Worst-case evaluation complexity of a derivative-free quadratic regularization method

Geovani Nunes Grapiglia*

July 21, 2022

Abstract

This short paper presents a derivative-free quadratic regularization method for unconstrained minimization of a smooth function with Lipschitz continuous gradient. At each iteration, trial points are computed by minimizing a quadratic regularization of a local model of the objective function. The models are based on forward finite-difference gradient approximations. By using a suitable acceptance condition for the trial points, the accuracy of the gradient approximations is dynamically adjusted as a function of the regularization parameter used to control the step-sizes. Worst-case evaluation complexity bounds are established for the new method. Specifically, for nonconvex problems, it is shown that the proposed method needs at most $\mathcal{O}\left(n\epsilon^{-2}\right)$ function evaluations to generate an ϵ-approximate stationary point, where n is the problem dimension. For convex problems, an evaluation complexity bound of $\mathcal{O}\left(n\epsilon^{-1}\right)$ is obtained, which is reduced to $\mathcal{O}\left(n\log(\epsilon^{-1})\right)$ under strong convexity. Numerical results illustrating the performance of the proposed method are also reported.

Keywords: derivative-free optimization; black-box optimization; zeroth-order optimization; worst-case complexity

1 Introduction

In many practical optimization problems, the gradients of the functions involved are not readily available. Examples include computer-aided molecular design problems [20], aerodynamic shape optimization [10], tuning of algorithmic parameters [2], model calibration [19], and optimization of cardiovascular geometries [13, 18]. These problems can be addressed with Derivative-Free Optimization (DFO) methods, i.e., methods that rely only on function evaluations (see. e.g., [5, 3, 12]). Very often, the evaluation of the objective function is computationally expensive. Therefore, one of the main concerns in DFO is the development of methods with a low worst-case complexity in terms of function evaluations. In [9], a derivative-free quadratic regularization method based on forward finite-difference gradient approximations has been proposed for the unconstrained minimization of a function $f:\mathbb{R}^n\to\mathbb{R}$ (potentially nonconvex) with Lipschitz continuous gradient. At its kth iteration, this method builds a forward finite-difference gradient approximation g_k aiming an error bound of the form

$$\|g_k-\nabla f(x_k)\|\leq\kappa_g\|x_k-x_{k-1}\|,$$

*Université catholique de Louvain, ICTEAM/INMA, Avenue Georges Lemaître, 4-6/ L4.05.01, B-1348, Louvain-la-Neuve, Belgium (geovani.grapiglia@uclouvain.be).

where $\{x_k\}$ is the sequence of iterates and $\kappa_g > 0$ is a certain constant. It was shown that the referred method needs at most $\mathcal{O}\left(n\epsilon^{-2}\right)$ function evaluations to find an ϵ-approximate stationary point.

In the present paper, a new derivative-free quadratic regularization method is presented. At its kth iteration, a trial point is computed by minimizing a quadratic regularization of a local model of the objective function. This model is also defined by a forward finite-difference gradient approximation g_k, but, in contrast to [9], the new method builds g_k aiming an error bound of the form

$$\|g_k - \nabla f(x_k)\| \leq \kappa_g \epsilon, \tag{1}$$

assuming that the goal is to find $\bar{x}$ such that $\|\nabla f(\bar{x})\| \leq \epsilon$. By using an acceptance condition for the trial points derived from (1), the accuracy of the gradient approximations is dynamically adjusted as a function of the regularization parameter. It is shown that the proposed method needs at most $\mathcal{O}\left(n\epsilon^{-2}\right)$ to find an ϵ-approximate stationary point when the objective function is nonconvex. In terms of n and ϵ, this bound agrees with the bound established in [9]. However, the use of (1) allows the derivation of additional complexity bounds under convexity. For convex functions, it is shown that the new method needs at most $\mathcal{O}\left(n\epsilon^{-1}\right)$ function evaluations to find an ϵ-approximate stationary point, while for strongly convex functions, a bound of $\mathcal{O}\left(n\log(\epsilon^{-1})\right)$ is obtained. To the best of this author's knowledge, this is the first time that evaluation complexity bounds with linear dependence in n are obtained for a *deterministic* DFO method in the context of convex and strongly convex objective functions with Lipschitz continuous gradients[1].

The paper is organized as follows. Section 2 contains the main preliminary results. In Section 3, the new method is described and its worst-case complexity is analyzed. Finally, in Section 4, preliminary numerical results are reported.

2 Auxiliary Results

The problem class considered in this work is specified by the following assumptions:

A1. The gradient of f is L-Lipschitz continuous, i.e.,

$$\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|, \quad \forall x, y \in \mathbb{R}^n.$$

A2. There exists $f_{low} \in \mathbb{R}$ such that $f(x) \geq f_{low}$ for all $x \in \mathbb{R}^n$.

In the proposed method, given $x \in \mathbb{R}^n$, trial points are computed by (approximately) minimizing quadratic models of the form

$$M_{x,\sigma}(y) := f(x) + \langle g, y - x\rangle + \frac{1}{2}\langle B(y - x), y - x\rangle + \frac{\sigma}{2}\|y - x\|^2 \tag{2}$$

where $g \in \mathbb{R}^n$ is an approximation to $\nabla f(x)$, $B \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix, and $\sigma > 0$ is a regularization parameter.

