iterations with this radius is infinite. Consider $\mathcal{K}$ the set of the indices of the successful trust-region iterations with radius ρ. Note that, by the contradiction hypothesis and Lemma 3.2, the set $\mathcal{K}$ is infinite.

Assume initially that $\eta^{\widehat{a}} > 0$ for some iteration $\widehat{a}$ with this radius. Without loss of generality we can consider $\widehat{a}$ the first iteration with this property. Thus, by (8), $\eta^k \geq \eta^{\widehat{a}}$ for all $k \geq \widehat{a}$. Define $\widehat{c} = 0.1\gamma\eta^{\widehat{a}}$. Using (9) and (10), we have for all $k \in \mathcal{K}$ with $k \geq \widehat{a}$ that

$$F(x^k) - F(x^k + d^k) \geq \widehat{c}, \tag{66}$$

which contradicts (65).

Suppose now that η^k is null for all iterations with the radius ρ. Consider ν the first iteration with this radius and $\widehat{\nu} = \nu + 5$. Thus, for all $k \in \mathcal{K}$ with $k \geq \widehat{\nu}$, we have that the step d^k is an exact solution of (6) and $Q^k = Q^{\widehat{\nu}}$. Consequently,

$$g^k = \nabla Q^{\widehat{\nu}}(x^k) = g^{\widehat{\nu}} + G^k(x^k - x^{\widehat{\nu}}). \tag{67}$$

Let λ_ℓ be the smallest eigenvalue of the Hessian $G^{\widehat{\nu}}$ of the model $Q^{\widehat{\nu}}$. Now we consider three situations.

First suppose that $\lambda_\ell < 0$. Consider v_ℓ the eigenvector associated to λ_ℓ with $\|v_\ell\| = \rho$ and ${v_\ell}^T g^{\widehat{\nu}} \leq 0$. Since $\lambda_\ell < 0$, the model $Q^{\widehat{\nu}}$ is unbounded below. Consequently, for all $k \in \mathcal{K}$ with $k \geq \widehat{\nu}$, $\|d^k\| = \rho$ and

$$Q^k(x^k) - Q^k(x^k + d^k) \geq Q^k(x^k) - Q^k(x^k + v_\ell) = -{v_\ell}^T g^k - \frac{1}{2}\lambda_\ell \|v_\ell\|^2 \geq \frac{1}{2}|\lambda_\ell|\,\rho^2.$$

From this and (10), we obtain (66) with $\widehat{c} = \dfrac{1}{20}|\lambda_\ell|\,\rho^2$, for all $k \in \mathcal{K}$, $k \geq \widehat{\nu}$, which contradicts (65).

Now assume that $\lambda_\ell = 0$ and the model $Q^{\widehat{\nu}}$ is unbounded below. Then, by Lemma 5.1 of the Appendix there exists $\omega \in \mathcal{N}(G^{\widehat{\nu}})$ with $\|\omega\| = 1$ and $\omega^T g^{\widehat{\nu}} > 0$. Therefore, for all $k \in \mathcal{K}$ with $k \geq \widehat{\nu}$,

$$Q^k(x^k) - Q^k(x^k + d^k) \geq Q^k(x^k) - Q^k(x^k - \rho\omega) = \rho\omega^T g^k. \tag{68}$$

Since $Q^k = Q^{\widehat{\nu}}$, then (67) holds and $\omega^T g^k = \omega^T g^{\widehat{\nu}}$. Using this and (10) in (68), we obtain (66) with $\widehat{c} = 0.1\rho\omega^T g^{\widehat{\nu}}$, for all $k \in \mathcal{K}$, $k \geq \widehat{\nu}$, which contradicts (65).

