Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 11–11

# A SURROGATE MANAGEMENT FRAMEWORK USING RIGOROUS TRUST-REGIONS STEPS

S. GRATTON AND L. N. VICENTE

Abstract: Surrogate models and heuristics are frequently used in the optimization engineering community as convenient approaches to deal with functions for which evaluations are expensive or noisy, or lack convexity. These methodologies do not typically guarantee any type of convergence under reasonable assumptions and frequently render slow convergence.

In this paper we will show how to incorporate the use of surrogate models, heuristics, or any other process of attempting a function value decrease in trust-region algorithms for unconstrained derivative-free optimization, in a way that global convergence of the latter algorithms to stationary points is retained. Our approach follows the lines of search/poll direct-search methods and corresponding surrogate management frameworks, both in algorithmic design and in the form of organizing the convergence theory.

Keywords: Surrogate modeling, trust-region methods, search step, global convergence.
AMS Subject Classification (2010): 90C30, 90C56.

## 1. Introduction

Booker et al. [2] introduced in 1998 an algorithmic framework to incorporate the use of surrogates in direct-search methods. Since then this approach has been popular among optimizers and practitioners (see [1, 2, 6, 8, 9, 11, 12, 13]).

Part of the success of this approach relies on its simplicity. The iterations of direct-search methods (of directional type) have been organized in [2] around two major steps, a search step and a poll step. The search step is optional and not responsible for the main convergence properties of the overall direct-search method. It is required to evaluate the objective function at a finite number of points and the criterion to declare its success is simple. In fact, if global convergence of the direct-search method is ensured by using integer lattices and simple decrease, the search step is successful if it generates a point in the underlying mesh for which the objective function value is lower than

*Date*: April 15, 2011.
Support for this research was provided by FCT under the grant PTDC/MAT/098214/2008. Part of this work was developed while this author was visiting CERFACS.

the one at the current iterate [2]. If, on the other hand, global convergence is guaranteed by imposing a sufficient decrease condition based on a forcing function, all it is required from the search step is then to yield a sufficient decrease [7]. When the search step in unsuccessful, the method reverts to the poll step which can be viewed as a rigorous step, i.e., a step which must ensure some form of decrease for small step sizes at non-stationary points.

The purpose of this paper is to introduce a similar framework but when the rigorous steps are the trust-region ones, thus replacing the use of direct search by trust-region methods in the surrogate management framework. Given the type of scheme that ensures global convergence for trust regions, where no underlying mesh is available, the search step must be based on some form of sufficient decrease. As in the search/poll direct-search methods, the method reverts to the rigorous step (now a trust-region one) if the search step in not successful.

Another contribution of this paper is to rewrite the convergence of the overall trust-region method as a direct search one, by showing first that there is a subsequence of non successful iterates where the step size (in our case the trust-region radius) tends to zero. Such iterates correspond to non successful rigorous trust-region steps, where the size of the true gradient is of the order of the trust-region radius. Convergence of a subsequence of iterates to a stationary point can then be easily guaranteed by taking the limit when the trust-region radius goes to zero. We also study under what conditions can one establish that all limit points are stationary. It is important to note that such a rewriting of the convergence theory of trust-region methods is not allowed in derivative-based methods, where, in fact, it is possible to show under appropriate conditions that the trust-region radius is bounded away from zero. In trust-region methods for derivative-free optimization (DFO), the presence of a criticality step (taken when the model gradient is sufficiently small, and where the models are improved in a ball of appropriate radius) is essential to drive the trust-region radius to zero.

After this introduction the paper continues in Section 2 with a description of the type of surrogate management framework for trust-region methods that fits the above requirements. In Section 3 we show that such a framework enjoys global convergence to first-order stationary points. The paper is ended in Section 4 with some concluding remarks.

## 2. Surrogate management framework

We start by describing, at an abstract level, the surrogate management framework for incorporating a search step and a trust-region method.

**Algorithm 2.1. Surrogate Management Framework for TRM**

**Initialization:** Choose an initial point $x_0$ and an initial trust-region radius $\Delta_0 > 0$. Initialize all sample sets, models, constants, and tolerances for both the Search Step and the Rigorous TR Step. Set $k = 0$.