The next lemma gives sufficient conditions under which an approximate minimizer x^+ of $M_{x,\sigma}(\,\cdot\,)$ yields a decrease in the objective function that is at least of $\mathcal{O}\left(\|x^+ - x\|^2\right)$.

[1]Evaluation complexity bounds of $\mathcal{O}\left(n\epsilon^{-1}\right)$ and $\mathcal{O}\left(n\log(\epsilon^{-1})\right)$ (in the convex and strongly convex cases, respectively) were established in [16] and [6] for *randomized* DFO methods. They constitute upper bounds for the number of function evaluations that the corresponding methods need to find $\bar{x}$ such that $E[f(\bar{x})] - f^* \leq \epsilon$, where f^* is the optimal value of $f(\,\cdot\,)$ and $E[X]$ denotes the expected value of a random variable X. For deterministic direct search methods, bounds of $\mathcal{O}\left(n^2\epsilon^{-1}\right)$ and $\mathcal{O}\left(n^2\log(\epsilon^{-1})\right)$ were established in [7] and [11].

Lemma 2.1. *Suppose that A1 holds. Given $\epsilon > 0$, let $x \in \mathbb{R}^n$ and $g \in \mathbb{R}^n$ such that*

$$\|\nabla f(x)\| > \epsilon \tag{3}$$

and

$$\|g - \nabla f(x)\| \leq \frac{\epsilon}{5}. \tag{4}$$

Moreover, let $x^+ \in \mathbb{R}^n$ be a point such that

$$M_{x,\sigma}(x^+) \leq f(x) \quad \text{and} \quad \|\nabla M_{x,\sigma}(x^+)\| \leq \theta\sigma\|x^+ - x\|, \tag{5}$$

for some $\sigma > 0$ and $\theta \in [0, 1)$. If

$$\sigma \geq 2\left[2L + 3\|B\|\right](1 - \theta)^{-1}, \tag{6}$$

then

$$f(x) - f(x^+) \geq \frac{(1-\theta)\sigma}{8}\|x^+ - x\|^2. \tag{7}$$

Proof. Using A1, the Cauchy-Schwarz inequality, the first inequality in (5) and (4), we get

$$\begin{aligned}
f(x^+) &\leq f(x) + \langle \nabla f(x), x^+ - x\rangle + \frac{L}{2}\|x^+ - x\|^2 \\
&= f(x) + \langle g, x^+ - x\rangle + \frac{1}{2}\langle B(x^+ - x), x^+ - x\rangle + \frac{\sigma}{2}\|x^+ - x\|^2 \\
&\quad + \langle \nabla f(x) - g, x^+ - x\rangle - \frac{1}{2}\langle B(x^+ - x), x^+ - x\rangle + \frac{(L-\sigma)}{2}\|x^+ - x\|^2 \\
&= M_{x,\sigma}(x^+) + \langle \nabla f(x) - g, x^+ - x\rangle - \frac{1}{2}\langle B(x^+ - x), x^+ - x\rangle \\
&\quad + \frac{(L-\sigma)}{2}\|x^+ - x\|^2 \\
&\leq f(x) + \|\nabla f(x) - g\|\|x^+ - x\| + \frac{(\|B\| + L - \sigma)}{2}\|x^+ - x\|^2 \\
&\leq f(x) + \frac{\epsilon}{5}\|x^+ - x\| + \frac{(\|B\| + L - \sigma)}{2}\|x^+ - x\|^2. \qquad (8)
\end{aligned}$$

By (3) and (4) we have

$$\epsilon < \|\nabla f(x)\| \leq \|\nabla f(x) - g\| + \|g\| \leq \frac{\epsilon}{5} + \|g\|,$$

which implies that

$$\frac{\epsilon}{5} \leq \frac{\|g\|}{4}. \tag{9}$$

Moreover, by (2) and the second inequality in (5) we also have

$$\begin{aligned}
\|g\| &\leq \|\nabla M_{x,\sigma}(x^+)\| + \|g - \nabla M_{x,\sigma}(x^+)\| \\
&= \|\nabla M_{x,\sigma}(x^+)\| + \|(B + \sigma I)(x^+ - x)\| \\
&\leq \left[(\theta + 1)\sigma + \|B\|\right]\|x^+ - x\|. \qquad (10)
\end{aligned}$$

Now, combining (8), (9) and (10) it follows that

$$\begin{aligned} f(x^+) &\leq f(x) + \frac{[(\theta+1)\sigma + \|B\|]}{4}\|x^+ - x\|^2 + \frac{(\|B\| + L - \sigma)}{2}\|x^+ - x\|^2 \\ &= f(x) + \frac{[(\theta+1)\sigma + 3\|B\| + 2L - 2\sigma]}{4}\|x^+ - x\|^2 \\ &= f(x) + \frac{[(2L + 3\|B\|) - (1-\theta)\sigma]}{4}\|x^+ - x\|^2. \end{aligned} \tag{11}$$

Finally, by (6) and (11) we conclude that

$$f(x) - f(x^+) \geq \frac{[(1-\theta)\sigma - (2L + 3\|B\|)]}{4}\|x^+ - x\|^2 \geq \frac{(1-\theta)\sigma}{8}\|x^+ - x\|^2.$$

□

The next lemma suggests a way to construct g such that (4) holds for σ sufficiently large.

