Optimization Letters
https://doi.org/10.1007/s11590-023-01984-z

ORIGINAL PAPER

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

Geovani Nunes Grapiglia[1]

Received: 16 June 2022 / Accepted: 25 January 2023

Abstract
This manuscript 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 stepsizes. 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 [22], aerodynamic shape optimization [9], tuning of algorithmic parameters [3], model calibration [21], and optimization of cardiovascular geometries [13,

G. N. Grapiglia was partially supported by CNPq (Grant 312777/2020-5).

✉ Geovani Nunes Grapiglia
geovani.grapiglia@uclouvain.be

1 ICTEAM/INMA, Université catholique de Louvain, Avenue Georges Lemaître, 4-6/ L4.05.01, B-1348, Louvain-la-Neuve, Belgium

Published online: 09 February 2023

20]. These problems can be addressed with Derivative-Free Optimization (DFO) methods, i.e., methods that rely only on function evaluations (see. e.g., [2, 5, 11]). 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 [8], a derivative-free quadratic regularization method based on forward finite-difference gradient approximations has been proposed for the unconstrained minimization of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (potentially nonconvex) with Lipschitz continuous gradient. At its kth iteration, this method builds a forward finite-difference gradient approximation g_k attempting to satisfy an error bound of the form

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

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, i.e., a point $\bar{x}$ such that $\|\nabla f(\bar{x})\| \leq \epsilon$.

In the present manuscript, 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 [8], the new method builds g_k attempting to satisfy an error bound of the form

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

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.[1] In terms of n and ϵ, this bound agrees with the bound established in [8]. 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 our 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[2].

[1] In the context of nonconvex problems, evaluation complexity bounds of $\mathcal{O}\left(n^2\epsilon^{-2}\right)$ were obtained by Vicente [23] and by Konecny and Richtárik [10] for direct search methods, and also by Garmanjani, Júdice and Vicente [7] for a derivative-free trust-region method.

[2] 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 [4, 18] 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 [6, 10].

1.1 Contents

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

1.2 Notation

The symbol $\|\cdot\|$ denotes the 2-norm for vectors or matrices (depending on the context). The Euclidean inner product between $x, y \in \mathbb{R}^n$ is denoted by $\langle x, y\rangle$. For $j = 1, 2, \dots, n$, $e_j \in \mathbb{R}^n$ is the j-th vector of the canonical basis of $\mathbb{R}^n$. Given $a \in \mathbb{R}$, $\lfloor a \rfloor$ is the largest integer less than or equal to a.

2 Problem formulation and auxiliary results

The problem under consideration is the unconstrained minimization of a real-valued function stated as

$$\min_{x\in\mathbb{R}^n} f(x). \tag{2}$$

The problem class is specified by the following assumptions[3]:

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

$$M_{x,\sigma}(y) := f(x) + \langle g, y - x\rangle + \frac{1}{2}\langle B(y - x), y - x\rangle + \frac{\sigma}{2}\|y - x\|^2 \tag{3}$$

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 $\mathcal{O}\big(\|x^+ - x\|^2\big)$. This result is the basis of the step search procedure used in the proposed method.

[3] Assumptions A1 and A2 are the usual assumptions for the analysis of first-order and derivative-free methods (see, e.g., Section 1.2.3 of [16]). In particular, any twice continuously differentiable function with uniformly bounded Hessian satisfies A1.

Lemma 1 *Suppose that A1 holds. Given* $\epsilon > 0$, *let* $x \in \mathbb{R}^n$ *and* $g \in \mathbb{R}^n$ *be such that* $\|\nabla f(x)\| > \epsilon$ *and*

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

Moreover, given $\theta \in [0, 1)$, *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 $\sigma > 0$. *If*

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

then

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

Proof Assumption A1 implies that

$$f(x^+) \leq f(x) + \langle \nabla f(x), x^+ - x\rangle + \frac{L}{2}\|x^+ - x\|^2,$$

see, e.g., [2, Lemma 9.4] and its proof. Then, by the Cauchy–Schwarz inequality, the first inequality in (5) and (4), one obtains

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

Since $\|\nabla f(x)\| > \epsilon$, it follows from (4) that

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

which implies that

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

Moreover, (3) and the second inequality in (5) imply that

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

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

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

Using (11) and assuming that (6) holds, lead to

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

□

The next lemma suggests a way to construct g such that (4) holds for σ sufficiently large. This can be done using forward finite-differences with a carefully selected stepsize parameter $h > 0$.