**Search Step:** Try to compute a point $x$ with $f(x) \leq f(x_k) - \rho(\Delta_k)$ by evaluating the function $f$ (at a finite number of points).
If such a point is found, then set $x_{k+1} = x$, declare the iteration and the Search Step successful, maintain or increase the trust-region radius ($\Delta_{k+1} \geq \Delta_k$), increment $k$ by one, and skip the Rigorous TR Step.

**Rigorous Trust-Region Step:** Apply a step of a trust-region method (including setting the trust-region radius $\Delta_{k+1}$), increment $k$ by one, and return to the Search Step.

Now we choose the derivative-free trust-region method from [4, Section 4] (see also [5, Section 10.3]) to concretize an example of the above surrogate management framework. A rigorous definition of fully linear model will be given later. For the moment, one can think of a quadratic model with similar accuracy properties of a first-order expansion Taylor model.

**Algorithm 2.2. Surrogate Management Framework for TRM (a concrete example)**

**Initialization:** Choose an initial point $x_0$ and an initial trust-region radius $\Delta_0 \in (0, \Delta_{max}]$ for some $\Delta_{max} > 0$. Initialize all sample sets, models, constants, and tolerances for the Search Step.

**For the TR step:** Choose an initial model $m_0(x_0 + s)$. The constants $\eta_0$, $\eta_1$, $\gamma$, $\gamma_{inc}$, $\epsilon_c$, $\mu$, and $\beta$ should also be chosen such that $0 \leq \eta_0 \leq \eta_1 < 1$ (with $\eta_1 \neq 0$), $0 < \gamma < 1 < \gamma_{inc}$, $\epsilon_c > 0$, and $\mu > \beta > 0$. Set $k = 0$.

**Search Step:** Try to compute a point $x$ with $f(x) \leq f(x_k) - \rho(\Delta_k)$ by evaluating the function $f$ (at a finite number of points).
If such a point is found, then set $x_{k+1} = x$, declare the iteration and the Search Step successful, maintain or increase the trust-region radius

($\Delta_{k+1} \in [\Delta_k, \min\{\gamma_{inc}\Delta_k, \Delta_{max}\}]$), increment $k$ by one, and skip the Rigorous TR Step.

**TR Step 1 (criticality step):** Apply some criticality step when $\|g_k\| \leq \epsilon_c$ yielding a new model $m_k(x_k+s)$ (i.e., a new gradient model $g_k$ and a new Hessian model $H_k$) and a new trust-region radius $\Delta_k$ such that $\Delta_k \leq \mu\|g_k\|$ and $m_k$ is fully linear on $B(x_k;\Delta_k)$, and such that, if $\Delta_k$ is reduced, one has $\beta\|g_k\| \leq \Delta_k$.

**TR Step 2 (step calculation):** Compute a step $s_k$ that sufficiently reduces the model $m_k$, in the sense of

$$m_k(x_k) - m_k(x_k+s_k) \;\geq\; \eta_1 \frac{\kappa_{fcd}}{2}\|g_k\| \min\left\{\frac{\|g_k\|}{\|H_k\|}, \Delta_k\right\} \qquad (1)$$

(with $\kappa_{fcd} \in (0,1]$), and such that $x_k + s_k \in B(x_k;\Delta_k)$.

**TR Step 3 (acceptance of the trial point):** Compute $f(x_k+s_k)$ and define

$$\rho_k \;=\; \frac{f(x_k) - f(x_k+s_k)}{m_k(x_k) - m_k(x_k+s_k)}.$$

If $\rho_k \geq \eta_1$ or if both $\rho_k \geq \eta_0$ and the model is fully linear on $B(x_k;\Delta_k)$, then $x_{k+1} = x_k+s_k$ and the model is updated to take into consideration the new iterate, resulting in a new model $m_{k+1}(x_{k+1}+s)$; otherwise the model and the iterate remain unchanged ($m_{k+1} = m_k$ and $x_{k+1} = x_k$).

**TR Step 4 (model improvement):** If $\rho_k < \eta_1$ use a model-improvement algorithm to attempt to certify that $m_k$ is fully linear on $B(x_k;\Delta_k)$ (if such a certificate is not obtained, one makes one or more suitable improvement steps). Define $m_{k+1}(x_k+s)$ to be the (possibly improved) model.