Lemma 2.2. *Suppose that A1 holds and assume that x^+ satisfies (5) for some $x \in \mathbb{R}^n$, $\sigma > 0$ and $\theta \in [0,1)$. Moreover, suppose that x satisfies (3) for some $\epsilon > 0$, and that the vector g in $M_{x,\sigma}(\,\cdot\,)$ is defined by*

$$g_j = \frac{f(x + he_j) - f(x)}{h}, \quad j = 1, \ldots, n. \tag{12}$$

with

$$0 < h \leq \frac{2\epsilon}{5\sigma\sqrt{n}}. \tag{13}$$

If σ satisfies (6) then the point x^+ satisfies (7). Moreover,

$$\|g\| \geq \frac{4\epsilon}{5}. \tag{14}$$

Proof. By A1, (12), (13) and (6) we have

$$\|\nabla f(x) - g\| \leq \frac{\sqrt{n}L}{2}h \leq \frac{L\epsilon}{5\sigma} \leq \frac{\epsilon}{5}. \tag{15}$$

Then, in view of (3), (15), (5) and (6), it follows from Lemma 2.1 that x^+ satisfies (7). Finally, assume by contradiction that (14) is not true, i.e.,

$$\|g\| < \frac{4\epsilon}{5}. \tag{16}$$

In this case, combining (15) and (16) we would have

$$\|\nabla f(x)\| \leq \|\nabla f(x) - g\| + \|g\| < \frac{\epsilon}{5} + \frac{4\epsilon}{5} = \epsilon,$$

which contradicts (3). Thus, (14) also must be true. □

3 Derivative-Free Quadratic Regularization Method

Consider now the following Derivative-Free Quadratic Regularization Method (DFQRM):

Algorithm 1. DFQRM

Step 0. Given $x_0 \in \mathbb{R}^n$, a symmetric positive semidefinite matrix $B_0 \in \mathbb{R}^{n\times n}$, $\sigma_0 > 0$, $\epsilon > 0$, and $\theta \in [0,1)$, set $k := 0$.

Step 1. Find the smallest integer $i \geq 0$ such that $2^i\sigma_k \geq 2\sigma_0$.

Step 1.1. For

$$h_i = \frac{2\epsilon}{5(2^i\sigma_k)\sqrt{n}}, \tag{17}$$

compute $g_{k,i} \in \mathbb{R}^n$ by

$$[g_{k,i}]_j = \frac{f(x_k + h_i e_j) - f(x_k)}{h_i}, \quad j = 1, \ldots, n. \tag{18}$$

Step 1.2. If

$$\|g_{k,i}\| \geq \frac{4\epsilon}{5} \tag{19}$$

go to Step 1.3. Otherwise, set $i := i + 1$ and go to Step 1.1.

Step 1.3. Consider the quadratic model

$$M_{x_k,2^i\sigma_k}(y) := f(x_k) + \langle g_{k,i}, y - x_k\rangle + \frac{1}{2}\langle B_k(y - x_k), y - x_k\rangle + \frac{2^i\sigma_k}{2}\|y - x_k\|^2,$$

and compute an approximate solution $x_{k,i}^+$ of the subproblem

$$\min_{y\in\mathbb{R}^n} M_{x_k,2^i\sigma_k}(y), \tag{20}$$

such that

$$M_{x_k,2^i\sigma_k}(x_{k,i}^+) \leq f(x_k) \quad \text{and} \quad \|\nabla M_{x_k,2^i\sigma_k}(x_{k,i}^+)\| \leq \theta(2^i\sigma_k)\|x_{k,i}^+ - x_k\|. \tag{21}$$

Step 1.4. If

$$f(x_k) - f(x_{k,i}^+) \geq \frac{(1-\theta)(2^i\sigma_k)}{8}\|x_{k,i}^+ - x_k\|^2 \tag{22}$$

holds, set $i_k = i$, $g_k = g_{k,i_k}$ and go to Step 2. Otherwise, set $i := i + 1$ and go to Step 1.1.

Step 2. Set $x_{k+1} = x_{k,i_k}^+$, $\sigma_{k+1} = 2^{i_k-1}\sigma_k$, choose a symmetric positive semidefinite matrix $B_{k+1} \in \mathbb{R}^{n\times n}$, set $k := k + 1$, and go to Step 1.

The analysis of Algorithm 1 will be carried out with the following additional assumption:

A3. There exists $M \geq 0$ such that $\|B_k\| \leq M$ for all k.

The next lemma gives a lower bound and an upper bound for the sequence of regularization parameters.