Lemma 2 *Suppose that A1 holds and assume that, given $\theta \in [0, 1)$, x^{+} satisfies (5) for some $x \in \mathbb{R}^{n}$ and $\sigma > 0$. Moreover, suppose that $\|\nabla f(x)\| > \epsilon$ for some $\epsilon > 0$, and that the vector g in $M_{x,\sigma}(\cdot)$ is defined by*

$$
g_{j} = \frac{f(x + he_{j}) - f(x)}{h}, \quad j = 1, 2 \ldots, n. \tag{12}
$$

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),

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

Then, in view of (15), (5) and (6), it follows from Lemma 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, (15) and (16) imply that

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

which contradicts the assumption $\|\nabla f(x)\| > \epsilon$. Thus, (14) also must be true. □

3 Derivative-free quadratic regularization method

The proposed method works as follows. At the beginning of the kth iteration, one has an estimate x_k for the solution of (2), a symmetric positive semidefinite matrix B_k, $\epsilon > 0$, $\theta \in [0, 1)$ and a regularization parameter σ_k. An approximation $g_k \in \mathbb{R}^n$ for $\nabla f(x_k)$ is computed using forward finite-differences (as described in (12)) with stepsize

$$h_k = \frac{2\epsilon}{5\sigma_k\sqrt{n}}.$$

This calculation requires $n + 1$ evaluations of $f(\cdot)$ when $k = 0$ and n evaluations of $f(\cdot)$ when $k > 0$. Once g_k is obtained, a trial point x_k^+ is computed as an approximate solution of the auxiliary problem

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

with $M_{x_k,\sigma_k}(\cdot)$ defined in (3). If

$$\|g_k\| \geq \frac{4\epsilon}{5} \quad \text{and} \quad f(x_k) - f(x_k^+) \geq \frac{(1-\theta)\sigma_k}{8}\|x_k^+ - x_k\|^2$$

hold, then x_k^+ is accepted and the method sets $x_{k+1} = x_k^+$. Otherwise, the constant σ_k is multiplied by two[4] until the corresponding point x_k^+ is accepted. For the next iteration, a new matrix B_{k+1} is computed and the regularization parameter is updated. This method is detailed below.

Algorithm 1. Derivative-Free Quadratic Regularization Method (DFQRM)

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

Step 1. Set $i := 0$.

Step 1.1. For

4 In fact, for the theoretical guarantees, any other factor bigger than one can be used.

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

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

$$[g_{k,i}]_j = \frac{f(x_k + h_{k,i}e_j) - f(x_k)}{h_{k,i}}, \quad j = 1, 2, \ldots, n. \tag{18}$$

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

Remark 1 DFQRM described above is inspired by Algorithm 1 of [8]. They differ in the choice of $h_{k,i}$ and in the acceptance conditions for $x^+_{k,i}$.

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.

Thanks to A3 and the definition of σ_{k+1}, it is possible to obtain uniform lower and upper bounds for a relevant portion of the sequence $\left\{\sigma_k\right\}_{k\geq 0}$ of regularization parameters, as demonstrated next.

Lemma 3 *Suppose that A1 and A3 hold and let* $\{x_k\}$ *be a sequence generated by Algorithm* 1. *Given* $\epsilon > 0$, *let* $T(\epsilon)$ *be the first iteration index such that* $\|\nabla f(x_{T(\epsilon)})\| \leq \epsilon$. *Then*

$$\sigma_{\min} \leq \sigma_k \leq 2\max\left\{\sigma_0, [2L + 3M](1-\theta)^{-1}\right\} := \sigma_{\max}, \tag{23}$$

for all $k \in \{0, \dots, T(\epsilon)\}$.

