TWMS J. App. Eng. Math. V.5, N.1, 2015, pp. 46-60.

DERIVATIVE FREE MULTILEVEL OPTIMIZATION

B. KARASÖZEN[1], §

Abstract. Optimization problems with different levels arise by discretization of ordinary and partial differential equations. We present a trust-region based derivative-free multilevel optimization algorithm. The performance of the algorithm is shown on a shape optimization problem and global convergence to the first order critical point is proved.

Keywords: Derivative-free optimization; multilevel optimization; shape optimization; trust-region methods.

AMS Subject Classification: 90C30, 65K05, 90C26, 90C06, 90C56 ; 65D05.

1. Introduction

Discretization of infinite dimensional optimization problems, such as optimal control problems with partial differential equations(PDEs) lead to large-scale finite dimensional optimization problems. This kind of problems can be solved for a dicretization level by the existing large scale numerical optimization packages. But this approach does not exploit the structure of the underlying infinite dimensional optimization problem which can be dicretized at different levels. There exist several methods, which make use of the discretization of infinite dimensional problems at different levels. The simplest approach is to use coarse grids in order to compute approximate solutions for the starting points on a finer grid. Efficient optimizations methods were developed recently within the framework of multi-grid methods [2, 11, 13]. In [11], a recursive line-search based truncated Newton method was developed for efficient solution of large-scale convex optimization problems arising from the discretization of partial differential equations (PDEs). This approach is then extended to nonconvex problems using the trust-region approach in a series of papers [7, 8, 9, 17], known as recursive multilevel (multiscale) trust-region (RMTR) methods. The recursive multilevel optimization methods use a model of the objective function on the coarse grid at the lower level for optimization on the fine grid at the higher level .

For continuously differentiable functions usually gradient based methods are used the find the local minimum of the objective function. However, for many practical problems the derivatives of the objective function are either not available or costly to evaluate. Derivative free optimization methods build models of function based on sample function value or directly exploit a sample set of function values without building an explicit model [6, 10]. Among them are the well known trust-region based methods by modeling

[1]Department of Mathematics & Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey.
e-mail: bulent@metu.edu.tr;

§ Manuscript received: March 01, 2014.

the objective function by multivariate interpolation in combination with the trust-region techniques [3, 4, 5].

In this paper we have developed a trust-region based derivative-free recursive multi-level optimization (DFRMTR) method for solving the finite-dimensional optimization problem at different levels. In the next section, the DFRMTR algorithm is described. In Section 3 we show the implementation of our algorithm for a shape optimization problem and the convergence analysis of the method is given in Section 4.

2. Derivative free multilevel optimization

Consider the problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a smooth function and bounded below. Trust-region methods compute a sequence of iterates x_k, starting from an initial guess x_0, converging to the solution of the problem (1). At each iteration k, a linear or quadratic model Q_k of f is constructed in the neighborhood of the current iterate x_k, $\mathcal{B}_k=\{x_k+s : s\in\mathbb{R}^n \text{ and } ||s||\leq \Delta_k\}$, with the trust-region radius Δ_k, here $||\cdot||$ represents the Euclidean norm. At the kth step, the model function within the trust-region $\mathcal{B}_k$, is given as

$$Q_k\,(x_k+s)=Q_k\,(x_k)+<s,g_k>+\frac{1}{2}<s,H_k s> \tag{2}$$

for some $g\in\mathbb{R}^n$ and some symmetric $n\times n$ matrix H, where $<\cdot,\cdot>$ denotes the inner product . The vector g and the matrix H do not necessarily correspond to the first and second derivatives of the objective function f. They are determined by requiring that the model (2) interpolates the function f at a set $Y=\{y_i\}$ of points containing the current iterate x_k, $f(y_i)=m_k(y_i)$ for all $y_i\in Y$. Here, Y denotes the set of interpolation points, which is a subset of the set of points at which the values of f is known, including the current iterate. Building the full quadratic model in (2) requires the determination of $f(x_k)$, the components of the vector g_k and the entries of the matrix H_k; so that the cardinality of Y must be equal to $p=\frac{1}{2}(n+1)(n+2)$.

At each iteration step k the trust-region subproblem defining s_k

$$\min_{s\in\mathcal{B}(0;\Delta_k)} Q_k(x_k+s), \tag{3}$$

has to be solved. where $\mathcal{B}(0;\Delta_k)$ is the trust-region of radius Δ_k centered at 0 and $s=x-x_k$.

The RMTR methods developed by [7, 9, 17], consider a collection of twice-continuously differentiable functions $f_i:\mathbb{R}^{n_i}\to\mathbb{R}, i=1,\ldots,r$ with $n_i>n_{i-1}$ and $f_r(x)=f(x)$. Because f_i represents a finer discretization of the infinite dimensional objective function, it has more variables than f_{i-1} and therefore it is more costly to minimize f_i than f_{i-1} The main idea of the multilevel optimization is then to use f_{r-1} to construct a cheaper model h_r in the neighborhood of the current iterate than the quadratic model for f_r in order to define the step in the trust-region algorithm. If more than one level is available $(r>1)$, this can be done recursively. At the level $i=0$, always the quadratic model is used. The linear and quadratic models in [7, 9, 17] use the gradient and the Hessian of $f(x)$ and the relation between this models is by the restriction and prolongation between operators between a fine and a coarse grid as in multigrid algorithms.

We construct the lower model h_i based on Q_{i-1} which results from the interpolation of $f(x)$ at the level $i-1$, similar to the recursive multilevel optimization methods in [7, 9, 17]. The lower level model is defined as

$$h_{i,k}(x_{i,k}) = Q_{i-1}(x_{i,k}) + (\nabla Q_{i,k}(x_{\min}) - \nabla Q_{i-1,k}(x_{\min}))^T (x_{i,k} - x_{\min}), \tag{4}$$

where $x_{\min}$ is the minimum of value of $f(x)$ at the $(i-1)$th level and Q_{i-1} is a fixed function for every iteration k at ith level. In the derivative-free case, the prolongation and restriction operators can not be used, because the gradient and Hessian of the function $f(x)$ are not available. Instead of this, we use either the minimum point in the interpolation set of $(i-1)$th level to construct the new interpolation set at the level i, or we take the interpolation set from the previous level $i-1$. In the following, i $(1 \leq i \leq r)$ denotes the level index and the k, the the current iteration at the level i. The lower level model corresponds to the modification of the model Q_{i-1} at i'th level. The lower level model can not always be useful since $\nabla Q_{i-1,k}$ can be close to zero with respect to $\nabla Q_{i,k}$. In this case the current iterate appears to be first-order critical for $(i-1)$th level while it is not for i'th level. Therefore the lower model is useful only if $\nabla Q_{i-1,k}$ is large enough compared to $\nabla Q_{i,k}$ and $\nabla Q_{i,k}$ is greater than a constant ϵ_Q, which is given by the following condition (see also [7, 9, 17]):

$$\|\nabla Q_{i,k}\| \geq \kappa_Q \|\nabla Q_{i-1,k}\| \text{ and } \|\nabla Q_{i,k}\| \geq \epsilon_Q, \text{ where } \kappa_Q, \epsilon_Q \in (0,1). \tag{5}$$

When the conditions above are satisfied, for kth iteration at the ith level the lower level model $h_{i,k}$ is used and the trust-region subproblem becomes

$$\min_{\|s_{i,k}\| \leq \Delta_{i,k}} h_{i,k}(x_{\min} + s_{i,k}) \tag{6}$$

where $s_{i,k} = x_{i,k} - x_{\min}$.

