Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-024-02519-x

Worst Case Complexity Bounds for Linesearch-Type Derivative-Free Algorithms

Andrea Brilli[1] · **Morteza Kimiaei**[2] · **Giampaolo Liuzzi**[1] · **Stefano Lucidi**[1]

Received: 17 December 2023 / Accepted: 13 August 2024

Abstract

This paper is devoted to the analysis of worst case complexity bounds for linesearch-type derivative-free algorithms for the minimization of general non-convex smooth functions. We consider a derivative-free algorithm based on a linesearch extrapolation technique. First we prove that it enjoys the same complexity properties which have been proved for pattern and direct search algorithms, namely that the number of iterations and of function evaluations required to drive the norm of the gradient of the objective function below a given threshold ϵ for the first time is $\mathcal{O}(\epsilon^{-2})$ in the worst case. This is the first contribution proving worst-case complexity properties for derivative-free linesearch-type algorithms. Then we show that the lineasearch approach used by the described algorithm allows us to guarantee the further property that the number of iterations such that the norm of the gradient is bigger than ϵ is $\mathcal{O}(\epsilon^{-2})$ in the worst case.

Communicated by Lam M. Nguyen.

Morteza Kimiaei acknowledges the financial support of the Doctoral Program Vienna Graduate School on Computational Optimization (VGSCO) funded by the Austrian Science Foundation under Project No W1260-N35.

✉ Giampaolo Liuzzi
liuzzi@diag.uniroma1.it

Andrea Brilli
brilli@diag.uniroma1.it

Morteza Kimiaei
kimiaeim83@univie.ac.at

Stefano Lucidi
lucidi@diag.uniroma1.it

[1] Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto, 25, 00185 Rome, Italy

[2] Fakultät für Mathematik, Universität Wien, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria

Published online: 04 October 2024

Keywords Derivative-free optimization · Unconstrained optimization · Line search · Worst case complexity

Mathematics Subject Classification 90C30 · 90C56

1 Introduction

In this paper we consider the following unconstrained minimization problem

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

We assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a black-box function which is known by means of an oracle that only outputs function values. Hence, derivatives of f can neither be approximated nor computed explicitly.

1.1 Literature Review

Over the past decade, the analysis of worst case complexity for optimization algorithms has gained more and more interest and attracted many researchers [1, 5]. Specifically for derivative-free algorithms, in [7, 18] worst case complexity bounds have been derived for direct search methods using sufficient decrease in f. In particular, it has been proved that direct search methods (based on a search step and a poll step and using sufficient decrease acceptability) require at most $\mathcal{O}(n\epsilon^{-2})$ iterations and $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations to find a point x_k such that $\|\nabla f(x_k)\| \leq \epsilon$.

In [4] an adaptive cubic regularization algorithm has been proposed which is based on gradient estimation via finite differences. The algorithm has a worst case complexity of $\mathcal{O}(n^2\epsilon^{-3/2})$ which is better than the complexity obtained for direct search methods.

In [10], the complexity of a smoothing technique for the optimization of nonsmooth functions has been studied. It has been shown that the smoothing algorithm has a worst case complexity of $\mathcal{O}(\epsilon^{-3})$ to achieve ϵ-stationarity.

Analogous results for linesearch-based derivative algorithms (see e.g. [8, 14]) have not yet been established. The latter algorithms typically have stronger asymptotic convergence properties which are tied to the use of suitable though more complex extrapolation techniques.

1.2 Our Contribution

In this paper we show that derivative free algorithms which are based on a linesearch-type extrapolation technique with sufficient decrease have the same worst case complexity proved for direct search methods. Proving this complexity result heavily depends on showing that the algorithm is able to produce sufficient decrease of an auxiliary function regardless of what type of iteration we are considering, either a success or failure one. Furthermore, thanks to the linesearch approach with sufficient decrease, they also have the property that the number of iterations (in the worst case) for which $\|\nabla f(x_k)\| \geq \epsilon$ is of the order of ϵ^{-2}. This last property considerably enriches the worst case analysis of derivative-free algorithm and, to the best of our knowledge, is new in this context. We stress that the property characterizes the behaviour of the derivative-free algorithm better than the usual complexity result. Indeed, typical complexity results give the number of iteration required to drive the norm of the gradient below a prefixed tolerance for the first time. If we let the method run, the norm of the gradient might well rise above the tolerance again. The property we prove in this paper indicates that the total number of iterations with a gradient norm above a specified tolerance is bounded by a constant that depends on ϵ^{-2}.

1.3 Organization of the Paper

The paper is organized as follows. In Sect. 2 we describe an algorithm model which is based on linesearch techniques. In Sect. 3 we prove the asymptotic convergence properties of the algorithm model. Namely we prove that every limit point of the sequence generated by the algorithm is stationary. In Sect. 4 we derive the worst case complexity bounds for the number of iterations and function evaluations required by the algorithm to drive the norm of the gradient below a preset tolerance. Furthermore, we also prove that the number of iterations with a norm of the gradient above a preset tolerance can be bounded by a constant in the worst case. In Sect. 5, we present two more algorithms which are still based on a linesearch technique and that have slightly different worst case complexity results. The theoretical results concerning these two methods are similar to (though not trivially obtainable from) the preceding ones. For that reason, the proofs of these latter results are reported in two appendices at the end of the paper. In Sect. 6, we report the results of a numerical comparison between the proposed methods and a well-established derivative-free method on a small set of test problems.

Finally, in Sect. 7 we draw some conclusions.

2 A Linesearch-Type Algorithm

In the following section we introduce a derivative-free algorithm which is based on an extrapolation technique.

Linesearch Algorithm Model (LAM)

Data: $c \in (0,1), \theta \in (0,1), x_0 \in \mathbb{R}^n, \tilde{\alpha}_0^i > 0, i \in \{1,\dots,n\}$, and set $d_0^i = e^i$, for $i = 1,\dots,n$.

For $k = 0, 1, \dots$

Set $y_k^1 = x_k$.

For $i = 1,\dots,n$

Let $\bar{\alpha}_k^i = \max\{\tilde{\alpha}_k^i, c \max_{j=1,\dots,n}\{\tilde{\alpha}_k^j\}\}$.

Compute α and d by the `DF-Linesearch`$(\bar{\alpha}_k^i, y_k^i, d_k^i; \alpha, d)$.

Set $y_k^{i+1} = y_k^i + \alpha d, \quad d_{k+1}^i = d, \quad \alpha_k^i = \alpha$.

End For

Set $x_{k+1} = y_k^{n+1}$.

If $x_{k+1} = x_k$ **then** $\tilde{\alpha}_{k+1}^i = \theta\bar{\alpha}_k^i$ for all $i = 1,\dots,n$,

else set $\tilde{\alpha}_{k+1}^i = \max\{\bar{\alpha}_k^i, \alpha_k^i\}$ for all $i = 1,\dots,n$.

End For

As we can see, at each iteration k, LAM performs an exploration of the space around the current iterate x_k using the coordinate directions and producing the points y_k^i, $i = 1,\dots,n+1$ (note that $y_k^1 = x_k$).

More in particular, points y_k^i, for $i = 2,\dots,n+1$, are computed by means of a suitable derivative-free linesearch, namely the `DF-Linesearch` procedure. The linesearch is invoked by passing a tentative step size, i.e.

$$\bar{\alpha}_k^i = \max\{\tilde{\alpha}_k^i, c \max_{j=1,\dots,n}\{\tilde{\alpha}_k^j\}\}.$$

It is worth noting that the initial step size $\bar{\alpha}_k^i$ is quite unusual for linesearch-type derivative-free algorithms. In this work, this initialization is of paramount importance since it allows us to prove that in each iteration the method achieves sufficient decrease of an auxiliary function that is used to derive the worst-case complexity bound.

Then, an actual step size, i.e. α is produced, which can either be 0 or strictly greater than zero. When the exploration of the neighbourhood around the current iterate x_k "fails", namely when $x_{k+1} = x_k$, all the tentative steps for the next iteration are reduced; when $x_{k+1} \neq x_k$, the tentative steps for the next iteration are updated and possibly augmented.

In the following, we report the `DF-Linesearch` procedure.

DF-Linesearch ($\bar{\alpha}, y, d; \alpha, \hat{d}$).

Input. $\bar{\alpha}, y, d$. **Output.** $\alpha, \hat{d}$
Data. $\gamma > 0$, $\delta \in (0, 1)$.
Step 1. Set $\alpha = \bar{\alpha}$, $\hat{d} \leftarrow d$.
Step 2. If $f(y + \alpha\hat{d}) \leq f(y) - \gamma\alpha^2$ **then** go to Step 5.
Step 3. If $f(y - \alpha\hat{d}) \leq f(y) - \gamma\alpha^2$ **then** set $\hat{d} \leftarrow -d$ and go to Step 5.
Step 4. Set $\alpha = 0$ and **return** $(\alpha, \hat{d})$.
Step 5. While $f\left(y + \frac{\alpha}{\delta}\hat{d}\right) \leq f(y + \alpha\hat{d}) - \gamma\left(\left(\frac{1}{\delta} - 1\right)\alpha\right)^2$
$\alpha \leftarrow \alpha/\delta$.
Step 6. Return $(\alpha, \hat{d})$

The DF-Linesearch procedure employs a sufficient decrease criterion, which is evaluated between successive points, i.e.

$$f\left(y + \frac{\alpha}{\delta}\hat{d}\right) \leq f(y + \alpha\hat{d}) - \gamma\left(\left(\frac{1}{\delta} - 1\right)\alpha\right)^2.$$

Regarding the linesearch (expansion) procedure, its primary distinction from other linesearch techniques (see e.g. [13, 14]) lies in its continual expansion of the step while maintaining adequate decrease between successive points. We note that other derivative-free linesearch approaches employ the following sufficient decrease criterion

$$f\left(y + \frac{\alpha}{\delta}\hat{d}\right) \leq f(y) - \gamma\left(\frac{\alpha}{\delta}\right)^2.$$

As we can see the above formula uses as reference point the initial point y whereas the criterion in our proposed method uses as reference point $y + \alpha d$.

Let it be highlighted that the adopted sufficient decrease criterion allows us to obtain the complexity bound on the number of function evaluations in a more straightforward way than the usual criterion. Needless to say, both the criteria are able to convey the method the same asymptotic global convergence properties.

3 Asymptotic Convergence Analysis for LAM

In order to carry out the convergence analysis for LAM, we need the following standard assumptions.

Assumption 3.1 The function f is continuously differentiable on $\mathbb{R}^n$, and its gradient is Lipschitz continuous with Lipschitz constant L, i.e. for all $x, y \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Assumption 3.2 The function f is bounded from below, i.e. $f_{\min} \in \mathbb{R}$ exists such that

$$f_{\min} \leq f(x), \qquad \text{for all } x \in \mathbb{R}^n.$$

First of all, we derive an upper bound on the norm of $\nabla f(x_k)$ at each iteration k.