Proof The proof is done using induction over k. Notice that (23) holds for $k = 0$. Assume that $T(\epsilon) \geq 1$ and that (23) is true for some $k \in \{0, \dots, T(\epsilon) - 1\}$. It follows from Step 2 of Algorithm 1 that

$$\sigma_{k+1} = \max\left\{2^{i_k - 1}\sigma_k, \sigma_{\min}\right\} \geq \sigma_{\min}. \tag{24}$$

Moreover,

$$2^{i_k - 1}\sigma_k \leq \sigma_{\max}. \tag{25}$$

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

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

Suppose that $i_k \geq 1$. Then, assuming that (25) is false one would get

$$2^{i_k - 1}\sigma_k > \sigma_{\max} > 2\left[2L + 3\|B_k\|\right](1-\theta)^{-1},$$

where the last inequality is due to A3. Moreover, by the definition of $T(\epsilon)$, $\|\nabla f(x_k)\| > \epsilon$. In this case, by Lemma 2, inequality (22) would have been satisfied for $i \leq i_k - 1$, contradicting the definition of i_k. Thus, in view of (24) and (25),

$$\sigma_{\min} \leq \sigma_{k+1} = \max\left\{2^{i_k - 1}\sigma_k, \sigma_{\min}\right\} \leq \max\left\{\sigma_{\max}, \sigma_{\min}\right\} = \sigma_{\max},$$

which concludes the proof. □

The next lemma establishes a lower bound of $\mathcal{O}\left(\|\nabla f(x_k)\|^2\right)$ for the difference $f(x_k) - f(x_{k+1})$ for $k = 0, \dots, T(\epsilon) - 1$. Later, it will be crucial to establish upper bounds for $T(\epsilon)$.

Lemma 4 *Suppose that A1 and A3 hold and let* $\{x_k\}$ *be a sequence generated by Algorithm* 1. *Given* $\epsilon > 0$, *let* $T(\epsilon)$ *be the first iteration index such that* $\|\nabla f(x_{T(\epsilon)})\| \leq \epsilon$. *If* $T(\epsilon) \geq 1$, *then*

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2C_f}\|\nabla f(x_k)\|^2, \quad \text{for} \quad k = 0, \dots, T(\epsilon) - 1, \tag{26}$$

where

$$C_f := \frac{81\left[2(\theta+1)+M\sigma_{\min}^{-1}\right]^2}{32(1-\theta)} \max\left\{\sigma_{\max}, L^2\sigma_{\min}^{-1}\right\}, \tag{27}$$

with $\sigma_{\max}$ is defined in (23).

Proof Given $k \in \{0, \ldots, T(\epsilon)-1\}$, 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,

$$\left(\frac{8\sigma_{k+1}}{L+8\sigma_{k+1}}\right)\|\nabla f(x_k)\| \leq \|g_k\|. \tag{28}$$

In view of Lemma 3,

$$M = \left(M\sigma_{\min}^{-1}\right)\sigma_{\min} \leq \left(M\sigma_{\min}^{-1}\right)\sigma_{k+1}. \tag{29}$$

Then, combining the second inequality in (21) with (29) and A3, it follows that

$$\begin{aligned}
\|g_k\| &\leq \|g_k + (B_k + 2^{i_k}\sigma_k I)(x_{k+1}-x_k)\| + \|(B_k + 2^{i_k}\sigma_k I)(x_{k+1}-x_k)\| \\
&\leq \|\nabla M_{x_k,2\sigma_{k+1}}(x_{k+1})\| + (\|B_k\| + 2\sigma_{k+1})\|x_{k+1}-x_k\| \\
&\leq (2\theta\sigma_{k+1} + 2\sigma_{k+1} + M)\|x_{k+1}-x_k\| \\
&\leq \left[2(\theta+1) + M\sigma_{\min}^{-1}\right]\sigma_{k+1}\|x_{k+1}-x_k\|.
\end{aligned} \tag{30}$$

Combining (28) and (30) it follows that

$$\|x_{k+1}-x_k\| \geq \frac{8\|\nabla f(x_k)\|}{\left[2(\theta+1)+M\sigma_{\min}^{-1}\right](L+8\sigma_{k+1})}. \tag{31}$$

The rest of the proof is divided in two cases.

Case I $\sigma_{k+1} \geq L$.

In this case, (22), (31), Lemma 3 and (27), imply that

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

that is, (26) holds.

Case II $\sigma_{k+1} < L$.