Otherwise, the model (2) is used and trust-region subproblem at the Taylor step becomes

$$\min_{\|s_{i,k}\| \leq \Delta_{i,k}} Q_{i,k}(x_{i,k} + s_{i,k}) + \nabla Q_{i,k}(x_{i,k} + s_{i,k})^T s_{i,k} + s_{i,k}^T \nabla^2 Q_{i,k}(x_{i,k} + s_{i,k}) s_{i,k}. \tag{7}$$

We can now define the derivative-free multilevel optimization (DFRMTR) algorithm.

Step 0: Initialization

Given $x^0 \in \mathbb{R}^n$, the initial guess, Δ_0, the initial trust-region radius, construct a well-poised interpolation set Y around $x^0 \in Y \subset \mathbb{R}^n$ within the initial trust-region. and build the quadratic interpolation model Q_i on the interpolation set Y.

Step 1: Model Choice

If $i = 1$ or if the conditions: $\|\nabla Q_i\| \geq \kappa_Q \|\nabla Q_{i-1}\|$ and $\|\nabla Q_i\| > \epsilon_Q$ fail, go to Step 3 (Taylor step), otherwise go to Step 2 (lower level model).

Step 2: Lower level model computation

Solve the trust-region subproblem $\min_{\|s\| < \Delta_{i,k}} h_{i,k}(x_{\min} + s)$.

Step 3: Taylor step computation

Solve the model problem

$$\min_{\|s\| \leq \Delta_{i,k}} Q_i(x) + \nabla Q_{i,k}^T(x)s + \frac{1}{2}s^T \nabla^2 Q_{i,k}(x)s.$$

Step 4: Updating the interpolation set

Compute the ratio

$$\rho_{i,k} = (f_i(x_k) - f_i(\hat{x}_k))/\delta_{i,k}, \quad \text{with } \hat{x}_k = \hat{x}_{i,k} = x_{i,k} + s_{i,k},$$

- $\delta_{i,k} = Q_i(x_k) - Q_i(\hat{x}_k)$ (if the model (2) is used),
- $\delta_{i,k} = h_i(x_k) - h_i(\hat{x}_k)$ (if the lower level model is used).
- *Successful step:* If $\rho_{i,k} \geq \eta_0$, include $\hat{x}_k$ in Y by dropping one of the existing interpolation points.
- *Unsuccessful step:* If $\rho_{i,k} < \eta_0$ and Y is inadequate in $x \in \mathcal{B}_k$, improve the geometry of the interpolation set.

Step 5: Updating the current iterate

Determine $\bar{x}_k$ with the best objective function value $f(\bar{x}_k) = \min_{x_j \in Y, x_j \neq x_k} f(x_j)$. If the improvement is sufficient $\bar{\rho}_{i,k} = (f_i(x_k) - f_i(\bar{x}_k))\delta_{i,k} \geq \eta_0$, set $x_{k+1} = \bar{x}_k$, otherwise $x_{k+1} = x_k$ where $x_k = x_{i,k}$.

Step 6: Trust-region radius update

- if $\rho_{i,k} \geq \eta_1$, increase the trust-region radius, $\Delta_{i,k+1} \in [\Delta_{i,k}, \gamma_2 \Delta_{i,k}]$
- if $\rho_{i,k} < \eta_0$ and the cardinality of $Y \cap B_k$ was less than $n+1$ when $\hat{x}_k$ was computed, reduce the trust-region radius, $\Delta_{i,k+1} \in [\gamma_0 \Delta_{i,k}, \gamma_1 \Delta_{i,k+1}]$
- otherwise set $\Delta_{i,k+1} = \Delta_{i,k}$.

Step 7: Termination

The algorithm is terminated when one of the following three criteria are satisfied:

- The radius of trust-region is small enough, such that. $\Delta \leq \epsilon_\Delta$
- Final interpolation point set has the ‘ *good geometry*’ property.
- Maximum function evaluations or number of maximum iterations are reached.

Increment k by one and go to Step 0.

In the algorithm, some constants and parameters are used: ϵ_Δ denotes the minimum value for the trust-region radius, $0 < \eta_0 < \eta_1 < 1$ are parameters to improve quality of interpolation set, $0 < \gamma_0 \leq \gamma_1 < \gamma_2 \leq 1$ are constant which monitor the trust-region radius, $\kappa_Q \in (0, 1)$ is a constant which is used for the model choice, $\epsilon_Q \in (0, 1)$ is the tolerance for the gradient norm and ϵ_{fun} is the tolerance for the function reduction.

There exist various implementations derivative free methods. The oldest one is the DFO (Derivative Free Optimization) [3, 4]. The package CONDOR (Constrained, Nonlinear, Direct, parallel Optimization using trust-region method for high-computing load function) [1] is based on the UOBYQA algorithm [14]. The DFO uses Newton polynomials while CONDOR uses Lagrange polynomials as the basis for the interpolation polynomial. The linear or quadratic models are minimized in DFO package by applying a standard optimization procedure, e.g., sequential quadratic programming (SQP), using the IPOPT package. CONDOR uses the Moré and Sorenson algorithm [12] for the computation of the trust-region radius and the minimization of the quadratic model. The trust-region subproblem for the lower level problem (6) is solved by CONDOR using the Moré and Sorensen algorithm [12]. For solution of the trust-region subproblem (7) withe DFO, we use either, *trust* with the full eigenvalue decomposition, based on the secular equation $\frac{1}{\Delta} - \frac{1}{\|s\|} = 0$ [16] or *lmlib* or the Levenberg-Marquardt algorithm with the Moré and Sorensen technique. Updating and improving the interpolation set are explained in detail in [3, 4] and they are implemented in different ways in the DFO and CONDOR. The geometry of the interpolation set has to maintained at every iteration. There must be at least $n+1$ points are in the trust-region. If there are less than $n+1$ points, the farthest point from the current trust-region center by $\tilde{x}$ is replaced with a point on the boundary of the current trust-region, so that the interpolation set is well poised [3, 4].

3. Numerical Example

We consider a *shape optimization problem* over a rectangular region with a rectangular hole $\Omega(u)$ parameterized with the coordinates $u = (P9, P12)$, as in Figure 1.

$$\min \frac{1}{2} \int_{\Omega(u)} (y(x_1, x_2) - y_d)^2 \, dx$$

such that

$$\Delta y = 1 \text{ in } \Omega(u), \quad y = 0 \text{ on } \partial\Omega(u).$$

Here, $y(x_1, x_2)$ is a state variable, y_d is the desired state, and u is the control variable. Data y_d is designed in such a way, that $y(u^*)$ is the global minimum for the optimal control u^*.