Proposition 3.1 *Suppose that Assumption 3.1 holds. Let $\{x_k\}$ be the sequence produced by the LAM framework. Then, for each k such that $x_{k+1} \neq x_k$*

$$\|\nabla f(x_k)\| \leq \sqrt{n}\left(\frac{\gamma + L(\sqrt{n}+1)}{\delta}\right) \max_{i=1,\ldots,n} \{\tilde{\alpha}^i_{k+1}\}, \tag{1}$$

whereas, for each k such that $x_{k+1} = x_k$

$$\|\nabla f(x_k)\| \leq \sqrt{n}\frac{\gamma + L}{\theta} \max_{i=1,\ldots,n} \{\tilde{\alpha}^i_{k+1}\}, \tag{2}$$

where the constants γ, δ, θ are those defined in the LAM algorithm and the DF-Linesearch procedure.

Proof For each iteration k such that $x_{k+1} \neq x_k$ and every index $i = 1, \ldots, n$, one of two cases can occur:

Case (i), $\alpha^i_k = 0$. By $\alpha^i_k = 0$, and $\tilde{\alpha}^i_{k+1} = \bar{\alpha}^i_k$, we have:

$$f(y^i_k + \bar{\alpha}^i_k e_i) > f(y^i_k) - \gamma(\bar{\alpha}^i_k)^2,$$
$$f(y^i_k - \bar{\alpha}^i_k e_i) > f(y^i_k) - \gamma(\bar{\alpha}^i_k)^2.$$

Then we get from the Mean Value Theorem

$$\nabla f(u^i_k)^T e_i > -\gamma\bar{\alpha}^i_k, \tag{3}$$
$$\nabla f(v^i_k)^T e_i < \gamma\bar{\alpha}^i_k, \tag{4}$$

where $u^i_k = y^i_k + \lambda^i_k\bar{\alpha}^i_k e_i$ and $v^i_k = y^i_k - \mu^i_k\bar{\alpha}^i_k e_i$ with $\lambda^i_k, \mu^i_k \in (0, 1)$. From (3) and (4) and the Lipschitz continuity of ∇f, we have that

$$\nabla f(x_k)^T e_i > -\gamma\bar{\alpha}^i_k - L\|x_k - u^i_k\| > -\gamma\bar{\alpha}^i_k - L\|x_k - y^i_k\| - L\bar{\alpha}^i_k,$$
$$\nabla f(x_k)^T e_i < \gamma\bar{\alpha}^i_k + L\|x_k - v^i_k\| < \gamma\bar{\alpha}^i_k + L\|x_k - y^i_k\| + L\bar{\alpha}^i_k.$$

Hence

$$
\begin{aligned}
|\nabla f(x_k)^T e_i| &< (\gamma + L)\bar{\alpha}_k^i + L\|x_k - y_k^i\| \leq (\gamma + L)\bar{\alpha}_k^i + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\} \\
&= (\gamma + L)\,\tilde{\alpha}_{k+1}^i + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\} \\
&\leq (\gamma + L) \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\} + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\},
\end{aligned}
$$

so that

$$
|\nabla f(x_k)^T e_i| \leq \Big(\gamma + L(\sqrt{n} + 1)\Big) \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}.
$$

Case (ii). From $\alpha_k^i = \alpha$, and $\tilde{\alpha}_{k+1}^i = \alpha \geq \bar{\alpha}_k^i$, it results either

$$
\begin{aligned}
&f\left(y_k^i + \frac{\tilde{\alpha}_{k+1}^i}{\delta} e_i\right) > f(y_k^i + \tilde{\alpha}_{k+1}^i e_i) - \gamma\left(\frac{1}{\delta} - 1\right)^2 (\tilde{\alpha}_{k+1}^i)^2, \\
&f(y_k^i + \delta\tilde{\alpha}_{k+1}^i e_i) \geq f(y_k^i + \tilde{\alpha}_{k+1}^i e_i) + \gamma(1-\delta)^2(\tilde{\alpha}_{k+1}^i)^2
\end{aligned}
$$

or

$$
\begin{aligned}
&f\left(y_k^i - \frac{\tilde{\alpha}_{k+1}^i}{\delta} e_i\right) > f(y_k^i - \tilde{\alpha}_{k+1}^i e_i) - \gamma\left(\frac{1}{\delta} - 1\right)^2 (\tilde{\alpha}_{k+1}^i)^2, \\
&f(y_k^i - \delta\tilde{\alpha}_{k+1}^i e_i) \geq f(y_k^i - \tilde{\alpha}_{k+1}^i e_i) + \gamma(1-\delta)^2(\tilde{\alpha}_{k+1}^i)^2.
\end{aligned}
$$

Then, we get,

$$
\nabla f(\bar{u}_k^i)^T e_i > -\gamma\left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i, \quad -\nabla f(\hat{u}_k^i)^T e_i \geq \gamma(1-\delta)\tilde{\alpha}_{k+1}^i, \tag{5}
$$

or

$$
\nabla f(\bar{v}_k^i)^T e_i < \gamma\left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i, \quad -\nabla f(\hat{v}_k^i)^T e_i \leq -\gamma(1-\delta)\tilde{\alpha}_{k+1}^i, \tag{6}
$$

where $\bar{u}_k^i = y_k^i + \bar{\lambda}_k^i \left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i e_i$, $\hat{u}_k^i = y_k^i + \hat{\lambda}_k^i(1-\delta)\tilde{\alpha}_{k+1}^i e_i$, $\bar{v}_k^i = y_k^i - \bar{\mu}_k^i\left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i e_i$, and $\hat{v}_k^i = y_k^i - \hat{\mu}_k^i(1-\delta)\tilde{\alpha}_{k+1}^i e_i$, with $\bar{\lambda}_k^i, \hat{\lambda}_k^i, \bar{\mu}_k^i, \hat{\mu}_k^i \in (0, 1)$.

When (5) holds, from $\nabla f(\bar{u}_k^i)^T e_i > -\gamma\left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i$ we can write

$$
[\nabla f(\bar{u}_k^i) - \nabla f(x_k) + \nabla f(x_k)]^T e_i > -\gamma\left(\frac{1-\delta}{\delta}\right)\tilde{\alpha}_{k+1}^i,
$$

so that we obtain

$$\begin{aligned}\nabla f(x_k)^T e_i &> -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}^i_{k+1} - L\|x_k - \bar{u}^i_k\| \\ &> -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}^i_{k+1} - L\|x_k - y^i_k\| - L\left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}^i_{k+1}.\end{aligned} \tag{7}$$

From $\nabla f(\hat{u}^i_k)^T e_i \leq -\gamma(1-\delta)\tilde{\alpha}^i_{k+1}$ in (5), we can write

$$[\nabla f(\hat{u}^i_k) - \nabla f(x_k) + \nabla f(x_k)]^T e_i \leq -\gamma(1-\delta)\tilde{\alpha}^i_{k+1},$$

so that, in this case, we obtain

$$\begin{aligned}\nabla f(x_k)^T e_i &\leq -\gamma(1-\delta)\tilde{\alpha}^i_{k+1} + L\|x_k - \hat{u}^i_k\| \\ &\leq \gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}^i_{k+1} + L\|x_k - y^i_k\| + L\left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}^i_{k+1}.\end{aligned} \tag{8}$$

Now, considering (7) and (8), we get

$$|\nabla f(x_k)^T e_i| \leq \left(\frac{\gamma + L(\sqrt{n}+1)}{\delta}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}^i_{k+1}\}. \tag{9}$$

The same bound can be obtained when (6) holds. Thus, finally, we obtain

$$\|\nabla f(x_k)\| \leq \sqrt{n} \left(\frac{\gamma + L(\sqrt{n}+1)}{\delta}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}^i_{k+1}\}.$$

On the other hand, for each iteration k such that $x_{k+1} = x_k$, i.e. $y^i_k = x_k$ for all $i = 1, \dots, n+1$, we have for every index $i = 1, \dots, n$

$$f\left(x_k + \frac{\tilde{\alpha}^i_{k+1}}{\theta} e_i\right) > f(x_k) - \gamma \left(\frac{\tilde{\alpha}^i_{k+1}}{\theta}\right)^2,$$

$$f\left(x_k - \frac{\tilde{\alpha}^i_{k+1}}{\theta} e_i\right) > f(x_k) - \gamma \left(\frac{\tilde{\alpha}^i_{k+1}}{\theta}\right)^2.$$

Then we get from the Mean Value Theorem

$$\nabla f(u^i_k)^T e_i > -\gamma \frac{\tilde{\alpha}^i_{k+1}}{\theta}, \tag{10}$$

$$\nabla f(v^i_k)^T e_i < \gamma \frac{\tilde{\alpha}^i_{k+1}}{\theta}, \tag{11}$$

where $u_k^i = x_k + \lambda_k^i \frac{\tilde{\alpha}_{k+1}^i}{\theta} e_i$ and $v_k^i = x_k - \mu_k^i \frac{\tilde{\alpha}_{k+1}^i}{\theta} e_i$ with $\lambda_k^i, \mu_k^i \in (0, 1)$. From (10) and (11) and the Lipschitz continuity of ∇f, we have that

$$\nabla f(x_k)^T e_i > -\gamma \frac{\tilde{\alpha}_{k+1}^i}{\theta} - L\|x_k - u_k^i\| > -\gamma \frac{\tilde{\alpha}_{k+1}^i}{\theta} - L \frac{\tilde{\alpha}_{k+1}^i}{\theta},$$
$$\nabla f(x_k)^T e_i < \quad \gamma \frac{\tilde{\alpha}_{k+1}^i}{\theta} + L\|x_k - v_k^i\| < \quad \gamma \frac{\tilde{\alpha}_{k+1}^i}{\theta} + L \frac{\tilde{\alpha}_{k+1}^i}{\theta}.$$

Hence

$$|\nabla f(x_k)^T e_i| < (\gamma + L) \frac{\tilde{\alpha}_{k+1}^i}{\theta},$$

so that we can write

$$\|\nabla f(x_k)\| \leq \sqrt{n} \frac{\gamma + L}{\theta} \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k+1}^i\},$$

concluding the proof. □

Drawing inspiration from [3, 6, 12], we now introduce the following function:

$$\Phi_k = f(x_k) + \eta \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2, \tag{12}$$

where η satisfies:

$$0 < \eta < \gamma(1 - \delta)^2. \tag{13}$$

Note that, whenever Assumption 3.2 holds, we have

$$\Phi_k \geq f_{\min}. \tag{14}$$

Function Φ_k allows us to state the following result which characterizes the evolution of the algorithm at each iteration.

Proposition 3.2 *Let $\{x_k\}$ and $\{\tilde{\alpha}_k^i\}$, $i = 1, \ldots, n$, be the sequences produced by LAM. Then for all $k = 0, 1, \ldots$:*

$$\Phi_{k+1} - \Phi_k \leq -\tilde{c}_{\scriptscriptstyle LAM} \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k+1}^i\}^2, \tag{15}$$

where

$$\tilde{c}_{\scriptscriptstyle LAM} = \min \left\{ \eta \left(\frac{1 - \theta^2}{\theta^2} \right), \gamma c^2, \gamma(1 - \delta)^2 - \eta \right\}, \tag{16}$$

η is a parameter satisfying (13) and $c, \gamma, \delta, \theta$ are the constants defined in the LAM algorithm and the DF-Linesearch procedure.