For the third situation consider the case in which $\lambda_\ell \geq 0$ and the model is bounded below. Let $\lambda_{\widehat{\ell}} \geq \lambda_\ell$ be the smallest positive eigenvalue of $G^{\widehat{\nu}}$. Now we consider two cases. First, suppose that

$$\left\|g^k\right\| \geq \rho\lambda_{\widehat{\ell}} \tag{69}$$

for infinite indices $k \in \mathcal{K}$, $k \geq \widehat{\nu}$. For each one of these indices, take $d = -\dfrac{g^k}{\|g^k\|}$ and $\bar{t} = \dfrac{\rho\lambda_{\widehat{\ell}}}{M} \leq \rho$, where M is given by Hypothesis H3. Thus $\|\bar{t}d\| \leq \rho$ and

$$Q^k(x^k) - Q^k(x^k + d^k) \geq Q^k(x^k) - Q^k(x^k + \bar{t}d) = -\frac{1}{2}\bar{t}d^T g^k - \frac{1}{2}\bar{t}d^T\left(g^k + \bar{t}G^k d\right).$$

Using the definitions of d and $\bar{t}$, the triangle inequality and (69),

$$\bar{t}d^T\left(g^k + \bar{t}G^k d\right) \leq 0.$$

Using this, (10) and (69)

$$F(x^k) - F(x^k + d^k) \geq 0.1 \left(Q^k(x^k) - Q^k(x^k + d^k) \right) \geq -\frac{1}{20} \bar{t} d^T g^{\widehat{\nu}} \geq \frac{(\rho \lambda_{\widehat{\ell}})^2}{20M}.$$

Thus, we obtain (66) with $\widehat{c} = \frac{(\rho \lambda_{\widehat{\ell}})^2}{20M}$, for infinite indices $k \in \mathcal{K}$, $k \geq \widehat{\nu}$, which contradicts (65).

Finally, we assume that $\|g^k\| < \lambda_{\widehat{\ell}} \rho$ holds for all $k \in \mathcal{K}$, $k \geq \widehat{\nu}$, sufficiently large. By (67) and Lemma 5.2 of the Appendix, there exists $\bar{d}^k \in \mathbb{R}^n$ such that

$$\nabla Q^{\widehat{\nu}}(x^k + \bar{d}^k) = g^k + G^{\widehat{\nu}} \bar{d}^k = 0,$$

and $\|g^k\| = \|G^{\widehat{\nu}} \bar{d}^k\| \geq \lambda_{\widehat{\ell}} \|\bar{d}^k\|$, for each k sufficiently large. Consequently, $\|\bar{d}^k\| < \rho$. As d^k is an exact solution of the subproblem, then $d^k = \bar{d}^k$, $Q^{\widehat{\nu}}(x^k + d^k) \leq Q^{\widehat{\nu}}(x)$, for all $x \in \mathbb{R}^n$, and $x^{k+1} = x^k + d^k$. Since $Q^k = Q^{\widehat{\nu}}$ and $\eta^k = 0$ for all $k \geq \widehat{\nu}$, the first inequality in (9) does not hold for the next iterations with this radius, which contradicts the hypothesis that $\mathcal{K}$ is infinite, completing the proof. □

Now we present an useful result particularly relevant in the derivative-free context.

Corollary 4.7 *The sequence of the trust-region radius generated by the algorithm converges to zero.*

Proof. The result follows from the last lemma and the fact that the sequence of the trust-region radius is bounded below by zero and monotone nonincreasing. □

The next result establishes the weak convergence of the algorithm.

Theorem 4.8 *The sequence (x^k) generated by the algorithm has a stationary accumulation point.*

Proof. Consider $\mathbb{N}' \subset \mathbb{N}$ the set of the indices of the unsuccessful trust-region iterations. Let $C_8 > 0$ be the constant given by (59). By Lemma 4.5, for all $k \in \mathbb{N}'$,

$$\left\| \nabla F(x^k) \right\| \leq C_8 \rho.$$

Using the last corollary, we have that

$$\liminf_{k \to \infty} \left\| \nabla F(x^k) \right\| = 0$$

which concludes the proof. □

4.2 Strong convergence.

In this subsection we prove the global convergence of the algorithm.