In this case, it follows from (22), (31), Lemma 3 and (27) that

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

that is, (26) also holds. □

Remark 2 By (31) and Lemma 3, if

$$
\|x_{k+1} - x_k\| \leq \epsilon \tag{32}
$$

then

$$
\|\nabla f(x_k)\| \leq \frac{\left[2(\theta+1) + M\sigma_{\min}^{-1}\right](L + 8\sigma_{\max})}{8}\epsilon = \mathcal{O}(\epsilon).
$$

Thus, condition (32) is a reasonable stopping criterion for Algorithm 1.

In view of Lemma 4, 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 possibly nonconvex problem. Its proof is based on the analysis of the gradient method presented at Section 1.2.3 of [16].

Theorem 1 *Suppose that A1–A3 hold and let $\left\{x_k\right\}$ be a sequence generated by Algorithm* 1. *Given $\epsilon > 0$, let $T(\epsilon)$ be the first iteration index such that $\|\nabla f(x_{T(\epsilon)})\| \leq \epsilon$. If $T(\epsilon) \geq 1$, then*

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

where C_f is defined in (27). *Consequently,*

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

Proof By Lemma 4,

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

Summing up these inequalities and using A2, one obtains

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

which implies that (33) holds. Finally, combining (33) and the definition of $T(\epsilon)$, it follows that (34) is true. □

Remark 3 More specifically, Theorem 1 gives a complexity bound of $\mathcal{O}(C_f\epsilon^{-2})$ iterations. In general, it follows from (27) and (23) that $C_f=\mathcal{O}(L^2)$. However, if L is known, one could take $\sigma_{\min}=L$ and $B_k=0$ for all k. In this case, the corresponding constant C_f depends linearly on L, and the complexity bound given by Theorem 1 reduces to $\mathcal{O}\left(L\epsilon^{-2}\right)$.

Consider the additional assumption:

A4 $f:\mathbb{R}^n\to\mathbb{R}$ is convex and the sublevel set

$$\mathcal{L}_f(x_0)=\left\{x\in\mathbb{R}^n\ :\ f(x)\leq f(x_0)\right\}$$

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. Its proof is inspired by the proof of Theorem 2.1.14 in [16], and by a reasoning described in [15].

Theorem 2 *Suppose that A1, A3 and A4 hold and let $\left\{x_k\right\}$ be a sequence generated by Algorithm* 1. *Given $\epsilon>0$, let $T(\epsilon)$ be the first iteration index such that $\|\nabla f(x_{T(\epsilon)})\|\leq\epsilon$. If $T(\epsilon)\geq 2$, then*

$$\min_{k=0,\ldots,2s(\epsilon)-1}\|\nabla f(x_k)\| \leq \frac{2C_fD_0}{s(\epsilon)}, \tag{36}$$

where $s(\epsilon)=\lfloor T(\epsilon)/2\rfloor$, C_f is defined in (27) *and*

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

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

$$T(\epsilon) \leq 2s(\epsilon) + 1 < 4C_f D_0 \epsilon^{-1} + 1. \tag{38}$$

Proof Since $\mathcal{L}_f(x_0)$ is compact (by A4) and $f(\cdot)$ is continuous, 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 $k \in \{0, \ldots, 2s(\epsilon) - 1\}$, and so

$$\|x_k - x^*\| \leq D_0, \quad \forall k \in \{0, \ldots, 2s(\epsilon) - 1\}.$$

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

$$\|\nabla f(x_k)\| \geq \frac{f(x_k) - f(x^*)}{D_0}, \quad k = 0, \ldots, 2s(\epsilon) - 1. \tag{39}$$

Combining (35) and (39) one obtains

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

Define

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

Then, given $j \in \{1, \ldots, 2s(\epsilon) - 1\}$, it follows from (40) that

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

and so

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

Summing up these inequalities it follows that

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

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

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

In particular, for $j = s(\epsilon)$, it follows from (41) and (35) that

$$\begin{aligned}\frac{2C_fD_0^2}{s(\epsilon)} &\geq f(x_{s(\epsilon)})-f(x^*)=f(x_{2s(\epsilon)})-f(x^*)+\sum_{k=s(\epsilon)}^{2s(\epsilon)-1}\big(f(x_k)-f(x_{k+1})\big)\\ &\geq \sum_{k=s(\epsilon)}^{2s(\epsilon)-1}\big(f(x_k)-f(x_{k+1})\big)\geq \frac{1}{2C_f}\sum_{k=s(\epsilon)}^{2s(\epsilon)-1}\|\nabla f(x_k)\|^2\\ &\geq \frac{s(\epsilon)}{2C_f}\min_{k=s(\epsilon),\dots,2s(\epsilon)-1}\|\nabla f(x_k)\|^2.\end{aligned}$$