Lemma 3.1. *Suppose that A1 and A3 hold and let $\{x_k\}$ be a sequence generated by Algorithm 1 such that*

$$\|\nabla f(x_k)\| > \epsilon, \quad \textit{for} \quad k = 0, \ldots, T. \tag{23}$$

Then, the sequence of regularization parameters $\{\sigma_k\}$ *in Algorithm 1 satisfies*

$$\sigma_0 \leq \sigma_k \leq 2 \max\left\{\sigma_0, [2L + 3M](1-\theta)^{-1}\right\} := \sigma_{\max}, \tag{24}$$

for all $k \in \{0, \ldots, T+1\}$.

Proof. Notice that (24) holds for $k = 0$. Assume that (24) is true for some $k \in \{0, \ldots, T\}$. It follows from Step 1 of Algorithm 1 that

$$\sigma_{k+1} = 2^{i_k - 1}\sigma_k = \frac{1}{2}\left(2^{i_k}\sigma_k\right) \geq \frac{1}{2}(2\sigma_0) = \sigma_0. \tag{25}$$

Moreover, we also have

$$\sigma_{k+1} \leq \sigma_{\max}. \tag{26}$$

Indeed, if $i_k = 0$, it follows from the induction assumption that

$$\sigma_{k+1} = 2^{i_k - 1}\sigma_k = \frac{1}{2}\sigma_k < \sigma_k \leq \sigma_{\max}.$$

Suppose that $i_k \geq 1$. Then, assuming that (26) is false we would have

$$2^{i_k-1}\sigma_k = \sigma_{k+1} > 2\sigma_0 \quad \text{and} \quad 2^{i_k-1}\sigma_k = \sigma_{k+1} > 2\left[2L + 3\|B\|\right](1-\theta)^{-1},$$

where the last inequality is due to A3. In this case, by Lemma 2.2, inequality (22) would have been satisfied for $i \leq i_k - 1$, contradicting the definition of i_k. Thus, in view of (25) and (26), we have

$$\sigma_0 \leq \sigma_{k+1} \leq \sigma_{\max},$$

which concludes the proof. □

The next lemma establishes a lower bound of $\mathcal{O}\left(\|\nabla f(x_k)\|^2\right)$ for the difference $f(x_k) - f(x_{k+1})$.

Lemma 3.2. *Suppose that A1 and A3 hold and let* $\{x_k\}$ *be a sequence generated by Algorithm 1 such that (23) holds for some* $T \geq 1$. *Then*

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2C_f}\|\nabla f(x_k)\|^2, \quad \textit{for} \quad k = 0, \ldots, T-1, \tag{27}$$

where

$$C_f := \frac{81\left[2(\theta+1) + M\sigma_0^{-1}\right]^2}{32(1-\theta)} \max\left\{\sigma_{\max}, L^2\sigma_0^{-1}\right\}, \tag{28}$$

where $\sigma_{\max}$ *is defined in (24).*

Proof. Given $k \in \{0, \ldots, T-1\}$, it follows from (17), (18), A1 and (19) that

$$\|\nabla f(x_k) - g_k\| \leq \frac{\sqrt{n}L}{2} h_{i_k} = \frac{L\epsilon}{5(2^{i_k}\sigma_k)} \leq \frac{L}{8\sigma_{k+1}}\|g_k\|.$$

Consequently,

$$\|\nabla f(x_k)\| \leq \|\nabla f(x_k) - g_k\| + \|g_k\| \leq \left(\frac{L + 8\sigma_{k+1}}{8\sigma_{k+1}}\right)\|g_k\|.$$

Thus,

$$\left(\frac{8\sigma_{k+1}}{L+8\sigma_{k+1}}\right)\|\nabla f(x_k)\| \leq \|g_k\|. \tag{29}$$

In view of Lemma 3.1 we have

$$M = \left(M\sigma_0^{-1}\right)\sigma_0 \leq \left(M\sigma_0^{-1}\right)\sigma_{k+1}. \tag{30}$$

Then, combining the second inequality in (21) with (30) and A3, we obtain

$$\begin{aligned}
\|g_k\| &\leq \|g_k + (B_k + 2^{i_k}\sigma_k I)(x_{k+1}-x_k)\| + \|(B_k + 2^{i_k}\sigma_k I)(x_{k+1}-x_k)\| \\
&= \|\nabla M_{x_k,2\sigma_{k+1}}(x_{k+1})\| + (\|B\| + 2\sigma_{k+1})\|x_{k+1}-x_k\| \\
&\leq (2\theta\sigma_{k+1} + 2\sigma_{k+1} + M)\|x_{k+1}-x_k\| \\
&\leq \left[2(\theta+1) + M\sigma_0^{-1}\right]\sigma_{k+1}\|x_{k+1}-x_k\|.
\end{aligned} \tag{31}$$

Combining (29) and (31) it follows that

$$\|x_{k+1}-x_k\| \geq \frac{8\|\nabla f(x_k)\|}{\left[2(\theta+1)+M\sigma_0^{-1}\right](L+8\sigma_{k+1})}. \tag{32}$$