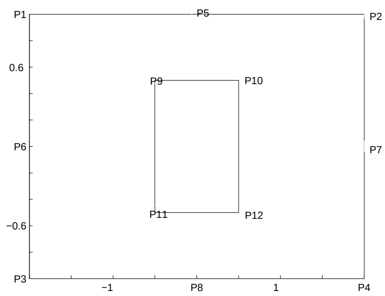

FIGURE 1. Shape optimization problem

We have used the parameters as in the MATLAB version of the DFO algorithm [15]:

$$\epsilon_{trust} = 0.01, \epsilon_{det} = 1e-12, \Delta_0 = 0.2, \eta_0 = 0.45, \eta_1 = 0.75,$$

$$\gamma_1 = 0.3, \gamma_2 = 1, \epsilon_{dist} = 0.001, \epsilon_{fun} = 1e-8, \kappa_Q = 0.01, \epsilon_Q = 0.001.$$

In CONDOR [1] only the parameters $\rho_{start} = 0.2$, the initial distance between sample points, and $\rho_{end} = 0.001$ stopping criteria for the distance of the points, can be specified by the user and other parameters are fixed.

The initial trust-region radius was taken as $\Delta_0 = 0.2$ and $\epsilon_{fun} = 1.0e-8$. In all computations, we use $u^0 = [-0.5, 0.5, 0.5, -0.5]^T$ as starting value for the control variable and $\bar{u}$ denotes the optimal control computed at the point $(P9, P12)$. The exact value of the optimal control is not known.

We give first the numerical results in Table 1 and 2, for DFO and CONDOR without using the DFRMTR. The number of iterations were not available in CONDOR. From

Tables 1 and 2, we see that CONDOR requires smaller number of function evaluations and it gives more accurate results for the minimization of the cost function than the DFO.

TABLE 1. DFO without multilevel derivative-free optimization

level	# iterations	# func. eval.	func. value	$\bar{u}$
1	32	103	5.0677e-008	$-0.7108, 0.2386, -0.1397, -0.2385$
2	35	91	1.1911e-010	$-0.7501, 0.2499, -0.2499, -0.2499$
3	39	82	9.4293e-009	$-0.7504, 0.2498, -0.2491, -0.2495$
4	33	84	1.0900e-008	$-0.7501, 0.2495, -0.2485, -0.2499$
5	35	75	5.0694e-011	$-0.7500, 0.2500, -0.2501, -0.2500$

TABLE 2. CONDOR without multilevel derivative-free optimization

level	# func. eval.	# func. value	$\bar{u}$
1	131	3.0482e-014	$-0.7158, 0.2275, -0.0894, -0.2275$
2	82	3.2440e-014	$-0.7500, 0.2500, -0.2500, -0.2500$
3	60	1.0045e-012	$-0.7500, 0.2500, -0.2500, -0.2500$
4	52	1.6304e-011	$-0.7500, 0.2500, -0.2500, -0.2500$
5	61	2.8944e-012	$-0.7500, 0.2500, -0.2500, -0.2500$

In Tables 3 and 4, the numerical results obtained with the DFRMTR algorithm are shown. The interpolation set of Q_i is constructed by using the minimum point of the $(i-1)$th level. The trust-region subproblem is solved by *lmlib* or *trust* routines.

TABLE 3. DFRMTR with *lmlib*

level	# iterations	# func. eval.	# func. value	$\bar{u}$
1	32	103	5.0677e-08	$-0.7108, 0.2386, -0.1397, -0.2385$
2	17	52	2.1545e-07	$-0.7492, 0.2515, -0.2617, -0.2525$
3	17	56	1.3604e-09	$-0.7501, 0.2499, -0.2503, -0.2499$
4	14	51	1.4646e-09	$-0.7501, 0.2499, -0.2503, -0.2499$
5	14	51	1.4966e-09	$-0.7501, 0.2499, -0.2503, -0.2499$

TABLE 4. DFRMTR with *trust*

level	# iterations	# func. eval.	# func. value	$\bar{u}$
1	50	136	8.6367e-10	$-0.7152, 0.2180, -0.0380, -0.2181$
2	34	103	4.2532e-10	$-0.7499, 0.2501, -0.2502, -0.2500$
3	14	49	5.5573e-10	$-0.7499, 0.2501, -0.2502, -0.2500$
4	14	49	6.0052e-10	$-0.7499, 0.2501, -0.2502, -0.2500$
5	14	49	6.1392e-10	$-0.7499, 0.2501, -0.2502, -0.2500$

In Tables 4 and 5, the numerical results are given by constructing the initial interpolation set of Q_i by using last interpolation set of previous $(i-1)$th level with *lmlib* and *trust* routines.

Table 5. DFRMTR with *lmlib*

level	# iterations	# func. eval.	func. value	$\bar{u}$
1	32	103	5.0677e-08	$-0.7108, 0.2386, -0.1397, -0.2385$
2	21	49	4.9240e-09	$-0.7493, 0.2500, -0.2492, -0.2499$
3	5	21	1.3604e-09	$-0.7493, 0.2500, -0.2492, -0.2499$
4	4	19	9.3654e-09	$-0.7493, 0.2500, -0.2492, -0.2499$
5	4	19	9.7371e-09	$-0.7493, 0.2500, -0.2492, -0.2499$

Table 6. DFRMTR with *trust*

level	# iterations	# func. eval.	func. value	$\bar{u}$
1	50	136	8.6367e-10	$-0.7152, 0.2180, -0.0380, -0.2181$
2	17	55	2.8223e-09	$-0.7494, 0.2500, -0.2497, -0.2501$
3	4	23	4.3601e-09	$-0.7494, 0.2500, -0.2497, -0.2501$
4	4	19	4.8409e-09	$-0.7494, 0.2500, -0.2497, -0.2501$
5	4	19	4.9758e-09	$-0.7494, 0.2500, -0.2497, -0.2501$

We also obtained numerical results using $f(x)$ instead of the model function $Q(x)$. At the coarsest $i = 1$, the quadratic Taylor model (2) is solved with DFO. At all other levels levels, the algorithm chooses either the quadratic Taylor model (2) and or the lower level model

$$h_i(x) = f_{i-1}(x) + (\nabla Q_i(x_{i-1}) - \nabla Q_i(x_{i-1}))^T (x - x_i)$$

is solved, where x_i is the minimum point obtained from first level and Q_i is the last model function of level i.

The results are given in the Tables 7-10. In Table 7 and 8, the interpolation set of Q_i is constructed by using the minimum point of the $(i-1)$th level.