**TR Step 5 (trust-region radius update):** Set

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \min\{\gamma_{inc}\Delta_k, \Delta_{max}\}] & \text{if } \rho_k \geq \eta_1, \\ \{\gamma\Delta_k\} & \text{if } \rho_k < \eta_1 \text{ and } m_k \text{ is fully linear}, \\ \{\Delta_k\} & \text{if } \rho_k < \eta_1 \text{ and } m_k \\ & \text{is not certifiably fully linear.} \end{cases}$$

Increment $k$ by one and go to the Search Step.

The search step is either successful (and those iterations will be labeled by indices in $\mathcal{S}_{search}$) or not (in which case a rigorous TR step is executed). Note that the rigorous TR step of Algorithm 2.2 (composed by TR Steps 1–5) gives rise to four types of trust-region iterations:



[13] T. Zhang, K. K. Choi, S. Rahman, K. Cho, P. Baker, M. Shakil, and D. Heitkamp. A hybrid surrogate and pattern search optimization method and application to microelectronics. *Structural and Multidisciplinary Optimization*, 32:327–345, 2006.

S. Gratton
ENSEEIHT, INPT, rue Charles Camichel, B.P. 7122 31071, Toulouse Cedex 7, France (serge.gratton@enseeiht.fr).

L. N. Vicente
CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt).

used in the search step could certainly be less locally based. The specifics of such a sample set communication are application dependent and out of the scope of this paper.

We have chosen as a rigorous trust-region method the one from Conn, Scheinberg, and Vicente [4] due to its high level of abstraction and applicability, but our choice could have also contemplated the more recent self-correcting geometry method of Scheinberg and Toint [10], which dispenses with the model-improving iterations by judiciously updating the sample set with the incoming solution of the trust-region subproblem. It is also important to remark that such a form of surrogate management framework using rigorous trust-regions steps is not at all restricted to optimization without derivatives. In fact, the principle of a search or oracle step can also be applied to most derivative-based trust-region methods described in [3].

## References


[1] C. Audet and D. Orban. Finding optimal algorithmic parameters using derivative-free optimization. *SIAM J. Optim.*, 17:642–664, 2006.

[2] A. J. Booker, J. E. Dennis Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization*, 17:1–13, 1998.

[3] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[4] A. R. Conn, K. Scheinberg, and L. N. Vicente. Global convergence of general derivative-free trust-region algorithms to first and second order critical points. *SIAM J. Optim.*, 20:387–415, 2009.

[5] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[6] A. L. Custódio, H. Rocha, and L. N. Vicente. Incorporating minimum Frobenius norm models in direct search. *Comput. Optim. Appl.*, 46:265–278, 2010.

[7] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45:385–482, 2003.

[8] A. L. Marsden, M. Wang, J. E. Dennis Jr., and P. Moin. Optimal aeroacoustic shape design using the surrogate management framework. *Optim. Eng.*, 5:235–262, 2004.

[9] G. Nicosia and G. Stracquadanio. Generalized pattern search algorithm for peptide structure. *Biophysical Journal*, 95:4988–4999, 2008.

[10] K. Scheinberg and Ph. L. Toint. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. *SIAM J. Optim.*, 20:3512–3532, 2010.

[11] V. Torczon and M. W. Trosset. Using approximations to accelerate engineering design optimization. In *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, Missouri, September 2-4*, 1998.

[12] A. I. F. Vaz and L. N. Vicente. A particle swarm pattern search method for bound constrained global optimization. *J. Global Optim.*, 39:197–219, 2007.




(1) **Successful iterations** (indices in $\mathcal{S}_{tr}$), when $\rho_k \geq \eta_1$ (the new iterate is accepted and the trust-region radius is retained or increased).
(2) **Acceptable iterations**, when $\eta_1 > \rho_k \geq \eta_0$ and $m_k$ is fully linear (new iterate is accepted and the trust-region radius is decreased). Note that there are no acceptable iterations when $\eta_0 = \eta_1 \in (0, 1)$.
(3) **Model-improving**, when $\eta_1 > \rho_k$ and $m_k$ is not certifiably fully linear (the model is improved and the new point might be included in the sample set but is not accepted as a new iterate).
(4) **Unsuccessful iterations**, when $\rho_k < \eta_0$ and $m_k$ is fully linear (the trust-region radius is reduced and nothing else changes). Note that this is the case when no (acceptable) decrease was obtained and there is no need to improve the model.

The successful iterations of the overall algorithmic framework will be those corresponding to either successful search steps or successful rigorous TR steps:

$$\mathcal{S} \;=\; \mathcal{S}_{search} \cup \mathcal{S}_{tr}.$$

It is also important to note that unsuccessful iterations can only occur in the rigorous TR step.