Suppose that $\sigma_{k+1} \geq L$. By (22), (32), Lemma 3.1 and (28), we get

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\geq \frac{2(1-\theta)\sigma_{k+1}}{8}\|x_{k+1}-x_k\|^2 \\
&\geq \frac{2(1-\theta)\sigma_{k+1}}{8}\left(\frac{8^2\|\nabla f(x_k)\|^2}{\left[2(\theta+1)+M\sigma_0^{-1}\right]^2 9^2\sigma_{k+1}^2}\right) \\
&\geq \frac{16(1-\theta)}{81\left[2(\theta+1)+M\sigma_0^{-1}\right]^2\sigma_{\max}}\|\nabla f(x_k)\|^2 \\
&\geq \frac{1}{2C_f}\|\nabla f(x_k)\|^2,
\end{aligned}$$

that is, (27) holds. Now, suppose that $\sigma_{k+1} < L$. In this case, it follows from (22), (32), Lemma 3.1 and (28) that

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\geq \frac{2(1-\theta)\sigma_{k+1}}{8}\|x_{k+1}-x_k\|^2 \\
&\geq \frac{2(1-\theta)\sigma_{k+1}}{8}\left(\frac{8^2\|\nabla f(x_k)\|^2}{\left[2(\theta+1)+M\sigma_0^{-1}\right]^2 9^2 L^2}\right) \\
&\geq \frac{16(1-\theta)\sigma_0}{81\left[2(\theta+1)+M\sigma_0^{-1}\right]^2 L^2}\|\nabla f(x_k)\|^2 \\
&\geq \frac{1}{2C_f}\|\nabla f(x_k)\|^2,
\end{aligned}$$

that is, (27) also holds. □

In view of Lemma 3.2, complexity bounds can be obtained for Algorithm 1 under different scenarios. The next theorem gives an iteration complexity bound of $\mathcal{O}\left(\epsilon^{-2}\right)$ for Algorithm 1 applied to a nonconvex problem.

Theorem 3.3. *Suppose that A1-A3 hold and let $\{x_k\}$ be a sequence generated by Algorithm 1 such that (23) holds for some $T \geq 1$. Then*

$$\min_{k=0,\dots,T-1} \|\nabla f(x_k)\| \leq \frac{[2C_f(f(x_0) - f_{low})]^{\frac{1}{2}}}{\sqrt{T}}, \tag{33}$$

where C_f is defined in (28). Consequently,

$$T < 2C_f(f(x_0) - f_{low})\epsilon^{-2}. \tag{34}$$

Proof. By Lemma 3.2,

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2C_f}\|\nabla f(x_k)\|^2, \quad \text{for} \quad k = 0, \dots, T-1. \tag{35}$$

Summing up these inequalities and using A2 we get

$$\frac{1}{2C_f}\sum_{k=0}^{T-1}\|\nabla f(x_k)\|^2 \leq f(x_0) - f_{low},$$

which implies that (33) holds. Finally, combining (33) and (23) we get the bound (34).

□

Consider the additional assumption:

A4. $f : \mathbb{R}^n \to \mathbb{R}$ is convex and the sublevel set $\mathcal{L}_f(x_0) = \{x \in \mathbb{R}^n \, : \, f(x) \leq f(x_0)\}$ is compact.

The next theorem gives an iteration complexity bound of $\mathcal{O}\left(\epsilon^{-1}\right)$ for Algorithm 1 applied to a convex problem.

Theorem 3.4. *Suppose that A1, A3 and A4 holds and let $\{x_k\}$ be a sequence generated by Algorithm 1 such that (23) holds for $T = 3s$, with $s \geq 1$. Then*

$$\min_{k=0,\dots,T-1} \|\nabla f(x_k)\| \leq \frac{\sqrt{18}C_f D_0}{T}, \tag{36}$$

where C_f is defined in (28) and

$$D_0 = \sup_{x \in \mathcal{L}_f(x_0)} \|x - x^*\|, \tag{37}$$

with x^* being a minimizer of $f(\,\cdot\,)$. Consequently,

$$T < \sqrt{18}C_f D_0 \epsilon^{-1}. \tag{38}$$

Proof. Since $\mathcal{L}_f(x_0)$ is compact (by A4), it follows that $f(\,\cdot\,)$ has a minimizer x^* and $D_0 < +\infty$. In view of (35), we have $x_k \in \mathcal{L}_f(x_0)$ for all $k \in \{0, \dots, T\}$, and so

$$\|x_k - x^*\| \leq D_0, \quad \forall k \in \{0, \dots, T\}.$$

Then, it follows from the convexity of $f(\,\cdot\,)$ that

$$\|\nabla f(x_k)\| \geq \frac{f(x_k) - f(x^*)}{D_0}, \quad k = 0, \dots, T. \tag{39}$$

Combining (35) and (39) we obtain

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2C_f D_0^2}(f(x_k) - f(x^*))^2, \quad k = 0, \ldots, T-1. \tag{40}$$

Define

$$\delta_k = \frac{1}{2C_f D_0^2}(f(x_k) - f(x^*)).$$

Then, given $j \in \{1, \ldots, T-1\}$, it follows from (40) that

$$\delta_k - \delta_{k+1} \geq \delta_k^2, \quad k = 0, \ldots, j,$$

and so

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} \geq \frac{\delta_k - \delta_{k+1}}{\delta_{k+1}\delta_k} \geq \frac{\delta_k^2}{\delta_k^2} = 1, \quad k = 0, \ldots, j-1.$$