Proof We split the set of iteration indices $\{0, 1, 2, \dots\}$ into the two subsets K_1 and K_2, namely

(i) $k \in K_1$ when $x_k \neq x_{k-1}$;

(ii) $k \in K_2$ when $x_k = x_{k-1}$.

Let us first consider the case when $k \in K_2$, i.e. the $(k-1)$-th iteration is of failure, we have

(i) $f(x_k) = f(x_{k-1})$;
(ii) $\tilde{\alpha}_k^i = \theta \bar{\alpha}_{k-1}^i$, for every $i = 1, \dots, n$.

In particular, since, for every $i = 1, \dots, n$, $\bar{\alpha}_{k-1}^i = \max\{\tilde{\alpha}_{k-1}^i, c \max_{j=1,\dots,n} \{\tilde{\alpha}_{k-1}^j\}\}$, we have either

$$\tilde{\alpha}_k^i = \theta \tilde{\alpha}_{k-1}^i \leq \theta \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$$

or

$$\tilde{\alpha}_k^i = \theta c \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\} \leq \theta \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}.$$

Hence, in case of failure, we can write $\tilde{\alpha}_k^i \leq \theta \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$, so that $\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} \leq \theta \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$. Then, from

$$\Phi_k - \Phi_{k-1} = f(x_k) - f(x_{k-1}) + \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right)$$

and considering

$$-\frac{1}{\theta^2} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 \geq - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2, \quad \text{and } f(x_k) = f(x_{k-1}),$$

we obtain

$$\begin{aligned} \Phi_k - \Phi_{k-1} &= \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\ &\leq \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \frac{1}{\theta^2} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 \right), \end{aligned}$$

so that

$$\Phi_k - \Phi_{k-1} \leq -\eta \left(\frac{1-\theta^2}{\theta^2} \right) \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}. \tag{17}$$

Let us now consider the case when $k \in K_1$, i.e. the $(k-1)$-th iteration is of "success". In this case, we have $x_k \neq x_{k-1}$ and $f(x_k) < f(x_{k-1})$. Then, we consider the following two cases:

1. $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} = \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$;
2. $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} > \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$.

Case 1. Since the $(k-1)$-th iteration is of success, there is an index $\bar{j}$ such that:

$$f(y_{k-1}^{\bar{j}+1}) \le f(y_{k-1}^{\bar{j}}) - \gamma(\bar{\alpha}_{k-1}^{\bar{j}})^2$$

with

$$\bar{\alpha}_{k-1}^{\bar{j}} = \max\{\tilde{\alpha}_{k-1}^{\bar{j}}, c\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}\} \ge c\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\},$$

Then we have

$$f(y_{k-1}^{\bar{j}+1}) \le f(y_{k-1}^{\bar{j}}) - \gamma(\bar{\alpha}_{k-1}^{\bar{j}})^2 \le f(y_{k-1}^{\bar{j}}) - \gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2.$$

and, exploiting that we are in case 1,

$$f(y_{k-1}^{\bar{j}+1}) \le f(y_{k-1}^{\bar{j}}) - \gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2.$$

That implies

$$f(x_k) \le f(x_{k-1}) - \gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2.$$

Recalling, again, that we are in case 1, we obtain:

$$\Phi_k - \Phi_{k-1} = f(x_k) - f(x_{k-1}) \le -\gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2. \tag{18}$$

Case 2. Since $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} > \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$, we have that an index $\bar{j}$ exists such that a linesearch has been performed along the $\bar{j}$-th direction which determines a steplength $\alpha_{k-1}^{\bar{j}}$ satisfying

$$\alpha_{k-1}^{\bar{j}} = \tilde{\alpha}_k^{\bar{j}} = \max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}.$$

More specifically, we have

$$\begin{aligned} f(x_k) &\le f(y_{k-1}^{\bar{j}} + \tilde{\alpha}_k^{\bar{j}} d_k^{\bar{j}}) \\ &\le f(y_{k-1}^{\bar{j}} + \delta\tilde{\alpha}_k^{\bar{j}} d_k^{\bar{j}}) - \gamma(1-\delta)^2(\tilde{\alpha}_k^{\bar{j}})^2 \le f(x_{k-1}) - \gamma(1-\delta)^2(\tilde{\alpha}_k^{\bar{j}})^2 \\ &= f(x_{k-1}) - \gamma(1-\delta)^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2. \end{aligned} \tag{19}$$

Hence, we obtain:

$$\begin{aligned}\Phi_k - \Phi_{k-1} &= f(x_k) - f(x_{k-1}) + \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\ &\leq -\gamma(1-\delta)^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 + \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\ &\leq -\gamma(1-\delta)^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 + \eta \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 \\ &\leq -\left(\gamma(1-\delta)^2 - \eta\right) \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2. \end{aligned} \tag{20}$$

Finally (17), (18) and (20) conclude the proof. □

Exploiting the properties of the function Φ, we prove that the sequences of initial stepsizes $\tilde{\alpha}_k^i$, $i = 1, \dots, n$, are all convergent to zero.

Proposition 3.3 *Suppose that Assumption 3.2 holds. Then, the LAM framework produces sequences $\{\tilde{\alpha}_k^i\}$, $i = 1, \dots, n$, such that*

$$\lim_{k\to\infty} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} = 0.$$

Proof By Proposition 3.2, the sequence $\{\Phi_k\}$ is monotonically decreasing. Since $\Phi_k \geq f_{\min}$, we have that

$$\lim_{k\to\infty} \Phi_k = \bar{\Phi}.$$

Then, the proof is concluded recalling again Proposition 3.2 and the above limit. □

Corollary 3.1 *Suppose that Assumptions 3.1 and 3.2 hold. Then, LAM produces an infinite sequence $\{x_k\}$ such that*

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

Proof The proof easily follows recalling Proposition 3.1 and Proposition 3.3. □

Remark Note that the result of Proposition 3.3 is somewhat stronger than analogous results for GPS [17] and MADS-type [2] algorithms. Indeed, for those algorithms it is only possible to show that a subsequence of the stepsizes converges to zero. As a consequence, also the result of Corollary 3.1 is stronger in that it states that every limit point of the sequence of iterates is stationary.

4 Complexity Bounds for LAM

This section is devoted to the definition of the worst case complexity bounds for the LAM algorithm.

The next two propositions ensure that the algorithm model takes at most $\mathcal{O}(n\epsilon^{-2})$ iterations and $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations to produce a point x_k such that $\|\nabla f(x_k)\| \le \epsilon$.

Proposition 4.1 *Suppose that Assumptions 3.1 and 3.2 hold. Let $\{x_k\}$ be the sequence of points produced by LAM. Given any $\epsilon \in (0,1)$ and $\tilde{\alpha}_0^i \ge \epsilon$, for $i = 1, \dots, n$, assume that $\bar{j}_\epsilon + 1$ is the first iteration such that $\|\nabla f(x_{\bar{j}_\epsilon+1})\| \le \epsilon$, i.e. $\|\nabla f(x_k)\| > \epsilon$, for all $k = 0, 1, \dots, \bar{j}_\epsilon$. Then,*

$$\bar{j}_\epsilon \le \left\lceil \frac{n\, c_1^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n\epsilon^{-2}), \tag{21}$$

where $\tilde{c}_{LAM}$ is given by (16) and

$$c_1 = \frac{\gamma + L}{\theta}. \tag{22}$$

Proof Using the function Φ_k defined by (12) we can write:

$$\Phi_{\bar{j}+1} - \Phi_0 = (\Phi_{\bar{j}+1} - \Phi_{\bar{j}}) + (\Phi_{\bar{j}} - \Phi_{\bar{j}-1}) + \cdots + (\Phi_1 - \Phi_0)$$

and exploiting Proposition 3.2 we have:

$$\Phi_{\bar{j}+1} - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{k=1}^{\bar{j}+1} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 = -\tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2.$$

By recalling (14), we can write

$$f_{\min} - \Phi_0 \le \Phi_{\bar{j}} - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2. \tag{23}$$

As done previously the set of iteration indices $\{0, 1, 2, \dots\}$ can be divided into the two subsets K_1 and K_2, namely

(i) $k \in K_1$ when $x_k \ne x_{k-1}$;

(ii) $k \in K_2$ when $x_k = x_{k-1}$.

Furthermore Proposition 3.3 implies that K_2 is infinite.
Now it is possible to define the following two set of indices:

$$\begin{aligned} J_1 &= \{0, \dots, \bar{j}\} \cap K_1 \quad \text{succ. iterations up to } \bar{j}, \\ J_2 &= \{0, \dots, \bar{j}\} \cap K_2 \quad \text{unsucc. iterations up to } \bar{j}, \end{aligned}$$

and to rewrite (23):

$$f_{\min} - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{k \in J_1} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 - \tilde{c}_{\text{LAM}} \sum_{k \in J_2} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2. \tag{24}$$

For all $k \in J_1$ we associate an index $m(k)$ given by:

(i) if $J_2 \cap \{0, \dots, k-1\} \neq \emptyset$ then $m(k)$ in the biggest index of $J_2 \cap \{0, \dots, k-1\}$;

(ii) if $J_2 \cap \{0, \dots, k-1\} = \emptyset$ then $m(k) = -1$.

The steps of LAM ensure that, for all $k \in J_1$,

$$\max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \geq \max_{i=1,\dots,n} \{\tilde{\alpha}_{m(k)+1}^i\}^2. \tag{25}$$

Using these inequalities in (24), we obtain

$$f_{\min} - \Phi_0 \leq -\tilde{c}_{\text{LAM}} \sum_{k \in J_1} \max_{i=1,\dots,n} \{\tilde{\alpha}_{m(k)+1}^i\}^2 - \tilde{c}_{\text{LAM}} \sum_{k \in J_2} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2. \tag{26}$$

Now recalling (2) of Proposition 3.1, the choices for α_0^i, for $i = 1, \dots, n$ and that $\|\nabla f(x_k)\| > \epsilon$, for all $k = 0, 1, \dots, \bar{j}_\epsilon$, we obtain:

$$\Phi_0 - f_{\min} \geq (\bar{j} + 1)\tilde{c}_{\text{LAM}} \frac{\theta^2}{n(\gamma + L)^2} \epsilon^2.$$

Thus, the number $\bar{j}_\epsilon$ of iterations can be bounded from above by

$$\bar{j}_\epsilon \leq \left\lceil \frac{n(\gamma + L)^2(\Phi_0 - f_{\min})}{\tilde{c}_{\text{LAM}}\, \theta^2} \epsilon^{-2} \right\rceil = \mathcal{O}(n\epsilon^{-2})$$

which concludes the proof. □

Now, we prove the worst case complexity bound for the number of function evaluations.