## 3. Convergence to first-order stationarity

As is mentioned in [5, Chapter 10], it might be possible (especially at the early iterations) that the function $f$ is evaluated outside $L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ when considering sampling techniques used for modeling. If we assume that sampling is restricted to sets of the form $B(x_k; \Delta_k)$ and that $\Delta_k$ never exceeds the given positive constant $\Delta_{max}$, then the enlarged region where $f$ is sampled can be rigorously described as

$$L_{enl}(x_0) \;=\; L(x_0) \cup \bigcup_{x \in L(x_0)} B(x; \Delta_{max}) \;=\; \bigcup_{x \in L(x_0)} B(x; \Delta_{max}).$$

The derivation of convergence results for trust-region methods typically requires some form of continuous differentiability of the objective function. In the DFO context, one requires Lipschitz continuity of the gradient to be able to work with models which are fully linear.

**Assumption 3.1.** *Suppose $x_0$ and $\Delta_{max}$ are given. Assume that $f$ is continuously differentiable with Lipschitz continuous gradient in an open domain containing the set $L_{enl}(x_0)$.*

The following definition of fully linear models is taken verbatim from [4, Definition 3.1] (see also [5, Definition 10.3]).

**Definition 3.1.** *Let a function $f : \mathbb{R}^n \to \mathbb{R}$, that satisfies Assumption 3.1, be given. A set of model functions $\mathcal{M} = \{m : \mathbb{R}^n \to \mathbb{R}, m \in C^1\}$ is called a fully linear class of models if:*

*(1) There exist positive constants $\kappa_{ef}$, $\kappa_{eg}$, and $\nu_1^m$ such that for any $x \in L(x_0)$ and $\Delta \in (0, \Delta_{max}]$ there exists a model function $m(x+s)$ in $\mathcal{M}$, with Lipschitz continuous gradient and corresponding Lipschitz constant bounded by $\nu_1^m$, and such that*

- *the error between the gradient of the model and the gradient of the function satisfies*

$$\|\nabla f(x+s) - \nabla m(x+s)\| \le \kappa_{eg}\,\Delta, \quad \forall s \in B(0;\Delta), \tag{2}$$

*and*

- *the error between the model and the function satisfies*

$$|f(x+s) - m(x+s)| \le \kappa_{ef}\,\Delta^2, \quad \forall s \in B(0;\Delta). \tag{3}$$

*Such a model $m$ is called fully linear on $B(x;\Delta)$.*

*(2) For this class $\mathcal{M}$ there exists an algorithm, which we will call a 'model-improvement' algorithm, that in a finite, uniformly bounded (with respect to $x$ and $\Delta$) number of steps can*

- *either establish that a given model $m \in \mathcal{M}$ is fully linear on $B(x;\Delta)$ (we will say that a certificate has been provided),*
- *or find a model $\tilde{m} \in \mathcal{M}$ that is fully linear on $B(x;\Delta)$.*

As in the convergence of most trust-region methods, we need to assume that the objective function is bounded from below and the model Hessians are uniformly bounded.

**Assumption 3.2.** *Assume $f$ is bounded below on $L(x_0)$, that is there exists a constant $\kappa_*$ such that, for all $x \in L(x_0)$, $f(x) \ge \kappa_*$.*

**Assumption 3.3.** *There exists a constant $\kappa_{bhm} > 0$ such that, for all $x_k$ generated by the algorithm in the rigorous TR steps,*

$$\|H_k\| \le \kappa_{bhm}.$$

The first piece of the convergence theory concerns only the rigorous TR step, and is a restatement of [4, Lemma 5.2] (see also [5, Lemma 10.6]).



all $i$ sufficiently large,

$$\|x_{k_i} - x_{\ell_i}\| \le \sum_{\substack{j=k_i \\ j\in\mathcal{K}\cap\mathcal{S}}}^{\ell_i-1} \|x_j - x_{j+1}\| \le \sum_{\substack{j=k_i \\ j\in\mathcal{K}\cap\mathcal{S}}}^{\ell_i-1} \Delta_j \le C_2[f(x_{k_i}) - f(x_{\ell_i})].$$

Since the sequence $\{f(x_k)\}$ is bounded below (Assumption 3.2) and monotonic decreasing, we see that the right-hand side of this inequality must converge to zero, and we therefore obtain that $\lim_{i\to+\infty} \|x_{k_i} - x_{\ell_i}\| = 0$.