Summing up these inequalities we get

$$\frac{1}{\delta_j} - \frac{1}{\delta_0} \geq j,$$

which gives $\delta_j \leq 1/j$. Thus,

$$f(x_j) - f(x^*) \leq \frac{2C_f D_0^2}{j}, \quad \forall j \in \{1, \ldots, T-1\}. \tag{41}$$

In particular, for $j = 2s$, it follows from (41), (35) and $T = 3s$ that

$$\begin{aligned}
\frac{2C_f D_0^2}{2s} &\geq f(x_{2s}) - f(x^*) = f(x_T) - f(x^*) + \sum_{k=2s}^{T-1} f(x_k) - f(x_{k+1}) \\
&\geq \sum_{k=2s}^{T-1} f(x_k) - f(x_{k+1}) \geq \frac{1}{2C_f} \sum_{k=2s}^{T-1} \|\nabla f(x_k)\|^2 \\
&\geq \frac{s}{2C_f} \min_{k=0,\ldots,T-1} \|\nabla f(x_k)\|^2.
\end{aligned}$$

Therefore, using $s = T/3$ it follows that

$$\min_{k=0,\ldots,T-1} \|\nabla f(x_k)\|^2 \leq \frac{2C_f^2 D_0^2}{s^2} = \frac{18 C_f^2 D_0^2}{T^2},$$

which gives (36). Finally, combining (36) and (23) we get the bound (38). □

Now, instead of A4, consider the assumption:

A5. $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex.

The next theorem gives an iteration complexity bound of $\mathcal{O}\left(\log(\epsilon^{-1})\right)$ for Algorithm 1 applied to a strongly convex function.

Theorem 3.5. *Suppose that A1, A3 and A5 hold, and let $\{x_k\}$ be a sequence generated by Algorithm 1 such that (23) holds for some $T \geq 1$. Then*

$$\|\nabla f(x_k)\| \leq \sqrt{2C_f(f(x_0) - f(x^*))} \left(1 - \frac{\mu}{C_f}\right)^{\frac{k}{2}}, \quad k = 0, \ldots, T-1, \tag{42}$$

where C_f is defined in (28) and x^* is the minimizer of $f(\cdot)$. Consequently,

$$T \leq 1 + \frac{2}{\left|\log\left(1 - \frac{\mu}{C_f}\right)\right|} \log\left(\sqrt{2C_f(f(x_0) - f(x^*))}\epsilon^{-1}\right). \tag{43}$$

Proof. Let $k \in \{0, \ldots, T-1\}$. By A5 we have

$$\|\nabla f(x_k)\|^2 \geq 2\mu(f(x_k) - f(x^*)). \tag{44}$$

Combining (35) and (44) we get

$$(f(x_k) - f(x^*)) - (f(x_{k+1}) - f(x^*)) = f(x_k) - f(x_{k+1}) \geq \frac{\mu}{C_f}\left(f(x_k) - f(x^*)\right),$$

which gives

$$f(x_{k+1}) - f(x^*) \leq \left(1 - \frac{\mu}{C_f}\right)\left(f(x_k) - f(x^*)\right).$$

Therefore,

$$f(x_k) - f(x^*) \leq \left(1 - \frac{\mu}{C_f}\right)^k (f(x_0) - f(x^*)), \quad k = 0, \ldots, T. \tag{45}$$

Now, combining (35) and (45), it follows that

$$\|\nabla f(x_k)\|^2 \leq 2C_f\left(1 - \frac{\mu}{C_f}\right)^k (f(x_0) - f(x^*)), \quad k = 0, \ldots, T-1,$$

which gives (42). Finally, combining (23) and (42) we obtain the bound (43). □

The next lemma gives an upper bound for the total number of function evaluations that Algorithm 1 performs until it finds an ϵ-approximate stationary point of the objective function.

Lemma 3.6. *Suppose that A1 and A3 hold. Let $T(\epsilon)$ be the first iteration index such that*

$$\|\nabla f(x_{T(\epsilon)+1})\| \leq \epsilon,$$

and let $FE(\epsilon)$ be the total number of function evaluations performed by Algorithm 1 up to the $T(\epsilon)$th iteration. Then,

$$FE(\epsilon) \leq 1 + (n+1)\left[2 + 2T(\epsilon) + \log_2(\sigma_{\max}) - \log_2(\sigma_0)\right],$$

where $\sigma_{\max}$ is defined in (24).