Proposition 4.2 *Suppose that Assumptions 3.1 and 3.2 hold. Let $\{x_k\}$ be the sequence of point produced by LAM. Given any $\epsilon \in (0, 1)$ and $\tilde{\alpha}_0^i \geq \epsilon$, assume that $\bar{j}_\epsilon + 1$ is the first iteration such that $\|\nabla f(x_{\bar{j}_\epsilon+1})\| \leq \epsilon$, i.e. $\|\nabla f(x_k)\| > \epsilon$, for all $k = 0, 1, \dots, \bar{j}_\epsilon$. Then, the number of function evaluations Nf_ϵ required by LAM until the $\bar{j}_\epsilon$-th iteration are in the worst case such that*

$$Nf_\epsilon \leq 2n \left\lceil \frac{nc_1^2(\Phi_0 - f_{\min})}{\tilde{c}_{LAM}} \epsilon^{-2} \right\rceil + \left\lceil \frac{nc_1^2(f(x_0) - f_{\min})}{\gamma c^2} \frac{\delta^2}{(1-\delta)^2} \epsilon^{-2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}),$$

where c_1 and $\tilde{c}_{LAM}$ are defined in (22) and (16), respectively.

Proof By assumption, for all $k = 0, 1, \dots, \bar{j}$, we have that

$$\epsilon < \|\nabla f(x_k)\|. \tag{27}$$

Let $U_{\bar{J}_\epsilon}$ and $S_{\bar{J}_\epsilon}$ be the index sets of unsuccessful and successful iterations until the iteration $\bar{J}_\epsilon$.
Then, for every iteration k, if $k \in U_{\bar{J}_\epsilon}$, the algorithm performs

$$Nf_k^u = 2n$$

function evaluations.

On the other hand, if $k \in S_{\bar{J}_\epsilon}$, we can distinguish the function evaluations performed by the algorithm in those producing a sufficient decrease in the objective function value and those producing a failure, i.e. the last function evaluation performed by the DF-Linesearch procedure. Hence, in that case we can further distinguish the function evaluations as Nf_k^s and $\overline{Nf}_k^s$. As concerns $\overline{Nf}_k^s$ we have

$$\overline{Nf}_k^s \leq 2n.$$

Concerning Nf_k^s, every time that such a function evaluation is performed, we have by the instructions of the DF-Linesearch procedure that

$$f(y_k^i + \alpha_j \hat{d}) - f(y_k^i + \alpha_j/\delta \hat{d}) \geq \gamma \left(\frac{1-\delta}{\delta}\right)^2 \alpha_j^2 \geq \gamma \left(\frac{1-\delta}{\delta}\right)^2 c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2.$$

Now, for all $k \in S_{\bar{J}_\epsilon}$ we can define an index $\tilde{m}(k)$ given by:

(i) if $k > 0$ and $U_{\bar{J}_\epsilon} \cap \{0, \dots, k-1\} \neq \emptyset$ then $\tilde{m}(k)$ in the biggest index of $U_{\bar{J}_\epsilon} \cap \{0, \dots, k-1\}$;

(ii) if $k = 0$ or $U_{\bar{J}_\epsilon} \cap \{0, \dots, k-1\} = \emptyset$ then $\tilde{m}(k) = 0$.

Then, we obtain:

$$\begin{aligned} f(y_k^i + \alpha_j \hat{d}) - f(y_k^i + \alpha_j/\delta \hat{d}) &\geq \gamma \left(\frac{1-\delta}{\delta}\right)^2 c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_{m(k)}^i\}^2 \\ &\geq \gamma \left(\frac{1-\delta}{\delta}\right)^2 c^2 \frac{\epsilon^2}{nc_1^2}. \end{aligned}$$

Then, recalling that f is bounded from below by $f_{\min}$, summing the above relation over all such function evaluations, we obtain

$$f_0 - f_{\min} \geq Nf_k^s \gamma \left(\frac{1-\delta}{\delta}\right)^2 c^2 \frac{\epsilon^2}{nc_1^2},$$

so that

$$Nf_k^s \leq \frac{c_1^2 n (f_0 - f_{\min})}{\gamma c^2 \epsilon^2} \frac{\delta^2}{(1-\delta)^2}.$$

Finally, recalling the the number of iterations performed by the algorithm is bounded by $\mathcal{O}(\epsilon^{-2})$, and denoting by Nf_ϵ the total number of function evaluations performed by the algorithm, we can write in the worst case

$$Nf_\epsilon \le 2n \left\lceil \frac{nc_1^2(\Phi_0 - f_{\min})}{\tilde{c}_{\text{LAM}}} \epsilon^{-2} \right\rceil + \left\lceil \frac{nc_1^2(f_0 - f_{\min})}{\gamma c^2 \epsilon^2} \frac{\delta^2}{(1-\delta)^2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}), \tag{28}$$

where c_1 and $\tilde{c}_{\text{LAM}}$ are defined in (22) and (16), respectively. This concludes the proof. □

The previous results show that the linesearch DF approach is also able to propose algorithms with exactly the same complexity bounds of those obtained in [18] for direct search methods.
Instead, the next proposition shows that the use of the linesearch technique allows us to guarantee to a DF algorithm the further property that the number of iterations such that $\|\nabla f(x_k)\| > \epsilon$ is of the order of ϵ^{-2}. In particular, for the LAM algorithm, the number of such iterations is $\mathcal{O}(n^2\,\epsilon^{-2})$ in the worst case.

Proposition 4.3 *Suppose that Assumptions 3.1 and 3.2 hold. Let $\{x_k\}$ be the sequence of point produced by LAM. Given any $\epsilon \in (0,1)$, consider the following subset of indices:*

$$K_\epsilon = \{\, k = 1, \dots :\ \|\nabla f(x_k)\| > \epsilon \,\}. \tag{29}$$

Then,

$$|K_\epsilon| \le \left\lceil \frac{n\, c_2^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}), \tag{30}$$

where $\tilde{c}_{LAM}$ is given by (16) and

$$c_2 = \max\left\{ \frac{\gamma + L(\sqrt{n}+1)}{\delta},\ \frac{\gamma + L}{\theta} \right\}. \tag{31}$$

Proof Proposition 3.2 shows that the sequence $\{\Phi_k\}$ is not increasing and that, for every k, we have:

$$\Phi_k - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{\tilde{k}=1}^{k} \max_{i=1,\dots,n} \{\tilde{\alpha}_{\tilde{k}}^i\}^2 = -\tilde{c}_{\text{LAM}} \sum_{\tilde{k}=0}^{k} \max_{i=1,\dots,n} \{\tilde{\alpha}_{\tilde{k}+1}^i\}^2. \tag{32}$$

Since the sequence $\{\Phi_k\}$ is bounded from below, Φ^* exists such that:

$$\lim_{k\to\infty} \Phi_k = \Phi^* \ge f_{\min}.$$

Taking the limit for $k \to \infty$ in (32) we obtain:

$$\Phi_0 - \Phi^* \geq \tilde{c}_{\text{LAM}} \sum_{k=0}^{\infty} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \geq \tilde{c}_{\text{LAM}} \sum_{k \in K_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2,$$

which, together with the definition of K_ϵ and Proposition 3.1, gives

$$\Phi_0 - f_{\min} \geq \tilde{c}_{\text{LAM}} \sum_{k \in K_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \geq |K_\epsilon| \tilde{c}_{\text{LAM}} \frac{\epsilon^2}{n c_2^2}.$$

Thus

$$|K_\epsilon| \leq \left\lceil \frac{n c_2^2 (\Phi_0 - f_{\min})}{\tilde{c}_{\text{LAM}}} \epsilon^{-2} \right\rceil = \mathcal{O}(n^2 \epsilon^{-2})$$

and the proof is concluded. □

5 Other Types of Linesearch Methods

In this section we propose two variants of the LAM algorithm which enjoy similar complexity bounds as those of LAM. In particular, the new algorithms have complexity bounds which are of the order of ϵ^{-2} and differ only on the dependence on the size n of the problem. These differences derive from the particular strategies used by the algorithms for updating the tentative steps $\tilde{\alpha}_k^i$, for $i = 1, \dots, n$ and the current point x_k. As seen before, LAM reduces the tentative steps only when all of them fail to produce a sufficient decrease of the objective function along the corresponding directions. Regarding the points, we recall that LAM accepts every point y_k^{i+1}, provided that a sufficient decrease of the objective function has been obtained, as a possible candidate to become the next point x_{k+1}. Roughly speaking, we can say that LAM updates the point by exploiting the information obtained along a single direction while it updates the tentative steps only when it has obtained information on all directions.
Of course, different approaches can be used to define derivative-free linesearch-type algorithms. As examples, we describe two algorithms which, differently from LAM, use the same strategy to update the trial steps and the current point. In particular, LAM1 differs from LAM since it updates every tentative step on the basis of information obtained only on the corresponding direction. Whereas, LAM2 differs from LAM since it exploits the information obtained from all the direction to compute the new current point.

In the following, for LAM1 and LAM2, we report the theoretical results concerning the complexity of the two algorithms. Since the proofs of these results are very similar to the analogous results of LAM, we report them in the appendix for the sake of completeness.

Linesearch Algorithm Model 1 (LAM1)

Data. $c \in (0,1), \theta \in (0,1), x_0 \in \mathbb{R}^n, \tilde{\alpha}_0^i > 0, i \in \{1,\dots,n\}$, and set $d_0^i = e^i$, for $i = 1,\dots,n$.

For $k = 0, 1, \dots$
  Set $y_k^1 = x_k$.
  For $i = 1,\dots,n$
    Let $\bar{\alpha}_k^i = \max\{\tilde{\alpha}_k^i, c \max_{j=1,\dots,n} \{\tilde{\alpha}_k^j\}\}$.
    Compute α and d by the `DF-Linesearch`$(\bar{\alpha}_k^i, y_k^i, d_k^i; \alpha, d)$.
    If $\alpha = 0$ **then** set $\alpha_k^i = 0$ and $\tilde{\alpha}_{k+1}^i = \theta\bar{\alpha}_k^i$.
    else set $\alpha_k^i = \alpha$, $\tilde{\alpha}_{k+1}^i = \alpha$.
    Set $d_{k+1}^i = d$, $y_k^{i+1} = y_k^i + \alpha_k^i d_{k+1}^i$.
  End For
  Set $x_{k+1} = y_k^{n+1}$.
End For

We remark that tentative step α_k^i is decreased as soon as it fails to sufficiently decrease the objective function along the direction d_k^i.
By repeating the same analysis made for LAM it is possible to state the following proposition.

Proposition 5.1 *Suppose that Assumption 3.1 holds. Let $\{x_k\}$ be the sequence produced by LAM1. Then, for each k such that $x_{k+1} \neq x_k$*

$$\|\nabla f(x_k)\| \le \sqrt{n}\left(\frac{\gamma + L(\sqrt{n}+1)}{\min\{\theta, \delta\}}\right)\max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\},$$

whereas, for each k such that $x_{k+1} = x_k$

$$\|\nabla f(x_k)\| \le \sqrt{n}\frac{\gamma + L}{\theta}\max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}.$$

Proof See "Appendix A". □

Next we prove that we can bound the difference $\Phi_{k+1} - \Phi_k$ with the quantity $\max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}^2$.

Proposition 5.2 *Let $\{x_k\}$ and $\{\alpha_k^i\}$, $i = 1,\dots,n$, be the sequences produced by LAM1. Then for all $k = 0, 1, \dots$,:*

$$\Phi_{k+1} - \Phi_k \le -\tilde{c}_{LAM}\max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}^2, \tag{33}$$

where $\tilde{c}_{LAM}$ is defined in (16).