Therefore,

$$\min_{k=0,\dots,2s(\epsilon)-1}\|\nabla f(x_k)\|^2\leq \frac{4C_f^2D_0^2}{s(\epsilon)^2},$$

which gives (36). By the definition of $T(\epsilon)$, the left-hand side in (36) is bigger than ϵ. Then, it follows that $s(\epsilon)<2C_fD_0\epsilon^{-1}$. Finally, using the definition of $s(\epsilon)$, one concludes that (38) is true. □

Now, instead of A4, consider the assumption:

A5 $f:\mathbb{R}^n\to\mathbb{R}$ is μ-strongly convex, that is, for all $x,y\in\mathbb{R}^n$ and for all $\alpha\in[0,1]$,

$$f(\alpha x+(1-\alpha)y)\leq \alpha f(x)+(1-\alpha)f(y)-\frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2.$$

The next theorem gives an iteration complexity bound of $\mathcal{O}\big(\log(\epsilon^{-1})\big)$ for Algorithm 1 applied to a strongly convex function.[5] Its proof is a direct consequence of (35) and the Polyak–Lojasiewics (PL) inequality [12, 19] satisfied by strongly convex functions.

Theorem 3 *Suppose that A1, A3 and A5 hold, and let* $\{x_k\}$ *be a sequence generated by Algorithm* 1. *Given* $\epsilon>0$, *let* $T(\epsilon)$ *be the first iteration index such that* $\|\nabla f(x_{T(\epsilon)})\|\leq\epsilon$. *If* $T(\epsilon)\geq 1$, *then*

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

where C_f *is defined in* (27) *and* x^* *is the minimizer of* $f(\cdot)$. *Consequently,*

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

[5] If assumptions A1 and A5 hold, then $\mu\leq L$. Since $L<C_f$, under these two assumptions it follows that $\frac{\mu}{C_f}\in(0,1)$.

Proof Let $k \in \{0, \dots, T(\epsilon) - 1\}$. By A5, $f(\cdot)$ satisfies the PL inequality (see, e.g., Section 4.2 of [17]):

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

Combining (35) and (44) one gets

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

which gives

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

Therefore,

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

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

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

which gives (42). Finally, applying (42) for $k = T(\epsilon) - 1$ and using the inequality $\|\nabla f(x_{T(\epsilon)-1})\| > \epsilon$, it follows that (43) holds. □

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. Its proof is an adaptation of the proof of Corollary 1 in [8].

Lemma 5 *Suppose that A1 and A3 hold. Let $T(\epsilon)$ be the first iteration index such that $\|\nabla f(x_{T(\epsilon)})\| \leq \epsilon$ and let $FE(\epsilon)$ be the total number of function evaluations performed by Algorithm* 1 *up to the* $(T(\epsilon) - 1)$*th iteration. Then,*

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

where $\sigma_{\max}$ is defined in (23).

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 $\sigma_{k+1} = 2^{i_k - 1}\sigma_k$, it follows that

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

Therefore,

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

where the last inequality is due to Lemma 3. □

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

Theorem 4 *Suppose that A1 and A3 hold, and let $FE(\epsilon)$ be defined as in Lemma 5. Then*

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

Proof Let $T(\epsilon)$ be defined as in Lemma 5. By Theorems 1, 2 and 3,

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

Then, combining Lemma 5 and (47), it follows that (46) is true. □

4 Numerical experiments

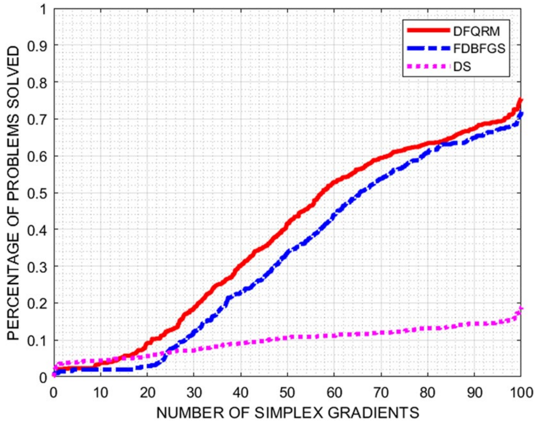
To investigate the practical performance of Algorithm 1, numerical experiments were carried out considering the following MATLAB implementations:

- *DFQRM:* Algorithm 1 with $\sigma_{\min} = 10^{-2}$, $\sigma_0 = 1$, $\epsilon = 10^{-5}$, $\theta = 0$ and B_k updated by

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

  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 [8].