For the next results, given $\varepsilon > 0$, consider k_ε the first iteration such that the current trust-region radius ρ satisfies

$$\rho \leq \frac{\varepsilon}{C_8}, \tag{70}$$

where $C_8 > 0$ is the constant defined by (59). The Corollary 4.7 guarantees the existence of such index. Note that for each iteration $k \geq k_\varepsilon$, the current radius also satisfies (70).

Lemma 4.9 *[19, (7.5)] Given $\varepsilon > 0$. Then, for all $k \geq k_\varepsilon$,*

$$\left\| \nabla F(x^{k+1}) - \nabla F(x^k) \right\| < \varepsilon.$$

Proof. Using the Mean Value Theorem, the Hypothesis H2, (70) and the definition of the constant C_8, we have for all $k \geq k_\varepsilon$ that

$$\left\| \nabla F(x^{k+1}) - \nabla F(x^k) \right\| \leq N\rho \leq \frac{N}{C_8}\varepsilon < \varepsilon.$$

□

Lemma 4.10 *Given $\varepsilon > 0$, consider $k \geq k_\varepsilon$ an iteration that performs a trust-region attempt with radius $\rho > 0$. If*

$$\left\| \nabla F(x^k) \right\| > \varepsilon, \tag{71}$$

then k is a successful iteration and

$$F(x^k) - F(x^k + d^k) > \frac{1}{15}\rho\varepsilon.$$

Proof. Consider $k \geq k_\varepsilon$ an iteration satisfying the hypotheses. Using (70) and (71),

$$\left\| \nabla F(x^k) \right\| > C_8\rho.$$

From Lemma 4.5, k is a successful trust-region iteration. Using (7), the Cauchy-Schwarz inequality and the Hypothesis H3 we have that

$$Q^k(x^k + d^k) \leq Q^k(x^k + d_c^k) \leq Q^k\left(x^k - \rho\frac{g^k}{\|g^k\|}\right) \leq Q^k(x^k) - \rho\left\|g^k\right\| + \frac{1}{2}\rho^2 M. \tag{72}$$

On the other hand, using the triangle inequality, (71) and Lemma 4.2,

$$\left\|g^k\right\| \geq \left\| \nabla F(x^k) \right\| - \left\| g^k - \nabla F(x^k) \right\| > \varepsilon - C_7\rho.$$

Applying this and (70) in (72) and using the fact that $C_8 > 3(C_7 + M/2)$, we have that

$$Q^k(x^k) - Q^k(x^k + d^k) > \left[1 - \frac{1}{C_8}\left(C_7 + \frac{1}{2}M\right)\right]\rho\varepsilon > \frac{2}{3}\rho\varepsilon.$$

Combining this with the condition (10), we complete the proof. □

Theorem 4.11 *[19, Theorem 1] All accumulation point of the sequence (x_k) is stationary, i.e.,*

$$\lim_{k\to\infty} \nabla F(x^k) = 0.$$

Proof. Suppose by contradiction that there exists $\varepsilon > 0$ such that the set

$$\mathcal{K} = \left\{ k \in \mathbb{N} \mid \|\nabla F(x^k)\| > 4\varepsilon \right\}$$

is infinite. Given $k \in \mathcal{K}$, $k \geq k_\varepsilon + 2$, consider $\ell_k > k$ the first index such that $\left\|\nabla F(x^{\ell_k+1})\right\| \leq \varepsilon$. The existence of ℓ_k is guaranteed by Theorem 4.8. Using triangle inequality and Lemma 4.9,

$$\left\|\nabla F(x^{k-1})\right\| \geq \left\|\nabla F(x^k)\right\| - \left\|\nabla F(x^k) - \nabla F(x^{k-1})\right\| > 3\varepsilon.$$

Analogously, $\left\|\nabla F(x^{k-2})\right\| > 2\varepsilon$. Therefore, from last lemma, all the iterations of indices in the interval $[\max\{\nu;\ k-2\}, \ell_k]$ that perform a trust-region attempt are successful trust-region iterations. Consequently, all iterations to k from ℓ_k inclusive use the same trust-region radius.