Proof. The number of function evaluations performed at the kth iteration of Algorithm 1 is bounded from above by $1+(n+1)(i_k+1)$ if $k=0$, and by $(n+1)(i_k+1)$ when $k>0$. Since $\sigma_{k+1}=2^{i_k-1}\sigma_k$, we have

$$(n+1)(i_k+1)=(n+1)\left[2+\log_2(\sigma_{k+1})-\log_2(\sigma_k)\right].$$

Therefore,

$$\begin{aligned} FE(\epsilon) &\leq 1+\sum_{k=0}^{T(\epsilon)}(n+1)(i_k+1)=1+(n+1)\left[2(T(\epsilon)+1)+\log_2(\sigma_{T(\epsilon)+1})-\log_2(\sigma_0)\right] \\ &\leq 1+(n+1)\left[2+2T(\epsilon)+\log_2(\sigma_{\max})-\log_2(\sigma_0)\right], \end{aligned}$$

where the last inequality is due to Lemma 3.1. □

The theorem below combines the previous results and establishes worst-case evaluation complexity bounds for Algorithm 1.

Theorem 3.7. *Suppose that A1 and A3 hold, and let $FE(\epsilon)$ be defined as in Lemma 3.6. Then*

$$FE(\epsilon)\leq\begin{cases}\mathcal{O}\left(n\epsilon^{-2}\right), & \text{if A2 holds (f is nonconvex),}\\ \mathcal{O}\left(n\epsilon^{-1}\right), & \text{if A4 holds (f is convex),}\\ \mathcal{O}\left(n\log\left(\epsilon^{-1}\right)\right), & \text{if A5 holds (f is strongly convex).}\end{cases} \tag{46}$$

Proof. Let $T(\epsilon)$ be the first iteration index such that $\|\nabla f(x_{T(\epsilon)+1})\|\leq\epsilon$. By Theorems 3.3, 3.4 and 3.5, we have

$$T(\epsilon)\leq\begin{cases}\mathcal{O}\left(\epsilon^{-2}\right), & \text{if A2 holds,}\\ \mathcal{O}\left(\epsilon^{-1}\right), & \text{if A4 holds,}\\ \mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right), & \text{if A5 holds.}\end{cases} \tag{47}$$

Then, combining Lemma 3.6 and (47) we get (46). □

4 Numerical Experiments

To investigate the practical performance of Algorithm 1, numerical experiments were carried out on the set of 102 problems from the Andrei's collection [1]. The following MATLAB implementations were compared:

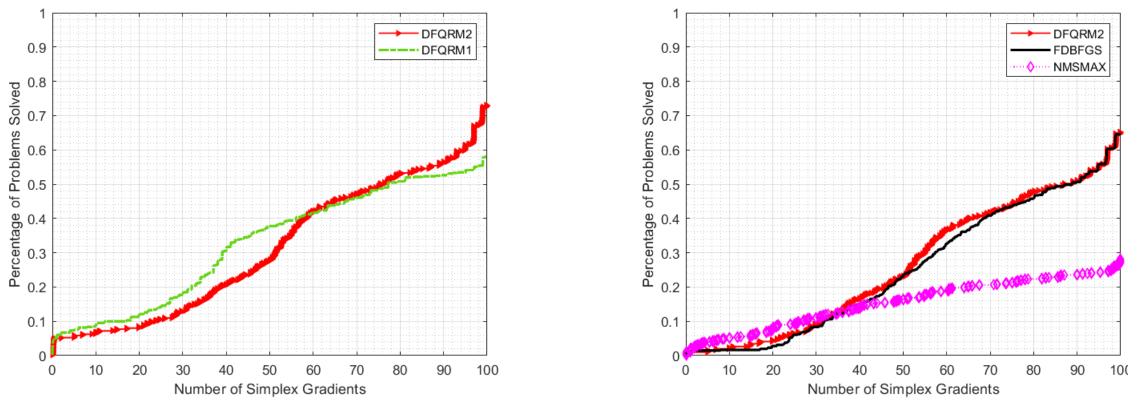
- **DFQRM1:** Algorithm 1 with $\sigma_0=10^{-2}$, $\epsilon=10^{-5}$, $\theta=0$ and $B_k=0$ for all k.
- **DFQRM2:** Algorithm 1 with $\sigma_0=10^{-2}$, $\epsilon=10^{-5}$, $\theta=0$ and B_k updated by

$$B_{k+1}=\begin{cases}B_k+\dfrac{y_ky_k^T}{s_k^Ty_k}-\dfrac{B_ks_ks_k^TB_k}{s_k^TB_ks_k}, & \text{if } s_k^Ty_k>0,\\ B_k, & \text{otherwise,}\end{cases}$$

  with $B_0=I$, $s_k=x_{k+1}-x_k$ and $y_k=g(x_{k+1})-g(x_k)$, where $g(x_k)=g_k$ and $g(x_{k+1})$ is the approximation to $\nabla f(x_{k+1})$ obtained by forward finite-differences with $h=h_{i_k}$.
- **FDBFGS:** the code described in Section 5 of [9].

- **NMSMAX:** an implemenation of the Nelder-Mead method [15][2].

For each problem, three choices of the dimension n and two choices of starting points were considered, resulting in 612 instances. Specifically, the experiments were performed with $n = 12, 24, 48$ and $x_0 = 10^s \bar{x}$, $s = 0, 1$, where $\bar{x}$ is the starting point provided in [1]. For each problem, a budget of 4900 function evaluations was allowed to each code (i.e., at least 100 simplex gradients). Figure 1 shows the corresponding data profiles [14][3]. As it can be seen, DFQRM2 (the quasi-Newton version of Algorithm 1) outperformed DFQRM1 (Fig. 1(a)). Moreover, DFQRM2 was competitive against FDBFGS, with both codes solving more problems than NMSMAX using the same budget of function evaluations (Fig. 1(b)).