Fig. 1 Data profiles for the precision 10^{-7}



- *DS*: an implementation of the simplified direct search method in [10] with $\alpha = 10$, $c = 10^{-3}$ and $D = \left\{\pm e_i \,|\, i = 1, \ldots, n\right\}$.

In the first experiment, the three codes were applied to the set of 102 problems from the Andrei's collection [1]. For each problem, four choices of the dimension n and two choices of starting points were considered, resulting in 816 instances. Specifically, the experiments were performed with $n = 6, 12, 24, 48$ 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 data profiles [14].[6] As it can be seen, DFQRM is sligthly superior than FDBFGS, with both codes solving more problems than DS using the same budget of function evaluations.

In the second experiment the codes were applied to ℓ_2-regularized logistic regression problems of the form

$$\min_{x \in \mathbb{R}^n} f_\mu(x) := -\sum_{i=1}^{m} \left[b^{(i)} \log(c_x(a^{(i)})) + (1 - b^{(i)}) \log(1 - c_x(a^{(i)}))\right] + \frac{\mu}{2}\|x\|_2^2,$$

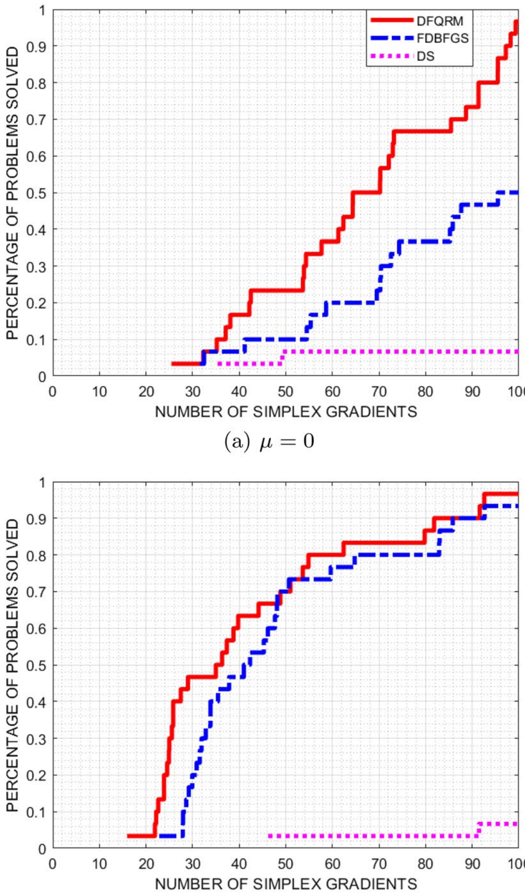
where $\left\{(a^{(i)}, b^{(i)})\right\}_{i=1}^{m} \subset \mathbb{R}^n \times \{0, 1\}$ is the dataset (with $a_1^{(i)} = 1$ for $i = 1, \ldots, m$), $c_x(a) := 1/(1 + e^{-\langle a, x\rangle})$ is the logistic model, and $\mu \geq 0$ is the regularization parameter. When $\mu > 0$, then $f_\mu(\cdot)$ is μ-strongly convex. Experiments were performed with ten datasets[7] and two choices of μ ($\mu \in \{0, 10\}$). For each μ, the following starting points were tested for all datasets:

$$x_0^{(1)} = \left[-1\ -1\ \ldots\ -1\right]^T, \quad x_0^{(2)} = \left[0\ 0\ \ldots\ 0\right]^T \quad \text{and} \quad x_0^{(2)} = \left[1\ 1\ \ldots\ 1\right]^T,$$

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

[7] Namely, the datasets `Iris`, `Breast Cancer Wisconsin`, `Wine`, `Sonar`, `Phishing`, `Ionosphere`, `Diabetes`, `Musk`, `Seeds`, `Bank Note Authentication`, freely available in the website http://archive.ics.uci.edu/ml/index.php.