Proof See "Appendix A". □

Finally, the following proposition gives the asymptotic convergence property for LAM1 along with the worst case complexity bounds. Specifically, we give complexity bounds in terms of iterations and function evaluations to achieve a norm of the gradient below a given threshold and number of iterations for which the norm of the gradient is above a given threshold.

Proposition 5.3 *Suppose that Assumptions 3.1 and 3.2 hold. Then, LAM1 produces an infinite sequence $\{x_k\}$ such that*

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

Given any $\epsilon \in (0,1)$ let $\bar{j}_\epsilon + 1$ be the first iteration such that $\|\nabla f(x_{\bar{j}_\epsilon+1})\| \le \epsilon$ and K_ϵ be the set

$$K_\epsilon = \{\, k = 1, \dots : \; \|\nabla f(x_k)\| > \epsilon \,\}.$$

Then:

$$\begin{aligned}
\bar{j}_\epsilon \; &\le \left\lceil \frac{n\, c_3^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}),\\
|K_\epsilon| &\le \left\lceil \frac{n\, c_3^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}),\\
Nf_\epsilon &\le 2n \left\lceil \frac{nc_3^2(\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\epsilon^{-2} \right\rceil + \left\lceil \frac{nc_3^2(f(x_0) - f_{\min})}{\gamma c^2} \frac{\delta^2}{(1-\delta)^2}\epsilon^{-2} \right\rceil = \mathcal{O}(n^3\epsilon^{-2}),
\end{aligned}$$

where Nf_ϵ is the number of function evaluations required by LAM1 until the $\bar{j}_\epsilon$-th iteration,

$$c_3 = \max\left\{ \left(\frac{\gamma + L(\sqrt{n}+1)}{\min\{\theta,\delta\}}\right), \left(\frac{\gamma + L}{\theta}\right) \right\}, \tag{34}$$

and $\tilde{c}_{LAM}$ is given by (16).

Proof See "Appendix A". □

We note that the complexity bounds of LAM1 are worse than those obtained for LAM. In fact LAM1 achieves a norm of the gradient less than ϵ in at most $\mathcal{O}(n^2\epsilon^{-2})$ iterations and $\mathcal{O}(n^3\epsilon^{-2})$ function evaluations in the worst case. Therefore a larger freedom of movement of LAM1 can positively impact the efficiency of the overall scheme but, on the other hand, it makes the theoretical analysis of the algorithm more complex. In particular, it requires bounding the norm of the gradient even in the successful iterations and this introduces a $(\sqrt{n})^2$ rather than $\sqrt{n}$ in the coefficient used in the bound of the gradient.

Next, we present a further algorithm model, namely LAM2.

Linesearch Algorithm Model 2 (LAM2)

Data. $c \in (0,1), \theta \in (0,1), x_0 \in \mathbb{R}^n, \tilde{\alpha}_0^i > 0, i \in \{1,\dots,n\}$, and set $d_0^i = e^i$, for $i = 1,\dots,n$.

For $k = 0, 1, \dots$
  For $i = 1,\dots,n$
    Let $\bar{\alpha}_k^i = \max\{\tilde{\alpha}_k^i, c \max_{j=1,\dots,n} \{\tilde{\alpha}_k^j\}\}$.
    Compute α and d by the DF-Linesearch$(\bar{\alpha}_k^i, x_k, d_k^i; \alpha, d)$.

    If $\alpha = 0$ **then** set $\alpha_k^i = 0$ and $\tilde{\alpha}_{k+1}^i = \theta\bar{\alpha}_k^i$.
    else set $\alpha_k^i = \alpha$, $\tilde{\alpha}_{k+1}^i = \alpha$.

    Set $d_{k+1}^i = d$.

  End For
  Set $x_{k+1} = \operatorname{argmin}_{i=1,\dots,n}\{f(x_k + \alpha_k^i d_{k+1}^i)\}$.
End For

LAM2 always explores all the search directions starting from x_k, and then chooses the best point to define the new iterate. The theoretical analysis of LAM2 can be done in a similar way to that of LAM. In particular we can state the following propositions.

For LAM2, in the next proposition we show that the norm of the gradient can be bounded in each iteration.

Proposition 5.4 *Suppose that Assumption 3.1 holds. Let $\{x_k\}$ be the sequence produced by the LAM2 framework. Then, for each k*

$$\|\nabla f(x_k)\| \le \sqrt{n}\left(\frac{\gamma + L}{\min\{\delta, \theta\}}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}. \tag{35}$$

Proof See "Appendix B". □

Next, we show that the difference $\Phi_{k+1} - \Phi_k$ can be bounded with the maximum of the tentative stepsizes.

Proposition 5.5 *Let $\{x_k\}$ and $\{\tilde{\alpha}_k^i\}$, $i = 1,\dots,n$, be the sequences produced by LAM2. Then for all $k = 0, 1, \dots$*

$$\Phi_{k+1} - \Phi_k \le -\tilde{c}_{LAM} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2, \tag{36}$$

where $\tilde{c}_{LAM}$ is defined in (16).

Proof See "Appendix B". □

Proposition 5.6 *Suppose that Assumptions 3.1 and 3.2 hold. Then, LAM2 produces an infinite sequence* $\{x_k\}$ *such that*

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

Given any $\epsilon \in (0,1)$ *let* $\bar{j}_\epsilon + 1$ *be the first iteration such that* $\|\nabla f(x_{\bar{j}_\epsilon+1})\| \le \epsilon$ *and* K_ϵ *be the set*

$$K_\epsilon = \{\, k = 1, \dots :\ \|\nabla f(x_k)\| > \epsilon \,\}.$$

Then:

$$\begin{aligned}
\bar{j}_\epsilon \ &\le \left\lceil \frac{n\, c_4^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n\epsilon^{-2}),\\
|K_\epsilon| &\le \left\lceil \frac{n\, c_4^2\, (\Phi_0 - f_{\min})}{\tilde{c}_{LAM}}\, \epsilon^{-2} \right\rceil = \mathcal{O}(n\epsilon^{-2}),\\
Nf_\epsilon &\le 2n \left\lceil \frac{nc_4^2(\Phi_0 - f_{\min})}{\tilde{c}_{LAM}} \epsilon^{-2} \right\rceil + \left\lceil \frac{nc_4^2(f(x_0) - f_{\min})}{\gamma c^2} \frac{\delta^2}{(1-\delta)^2} \epsilon^{-2} \right\rceil = \mathcal{O}(n^2\epsilon^{-2}),
\end{aligned}$$

where Nf_ϵ *is the number of function evaluations required by LAM2 until the* $\bar{j}_\epsilon$*-th iteration and* $\tilde{c}_{LAM}$ *is given by* (16) *and*

$$c_4 = \frac{\gamma + L}{\min\{\theta, \delta\}}. \tag{37}$$

Proof See "Appendix B". □

6 Numerical Experiments

In this section we report the results of a brief numerical comparison between the proposed algorithms and another well-established derivative-free method. In particular, we consider the variant of the Nelder–Mead algorithm [9, 16] implemented in the Python library `scipy`. We acknowledge that the comparison is very partial and preliminary but the main aim of the paper is the theoretical analysis of linesearch based derivative-free algorithms and the derivation of worst complexity bounds. The numerical experiments are inserted with the sole purpose of showing that the studied methods have numerical performances that align with other commonly used derivative-free algorithms. A thorough and extensive numerical comparison with other state-of-the-art derivative-free solvers is out of the scope of the present paper.

To carry out the numerical experiments, we selected a small set of problems from the CUTEst collection [11]. In particular, we extracted the unconstrained problems with more than 10 variables and such that the number of variables is not parametric. The set of problems with their dimensions is reported in Table 1.

LAM, LAM1 and LAM2 were run using the following values for the parameters.

$$c = 10^{-10},\ \theta = \delta = 0.5,\ \gamma = 10^{-6}, \quad \tilde{\alpha}_0^i = 1, \quad \forall\, i = 1, \dots, n.$$

Table 1 Set of CUTEst test problems used in the comparison

Problem	n
3PK	30
BA-L1LS	57
BA-L1SPLS	57
BQPGABIM	46
BQPGASIM	50
COATING	134
DECONVB	51
DECONVBNE	51
DECONVNE	51
DECONVU	51
DIAMON2DLS	66
DIAMON3DLS	99
DMN15103LS	99
DMN15332LS	66
DMN15333LS	99
DMN37142LS	66
DMN37143LS	99
HATFLDC	25
HATFLDCNE	25
HATFLDGLS	25
HOLMES	180
HYDC20LS	99
HYDCAR6LS	29
METHANB8LS	31
METHANL8LS	31
MINSURF	36
PARKCH	15
SANTALS	21
STRATEC	10
TOINTGOR	50
TOINTPSP	50
TOINTQOR	50

All the codes were run allowing a maximum of 10000 function evaluations. For the LAM's methods we also stop the codes when $\max_{i=1,\ldots,n}\{\tilde{\alpha}_k^i\} \leq 10^{-5}$.

The results are reported in terms of data profiles [15] with values of the precision parameter $\tau \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$.

From Fig. 1, we can see that all the LAM's algorithms dominate Nelder-Mead algorithm. Furthermore, from the data profiles, we can also notice that LAM1 is numerically the method of choice. The second-best method is LAM, even though at high precision levels, LAM and LAM2 are almost comparable.

7 Conclusions

In this paper we are concerned about the worst case complexity of linesearch-based derivative-free algorithms for the unconstrained optimization of a black-box objective function.

First, we considered a particular algorithm model, namely LAM, based on a suitable derivative-free linesearch procedure. We managed to show that the algorithm model takes at most $\mathcal{O}(n\epsilon^{-2})$ iterations and $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations to drive the norm of the gradient below ϵ and that produces at most $\mathcal{O}(n^2\epsilon^{-2})$ iterations where the norm of the gradient is above ϵ.

Then, to generalize the proposed analysis, we consider two other linesearch derivative-free algorithms LAM1 and LAM2. They differ from LAM because they use different strategies to exploit the information obtained by the samplings of the objective function along the coordinate axes. We prove that also LAM1 and LAM2 have similar theoretical properties. In the following table we summarize the complexity results for the three algorithms, namely LAM, LAM1 and LAM2.