Table 7. DFRMTR with *lmlib* using the function $f(x)$

level	# iterations	# func. eval.	func. value	$\bar{u}$
1	32	103	5.0677e-08	$-0.7108, 0.2386, -0.1397, -0.2385$
2	17	52	2.1545e-07	$-0.7492, 0.2515, -0.2617, -0.2525$
3	17	56	1.3604e-09	$-0.7501, 0.2499, -0.2503, -0.2499$
4	14	51	1.4646e-09	$-0.7501, 0.2499, -0.2503, -0.2499$
5	14	51	1.4966e-09	$-0.7501, 0.2499, -0.2503, -0.2499$

On coarse grids, *trust*: eigenvalue decomposition based on the secular equation requires more iterations and function evaluations than *lmlib*: the Levenberg-Marquardt algorithm with the More & Sorensen technique. On the finer grids, they produce almost the same number of iterations and function evaluations. The choice between the lower level model

Table 8. DFRMTR with *trust* using the function $f(x)$

level	# iterations	# func. eval.	func. value	$\bar{u}$
1	50	136	8.6367e-10	$-0.7152, 0.2180, -0.0380, -0.2181$
2	33	102	1.5754e-09	$-0.7499, 0.2500, -0.2507, -0.2502$
3	14	49	2.2922e-09	$-0.7499, 0.2500, -0.2507, -0.2502$
4	14	49	2.6092e-09	$-0.7499, 0.2500, -0.2507, -0.2502$
5	14	49	2.7059e-09	$-0.7499, 0.2500, -0.2507, -0.2502$

and Taylor model depends critically on parameters κ. Computations with *limlib* produced the most accurate results (see Table 8). The numerical results are affected by the construction of the interpolation set. When the interpolation set at level i is constructed around the minimum point of the level $i-1$, more function evaluations are required than by the construction of interpolation set using the last interpolation set of previous level (see Tables 3, 4 and Tables 7,8). The computation time (number of iterations and functions evaluations) increases when $f(x)$ is used instead of Q.

4. Convergence of the recursive multilevel derivative free method

In the following we show that, the solutions obtained by the derivative free multilevel algorithm in Section 2, converge to the first order critical points. The convergence analysis is based on minimization of derivative free trust-region subproblem in [6] and [16]. We consider the case that the interpolation set is constructed from the last iteration of previous level i. When the conditions (5) are satisfied the following trust-region subproblem (6) with the lower level model (4) will be solved We use the following assumptions for the trust-region method as in [6, 16]:

- A.1: The objective function f is twice continuously differentiable and its Hessian is uniformly bounded over $\mathbb{R}^n$, so that which there exists a positive constant κ_1 such that, for all $x_{i,k} \in \mathbb{R}^n$, $\left\|\nabla^2 f(x_{i,k})\right\| \leq \kappa_1$. where $\kappa_1 \geq 1$ as in [16].
- A.2: The objective function f is bounded below .
- A.3: The Hessian of the chosen model is uniformly bounded, that is there exists a constant $\kappa_2 > 0$ such that $1 + \|H_{i,k}\| \leq \kappa_2$.

In the following, we give some definitions, lemmas concerning the derivative free optimization [3].

An interpolation set Y is called *adequate* in $\mathcal{B}_{i,k}(\Delta_{i,k})$ whenever

- the cardinality of Y is at least $n+1$ within the trust-region $x^j \in \mathcal{B}_{i,k}(\Delta_{i,k})$ for all $x^j \in Y$,
- and $dist(x^j - x^c) < 2\Delta_{i,k}$ holds, where x^c denotes the center of the interpolation set and $\Delta_{i,k}$ is the trust-region radius at the level i and at kth iteration.

Theorem 4.1. *Assuming that A.1 - A.3 hold, then the relation between the the objective function and the model function (4) is given by*

$$|f_i(x_{i,k}) - h_i(x_{i,k})| \leq 3\kappa_{ay} \max[\Delta_{i,k}^2, \Delta_{i,k}^3]$$

for all $x_{i,k} \in \mathcal{B}_{i,k}(\Delta_{i,k})$ and some constant $\kappa_{ay} > 0$, where $h_i = h_{i,k}$.

Proof. As in the convergence analysis of the derivative free optimization [5], we can write

$$|f_{i-1}(x_{i-1,k}) - Q_{i-1}(x_{i-1,k})| \leq \breve{\kappa}_{md} \max[\Delta_{i,k}^2, \Delta_{i,k}^3] \quad (8)$$

$$\|\nabla f_{i-1}(x_{i-1,k}) - \nabla Q_{i-1}(x_{i-1,k})\| \leq \breve{\kappa}_{gd} \max[\Delta_{i,k}, \Delta_{i,k}^2] \tag{9}$$

for some constants $\breve{\kappa}_{md}, \breve{\kappa}_{gd} > 0$. Then it follows

$$|f_i(x_{i,k}) - Q_i(x_{i,k})| \leq \kappa_{md} \max\left[\Delta_{i,k}^2, \Delta_{i,k}^3\right] \tag{10}$$

for all $x_{i,k} \in \mathcal{B}_{i,k}(\Delta_{i,k})$ and some constant $\kappa_{md} > 0$.

If Taylor model is chosen with $\|\nabla Q_{i,k}(x_{min})\| \leq \Delta_{i,k}$ and $|Q_{i,k} - Q_{i-1,k}| \leq \Delta_{i,k}^2$, we obtain then

$$|f_i(x_{i,k}) - h_i(x_{i,k})| = |f_i(x_{i,k}) - Q_{i-1}(x_{i,k}) - (\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))\, s_{i,k}|\,. \tag{11}$$

$$\begin{aligned}
\|\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min})\| &\leq \|\nabla Q_{i,k}(x_{min})\| + \|\nabla Q_{i-1}(x_{min})\| \\
&\leq \|\nabla Q_{i,k}(x_{min})\| + \tfrac{1}{\kappa_Q}\|\nabla Q_{i,k}(x_{min})\| \\
&\leq (1 + \tfrac{1}{\kappa_Q})\|\nabla Q_{i,k}(x_{min})\| \\
&\leq \kappa_m \|\nabla Q_{i,k}(x_{min})\|\,,
\end{aligned} \tag{12}$$

where the condition $\|\nabla Q_{i,k}(x)\| \geq \kappa_Q \|\nabla Q_{i-1}(x)\|$ is used and $\kappa_m = 1 + \frac{1}{\kappa_Q} \geq 2$. Adding and subtracting $Q_i(x_{i,k})$ to (11), using the triangle inequality and Cauchy-Schwarz inequality, we obtain with (10) and the Assumption A.1:

$$\begin{aligned}
|f_i(x_{i,k}) - h_i(x_{i,k})| &\leq |f_i(x_{i,k}) + Q_i(x_{i,k}) - Q_i(x_{i,k}) - Q_{i-1}(x_{i,k})| \\
&+ \|(\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))\, s_{i,k}\| \\
&\leq |f_i(x_{i,k}) - Q_i(x_{i,k})| + |Q_i(x_{i,k}) - Q_{i-1}(x_{i,k})| \\
&+ \|(\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))\|\,\|s_{i,k}\| \\
&\leq \kappa_{md}\max[\Delta_{i,k}^2, \Delta_{i,k}^3] + \kappa_m \|\nabla Q_{i,k}(x_{min})\|\,\Delta_{i,k} + |Q_{i,k} - Q_{i-1,k}| \\
&\leq \kappa_{md}\max[\Delta_{i,k}^2, \Delta_{i,k}^3] + \kappa_m \Delta_{i,k}^2 + \Delta_{i,k}^2 \\
&\leq 3\kappa_{ay}\max[\Delta_{i,k}^2, \Delta_{i,k}^3],
\end{aligned}$$

where $\kappa_{ay} := \max[\kappa_{md}, \kappa_m, 1]$ and $\|s_{i,k}\| \leq \Delta_{i,k}$. □

Theorem 4.2. *Assuming that A.1 - A.3 hold and the lower level model (4) is chosen, we have*

$$\|\nabla f_i(x_{i,k}) - \nabla h_i(x_{i,k})\| \leq \kappa_{mat} \max[\Delta_{i,k}, \Delta_{i,k}^2]$$

for some constant κ_{mat} and for all $x_{i,k} \in \mathcal{B}_{i,k}(\Delta_{i,k})$.

