Gradient Convergence of Direct Search Based on Sufficient Decrease

Cunxin Huang[*] Zaikun Zhang[†]

February 9, 2026

Abstract

We analyze a direct-search method based on a sufficient decrease condition for unconstrained smooth optimization. Under standard assumptions, we improve the classical guarantee $\liminf_k \|\nabla f(x_k)\| = 0$ to the full limit $\lim_k \|\nabla f(x_k)\| = 0$. Using the same technique, we also show that if ∇f is only locally Lipschitz, then every accumulation point is stationary.

Keywords: derivative-free optimization, direct search, sufficient decrease, gradient convergence

1 Introduction

troduction⟩? We consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1.1}$$?eq:problem?

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is the objective function. In many applications, function values can be computed but derivative information is unavailable, unreliable, or too costly to obtain. This motivates derivative-free optimization (DFO) methods [2, 5, 10], among which direct-search methods [9] form a major class and are the focus of this paper.

Direct search and known convergence results. At each iteration, a direct-search method evaluates f along a finite set of directions and uses these evaluations to decide

[*]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (cun-xin.huang@connect.polyu.hk).

[†]School of Mathematics, Sun Yat-sen University, China (zhangzaikun@mail.sysu.edu.cn).

whether to accept a step and how to update the step size. These methods fall into two broad classes. The first is mesh-based direct search, where a trial point is accepted whenever it improves the current best function value (simple decrease). The direction sets in this class are tied to a mesh that refines over the iterations. Representative methods include generalized pattern search (GPS) [12] and mesh adaptive direct search (MADS) [1]. The second class replaces simple decrease with a *sufficient decrease* condition: a trial point is accepted only if the decrease in function value exceeds a threshold controlled by a forcing function (see (2.1)). Without the requirement of mesh geometry, the direction sets can be chosen more freely.

Under smoothness of f and a uniform quality condition on the direction sets, both classes guarantee

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

(see, e.g., [9, Theorem 3.11]), meaning that small gradients appear along a subsequence. Strengthening this to the full limit is harder and has been a long-standing open question in direct-search theory. Known results that achieve this stronger guarantee rely on additional mechanisms, such as complete polling [9, Theorem 3.14] or line-search-type globalization [11]. A recent survey and further references can be found in [7].

Recent developments. Beyond the classical setting, direct search has been extended to probabilistic variants [8], decentralized implementations [3], and optimization on Riemannian manifolds [4].

Contribution. We study a direct-search method based on sufficient decrease. Under standard assumptions, we strengthen the classical lim inf guarantee to

$$\lim_{k\to\infty} \|\nabla f(x_k)\| \;=\; 0. \tag{1.2}$$

eq:limit_conve

No additional mechanism beyond the sufficient decrease condition and the step-size dynamics is needed. Using the same proof technique, we also show that if ∇f is only locally Lipschitz continuous, then every accumulation point of the iterates is stationary.

Organization. Section 2 presents the algorithm and assumptions. Section 3 collects preliminary results, including the classical lim inf guarantee. Section 4 states and proves the main theorem (Theorem 4.1), which shows the full limit result (1.2). Section 5 discusses extensions to the locally Lipschitz continuous gradient setting (Theorem 5.1), and Section 6 offers concluding remarks.

2 Framework and assumptions of the algorithm

c:algorithm⟩ We begin by presenting Algorithm 2.1.

Algorithm 2.1 A simplified direct search based on sufficient decrease

---

rect-search⟩ Select $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\theta \in (0, 1)$, $\gamma \in (1, \infty)$, and a forcing function $\rho : (0, \infty) \to (0, \infty)$. For $k = 0, 1, 2, \ldots$, do the following.

1. Generate a set of nonzero vectors $\mathcal{D}_k \subset \mathbb{R}^n$.
2. Check whether there exists a $d \in \mathcal{D}_k$ such that

$$f(x_k) - f(x_k + \alpha_k d) \; > \; \rho(\alpha_k). \tag{2.1}$$

eq:sufficient_

3. If such a direction exists, set $x_{k+1} = x_k + \alpha_k d$, $\alpha_{k+1} = \gamma\alpha_k$; otherwise, set $x_{k+1} = x_k$ and $\alpha_{k+1} = \theta\alpha_k$.

In Algorithm 2.1, the inequality (2.1) is the sufficient decrease condition that governs whether a trial step is accepted and how to update the step size. There remain two components to be specified: the direction sets $\mathcal{D}_k$ and the forcing function ρ. Next, we state the assumptions on the direction sets. To do so, we recall the cosine measure [9], which quantifies how well a finite set of directions covers all possible descent directions.

ne_measure⟩? **Definition 2.1** (Cosine measure)**.** Let $\mathcal{D}$ be a finite set of nonzero vectors in $\mathbb{R}^n$. The cosine measure of $\mathcal{D}$ with respect to a nonzero vector v is

$$\operatorname{cm}(\mathcal{D}, v) \; = \; \max_{d \in \mathcal{D}} \frac{d^{\mathsf{T}} v}{\|d\| \|v\|}.$$