Finally,

$$\|\nabla f(x_{k_i})\| \le \|\nabla f(x_{k_i}) - \nabla f(x_{\ell_i})\| + \|\nabla f(x_{\ell_i}) - g_{\ell_i}\| + \|g_{\ell_i}\|.$$

The first term of the right-hand side tends to zero because of the Lipschitz continuity of the gradient of $f$ (Assumption 3.1), and is thus bounded by $\epsilon$ for $i$ sufficiently large. The explanation for the second term is twofold. For a TR step, we use the fact that from (9) and the mechanism of the criticality step (TR Step 1) at iteration $\ell_i$, the model $m_{\ell_i}$ is fully linear on $B(x_{\ell_i}; \mu\|g_{\ell_i}\|)$. So, using fully linearity and (10), we deduce for this step that the second term is bounded by $\kappa_\sigma\mu\epsilon$ (for $i$ sufficiently large). In the search step, this term is also bounded by $\kappa_{eg}\mu\epsilon$ for $i$ sufficiently large since the models are always fully linear and the trust-region radius converges to zero. The third term is bounded by $\epsilon$ by (10). As a consequence, we obtain from these bounds and (9) that

$$\|\nabla f(x_{k_i})\| \le (2 + \kappa_{eg}\mu)\epsilon \le \frac{1}{2}\epsilon_0$$

for $i$ large enough, which contradicts (8). Hence our initial assumption must be false and the theorem follows. ∎

## 4. Concluding remarks

Surrogate models can be used and managed in a variety of forms in the search step of the framework described in this paper, in particular using any of the ideas in Booker et al. [2] or in the review [5, Section 12.2]. Given a type of sample-based surrogate models chosen for the search step, it will then be of particular interest to consider the communication between this step and the TR rigorous one. In fact, not only could the rigorous TR step benefit from any new function evaluations made in the search step (as long as they correspond to points not too far from the current trust region), but the same could happen the other way round, in particular since the models

for some $\epsilon > 0$ and for all $i$ sufficiently large. The explanation for this is twofold. In the search step it results from Lemma 3.2 and the fact that the models are required to be fully linear. The explanation for a TR step comes from the fact that the true gradient goes to zero whenever the model one does (which can be seen from the proof of Theorem 3.1). Without loss of generality, we pick $\epsilon$ such that

$$\epsilon \ \leq \ \min \left\{ \frac{\epsilon_0}{2(2+\kappa_{eg}\mu)}, \epsilon_c \right\}. \tag{9}$$

Property (7) ensures the existence, for each $k_i$ in the subsequence, of a first iteration $\ell_i > k_i$ such that $\|g_{\ell_i}\| < \epsilon$. By removing elements from $\{k_i\}$, without loss of generality and without a change of notation, we thus obtain that there exists another subsequence indexed by $\{\ell_i\}$ such that

$$\|g_k\| \ \geq \ \epsilon \ \text{ for } \ k_i \leq k < \ell_i \ \text{ and } \ \|g_{\ell_i}\| \ < \ \epsilon, \tag{10}$$

for sufficiently large $i$.

We now restrict our attention to the set $\mathcal{K}$ corresponding to the subsequence of iterations whose indices are in the set

$$\cup_{i\in\mathbb{N}_0}\{k \in \mathbb{N}_0 : \ k_i \leq k < \ell_i\},$$

where $k_i$ and $\ell_i$ belong to the two subsequences defined above in (10).

We know that $\|g_k\| \geq \epsilon$ for $k \in \mathcal{K}$. From $\lim_{k\to+\infty} \Delta_k = 0$ and Lemma 3.1 we conclude that for any large enough $k \in \mathcal{K}$ the iteration $k$ is either successful or model improving.