Proof. Similar to (9) we obtain

$$\|\nabla f_i(x_{i,k}) - \nabla Q_{i,k}(x_{i,k})\| \leq \kappa_{gd} \max\left[\Delta_{i,k}, \Delta_{i,k}^2\right] \tag{13}$$

where $x_{i,k} \in \mathcal{B}_{i,k}(\Delta_{i,k})$ and $\kappa_{gd} > 0$ is a constant, and

$$\nabla h_{i,k} = \nabla Q_{i-1}(x_{i,k}) + \nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min})$$

Therefore,

$$\nabla f_{i,k} - \nabla h_{i,k} = \nabla f_i(x_{i,k}) - \nabla Q_{i-1}(x_{i,k}) - \nabla Q_{i,k}(x_{min}) + \nabla Q_{i-1}(x_{min}). \tag{14}$$

Adding and subtracting $\nabla Q_i(x_{i,k})$ to (14):

$$\begin{aligned}
\nabla f_i(x_{i,k}) - \nabla h_{i,k}(x_{i,k}) &= \nabla f_i(x_{i,k}) - \nabla Q_{i-1}(x_{i,k}) - \nabla Q_{i,k}(x_{min}) + \nabla Q_{i-1}(x_{min}) \\
&+ \nabla Q_{i,k}(x_{i,k}) - \nabla Q_{i,k}(x_{i,k}),
\end{aligned}$$

and then taking norm of (14) and using the triangle inequality, we obtain

$$\begin{aligned}
\|\nabla f_i(x_{i,k} - \nabla h_{i,k}(x_{i,k}))\| &\leq \|\nabla f_i(x_{i,k}) - \nabla Q_{i,k}(x_{i,k})\| \\
&+ \|\nabla Q_{i,k}(x_{i,k}) + \nabla Q_{i,k}(x_{min})\| \\
&+ \|\nabla Q_{i-1}(x_{i,k}) - \nabla Q_{i-1}(x_{min})\|\,.
\end{aligned}$$

Assuming that ∇Q is Lipschitz continuous at all levels, we obtain the following bound:

$$\|\nabla Q_{i-1}(x_{i,k}) - \nabla Q_{i-1}(x_{min})\| \leq \kappa \|x_{i,k} - x_{min}\| = \kappa \|s\| \leq \kappa\Delta_{i,k}$$
$$\|\nabla Q_{i,k}(x_{i,k}) - \nabla Q_{i,k}(x_{min})\| \leq \bar{\kappa} \|s\| \leq \bar{\kappa}\Delta_{i,k},$$

where κ and $\bar{\kappa}$ are constants independent of $\Delta_{i,k}$.
Thus, using last two inequalities and (13)

$$\begin{aligned} \|\nabla f_i(x_{i,k} - \nabla h_{i,k}(x_{i,k}))\| &\leq \kappa_{eg} \max[\Delta_{i,k}, \Delta_{i,k}^2] + \kappa\Delta_{i,k} + \acute{\kappa}\Delta_{i,k} \\ &\leq \kappa_{mat} \max[\Delta_{i,k}, \Delta_{i,k}^2], \end{aligned}$$

where $\kappa_{mat} := \kappa_{gd} + \kappa + \bar{\kappa}$. □

We denote $\mathcal{S} = \{k \mid \bar{\rho}_{i,k} \geq \eta_0\}$ as the index set of all successful iterations and $\mathcal{R} = \{k \mid \Delta_{i,k+1} < \Delta_{i,k}\}$ as the index set of all iterations where the trust-region radius is reduced.

Similar to Lemma 5 in the convergence proof of DFO in [5], we can write the following lemma for DFRMTR.

Lemma 4.1. (1) *For all k, if $\rho_{i,k} \geq \eta_0$, then $\bar{\rho}_{i,k} \geq \eta_0$ and thus iteration k is successful.*
(2) *If $k \in \mathcal{R}$, then Y is adequate in $\mathcal{B}_{i,k}(\Delta_{i,k})$.*
(3) *There are finite number of improvements of the geometry, unless $\nabla f_i(x_{i,k}) = 0$.*
(4) *There can only be a finite number of iterations such that $\rho_{i,k} < \eta_1$ before the trust-region radius is reduced in second item of Step 6 in DFRMTR.*

Proof. If $\rho_{i,k} \geq \eta_0$, $x_{i,k} + s_{i,k}$ is added to the interpolation set Y by Step 4. And it can be written by Step 5, $f(\bar{x}_{i,k}) \leq f(x_{i,k} + s_{i,k})$. Thus

$$\begin{aligned} \bar{\rho}_{i,k} &= \frac{f(x_{i,k}) - f(\bar{x}_{i,k})}{h_{i,k}(x_{i,k}) - h_{i,k}(x_{i,k} + s_{i,k})} \\ &\geq \frac{f(x_{i,k}) - f(x_{i,k} + s_{i,k})}{h_{i,k}(x_{i,k}) - h_{i,k}(x_{i,k} + s_{i,k})} = \rho_{i,k} \geq \eta_0. \end{aligned}$$

Then, $k \in \mathcal{S}$ and the k is successful. The proofs of (ii), (iii), (iv) can be done in the same way as n the convergence proof of DFO [5]. □

As in [16] we assume that for linear and quadratic models the "Cauchy point decrease condition"

$$Q_{i,k}(x_k) - Q_{i,k}(x_k + s_k) \geq \kappa_b \|g_{i,k}\| \min\left\{\frac{\|g_{i,k}\|}{1 + \|H_{i,k}\|}, \Delta_{i,k}\right\} \tag{15}$$

holds, where $g_{i,k} = \nabla_s Q_{i,k}(x_k), H_{i,k} = \nabla_{ss}^2 Q_{i,k}(x_k)$ and $\kappa_b \in (0,1)$.
In the following lemma we prove that the "Cauchy point condition" is valid for the lower level linear model in DFRMTR.

Lemma 4.2. *At every iteration k at level i, one has*

$$h_i(x_{i,k}) - h_i(x_{i,k} + s_{i,k}) \geq \frac{\kappa_Q \kappa_b}{2 + \kappa_Q} \|\nabla h_{i,k}\| \min\left[\frac{\kappa_Q \|\nabla h_{i,k}\|}{(2 + \kappa_Q)(1 + \|H_i\|)}, \Delta_{i,k}\right]$$

for some constant $\kappa_b \in (0,1)$ independent of k, and $H_i = \nabla^2 Q_i$ and $h_{i,k} = h_{i,k}(x_{i,k})$.