Fig. 2 Data profiles for the precision 10^{-3}



(a) $\mu = 0$

(b) $\mu = 10$

resulting in 30 instances (dataset, starting point) with dimensions ranging from 5 to 61. As it can be seen in Fig. 2, for both choices of μ, DFQRM outperformed DS by a large margin. Moreover, DFQRM was significantly better than FDBFGS in the case $\mu = 0$ (convex problems), while it was slightly better than FDBFGS in the case $\mu = 10$ (strongly convex problems).

5 Conclusion

This manuscript 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. Numerical results suggest that the new method compares favorably with the simplified direct search method from [10], and that it is competitive with the derivative-free method recently proposed in [8].

Acknowledgements The author is very grateful to the two anonymous referees, whose comments helped to improve the manuscript.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim. **10**(1), 147–161 (2008)
2. Audet, C., Hare, W.: Derivative-free and blackbox optimization. Springer, Berlin (2017)
3. Audet, C., Orban, D.: Finding optimal algorithmic parameters using derivative-free optimization. SIAM J. Optim. **17**(3), 642–664 (2006)
4. Bergou, E.H., Gorbunov, E., Richtárik, P.: Stochastic three points method for unconstrained smooth minimization. SIAM J. Optim. **30**(4), 2726–2749 (2020)
5. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. SIAM (2009)
6. Dodangeh, M., Vicente, L.N.: Worst-case complexity of direct search under convexity. Math. Program. **155**, 307–332 (2016)
7. Garmanjani, R., Júdice, D., Vicente, L.N.: Trust-region methods without using derivatives: worst case complexity and the nonsmooth case. SIAM J. Optim. **26**(4), 1987–2011 (2016)
8. Grapiglia, G.N.: Quadratic regularization methods with finite-difference gradient approximations. Comput. Optim. Appl. (2022). https://doi.org/10.1007/s10589-022-00373-z
9. Karbasian, H.R., Vermeire, B.C.: Gradient-free aerodynamic shape optimization using large Eddy simulation. Comput. Fluids **232**, 105185 (2022)
10. Konecny, J., Richtárik, P.: Simple complexity analysis of simplified direct search. arXiv:1410.0390 [math.OC] (2014)
11. Larson, J., Menickelly, M., Wild, S.M.: Derivative-free optimization. Acta Numer. **28**, 287–404 (2019)
12. Lojasiewicz, S.: A topological property of real analytic subsets (in French), Coll. du CNRS, Les équations aux dérivées partielles, pp. 87–89 (1963)
13. Marsden, A.L., Feinstein, J.A., Taylor, C.A.: A computational framework for derivative-free optimization of cardiovascular geometries. Comput. Methods Appl. Mech. Eng. **197**, 1890–1905 (2008)
14. Moré, J.J., Wild, S.M.: Benchmarking derivative-free optimization algorithms. SIAM J. Optim. **20**(1), 172–191 (2009)
15. Nesterov, Yu.: How to make gradients small. Optima **88**, 10–11 (2012)
16. Nesterov, Yu.: Lectures on Convex Optimization, 2nd edn. Springer, Berlin (2018)
17. Nesterov, Yu., Polyak, B.T.: Cubic regularization of Newton method and its global performance. Math. Program. **108**, 177–205 (2006)
18. Nesterov, Yu., Spokoiny, V.: Random gradient-free minimization of convex functions. Found. Comput. Math. **17**, 527–566 (2017)
19. Polyak, B.T.: Gradient methods for minimizing functionals (in Russian). Zh. Vychisl. Mat. Mat. Fiz., 643–653 (1963)
20. 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. Mater. Des. **209** (2021)

21. Sophy, O., Cartis, C., Kriest, I., Tett, S.F.B., Khatiwala, S.: A derivative-free optimisation method for global ocean biogeochemical models. Geosci. Model Dev. **15**(9), 3537–3554 (2022)
22. Sun, Y., Sahinidis, N.V., Sundaram, A., Cheon, M.-S.: Derivative-free optimization for chemical product design. Curr. Option Chem. Eng. **27**, 98–106 (2020)
23. Vicente, L.N.: Worst case complexity of direct search. EURO J. Comput. Optim. **1**, 143–153 (2013)