By the definition of k and ℓ_k, the triangle inequality and Lemma 4.9,

$$3\varepsilon < \left\|\nabla F(x^{\ell_k+1}) - \nabla F(x^k)\right\| = \left\| \sum_{i=1}^{\ell_k+1-k} \left(\nabla F(x^{k+i}) - \nabla F(x^{k+i-1})\right) \right\| < \mu\varepsilon, \tag{73}$$

where $\mu = \ell_k + 1 - k$. On the other hand, using the Mean Value Theorem, the Hypothesis H2 and the triangle inequality, we have

$$\left\|\nabla F(x^{\ell_k+1}) - \nabla F(x^k)\right\| \leq N \left\|x^{\ell_k+1} - x^k\right\| \leq N\mu\rho.$$

From this and (73),

$$\mu > \frac{3\varepsilon}{N\rho} \qquad \text{and} \qquad \mu > 3. \tag{74}$$

Now, consider ξ the cardinality of the set U of the trust-region iterations performed between k and ℓ_k. By the mechanism of the algorithm, every three consecutive iterations at least one of them performs a trust-region attempt. Moreover, all trust-region attempts performed between k and ℓ_k are successful. Using this, the fact that k and ℓ_k are not necessarily trust-region iterations and (74), we have

$$\xi \geq \frac{\mu - 2}{3} = \frac{\mu}{3}\left(1 - \frac{2}{\mu}\right) > \frac{\varepsilon}{3N\rho}. \tag{75}$$

Therefore, using the fact that $(F(x^k))$ is nonincreasing, the Lemma 4.10 and the definition of ξ and (75), we obtain that

$$F(x^k) - F(x^{\ell_k+1}) \geq \sum_{i \in U} F(x^i) - F(x^{i+1}) > \frac{\xi\rho\varepsilon}{15} > \frac{\varepsilon^2}{45N} > 0, \tag{76}$$

for all $k \in \mathcal{K}$, $k \geq k_\varepsilon + 2$. On the other hand, the sequence $(F(x^k))$ is nonincreasing and, by Hypothesis H1, unbounded below, hence it follows that $F(x^k) - F(x^{\ell_k+1}) \to 0$, contradicting (76). Therefore, the statement of the theorem is true. $\square$

5 Conclusion

We have presented an explicit algorithm for solve unconstrained derivative-free optimization based on [19]. The algorithm uses trust-region techniques with the models constructed by polynomial interpolation. The number of the interpolation points is fixed along the iterations and the objective function is evaluated at just one point in each iteration. Some properties of the algorithm are discussed. Moreover we prove in details that the algorithm is globally convergent. A promising direction of research is to extend the method for constrained problems by keeping the theoretical results.

Appendix

Consider the quadratic function

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

with $A \in \mathbb{R}^{n\times n}$ symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Lemma 5.1 *If A is positive semidefinite and f is unbounded below, then b does not belong to the subspace generated by the eigenvectors associated with the positive eigenvalues of A. Equivalently, there is a vector $\omega \in \mathcal{N}(\mathcal{A})$ such that $\|\omega\| = 1$ and $b^T\omega > 0$.*

Proof. Let $\{v_1, \dots, v_n\}$ be an orthonormal basis of eigenvectors of A such that $v_1, \dots, v_{\ell-1}$ are the eigenvectors corresponding to the null eigenvalues and $v_\ell, \dots, v_n$ the eigenvectors corresponding to the positive eigenvalues $\lambda_\ell, \dots, \lambda_n$. Then, for $i = 1, \dots, \ell-1$ and $j = \ell, \dots, n$, we have

$$Av_i = 0 \quad \text{and} \quad v_j = A\left(\frac{1}{\lambda_j}v_j\right).$$

Thus, $[v_1, \dots, v_{\ell-1}] = \mathcal{N}(\mathcal{A})$ and $[v_\ell, \dots, v_n] = \mathcal{R}(\mathcal{A})$. Since $Ax + b = \nabla f(x) \neq 0$, for all $x \in \mathbb{R}^n$, then $b \notin \mathcal{R}(\mathcal{A})$. Equivalently, $b \notin \mathcal{N}(\mathcal{A})^\perp$, which means that there is a vector $\omega \in \mathcal{N}(\mathcal{A})$ such that $b^T\omega \neq 0$. Normalizing and taking the opposite, if necessary, the proof is concluded. □