The cosine measure of $\mathcal{D}$ is

$$\operatorname{cm}(\mathcal{D}) \; = \; \min_{v \in \mathbb{R}^n \setminus \{0\}} \operatorname{cm}(\mathcal{D}, v).$$

We now propose two standard assumptions on the direction sets used in Algorithm 2.1, requiring that each direction set consists of unit vectors and has a uniformly positive cosine measure.

unit_length⟩ **Assumption 2.1.** *For each $k \ge 0$, the direction set $\mathcal{D}_k$ is a finite set of unit vectors in $\mathbb{R}^n$.*

⟨ass:D_k⟩ **Assumption 2.2.** *There exists a constant $\kappa > 0$ such that $\operatorname{cm}(\mathcal{D}_k) \ge \kappa$ for each $k \ge 0$.*

In the sufficient decrease condition (2.1), the forcing function ρ controls how much decrease is needed for a step to be accepted. The typical requirement on the forcing function is that $\rho : (0, \infty) \to (0, \infty)$ should be nondecreasing and satisfy $\rho(\alpha) = o(\alpha)$ as $\alpha \to 0^+$. In this paper, we adopt the standard polynomial form [6, 13] as stated in the following Assumption 2.3.

forcing_fun⟩ **Assumption 2.3.** *The forcing function ρ is of the form $\rho(\alpha) = c\alpha^p$ for some constant $c > 0$ and $p > 1$.*

We conclude this section with the following remark on Algorithm 2.1.

Remark 2.1. We make two remarks about Algorithm 2.1.

(a) When the sufficient decrease condition (2.1) is not satisfied, Algorithm 2.1 sets $x_{k+1} = x_k$. In fact, all results in this paper remain valid if we instead set x_{k+1} to any point satisfying $f(x_{k+1}) \le f(x_k)$. In other words, the sufficient decrease condition only needs to govern the step-size update; for the iterations where the sufficient decrease condition is not satisfied, any simple decrease step is allowed.

(b) The usual requirement on the expanding parameter γ is $\gamma \ge 1$. However, in this paper, we require $\gamma > 1$, as our proof technique relies on this.

3 Preliminary results

ing_results⟩ This section collects several existing results that will be used in the proofs that follow. We first make two standard assumptions on the objective function.

wer_bounded⟩ **Assumption 3.1.** *The objective function f is continuously differentiable and is bounded from below.*

g_lipschitz⟩ **Assumption 3.2.** *The gradient of the objective function ∇f is Lipschitz continuous with Lipschitz constant $L > 0$.*

The following Lemma 3.1 states that the step size converges to zero. This is a direct consequence of the lower boundedness assumption of f.

epsize_to_0⟩ **Lemma 3.1** ([9, Theorem 3.4])**.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.3, and 3.1, we have $\alpha_k \to 0$.*

The next lemma shows that when the step size is small enough relative to the gradient norm, at least one direction in $\mathcal{D}_k$ yields a decrease proportional to $\|\nabla f(x_k)\|\alpha_k$.

scent_lemma⟩ **Lemma 3.2.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, and 3.2, if*

$$\alpha_k \;\leq\; \kappa\|\nabla f(x_k)\|/L,$$

then we have

$$\max_{d\in\mathcal{D}_k}\{f(x_k)-f(x_k+\alpha_k d)\} \;\geq\; \frac{\kappa}{2}\|\nabla f(x_k)\|\alpha_k.$$

Proof. Since ∇f is Lipschitz continuous (Assumption 3.2), for each unit vector $d \in \mathcal{D}_k$ we have

$$f(x_k)-f(x_k+\alpha_k d) \;\geq\; \left[-\nabla f(x_k)^\top d\right]\alpha_k - \frac{L}{2}\alpha_k^2.$$

By Assumption 2.2, there exists $d^* \in \mathcal{D}_k$ with $-\nabla f(x_k)^\top d^* \geq \kappa\|\nabla f(x_k)\|$. Thus, using the assumption that $\alpha_k \leq \kappa L^{-1}\|\nabla f(x_k)\|$, we get

$$f(x_k)-f(x_k+\alpha_k d^*) \;\geq\; \kappa\|\nabla f(x_k)\|\alpha_k - \frac{L}{2}\alpha_k^2 \;\geq\; \frac{\kappa}{2}\|\nabla f(x_k)\|\alpha_k,$$

which completes the proof. □

Combining the two lemmas above yields the classical lim inf convergence guarantee, stated below as Theorem 3.1. We will strengthen it to a full limit in Section 4.

m:liminf_gk⟩ **Theorem 3.1** ([9, Theorem 3.11])**.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, we have*

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

4 Main results

ain_results⟩ This section contains the main results of this paper. In Section 4.1, we state the main theorem and outline the proof strategy. In Section 4.2, we introduce auxiliary level sets and stopping times, and establish several supporting lemmas. In Section 4.3, we combine these lemmas to complete the proof.