Moreover, for each $k \in \mathcal{K} \cap \mathcal{S}$ we have that either (TR step)

$$f(x_k) - f(x_{k+1}) \ \geq \ \eta_1[m_k(x_k) - m_k(x_k+s_k)] \ \geq \ \eta_1 \frac{\kappa_{fcd}}{2}\|g_k\| \min\left\{ \frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k \right\} \tag{11}$$

and for any such $k$ large enough, $\Delta_k \leq \frac{\epsilon}{\kappa_{bhm}}$, or (search step)

$$f(x_k) - f(x_{k+1}) \ \geq \ \rho(\Delta_k) \ = \ \gamma_\rho \Delta_k. \tag{12}$$

Hence, we have for $k \in \mathcal{K} \cap \mathcal{S}$ sufficiently large,

$$\Delta_k \ \leq \ \max\left( \frac{2}{\eta_1 \kappa_{fcd}\epsilon}, \frac{1}{\gamma_\rho} \right) [f(x_k) - f(x_{k+1})] \ := \ C_2[f(x_k) - f(x_{k+1})].$$

Since for any $k \in \mathcal{K}$ large enough the iteration is either successful or model improving and since for a model improving iteration $x_k = x_{k+1}$ we have, for



**Lemma 3.1.** *Consider an iteration $k$ corresponding to a rigorous TR step. If $m_k$ is fully linear on $B(x_k;\Delta_k)$ and the iteration is not successful (i.e. if it is acceptable or unsuccessful), then*

$$\|g_k\| \ \leq \ C_1 \Delta_k,$$

*where*

$$C_1 \ = \ \frac{1}{\min\left\{ \frac{1}{\kappa_{bhm}}, \frac{\kappa_{fcd}(1-\eta_1)}{4\kappa_{ef}} \right\}}.$$

We will now show that the trust-region radius converges to zero (this requires some modifications from [4, Lemma 5.5], see also [5, Lemma 10.9], to accommodate the search step).

**Lemma 3.2.** *If $\rho(\cdot)$ is chosen such that $\rho(\Delta) \to 0 \Rightarrow \Delta \to 0$, then*

$$\lim_{k\to+\infty} \Delta_k \ = \ 0. \tag{4}$$

*Proof*: The proof follows known arguments when the number of successful iterations is finite (see, e.g, the proof of [5, Lemma 10.7]). In this case, without loss of generality one can consider only iterations acceptable, model improvement or unsuccessful, where the trust-region radius is not increased. We then know that we can have only a finite (uniformly bounded, say by $N$) number of model-improvement iterations before the model becomes fully linear, which shows that there is an infinite number of iterations that are either acceptable or unsuccessful (and in either case a reduction occurs in the trust-region radius). Moreover, $\Delta_k$ is decreased at least once every $N$ iterations by a factor of $\gamma$. As a result, $\Delta_k$ converges to zero.

Let us now consider the case when $\mathcal{S}$ is infinite. Two types of successful iterations are possible (depending if they occur in the search step or in the rigorous TR one). In the former case, when $k \in \mathcal{S}_{search}$, we obtain

$$f(x_k) - f(x_{k+1}) \ \geq \ \rho(\Delta_k). \tag{5}$$

In the latter case, when $k \in \mathcal{S}_{tr}$ we have

$$f(x_k) - f(x_{k+1}) \ \geq \ \eta_1[m_k(x_k) - m_k(x_k+s_k)].$$

By using the bound on the fraction of Cauchy decrease (1), we have that

$$f(x_k) - f(x_{k+1}) \ \geq \ \eta_1 \frac{\kappa_{fcd}}{2}\|g_k\| \min\left\{ \frac{\|g_k\|}{\|H_k\|}, \Delta_k \right\}.$$

Due to the TR Step 1 of Algorithm 2.2 we have that $\|g_k\| \geq \min\{\epsilon_c, \mu^{-1}\Delta_k\}$, hence

$$f(x_k) - f(x_{k+1}) \;\geq\; \eta_1 \frac{\kappa_{fcd}}{2} \min\{\epsilon_c, \mu^{-1}\Delta_k\} \min\left\{ \frac{\min\{\epsilon_c, \mu^{-1}\Delta_k\}}{\|H_k\|}, \Delta_k \right\}. \quad (6)$$

Since $\mathcal{S}$ is infinite and $f$ is bounded from below, and by using Assumption 3.3 and the property assumed for $\rho(\cdot)$, the right-hand sides of the above expressions (5) and (6) have to converge to zero (whenever they occur an infinite number of times). Hence $\lim_{k\in\mathcal{S}} \Delta_k \;=\; 0$, and nothing else would remain to be proved if all iterations are successful. However, the trust-region radius can only be increased during a successful iteration, and it can only be increased by a ratio of at most $\gamma_{inc}$, which then completes the proof. ■