Lemma 5.2 *If f is bounded below, then there is a unique $x^* \in \mathcal{R}(A)$ such that $Ax^* + b = 0$. Moreover, if λ_ℓ is the smallest positive eigenvalue of A, then $\|Ax^*\| \geq \lambda_\ell \|x^*\|$.*

Proof. Since $\mathcal{R}(A^2) \subset \mathcal{R}(A)$ and $\dim(\mathcal{R}(A^2)) = \dim(\mathcal{R}(A^TA)) = \dim(\mathcal{R}(A))$, we have that $\mathcal{R}(A^2) = \mathcal{R}(A)$. As f is bounded below, $b \in \mathcal{R}(A) = \mathcal{R}(A^2)$. Thus, there is $u \in \mathbb{R}^n$ such that $A^2u = b$. This means that $A(-Au) + b = 0$, i.e., $x^* = -Au \in \mathcal{R}(A)$ and $Ax^* + b = 0$. To prove the uniqueness, note that if $x^*, \bar{x} \in \mathcal{R}(A)$ are such that $Ax^* + b = 0$ and $A\bar{x} + b = 0$, then $x^* - \bar{x} \in \mathcal{R}(A) = \mathcal{R}(A^T)$ and $A(x^* - \bar{x}) = 0$. But this means that $x^* - \bar{x} = 0$.

To establish the inequality, first consider the case where A is positive definite. Thus,

$$\|Ax^*\|^2 = (x^*)^T A^2 x^* \geq \lambda_\ell^2 \|x^*\|^2.$$

In the case where A has null eigenvalues, consider $\{v_1, \dots, v_n\}$ a orthonormal basis of eigenvectors such that $v_1, \dots, v_{\ell-1}$ are the eigenvectors corresponding to the null eigenvalues and $v_\ell, \dots, v_n$ the eigenvectors corresponding to the positive eigenvalues. Defining $P = (v_1 \dots v_n)$, $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, where λ_i is the eigenvalue associated to the eigenvector v_i, and $\widehat{c} = P^T b$, we have that if

$$z^* \in \mathcal{R}(D) \quad \text{and} \quad Dz^* + \widehat{c} = 0, \tag{77}$$

then $x^* = Pz^* \in \mathcal{R}(A)$ and $Ax^* + b = 0$. Indeed,

$$x^* = Pz^* = PDw^* = APw^* \in \mathcal{R}(A)$$

and

$$Ax^* + b = P(Dz^* + \widehat{c}) = 0.$$

Now we found z^* satisfying (77). Define $w^* \in \mathbb{R}^n$ by

$$w_i^* = \begin{cases} 0, & \text{if } i = 1, \dots, \ell-1 \\ \dfrac{-\widehat{c}_i}{\lambda_i^2}, & \text{if } i = \ell, \dots, n \end{cases}$$

and $z^* = Dw^*$. Since $b \in \mathcal{N}(A)^\perp = [v_1, \ldots, v_{\ell-1}]^\perp$, $\widehat{c}_i = v_i^T b = 0$, for $i = 1, \ldots, \ell - 1$ and, consequently, $Dz^* + \widehat{c} = 0$. To complete the proof, note that

$$\|z^*\|^2 = \sum_{i=\ell}^{n} \left(\frac{\widehat{c}_i}{\lambda_i}\right)^2 \leq \frac{1}{\lambda_\ell^2} \sum_{i=\ell}^{n} \widehat{c}_i^2 = \frac{1}{\lambda_\ell^2} \|\widehat{c}\|^2.$$