(a) (b)

Figure 1: Data profiles for the precision 10^{-7} and budget of 100 simplex gradients.

5 Conclusion

This paper presented a derivative-free quadratic regularization method for smooth unconstrained optimization in which finite-difference gradient approximations are employed. The accuracy of the gradient approximations and the regularization parameter are jointly adjusted by using an acceptance condition for the trial steps that forces $\{f(x_k)\}$ to be a decreasing sequence. For the class of differentiable functions of n variables that have Lipschitz continuous gradient, evaluation complexity bounds of $\mathcal{O}\left(n\epsilon^{-2}\right)$, $\mathcal{O}\left(n\epsilon^{-1}\right)$ and $\mathcal{O}\left(n\log(\epsilon^{-1})\right)$ were proved for the nonconvex, the convex, and the strongly convex cases, respectively. Preliminary numerical results suggest that the new method compares favorably with the Nelder-Mead method, and that it is competitive with the derivative-free method recently proposed in [9].

[2]The Nelder-Mead method was included in the comparison in view of its extensive use in several areas (see, e.g., [4, 17, 8]). The implementation NMSMAX is freely available in the Matrix Computation Toolbox (`https://www.maths.manchester.ac.uk/~higham/mctoolbox/`).

[3]The data profiles were generated using the code **data_profile.m** freely available in the website `https://www.mcs.anl.gov/~more/dfo/`.

References

[1] Andrei, N.: An unconstrained optimization test functions collection. Advanced Modeling and Optimization 10, 147–161 (2008)

[2] Audet, C., Orban, D.: Finding Optimal Algorithmic Parameters Using Derivative-Free Optimization. SIAM Journal on Optimization 17, 642–664 (2006)

[3] Audet, C., Hare, W.: Derivative-free and blackbox optimization. Springer, 2017.

[4] Arridge, C.S., Eggington, J.W.B.: Electromagnetic induction in the icy satellites of Uranus. Icarus 367, 114562 (2021)

[5] Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. SIAM, 2009.

[6] Bergou, E.H., Gorbunov, E., Richtárik, P.: Stochastic three points method for unconstrained smooth minimization. SIAM Journal on Optimization 30, 2726–2749 (2020)

[7] Dodangeh, M., Vicente, L.N.: Worst-case complexity of direct search under convexity. Mathematical Programming 155, 307–332 (2016)

[8] Gankin, Y., Nemira, A., Koniukhorskii, V., Chowell, G., Kirpich, A.: Investigating the first state of the Covid-19 pandemic in Ukraine using epidemiological and genomic data. Infection, Genomics and Evolution 95, 105087 (2021)

[9] Grapiglia, G.N.: Quadratic regularization methods with finite-difference gradient approximations. Computational Optimization and Applications (2022), https://doi.org/10.1007/s10589-022-00373-z

[10] Karbasian, H.R., Vermeire, B.C.: Gradient-free aerodynamic shape optimization using Large Eddy Simulation. Computers and Fluids 232, 105185 (2022)

[11] Konecny, J., Richtárik, P.: Simple complexity analysis of simplified direct search. arXiv:1410.0390 [math.OC] (2014)

[12] Larson, J., Menickelly, M., Wild, S.M.: Derivative-Free Optimization. Acta Numerica, 287–404 (2019)

[13] Marsden, A.L., Feinstein, J.A., Taylor, C.A.: A computational framework for derivative-free optimization of cardiovascular geometries. Computational Methods in Applied Mechanics and Engineering 197, 1890–1905 (2008)

[14] Moré, J.J., Wild, S.M.: Benchmarking derivative-Free optimization algorithms. SIAM Journal on Optimization 20, 172–191 (2009)

[15] Nelder, J.A., Mead, R.: A simplex method for function minimization. The Computer Journal 7, 308–313 (1965)

[16] Nesterov, Yu., Spokoiny, V.: Random gradient-free minimization of convex functions. Foundations of Computational Mathematics 17, 527–566 (2017)

[17] Reis, L.B., Gioria, R.S.: Optimization of liquid jet ejector geometry and its impact on flow fields. Applied Thermal Engineering 194, 117132 (2021)

[18] Russ, J.B., Li, R.L., Herschman, A.R., Haim, W., Vedula, V., Kysar, J.W., Kalfa, D.: Design optimization of a cardiovascular stent with application to a balloon expandable prosthetic heart valve. Materials & Design 209 (2021)

[19] Sophy, O., Cartis, C., Kriest, I., Tett, S.F.B., Khatiwala, S.: A derivative-free optimisation method for global ocean biogeochemical models. Geoscientific Model Development 15, 3537–3554 (2022)

[20] Sun, Y., Sahinidis, N.V., Sundaram, A., Cheon, M-S.: Derivative-free optimization for chemical product design. Current Option in Chemical Engineering 27, 98–106 (2020)