Proof.

$$\begin{aligned} h_i(x_{i,k}) - h_i(x_{i,k} + s) &= Q_{i-1}(x_{i,k}) + (\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))^T s - Q_{i-1}(x_{i,k} + s) \\ &- (\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))^T (x_{i,k} + s - x_{min}) \\ &= Q_{i-1}(x_{i,k}) - Q_{i-1}(x_{i,k} + s) - (\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))^T s. \end{aligned}$$

Taking the Taylor expansion, we obtain

$$Q_{i,k}(x_{min}) - Q_{i,k}(x_{min}+s) = -s\nabla Q_{i,k}(x_{min}) - 1/2s^2\nabla^2 Q_{i,k}(\xi_k), \tag{16}$$

where ξ_k is lying in the open stretch $(x_{min}, x_{min}+s)$. We assume that

$$-(\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min}))^T s \geq -\kappa_m \nabla Q_{i,k}(x_{min})^T s,$$
$$1/2s^2\nabla^2 Q_{i,k}(\xi_k) = 1/2s^T\nabla^2 Q_{i,k}(\xi_k)^T s \geq 0,$$
$$\|\nabla Q_{i,k}(x_{i,k})\| \leq \|\nabla Q_{i,k}(x_{min})\|.$$

Using the assumption of the Theorem 6.3.4 in [16], $Q_{i-1}(x_{i,k}) - Q_{i-1}(x_{i,k}+s) \geq 0$ we obtain

$$\begin{aligned} h_i(x_{i,k}) - h_i(x_{i,k}+s) &\geq (\nabla Q_{i-1}(x_{min}) - \nabla Q_{i,k}(x_{min}))^T s \\ &\geq -\kappa_m \nabla Q_{i,k}(x_{min})^T s \\ &= \kappa_m \left(Q_{i,k}(x_{min}) - Q_{i,k}(x_{min}+s) + 1/2s^2\nabla^2 Q_{i,k}(\xi_k)\right) \\ &\geq Q_{i,k}(x_{min}) - Q_{i,k}(x_{min}+s) \end{aligned}$$

where $\kappa_m \geq 2$ and (16) are used. Using (15), we can write

$$\begin{aligned} h_i(x_{i,k}) - h_i(x_{i,k}+s) &\geq Q_i(x_{min}) - Q_i(x_{min}+s) \\ &\geq \kappa_b \|\nabla Q_{i,k}(x_{min})\| \min\left[\tfrac{\|\nabla Q_{i,k}(x_{min})\|}{1+\|H_i\|}, \Delta_{i,k}\right], \end{aligned}$$

where $H_i = \nabla^2 Q_{i,k}(x_{min})$.

$$\begin{aligned} \|\nabla h_{i,k}\| &= \|\nabla Q_{i-1}(x_{i,k}) + \nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min})\| \\ &\leq \|\nabla Q_{i-1}(x_{i,k})\| + \|\nabla Q_{i,k}(x_{min}) - \nabla Q_{i-1}(x_{min})\| \\ &\leq \tfrac{1}{\kappa_Q}\|\nabla Q_{i-1}(x_{i,k})\| + \kappa_m \|\nabla Q_{i,k}(x_{min})\| \\ &\leq (\tfrac{1}{\kappa_Q} + \kappa_m)\|\nabla Q_{i,k}(x_{min})\|, \end{aligned} \tag{17}$$

where (5), (12) and the assumption (A.3) above are used. Putting $\kappa_m = 1 + 1/\kappa_Q$ into (17), we get

$$\|\nabla h_{i,k}\| \leq \left(\frac{1}{\kappa_Q} + 1 + \frac{1}{\kappa_Q}\right)\|\nabla Q_{i,k}(x_{min})\|, \text{ and } \frac{\kappa_Q}{\kappa_Q+2}\|\nabla h_{i,k}\| \leq \|\nabla Q_{i,k}(x_{min})\|, \tag{18}$$

$$\begin{aligned} h_i(x_{i,k}) - h_i(x_{i,k}+s) &\geq \kappa_b \|\nabla Q_{i,k}(x_{min})\| \min\left[\tfrac{\|\nabla Q_{i,k}(x_{min})\|}{(1+\|H_i\|)}, \Delta_{i,k}\right] \\ &\geq \kappa_b \tfrac{\kappa_Q}{2+\kappa_Q}\|\nabla h_{i,k}\| \min\left[\tfrac{\kappa_Q\|\nabla h_{i,k}\|}{(2+\kappa_Q)(1+\|H_i\|)}, \Delta_{i,k}\right], \end{aligned}$$

where $\kappa_b,\ \kappa_Q/(2+\kappa_Q) \in (0,1)$ since $\kappa_Q \in (0,1)$.

since

$$\left\|\nabla^2 h_{i,k}\right\| = \left\|\nabla^2 Q_{i-1}(x_k) + \nabla^2 Q_{i,k}(x_{min}) - \nabla^2 Q_{i-1}(x_{min})\right\|$$

$\left\|\nabla^2 h_i\right\| = \left\|\nabla^2 Q_i\right\| = \|H_{i,k}\|$, where $\nabla^2 Q_{i-1}(x_k) = \nabla^2 Q_{i-1}(x_{min})$ due to constant value result of second derivative of at most second degree function Q_{i-1}. □

Lemma 4.3. *Assuming that A.1 - A.3 and*

$$\Delta_{i,k} \leq \min\left[1, \frac{\kappa_Q\kappa_b\|\nabla h_{i,k}\|(1-\eta_0)}{(2+\kappa_Q)6\max[\kappa_h, \kappa_{ay}]}\right] \tag{19}$$

hold, then iteration k is successful and

$$\Delta_{i,k} \geq \Delta_{i,k-1}.$$

Proof. We note that $\eta_0, \kappa_b \in (0,1)$ and $\kappa_b(1-\eta_0) < 1$. Using κ_{ay}, κ_h and putting these in (19), we obtain

$$\frac{\|\nabla h_{i,k}\|}{6\max[\kappa_h, \kappa_{ay}]} \leq \frac{\|\nabla h_{i,k}\|}{6\kappa_h} \leq \frac{\|\nabla h_{i,k}\|}{6\kappa_2} \leq \frac{\|\nabla h_{i,k}\|}{\kappa_2} \leq \frac{\|\nabla h_{i,k}\|}{1+\|H_i\|}$$

where $\kappa_2 > 1 + \|H_i\|$ and $\kappa_h = \max[\kappa_1, \kappa_2]$.
Using these, we obtain

$$\Delta_{i,k} \leq \frac{\kappa_Q \kappa_b \|\nabla h_{i,k}\| (1-\eta_0)}{(2+\kappa_Q)6\max[\kappa_h, \kappa_{ay}]} \leq \frac{\kappa_Q \|\nabla h_{i,k}\|}{(2+\kappa_Q)(1+\|H_i\|)}$$

Combining this inequality with Lemma 4.4, at kth iteration, we get

$$\begin{aligned} h_i(x_{i,k}) - h_i(x_{i,k}+s_{i,k}) \quad &\geq \quad \tfrac{\kappa_Q\kappa_b}{2+\kappa_Q} \|\nabla h_{i,k}\| \min\left[\tfrac{\kappa_Q\|\nabla h_{i,k}\|}{(2+\kappa_Q)(1+\|H_i\|)}, \Delta_{i,k}\right] \\ &= \quad \tfrac{\kappa_Q\kappa_b\|\nabla h_{i,k}\|\Delta_{i,k}}{2+\kappa_Q} \end{aligned}$$