	$\bar{J}_\epsilon$	$\lvert K_\epsilon \rvert$	Nf_ϵ
LAM	$\mathcal{O}(n\epsilon^{-2})$	$\mathcal{O}(n^2\epsilon^{-2})$	$\mathcal{O}(n^2\epsilon^{-2})$
LAM1	$\mathcal{O}(n^2\epsilon^{-2})$	$\mathcal{O}(n^2\epsilon^{-2})$	$\mathcal{O}(n^3\epsilon^{-2})$
LAM2	$\mathcal{O}(n\epsilon^{-2})$	$\mathcal{O}(n\epsilon^{-2})$	$\mathcal{O}(n^2\epsilon^{-2})$

LAM

Name	Constant value	used in
	$\dfrac{\gamma + L(\sqrt{n}+1)}{\delta}$	(1)
	$\dfrac{\gamma + L}{\theta}$	(2)
	η	(12) and (13)
$\tilde{c}_{\text{LAM}}$	$\min\left\{\eta\left(\dfrac{1-\theta^2}{\theta^2}\right), \gamma c^2, \left(\gamma(1-\delta)^2 - \eta\right)\right\}$	(15) and (16)
c_1	$\dfrac{\gamma + L}{\theta}$	(21) and (22)
c_2	$\max\left\{\left(\dfrac{\gamma + L(\sqrt{n}+1)}{\delta}\right), \left(\dfrac{\gamma + L}{\theta}\right)\right\}$	(30) and (31)

LAM1

Name	Constant value	used in
c_3	$\max\left\{\left(\dfrac{\gamma + L(\sqrt{n}+1)}{\min\{\theta, \delta\}}\right), \left(\dfrac{\gamma + L}{\theta}\right)\right\}$	(34)

LAM2

Name	Constant value	used in
c_4	$\dfrac{\gamma + L}{\min\{\theta, \delta\}}$	(37)

We note that LAM2 has the same complexity bounds of LAM both for the number of iterations and the number of function evaluations needed to obtain a point where the norm of the gradient of the objective function is below a given threshold. LAM2

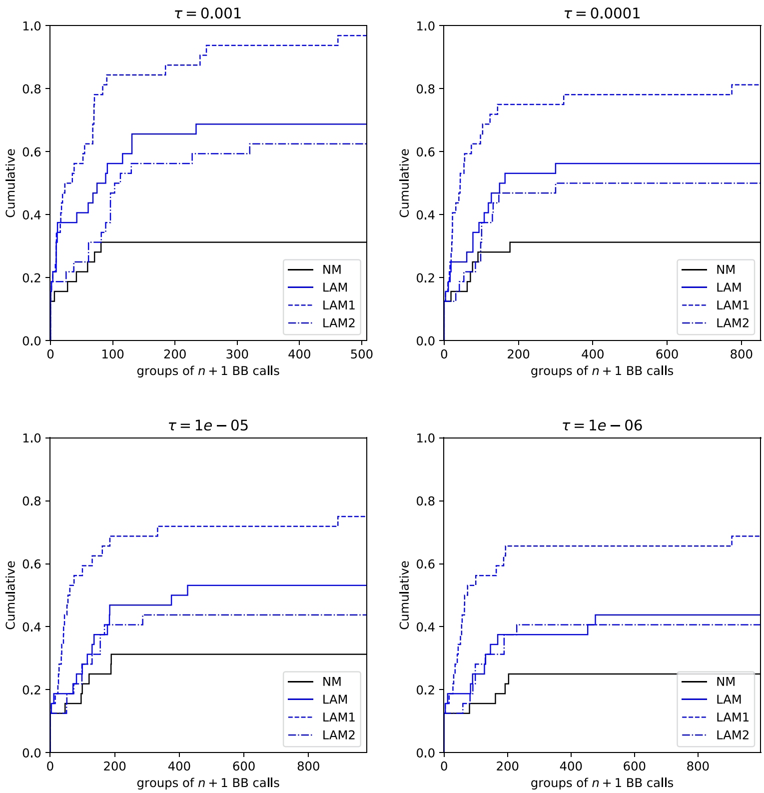

Fig. 1 Data profiles

is better in terms of bound of the number of iterations with norm of the gradient above a given threshold. This bound is $\mathcal{O}(n\epsilon^{-2})$ in the worst case.
Even though LAM2 is better than LAM in the worst case, it can be less computationally efficient than LAM on average. Indeed, the sampling technique of LAM2 is such that a complete exploration starting from the current iterate has to be performed before the selection of the new iterate. This, in many situations, can be too costly. On the contrary, LAM changes the iterate as soon as a sufficiently improving new point is detected by the linesearch.

We also report the results of a brief numerical experimentation on a small set of problems from the CUTEst collection [11]. On the basis of this experimentation, we can say that:

(i) the three versions of LAM compare favorably with a widely used derivative-free methods, i.e. the Nelder-Mead algorithm;
(ii) among the three versions of LAM, the most efficient one is LAM1, which was largely expected; LAM and LAM2 are very similar with LAM slightly dominating LAM2.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Data availability statement The problems used in this study are available as part of the CUTEst collection of problems [11]. Data sets, codes and results generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

A Proofs of Propositions for LAM1

Proof of Proposition 5.1 The proof is very similar to the proof of Proposition 3.1. The only difference being the case of an iteration k such that $x_{k+1} \neq x_k$ and for an index $i = 1, \dots, n$ such that $\alpha_k^i = 0$, and $\tilde{\alpha}_{k+1}^i = \theta \bar{\alpha}_k^i$. In this situation, we have:

$$f(y_k^i + \bar{\alpha}_k^i e_i) > f(y_k^i) - \gamma (\bar{\alpha}_k^i)^2,$$
$$f(y_k^i - \bar{\alpha}_k^i e_i) > f(y_k^i) - \gamma (\bar{\alpha}_k^i)^2.$$

Then we get from the Mean Value Theorem

$$\nabla f(u_k^i)^T e_i > -\gamma \bar{\alpha}_k^i, \tag{38}$$
$$\nabla f(v_k^i)^T e_i < \gamma \bar{\alpha}_k^i, \tag{39}$$

where $u_k^i = y_k^i + \lambda_k^i \bar{\alpha}_k^i e_i$ and $v_k^i = y_k^i - \mu_k^i \bar{\alpha}_k^i e_i$ with $\lambda_k^i, \mu_k^i \in (0, 1)$. From (38) and (39) and the Lipschitz continuity of ∇f, we have that

$$\nabla f(x_k)^T e_i > -\gamma \bar{\alpha}_k^i - L\|x_k - u_k^i\| > -\gamma \bar{\alpha}_k^i - L\|x_k - y_k^i\| - L\bar{\alpha}_k^i,$$
$$\nabla f(x_k)^T e_i < \gamma \bar{\alpha}_k^i + L\|x_k - v_k^i\| < \gamma \bar{\alpha}_k^i + L\|x_k - y_k^i\| + L\bar{\alpha}_k^i.$$

Hence

$$|\nabla f(x_k)^T e_i| < (\gamma + L)\bar{\alpha}_k^i + L\|x_k - y_k^i\| \leq (\gamma + L)\bar{\alpha}_k^i + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\} =$$
$$(\gamma + L)\frac{\tilde{\alpha}_{k+1}^i}{\theta} + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}$$
$$\leq \frac{(\gamma + L)}{\theta} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\} + L\sqrt{n} \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\},$$

so that we can finally write

$$|\nabla f(x_k)^T e_i| \leq \left(\frac{\gamma + L(\sqrt{n} + 1)}{\theta}\right) \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}. \tag{40}$$

Thus, recalling (40) and (9) (from the proof of Proposition 3.1 case (ii)), when $x_{k+1} \neq x_k$ we can write

$$\|\nabla f(x_k)\| \leq \sqrt{n}\left(\frac{\gamma + L(\sqrt{n} + 1)}{\min\{\theta, \delta\}}\right) \max_{i=1,\dots,n}\{\tilde{\alpha}_{k+1}^i\}.$$

The proof is then concluded noting that, given the definition of LAM1, when $x_{k+1} = x_k$ we can prove the bound on the norm of the gradient by the same reasonings as in the proof of Proposition 3.1. □

Proof of Proposition 5.2 If the $(k-1)$-th iteration is of failure, then

(i) $f(x_k) = f(x_{k-1})$;
(ii) $\tilde{\alpha}_k^i = \theta\bar{\alpha}_{k-1}^i$, for every $i = 1, \dots, n$.

In particular, since, for every $i = 1, \dots, n$, $\bar{\alpha}_{k-1}^i = \max\{\tilde{\alpha}_{k-1}^i, c \max_{j=1,\dots,n}\{\tilde{\alpha}_{k-1}^j\}\}$, we have either

$$\tilde{\alpha}_k^i = \theta\tilde{\alpha}_{k-1}^i \leq \theta \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$$

or

$$\tilde{\alpha}_k^i = \theta c \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\} \leq \theta \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}.$$

Hence, in case of failure, we can write $\tilde{\alpha}_k^i \leq \theta \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$, so that $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} \leq \theta \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$. Then, from

$$\Phi_k - \Phi_{k-1} = f(x_k) - f(x_{k-1}) + \eta\gamma\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)$$

and

$$-\frac{1}{\theta^2}\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 \geq -\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2$$

we obtain

$$\begin{aligned}\Phi_k - \Phi_{k-1} &= \eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &\leq \eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \frac{1}{\theta^2}\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2\right),\end{aligned}$$

so that

$$\Phi_k - \Phi_{k-1} \leq -\eta\left(\frac{1-\theta^2}{\theta^2}\right)\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}.$$

Let us now consider the case when $(k-1)$-th iteration is of "success", i.e. $x_k \neq x_{k-1}$ so that $f(x_k) < f(x_{k-1})$. Then, we consider the following three cases:

1. $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} = \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$;
2. $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} > \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$;
3. $\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\} < \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}$.

Case 1. Since the $(k-1)$-th iteration is of success, there is an index i such that the following holds

$$f(y_{k-1}^{i+1}) \leq f(y_{k-1}^i) - \gamma(\bar{\alpha}_{k-1}^i)^2$$

since the extrapolation cycle gets started when

$$f(y_{k-1}^i + \bar{\alpha}_{k-1}^i\hat{d}) \leq f(y_{k-1}^i) - \gamma(\bar{\alpha}_{k-1}^i)^2.$$

Then, considering that

$$\bar{\alpha}_{k-1}^i = \max\{\tilde{\alpha}_{k-1}^i, c\max_{j=1,\dots,n}\{\tilde{\alpha}_{k-1}^j\}\} \geq c\max_{j=1,\dots,n}\{\tilde{\alpha}_{k-1}^j\},$$

we have

$$f(y_{k-1}^{i+1}) \leq f(y_{k-1}^i) - \gamma(\bar{\alpha}_{k-1}^i)^2 \leq f(y_{k-1}^i) - \gamma c^2\max_{j=1,\dots,n}\{\tilde{\alpha}_{k-1}^j\}^2.$$

Then, recalling that we are in case 1,

$$f(y_{k-1}^{i+1}) \leq f(y_{k-1}^i) - \gamma c^2\max_{j=1,\dots,n}\{\tilde{\alpha}_k^j\}^2.$$

Moreover, since by definition $f(x_k) \leq f(y_{k-1}^{n+1})$ and $f(y_{k-1}^i) \leq f(x_{k-1})$, we can write

$$f(x_k) \leq f(y_{k-1}^{i+1}) \leq f(y_{k-1}^i) - \gamma c^2 \max_{j=1,\ldots,n} \{\tilde{\alpha}_k^j\}^2 \leq f(x_{k-1}) - \gamma c^2 \max_{j=1,\ldots,n} \{\tilde{\alpha}_k^j\}^2.$$

Then, we have

$$\Phi_k - \Phi_{k-1} = f(x_k) - f(x_{k-1}) \leq -\gamma c^2 \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2.$$