Moreover, since $x^* = Pz^*$, $Ax^* + b = 0$ and $\widehat{c} = P^T b$, we obtain that $\|x^*\| = \|z^*\|$ and $\|Ax^*\| = \|b\| = \|\widehat{c}\|$. Therefore,

$$\|x^*\|^2 \leq \frac{1}{\lambda_\ell^2} \|b\|^2 = \frac{1}{\lambda_\ell^2} \|Ax^*\|^2.$$

$\square$

References

[1] Arouxét, M.B., Echebest, N., Pilotta, E.A.: Active-set strategy in Powell's method for optimization without derivatives. Computational and Applied Mathematics **30**(1), 171–196 (2011)

[2] Bazaraa, M., Sherali, H., Shetty, C.: Nonlinear Programming Theory and Algorithms, 2 edn. Athena Scientific, Wiley, New York (2006)

[3] Bueno, L.F., Friedlander, A., Martínez, J.M., Sobral, F.N.C.: Inexact restoration method for derivative-free optimization with smooth constraints. SIAM Journal on Optimization **23**(2), 1189–1213 (2013)

[4] Conejo, P.D., Karas, E.W., Ribeiro, A.A., Pedroso, L.G., Sachine, M.: Global convergence of trust-region algorithms for convex constrained minimization without derivatives. Applied Mathematics and Computation **220**, 324–330 (September 2013)

[5] Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-Region Methods. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2000)

[6] Conn, A.R., Gould, N.I.M., Vicente, L.N.: Global convergence of general derivative-free trust-region algorithms to first and second order critical points. SIAM Journal on Numerical Analysis **20**, 387–415 (2009)

[7] Conn, A.R., Scheinberg, K., Toint, P.L.: On the convergence of derivative-free methods for unconstrained optimization. In A. Iserles and M. Buhmann, editors, Approximation Theory and Optimization: Tributes to M.J.D. Powell pp. 83–108 (1997)

[8] Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2009)

[9] Conn, A.R., Toint, P.L.: An Algorithm using Quadratic Interpolation for Unconstrained Derivative Free Optimization. In: G.D. Pillo, F. Gianessi (eds.) Nonlinear Optimization and Applications, pp. 27–47. Plenum (1996)

[10] Diniz-Ehrhardt, M.A., Martínez, J.M., Pedroso, L.G.: Derivative-free methods for nonlinear programming with general lower-level constraints. Computational and Applied Mathematics **30**, 19–52 (2011)

[11] Gumma, E.A.E., Hashim, M.H.A., Montaz Ali, M.: A derivative-free algorithm for linearly constrained optimization problems. Computational Optimization and Applications (2013). Doi: 10.1007/s10589-013-9607-y

[12] Hare, W.L., Lucet, Y.: Derivative-free optimization via proximal point methods. Journal of Optimization Theory and Applications (2013). Doi: 10.1007/s10957-013-0354-0

[13] Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer-Verlag (1999)

[14] Powell, M.J.D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal **7**, 155–162 (1964)

[15] Powell, M.J.D.: UOBYQA: unconstrained optimization by quadratic approximation. Math. Programming **92**, 555–582 (2002)

[16] Powell, M.J.D.: The NEWUOA software for unconstrained optimization without derivatives. In: G.D. Pillo, M. Roma (eds.) Large-Scale Nonlinear Optimization, pp. 255–297. Springer, New York (2006)

[17] Powell, M.J.D.: Developments of NEWUOA for minimization without derivatives. IMA Journal Numerical Analysis **28**, 649–664 (2008)

[18] Powell, M.J.D.: The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives. Tech. Rep. DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, Cambridge, England (2009)

[19] Powell, M.J.D.: On the convergence of trust region algorithms for unconstrained minimization without derivatives. Computational Optimization and Applications **53**, 527–555 (2012)

[20] Rios, L.M., Sahinidis, N.V.: Derivative-free optimization: a review of algorithms and comparison of software implementations. Journal of Global Optimization **56**, 1247–1293 (2013)