Now, we can write

$$\begin{aligned} |\rho_i - 1| \quad &= \quad \left|\frac{f_i(x_{i,k}) - f_i(x_{i,k}+s)}{h_i(x_{i,k}) - h_i(x_{i,k}+s)} - 1\right| \\ &\leq \quad \left|\frac{f_i(x_{i,k}+s) - h_i(x_{i,k}+s)}{h_i(x_{i,k}) - h_i(x_{i,k}+s)}\right| + \left|\frac{f_i(x_{i,k}) - h_i(x_{i,k})}{h_i(x_{i,k}) - h_i(x_{i,k}+s)}\right| \\ &\leq \quad 2\frac{3(2+\kappa_Q)\kappa_{ay}\max[\Delta_{i,k}^2, \Delta_{i,k}^3]}{\kappa_Q\kappa_b \|\nabla h_{i,k}\| \Delta_{i,k}} \\ &\leq \quad \frac{6(2+\kappa_Q)\kappa_{ay}\Delta_{i,k}}{\kappa_Q\kappa_b\|\nabla h_{i,k}\|} \\ &\leq \quad 1-\eta_0 \end{aligned}$$

Using the bounds

$$\frac{\kappa_b\kappa_Q(1-\eta_0)}{(2+\kappa_Q)} < 1, \ \frac{\|\nabla h\|}{6\max[\kappa_h, \kappa_{ay}]} \leq \frac{\|\nabla h\|}{\kappa_h}, \Delta_k \leq \min\left[1, \frac{\|\nabla h\|}{\kappa_h}\right]$$

we obtain

$$\Delta_{i,k} \leq \frac{\kappa_Q\kappa_b \|\nabla h_{i,k}\| (1-\eta_0)}{6(2+\kappa_Q)\max[\kappa_h, \kappa_{ay}]} \leq \frac{\kappa_Q\kappa_b \|\nabla h_{i,k}\| (1-\eta_0)}{6(2+\kappa_Q)\kappa_{ay}}.$$

Therefore, $\rho_i \geq \eta_0$ and the iteration is successful. Furthermore, at step 6 of the algorithm, $\Delta_{i,k} \geq \Delta_{i,k-1}$. □

Theorem 4.3. *Let's assume A.1 - A.3 hold and* $\|\nabla h_{i,k}\| \geq \kappa_c$*, then there exists a constant* $\kappa_d > 0$ *such that* $\Delta_{i,k} > \kappa_d$.

Proof. Assume that iteration k is the first k ($\rho_k < \eta_0$) such that

$$\Delta_{i,k+1} \leq \min\left[1, \frac{\gamma_0\kappa_Q\kappa_b\kappa_c(1-\eta_0)}{6(2+\kappa_Q)\max[\kappa_h, \kappa_{ay}]}\right]. \tag{20}$$

Then we have from the second item of Step 6: $\gamma_0\Delta_{i,k} \leq \Delta_{i,k+1}$ and hence

$$\Delta_{i,k} \leq \min\left[1, \frac{\kappa_Q\kappa_b\kappa_c(1-\eta_1)}{6(2+\kappa_Q)\max[\kappa_h, \kappa_{ay}]}\right].$$

The assumption on $\|\nabla h_{i,k}\| \geq \kappa_c$ implies that (19) holds and thus that kth is successful and $\Delta_{i,k+1} \geq \Delta_{i,k}$ satisfied. But this contradicts the fact that iteration is the first such that (20) holds, and initial assumption is therefore impossible and we get

$$\kappa_d = \gamma_0 \min\left[1, \frac{\kappa_Q \kappa_b \kappa_c (1-\eta_0)}{6(2+\kappa_Q)\max[\kappa_h, \kappa_{ay}]}\right].$$

□

Theorem 4.4. *We assume that A.1 - A.3 hold and there are only finitely many successful iterations. Then $x_{i,k} = x_{i,*}$ for k sufficiently large and $\nabla f_i(x_{i,*}) = 0$.*

Proof. For the details of the proof we refer to [5]

□

The next lemma shows that the trust-region radius converges to zero for DFRMTR as for the DFO in Lemma 10.9 in [6].

Lemma 4.4.

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

Proof. Assume that $\mathcal{S}$ is finite. We consider iterations after the last successful iteration. We know that there can be only a finite number of successful iterations before the model becomes fully linear and, hence there is an infinite number of iterations that are acceptable or unsuccessful and in either case the trust-region radius is reduced.

Since there are no more successful iterations, Δ_k is never increased for sufficiently large k. Moreover, Δ_k is decreased at least once every N iterations by a factor γ. Thus, Δ_k converges to zero.

Now let's consider the case when $\mathcal{S}$ is infinite. For any $k \in \mathcal{S}$ we have

$$f_i(x_k) - f_i(x_{k+1}) \geq \eta_0 \left[h_{i,k}(x_k) - h_{i,k}(x_k + s_k)\right].$$

Using the Cauchy decrease condition, we get

$$f_i(x_k) - f_i(x_{k+1}) \geq \eta_0 \frac{\kappa_Q \kappa_b}{2+\kappa_Q} \|\nabla h_{i,k}\| \min\left[\frac{\kappa_Q \|\nabla h_{i,k}\|}{(2+\kappa_Q)(1+\|H_i\|)}, \Delta_{i,k}\right].$$

Using (18) and the condition (5), and assuming $\|\nabla h_{i,k}\| \geq \frac{\epsilon_Q}{2}$, we obtain

$$\begin{aligned} f_i(x_k) - f_i(x_{k+1}) &\geq \eta_0 \tfrac{\kappa_Q \kappa_b}{2+\kappa_Q} \tfrac{\epsilon_Q}{2} \min\left[\tfrac{\kappa_Q \epsilon_Q}{2(1+\kappa_Q)(1+\|H_i\|)}, \Delta_{i,k}\right] \\ &\geq \eta_0 \tfrac{\kappa_Q \kappa_b}{2+\kappa_Q} \tfrac{\epsilon_Q}{2} \min\left[\tfrac{\kappa_Q \epsilon_Q}{2(1+\kappa_Q)\kappa_2}, \Delta_{i,k}\right], \end{aligned}$$

where $1 + \|H_i\| \leq \kappa_2$.

Since $\mathcal{S}$ is infinite and f is bounded from below, the right-hand side of the expression above converges to zero. Hence, $\lim\limits_{k\in\mathcal{S}} \Delta_k = 0$, and the proof is completed if all iterations are successful.

Recall that the trust-region radius can be increased only during a successful iteration, and it can be increased only by a ratio of at most γ_2 which is a constant in Step 6. If $k \notin \mathcal{S}$ be the index of an iteration after the first successful one, then $\Delta_k \leq \gamma_2 \Delta_{s_k}$, where s_k is the index of the last successful iteration before k. Since $\Delta_{s_k} \to 0$, then $\Delta_k \to 0$ for $k \notin \mathcal{S}$ and $k \to \infty$. □

Theorem 4.5. *In case of infinitely many successful iterations under assumptions A.1 - A.3,*

$$\lim_{k\to\infty} \inf \|\nabla h_{i,k}\| = 0$$

holds with $h_{i,k} = h_{i,k}(x_{i,k})$.