Case 2. Since $\max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\} > \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}$, we have that an index $\bar{j}$ exists such that

$$\max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\} = \tilde{\alpha}_k^{\bar{j}}$$

and a linesearch has been performed along the $\bar{j}$-th direction, so that (repeating the same reasonings as in the proof of Proposition 3.3) we have

$$f(y_{k-1}^{\bar{j}+1}) \leq f(y_{k-1}^{\bar{j}}) - \gamma(1-\delta)^2 (\tilde{\alpha}_k^{\bar{j}})^2.$$

Hence,

$$f(x_k) - f(x_{k-1}) \leq -\gamma(1-\delta)^2 (\tilde{\alpha}_k^{\bar{j}})^2,$$

so that

$$\begin{aligned}
\Phi_k - \Phi_{k-1} &= f(x_k) - f(x_{k-1}) + \eta \left(\max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\
&\leq -\gamma(1-\delta)^2 \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2 + \eta \left(\max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\
&\leq -\gamma(1-\delta)^2 \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2 + \eta \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2 \\
&\leq -(\gamma(1-\delta)^2 - \eta) \max_{i=1,\ldots,n} \{\tilde{\alpha}_k^i\}^2.
\end{aligned}$$

Case 3. In this case, we know that an index $\bar{\imath}$ exists such that

$$\tilde{\alpha}_k^{\bar{\imath}} \geq \bar{\alpha}_{k-1}^{\bar{\imath}} = \max\{\tilde{\alpha}_{k-1}^{\bar{\imath}}, c \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}\} \geq c \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}.$$

Then, we can write (recalling that the iteration is of success)

$$f(x_k) \leq f(x_{k-1}) - \gamma c^2 \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k-1}^i\}^2.$$

Hence, we have

$$
\begin{aligned}
\Phi_k - \Phi_{k-1} &= f(x_k) - f(x_{k-1}) + \eta \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \eta \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \\
&\le -\gamma c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 + \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\
&= -\gamma c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 + \gamma c^2 \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\
&\quad + \eta \left(\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}^2 \right) \\
&< -\gamma c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2,
\end{aligned}
$$

where the last inequality follows from the fact that we are in case 3. Hence, recalling the above three cases, for all k we can always write

$$
\Phi_k - \Phi_{k-1} \le -\tilde{c}_{\text{LAM}} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2,
$$

where $\tilde{c}_{\text{LAM}}$ is defined in (16), thus concluding the proof. □

Proof of Proposition 5.3 For all iterations k, by Proposition 5.2, we have

$$
\Phi_k - \Phi_{k-1} \le -\tilde{c}_{\text{LAM}} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2,
$$

where $\tilde{c}_{\text{LAM}}$ is defined in (16). Recalling the $\Phi_k \ge f_{\min}$, it results

$$
\lim_{k\to\infty} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} = 0.
$$

Hence, by Proposition 5.1, we obtain

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

For all iterations $k = 0, 1, \dots, \bar{j}_\epsilon$, by Proposition 5.1, we have that

$$
\epsilon < \|\nabla f(x_k)\| \le c_3 \sqrt{n} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}. \tag{41}
$$

Hence,

$$
\max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\} \ge \frac{\epsilon}{c_3 \sqrt{n}}, \tag{42}
$$

Then, considering that

$$
\Phi_{\bar{j}_\epsilon+1} - \Phi_0 = (\Phi_{\bar{j}_\epsilon+1} - \Phi_{\bar{j}_\epsilon}) + (\Phi_{\bar{j}_\epsilon} - \Phi_{\bar{j}_\epsilon-1}) + \cdots + (\Phi_1 - \Phi_0)
$$

and recalling (33) we can write

$$\Phi_{\bar{j}_\epsilon+1} - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{k=1}^{\bar{j}_\epsilon+1} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 = -\tilde{c} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2.$$

By recalling (14), we can write

$$f_{\min} - \Phi_0 \le \Phi_{\bar{j}_\epsilon} - \Phi_0 \le -\tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2. \tag{43}$$

Now, by (42), we have

$$\max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \ge \frac{\epsilon^2}{c_3^2 n}, \quad \text{for } k = 0, 1, \dots, \bar{j}_\epsilon,$$

and, from (43), we can write

$$\Phi_0 - f_{\min} \ge \tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \ge (\bar{j}_\epsilon + 1)\tilde{c}_{\text{LAM}} \frac{\epsilon^2}{n c_3^2}.$$

Thus, the number of iterations $\bar{j}_\epsilon$ can be bounded from above by

$$\bar{j}_\epsilon \le \left\lceil \frac{n c_3^2 (\Phi_0 - f_{\min})}{\tilde{c}_{\text{LAM}}} \epsilon^{-2} \right\rceil = \mathcal{O}(n^2 \epsilon^{-2}).$$

The proof is concluded noting that the bounds on $|K_\epsilon|$ and Nf_ϵ can be obtained by the same reasoning as those carried out in the proof of Propositions 4.3 and 4.2, respectively. □

B Proofs of Propositions for LAM2

Proof of Proposition 5.4 For each iteration k such that $x_{k+1} \neq x_k$ and every index $i = 1, \dots, n$, one of two cases can occur:

Case (i), $\alpha_k^i = 0$. By $\alpha_k^i = 0$, and $\tilde{\alpha}_{k+1}^i = \theta \bar{\alpha}_k^i$, we have:

$$f(x_k + \bar{\alpha}_k^i e_i) > f(x_k) - \gamma (\bar{\alpha}_k^i)^2,$$
$$f(x_k - \bar{\alpha}_k^i e_i) > f(x_k) - \gamma (\bar{\alpha}_k^i)^2.$$

Then we get from the Mean Value Theorem

$$\nabla f(u_k^i)^T e_i > -\gamma \bar{\alpha}_k^i, \tag{44}$$
$$\nabla f(v_k^i)^T e_i < \gamma \bar{\alpha}_k^i, \tag{45}$$

where $u_k^i = x_k + \lambda_k^i \bar{\alpha}_k^i e_i$ and $v_k^i = x_k - \mu_k^i \bar{\alpha}_k^i e_i$ with $\lambda_k^i, \mu_k^i \in (0, 1)$. From (44) and (45) and the Lipschitz continuity of ∇f, we have that

$$\begin{aligned} \nabla f(x_k)^T e_i &> -\gamma \bar{\alpha}_k^i - L\|x_k - u_k^i\| > -\gamma \bar{\alpha}_k^i - L\bar{\alpha}_k^i, \\ \nabla f(x_k)^T e_i &< \quad \gamma \bar{\alpha}_k^i + L\|x_k - v_k^i\| < \quad \gamma \bar{\alpha}_k^i + L\bar{\alpha}_k^i. \end{aligned}$$

Hence

$$|\nabla f(x_k)^T e_i| < (\gamma + L)\bar{\alpha}_k^i = (\gamma + L)\, \frac{\tilde{\alpha}_{k+1}^i}{\theta} \leq \left(\frac{\gamma + L}{\theta}\right) \max_{i=1,\ldots,n} \{\tilde{\alpha}_{k+1}^i\}. \quad (46)$$

Case (ii). From $\alpha_k^i = \alpha$, and $\tilde{\alpha}_{k+1}^i = \alpha \geq \bar{\alpha}_k^i$, we results in either

$$\begin{aligned} & f\left(x_k + \frac{\tilde{\alpha}_{k+1}^i}{\delta} e_i\right) > f(x_k + \tilde{\alpha}_{k+1}^i e_i) - \gamma \left(\frac{1}{\delta} - 1\right)^2 (\tilde{\alpha}_{k+1}^i)^2, \\ & f(x_k + \delta \tilde{\alpha}_{k+1}^i e_i) \geq f(x_k + \tilde{\alpha}_{k+1}^i e_i) + \gamma (1-\delta)^2 (\tilde{\alpha}_{k+1}^i)^2 \end{aligned}$$

or

$$\begin{aligned} & f\left(x_k - \frac{\tilde{\alpha}_{k+1}^i}{\delta} e_i\right) > f(x_k - \tilde{\alpha}_{k+1}^i e_i) - \gamma \left(\frac{1}{\delta} - 1\right)^2 (\tilde{\alpha}_{k+1}^i)^2, \\ & f(x_k - \delta \tilde{\alpha}_{k+1}^i e_i) \geq f(x_k - \tilde{\alpha}_{k+1}^i e_i) + \gamma (1-\delta)^2 (\tilde{\alpha}_{k+1}^i)^2. \end{aligned}$$

Then, we get,

$$\nabla f(\bar{u}_k^i)^T e_i > -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i, \quad -\nabla f(\hat{u}_k^i)^T e_i \geq \gamma (1-\delta) \tilde{\alpha}_{k+1}^i, \quad (47)$$

or

$$\nabla f(\bar{v}_k^i)^T e_i < \gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i, \quad -\nabla f(\hat{v}_k^i)^T e_i \leq -\gamma (1-\delta) \tilde{\alpha}_{k+1}^i, \quad (48)$$

where $\bar{u}_k^i = x_k + \bar{\lambda}_k^i \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i e_i$, $\hat{u}_k^i = x_k + \hat{\lambda}_k^i (1-\delta) \tilde{\alpha}_{k+1}^i e_i$, $\bar{v}_k^i = x_k - \bar{\mu}_k^i \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i e_i$, and $\hat{v}_k^i = x_k - \hat{\mu}_k^i (1-\delta) \tilde{\alpha}_{k+1}^i e_i$, with $\bar{\lambda}_k^i, \hat{\lambda}_k^i, \bar{\mu}_k^i, \hat{\mu}_k^i \in (0, 1)$.

When (47) holds, from $\nabla f(\bar{u}_k^i)^T e_i > -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i$ we can write

$$[\nabla f(\bar{u}_k^i) - \nabla f(x_k) + \nabla f(x_k)]^T e_i > -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i,$$

so that we obtain

$$\begin{aligned}\nabla f(x_k)^T e_i &> -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i - L\|x_k - \bar{u}_k^i\| \\ &> -\gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i - L(\frac{1-\delta}{\delta})\tilde{\alpha}_{k+1}^i.\end{aligned} \tag{49}$$

From $\nabla f(\hat{u}_k^i)^T e_i \leq -\gamma(1-\delta)\tilde{\alpha}_{k+1}^i$ in (47), we can write

$$[\nabla f(\hat{u}_k^i) - \nabla f(x_k) + \nabla f(x_k)]^T e_i \leq -\gamma(1-\delta)\tilde{\alpha}_{k+1}^i,$$

so that, in this case, we obtain

$$\begin{aligned}\nabla f(x_k)^T e_i &\leq -\gamma(1-\delta)\tilde{\alpha}_{k+1}^i + L\|x_k - \hat{u}_k^i\| \\ &\leq \gamma \left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i + L\left(\frac{1-\delta}{\delta}\right) \tilde{\alpha}_{k+1}^i.\end{aligned} \tag{50}$$

Now, considering (49) and (50), we get

$$|\nabla f(x_k)^T e_i| \leq \left(\frac{\gamma + L}{\delta}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}. \tag{51}$$

The same bound can be obtained when (48) holds. Thus, recalling (46) and (50), we obtain

$$\|\nabla f(x_k)\| \leq \sqrt{n} \left(\frac{\gamma + L}{\min\{\delta, \theta\}}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}.$$

On the other hand, for each iteration k such that $x_{k+1} = x_k$, i.e. $\alpha_k^i = 0$ for all $i = 1, \dots, n$. In this case, the proof exactly follows the proof of Proposition 3.1 for the LAM algorithm. Thus, in this case we can write

$$|\nabla f(x_k)^T e_i| < \left(\frac{\gamma + L}{\theta}\right) \tilde{\alpha}_{k+1}^i.$$

Hence, we can finally obtain

$$\|\nabla f(x_k)\| \leq \sqrt{n} \left(\frac{\gamma + L}{\min\{\delta, \theta\}}\right) \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\},$$

concluding the proof. □

Proof of Proposition 5.5 We split the set of iteration indices $\{0, 1, 2, \dots\}$ into the two subsets K_1 and K_2, namely

(i) $k \in K_1$ when $x_k \neq x_{k-1}$;

(ii) $k \in K_2$ when $x_k = x_{k-1}$.

and note that the sets K_1 and K_2 cannot be both finite.