Now we can state that there is a subsequence of iterates along which the true gradient goes to zero. The proof of this fact follows a new insight given by the fact that the trust-region radius is converging to zero. In fact, this behavior of the trust-region radius necessarily implies that there is an infinite number of iterations where it must be reduced. Also, the trust-region radius cannot possibly be reduced at search steps and thus we can focus on what happens in the rigorous TR ones. In more classical trust-region methods, one would immediately conclude that there is an infinite number of unsuccessful iterations. However, because of the more complex DFO setting, in particular the presence of the criticality step (main contributor for the convergence to zero of the trust-region radius) and the way simple decrease is handle (acceptable iterations), one has three rather than one type of situation responsible for a decrease in the trust-region radius. Fortunately, in all cases one has $\|g_k\| = \mathcal{O}(\Delta_k)$, allowing one to drive a subsequence of model gradients to zero, from which then the result stated below easily follows.

**Theorem 3.1.** *Let Assumptions 3.1, 3.2, and 3.3 hold. If $\rho(\cdot)$ is chosen such that $\rho(\Delta) \to 0 \Rightarrow \Delta \to 0$, then*

$$\liminf_{k\to+\infty} \|\nabla f(x_k)\| \;=\; 0.$$

*Proof*: From Lemma 3.2, we know that there must exist an infinite number of iterations where the trust-region radius is reduced (which must occur at rigorous TR steps). Thus, there is either an infinite number of criticality steps where the trust-region radius is reduced (and $\|g_k\| \leq \Delta_k/\beta$ holds) or an infinite number of either acceptable or unsuccessful iterations (where Lemma 3.1 applies), and let us denote all these iterations by the index sequence $\{\ell_i\}$. In any of these three cases, one has $\|g_{\ell_i}\| = \mathcal{O}(\Delta_{\ell_i})$ and by taking limits when $\Delta_{\ell_i}$ goes to zero, one obtains

$$\lim_{i\to+\infty} \|g_{\ell_i}\| \;=\; 0. \quad (7)$$

Also, in any of the cases, one has

$$\|\nabla f(x_{\ell_i}) - g_{\ell_i}\| \;\leq\; \kappa_{eg}\Delta_{\ell_i},$$

and, from (7) and $\Delta_{\ell_i} \to 0$, we derive $\|\nabla f(x_{\ell_i})\| \to 0$. ■



It is possible to extend this result to the whole sequence of iterates, establishing a result of the lim-type given in [4, Lemma 5.9] (see also [5, Lemma 10.13]). To do so, we need to impose that the search step $x - x_k = x_{k+1} - x_k$ stays in a trust region of radius proportional to $\Delta_k$ and to compute at this step a model which is fully linear in such a trust region. This observation is aligned with the generalization of liminf to lim in direct-search methods which requires the search step to essentially be empty (or to coincide with a complete poll step, which, note, can be seen as a way of imposing fully linearity); see [7] and also [5, Pages 132–133].

**Theorem 3.2.** *Let Assumptions 3.1, 3.2, and 3.3 hold, and $\gamma_{fac}, \gamma_\rho > 0$ be constants independent of the iteration counter. If $\rho(\Delta) = \gamma_\rho\Delta$, and if in the search step $x_{k+1} \in B(x_k; \gamma_{fac}\Delta_k)$ and a model $m_k(x_k + s)$ is formed, fully linear in $B(x_k; \gamma_{fac}\Delta_k)$, then*

$$\lim_{k\to+\infty} \nabla f(x_k) \;=\; 0.$$

*Proof*: The proof is classical and only requires a few adjustments. We will follow closely the presentation in [5, Theorem 10.13].

We have seen from Lemma 3.2 and Theorem 3.1 that in the case when $\mathcal{S}$ is finite the theorem holds. Hence, we will assume that $\mathcal{S}$ is infinite. Suppose, for the purpose of establishing a contradiction, that there exists a subsequence $\{k_i\}$ of successful iterations such that

$$\|\nabla f(x_{k_i})\| \;\geq\; \epsilon_0 \;>\; 0, \quad (8)$$

for some $\epsilon_0 > 0$ and for all $i$ (we can ignore model-improving iterations, since $x_k$ does not change during such iterations). Then, we obtain that

$$\|g_{k_i}\| \;\geq\; \epsilon \;>\; 0,$$