Proof. For the details of the proof we refer to [5, 6].

□

Lemma 4.5. *Provided the assumptions A.1 - A.3 hold and* $\{k_i\}$ *is a subsequence such that*

$$\lim_{i\to\infty} \|\nabla h_{i,k_i}(x_{i,k_i})\| = 0, \tag{21}$$

then

$$\lim_{i\to\infty} \|\nabla f_i(x_{i,k_i})\| = 0. \tag{22}$$

Proof. For the details of the proof we refer to [5, 6, 16]. □

The following two theorems are similar to those in the convergence of DFO, therefore the proofs are omitted.

Theorem 4.6. *Under the assumptions A.1 - A.3, there is at least one subsequence of successful iteration* $\{x_{i,k}\}$ *whose limit is a critical point*

$$\lim_{k\to\infty} \inf \|\nabla f_i(x_{i,k})\| = 0$$

Theorem 4.7. *Provided that the Assumptions A.1 - A.3 hold, then every limit point* $x_{i,*}$ *of the sequence* $\{x_{i,k}\}$ *is a critical point*

$$\nabla f_i(x_{i,*}) = 0.$$

We have investigated convergence of the model (4) in ith level at any iteration k. When after kth step or at the last iteration of ith level the Taylor model is chosen, then DFO convergence can be applied directly.

Since the Cauchy decrease condition can not be applied to nonlinear models, an equivalent condition is necessary for more general models. One way is to use a *backtracking algorithm* along the model steepest descent direction, where the backtracking is suggested from the boundary of the trust-region [6, 16]. We assume $Q_k(x_k + s)$ is not a quadratic function in s and we choose the smallest $j \geq 0$ such that

$$x_{k+1} = x_k + \beta^j s, \quad \text{where } s = -\frac{\Delta_k}{\|g_k\|} g_k \text{ and } \beta \in (0,1)\,. \tag{23}$$

By the way, a sufficient decrease takes the form

$$Q_k(x_{k+1}) \leq Q_k(x_k) + \kappa_c \beta^j s^T g_k, \qquad \kappa_c \in (0,1)\,, \tag{24}$$

which can, using (23), (24), equivalently be written as:

$$Q_k(x_{k+1}) - Q_k(x_k) \leq -\kappa_c \beta^j \Delta_k \|g_k\|\,. \tag{25}$$

Taylor expansion of the left hand side gives

$$-\beta^j \Delta_k \|g_k\| + \frac{1}{2}\frac{\beta^{2j}\Delta_k^2 g_k^T \nabla^2 Q_k\left(y_{k,j}\right) g_k}{\|g_k\|^2} \leq -\kappa_c \beta^j \Delta_k \|g_k\|\,,$$

for some $y_{k,j} \in \left[x_k, x_k + \beta^j s\right]$.

If $\left\|\nabla^2 Q_k(y_{k,j})\right\| \leq \kappa_{bhm}$ is assumed, then (25) is satisfied, when $\beta^j \Delta_k/||g_k|| \leq 2(1 - \kappa_c)/\kappa_{bhm}$. Thus, a j_k satisfying (25) can be found such that $\beta^{j_k} > \frac{2(1-\kappa_c)\beta\|g_k\|}{(\kappa_{bhm}\Delta_k)}$.

When $s_k^{AC} = \beta^{j_k} s$ is defined as the approximate Cauchy step, we obtain

$$Q_k(x_k) - Q_k(x_k + s_k^{AC}) \geq \kappa_c \beta^{j_k} \Delta_k \|g_k\|\,.$$

On the other hand, if the approximate Cauchy step is on the boundary, it can be derived from (4.20) that the decrease in the model exceeds or is equal to $\kappa_c \Delta_k \|g_k\|$, and

$$Q_k(x_k) - Q_k(x_k + s_k^{AC}) \geq \bar{\kappa}_c \|g_k\| \min \left\{ \frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k \right\}$$

for a suitably defined $\bar{\kappa}_c > 0$.

References

[1] Berghen, F.V. and Bersini, H., (2005), CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm, Journal of Computational and Applied Mathematics, 181, pp. 157-175.

[2] Borzi, A. and Schulz, V., (2009), Multigrid methods for PDE optimization, SIAM Review, 51, pp. 361-395.

[3] Conn, A.R. and Toint, Ph. L., (1996), An algorithm using quadratic interpolation for unconstrained derivative free optimization, in "Nonlinear Optimization and Applications", G. Di Pillo and F. Gianessi, eds, Plenum Publishing, New York, pp. 27-47.

[4] Conn, A.R., Scheinberg, K. and Toint, Ph. L, (1997), Recent progress in unconstrained nonlinear optimization without derivatives, Mathematical Programming, 79, pp. 397-414.

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

[6] Conn, A.R., Sheinberg, K. and Vicente, L.N., (2009), Introduction to Derivative-Free Optimization, SIAM Series on Optimization.

[7] Gratton, S., Sartenaer, A. and Toint, Ph. L, (2010), Numerical Experience with a recursive trust-region method for multilevel nonlinear optimization, Optimization Methods and Software, 25, pp. 359-386.

[8] Gratton, S., Sartenaer, A. and Toint, Ph. L, (2006), Second-order convergence properties of trustregion methods using incomplete curvature information, with an application to multigrid optimization, Journal of Computational and Applied Mathematics, 24, pp. 676-692.

[9] Gratton, S., Sartenaer, A. and Toint, Ph. L, (2008), Recursive Trust-Region Methods for Multiscale Nonlinear Optimization, SIAM Journal on Optimization, 19, pp. 414-444.

[10] Karasözen, B., (2007), Survey of Trust-Region Derivative Free Optimization Methods, Journal of Industrial and Management Optimization, 3, pp. 321-334.

[11] Lewis, R.M., and Nash, S.G., (2005), Model problems for the multigrid optimization of systems governed by differential equations, SIAM J. Sci. Comput., 26, pp. 18111837.

[12] Moré, J.J. and Sorenson, D.C., (1979), On the use of directions of negative curvature in a modified Newton Method. Mathematical Programming, 16, pp. 1-20.

[13] Nash, S.G., (2000), A multigrid approach to discretized optimization problems, Optim. Methods Softw., 14, pp. 99116.

[14] Powell, M.J.D., (2002), UOBYQA: Unconstrained Optimization By Quadratic Approximation, Mathematical Programming, 92, pp. 555-583.

[15] Terlaky, T., (2004), Algorithms for Continuous Optimization DFO-trust region Interpolation Algorithm. Computer and Software, McMaster University, Hamilton, Canada.

[16] Toint, Ph. L., Conn, A.R. and Gould, N.I.M., (2009), Trust-Region Methods, SIAM.

[17] Toint, Ph. L., Tomanos, D. and Mendonça, M. W., (2009), A multilevel algorithm for solving the trust-region subproblem. Optimization Methods and Software, 24, pp. 299-311.

Bülent Karasözen for the photography and short autobiography, see TWMS J. App. Eng. Math., V.1, N.2.