Let us first consider the case when $k \in K_2$, i.e. the $(k-1)$-th iteration is of failure, we have

(i) $f(x_k) = f(x_{k-1})$;
(ii) $\tilde{\alpha}_k^i = \theta \bar{\alpha}_{k-1}^i$, for every $i = 1, \dots, n$.

In this case, the same reasoning of Proposition 3.2 can be repeated thus leading us to obtain

$$\Phi_k - \Phi_{k-1} \le -\eta \left(\frac{1-\theta^2}{\theta^2} \right) \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}. \tag{52}$$

Let us now consider the case when $k \in K_1$, i.e. the $(k-1)$-th iteration is of "success". In this case, we have $x_k \neq x_{k-1}$ and $f(x_k) < f(x_{k-1})$. Then, we consider the following three cases:

1. $\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} = \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$;
2. $\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} > \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$;
3. $\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} < \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$.

Case 1. Since the $(k-1)$-th iteration is of success, there is an index i such that the following holds

$$f(x_{k-1} + \bar{\alpha}_{k-1}^i \hat{d}) \le f(x_{k-1}) - \gamma (\bar{\alpha}_{k-1}^i)^2.$$

Then, considering that $f(x_k) < f(x_{k-1})$ and

$$\bar{\alpha}_{k-1}^i = \max\{\tilde{\alpha}_{k-1}^i, c \max_{j=1,\dots,n} \{\tilde{\alpha}_{k-1}^j\}\} \ge c \max_{j=1,\dots,n} \{\tilde{\alpha}_{k-1}^j\},$$

we have

$$f(x_k) \le f(x_{k-1}) - \gamma (\bar{\alpha}_{k-1}^i)^2 \le f(x_{k-1}) - \gamma c^2 \max_{j=1,\dots,n} \{\tilde{\alpha}_{k-1}^j\}^2.$$

Then, we have

$$\Phi_k - \Phi_{k-1} = f(x_k) - f(x_{k-1}) \le -\gamma c^2 \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2.$$

Case 2. Since $\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} > \max_{i=1,\dots,n} \{\tilde{\alpha}_{k-1}^i\}$, we have that an index $\bar{j}$ exists such that

$$\max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} = \tilde{\alpha}_k^{\bar{j}}$$

and a linesearch has been performed along the $\bar{j}$-th direction, so that (repeating the same reasonings as in the proof of Proposition 3.3) we have

$$f(x_k) - f(x_{k-1}) \le -\gamma(1-\delta)^2(\tilde{\alpha}_k^{\bar{j}})^2.$$

Hence,

$$\begin{aligned}\Phi_k - \Phi_{k-1} &= f(x_k) - f(x_{k-1}) + \eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &\le -\gamma(1-\delta)^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 + \eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &\le -\gamma(1-\delta)^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 + \eta\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2\\ &\le -(\gamma(1-\delta)^2-\eta)\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2.\end{aligned}$$

Case 3. In this case, we know that an index $\bar{\imath}$ exists such that

$$\tilde{\alpha}_k^{\bar{\imath}} \ge \bar{\alpha}_{k-1}^{\bar{\imath}} = \max\{\tilde{\alpha}_{k-1}^{\bar{\imath}}, c\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}\} \ge c\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}.$$

Then, we can write (recalling that the iteration is of success)

$$f(x_k) \le f(x_{k-1}) - \gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2.$$

Hence, we have

$$\begin{aligned}\Phi_k - \Phi_{k-1} &= f(x_k) - f(x_{k-1}) + \eta\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \eta\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\\ &\le -\gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2 + \eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &= -\gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 + \gamma c^2\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &\quad +\eta\left(\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2 - \max_{i=1,\dots,n}\{\tilde{\alpha}_{k-1}^i\}^2\right)\\ &< -\gamma c^2\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2,\end{aligned}$$

where the last inequality follows from the fact that we are in case 3. Hence, recalling the above three cases, for all k we can always write

$$\Phi_k - \Phi_{k-1} \le -\tilde{c}_{\text{LAM}}\max_{i=1,\dots,n}\{\tilde{\alpha}_k^i\}^2,$$

where $\tilde{c}_{\text{LAM}}$ is defined in (16). □

Proof of Proposition 5.6 For all iterations k, by Proposition 5.5, we have

$$\Phi_k - \Phi_{k-1} \leq -\tilde{c}_{\text{LAM}} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2,$$

where $\tilde{c}_{\text{LAM}}$ is defined in (16). Recalling the $\Phi_k \geq f_{\min}$, it results

$$\lim_{k\to\infty} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\} = 0.$$

Hence, by Proposition 5.4, we obtain

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

For all iterations $k = 0, 1, \dots, \bar{j}_\epsilon$, by Proposition 5.4, we have that

$$\epsilon < \|\nabla f(x_k)\| \leq c_4\sqrt{n} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}. \tag{53}$$

Hence,

$$\max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\} \geq \frac{\epsilon}{c_4\sqrt{n}}, \tag{54}$$

Then, considering that

$$\Phi_{\bar{j}_\epsilon+1} - \Phi_0 = (\Phi_{\bar{j}_\epsilon+1} - \Phi_{\bar{j}_\epsilon}) + (\Phi_{\bar{j}_\epsilon} - \Phi_{\bar{j}_\epsilon-1}) + \cdots + (\Phi_1 - \Phi_0)$$

and recalling (36) we can write

$$\Phi_{\bar{j}_\epsilon+1} - \Phi_0 \leq -\tilde{c}_{\text{LAM}} \sum_{k=1}^{\bar{j}_\epsilon+1} \max_{i=1,\dots,n} \{\tilde{\alpha}_k^i\}^2 = -\tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2.$$

By recalling (14), we can write

$$f_{\min} - \Phi_0 \leq \Phi_{\bar{j}_\epsilon+1} - \Phi_0 \leq -\tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2. \tag{55}$$

Now, by (54), we have

$$\max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \geq \frac{\epsilon^2}{c_4^2 n}, \quad \text{for } k = 0, 1, \dots, \bar{j}_\epsilon,$$

and, from (55), we can write

$$\Phi_0 - f_{\min} \geq \tilde{c}_{\text{LAM}} \sum_{k=0}^{\bar{j}_\epsilon} \max_{i=1,\dots,n} \{\tilde{\alpha}_{k+1}^i\}^2 \geq (\bar{j}_\epsilon + 1)\tilde{c}_{\text{LAM}} \frac{\epsilon^2}{nc_4^2}.$$

Thus, the number of iterations $\bar{\jmath}$ can be bounded from above by

$$\bar{\jmath}_\epsilon \leq \left\lceil \frac{nc_4^2(\Phi_0 - f_{\min})}{\tilde{c}_{\text{LAM}}} \epsilon^{-2} \right\rceil = \mathcal{O}(n\epsilon^{-2}).$$

The proof is concluded noting that the bounds on $|K_\epsilon|$ and Nf_ϵ can be obtained by the same reasoning as those carried out in the proof of Propositions 4.2 and 4.3, respectively. □

References

1. Amaral, V.S., Andreani, R., Birgin, E.G., Marcondes, D.S., Martínez, J.M.: On complexity and convergence of high-order coordinate descent algorithms for smooth nonconvex box-constrained minimization. J. Glob. Optim. **84**(3), 527–561 (2022)
2. Audet, C., Dennis, J.E., Jr.: Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optim. **17**(1), 188–217 (2006)
3. Audet, C., Dzahini, K.J., Kokkolaras, M., Le Digabel, S.: Stochastic mesh adaptive direct search for blackbox optimization using probabilistic estimates. Comput. Optim. Appl. **79**(1), 1–34 (2021)
4. Cartis, C., Gould, N.I.M., Toint, P.L.: On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization. SIAM J. Optim. **22**(1), 66–86 (2012)
5. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Evaluation Complexity of Algorithms for Nonconvex Optimization: Theory, Computation and Perspectives. SIAM, Philadelphia (2022)
6. Chen, R., Menickelly, M., Scheinberg, K.: Stochastic optimization using a trust-region method and random models. Math. Program. **169**, 447–487 (2018)
7. Dodangeh, M., Vicente, L.N., Zhang, Z.: On the optimal order of worst case complexity of direct search. Optim. Lett. **10**(4), 699–708 (2016)
8. Fasano, G., Liuzzi, G., Lucidi, S., Rinaldi, F.: A linesearch-based derivative-free approach for nonsmooth constrained optimization. SIAM J. Optim. **24**(3), 959–992 (2014)
9. Gao, F., Han, L.: Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Comput. Optim. Appl. **51**(1), 259–277 (2012)
10. Garmanjani, R., Vicente, L.N.: Smoothing and worst-case complexity for direct-search methods in nonsmooth optimization. IMA J. Numer. Anal. **33**(3), 1008–1028 (2012)
11. Gould, N.I.M., Orban, D., Toint, P.L.: Cutest: a constrained and unconstrained testing environment with safe threads for mathematical optimization. Comput. Optim. Appl. **60**, 545–557 (2015)
12. Larson, J., Billups, S.C.: Stochastic derivative-free optimization using a trust region framework. Comput. Optim. Appl. **64**, 619–645 (2016)
13. Lucidi, S., Sciandrone, M.: A derivative-free algorithm for bound constrained optimization. Comput. Optim. Appl. **21**(2), 119–142 (2002)
14. Lucidi, S., Sciandrone, M.: On the global convergence of derivative-free methods for unconstrained optimization. SIAM J. Optim. **13**(1), 97–116 (2002)
15. Moré, J.J., Wild, S.M.: Benchmarking derivative-free optimization algorithms. SIAM J. Optim. **20**(1), 172–191 (2009)
16. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965)
17. Torczon, V.: On the convergence of pattern search algorithms. SIAM J. Optim. **7**(1), 1–25 (1997)
18. Vicente, L.N.: Worst case complexity of direct search. EURO J. Comput. Optim. **1**, 143–153 (2013)