4.1 Main theorem and proof strategy

ain_theorem⟩ We now state our main result.

convergence⟩

Theorem 4.1. *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, we have*

$$\lim_{k\to\infty}\|\nabla f(x_k)\| = 0. \tag{4.1}$$

?eq:limit_conve

Recall that the classical theory guarantees only $\liminf_k \|\nabla f(x_k)\| = 0$ (Theorem 3.1). To upgrade this to a full limit, we argue by contradiction and analyze the repeated crossings of the iterates between a region where $\|\nabla f\|$ is small and a region where it is bounded away from zero. The two main ingredients are (i) the Lipschitz continuity of ∇f, which ensures a positive separation between these two regions, and (ii) the step-size dynamics implied by the sufficient decrease mechanism.

4.2 Level sets of the gradient norm and stopping times

:level_sets⟩ For a given $\epsilon > 0$, we define two level sets of the gradient norm:

$$S_{\leq}^{\epsilon} \;=\; \{x \in \mathbb{R}^n : \|\nabla f(x)\| \leq \epsilon\}, \tag{4.2}$$ eq:S_lower

and

$$S_{>}^{\epsilon} \;=\; \{x \in \mathbb{R}^n : \|\nabla f(x)\| > 2\epsilon\}. \tag{4.3}$$ eq:S_greater

ist_g_level⟩ **Lemma 4.1.** *Under Assumption 3.2, for any $\epsilon > 0$, we have*

$$\mathrm{dist}(S_{\leq}^{\epsilon}, S_{>}^{\epsilon}) \;\geq\; \epsilon/L.$$

Proof. By Assumption 3.2, for any $x \in S_{\leq}^{\epsilon}$ and $y \in S_{>}^{\epsilon}$, we have

$$L\|x-y\| \;\geq\; \|\nabla f(x) - \nabla f(y)\| \;\geq\; \|\nabla f(y)\| - \|\nabla f(x)\| \;\geq\; \epsilon,$$

where the last inequality follows from the definitions of $S_{\leq}^{\epsilon}$ and $S_{>}^{\epsilon}$ in (4.2) and (4.3), respectively. Taking the infimum over all such x and y gives the desired result. □

cess_region⟩ **Lemma 4.2.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, for any $\epsilon > 0$, there exists $K^{\epsilon} \geq 0$ such that, for all $k \geq K^{\epsilon}$, if $x_k \notin S_{\leq}^{\epsilon}$, then the sufficient decrease condition* (2.1) *is satisfied, and hence $\alpha_{k+1} = \gamma\alpha_k$.*

Proof. By Lemma 3.2 and the definition of $S_{\leq}^{\epsilon}$ in (4.2), if $\alpha_k \leq \kappa\epsilon/L$ and $x_k \notin S_{\leq}^{\epsilon}$, then we have

$$\max_{d \in \mathcal{D}_k} \{f(x_k) - f(x_k + \alpha_k d)\} \;\geq\; \frac{\kappa\epsilon}{2}\alpha_k.$$

Since $\alpha_k \to 0$ (Lemma 3.1) and $\rho(\alpha) = c\alpha^p$ (Assumption 2.3), there exists $K^{\epsilon} \geq 0$ such that, for each $k \geq K^{\epsilon}$,

$$\alpha_k \;\leq\; \min\left\{\frac{\kappa\epsilon}{L}, \left(\frac{\kappa\epsilon}{2c}\right)^{\frac{1}{p-1}}\right\}$$

so that $\kappa\epsilon\alpha_k/2 \geq \rho(\alpha_k)$. This completes the proof. □

Given $\epsilon > 0$, we define the sequences of stopping times $\{m_j^\epsilon\}$ and $\{n_j^\epsilon\}$ inductively as follows. First, let

$$m_0^\epsilon \ = \ \inf\{k \geq K^\epsilon : x_k \in S_>^\epsilon\}, \tag{4.4}$$ eq:m0

representing the first index at which the iterates enter the large-gradient region $S_>^\epsilon$ after iteration K^ϵ. Then, for each $j \geq 0$, we define

$$n_j^\epsilon \ = \ \inf\{k > m_j^\epsilon : x_k \in S_\leq^\epsilon \text{ and } x_{k-1} \notin S_\leq^\epsilon\}, \tag{4.5}$$ eq:time_out

and

$$m_{j+1}^\epsilon \ = \ \inf\{k > n_j^\epsilon : x_k \in S_>^\epsilon \text{ and } x_{k-1} \notin S_>^\epsilon\}. \tag{4.6}$$ eq:time_in

Intuitively, the sequence $\{m_j^\epsilon\}$ records the iteration indices when the iterates $\{x_k\}$ enter the large gradient region $S_>^\epsilon$, while the sequence $\{n_j^\epsilon\}$ records the iteration indices when they enter the small gradient region $S_\leq^\epsilon$.

notes_mj_nj⟩ **Lemma 4.3.** *For each $j \geq 0$ and each $k \in [m_j^\epsilon, n_j^\epsilon - 1]$, we have*

$$\|\nabla f(x_k)\| \ > \ \epsilon \quad \textit{and} \quad \alpha_{k+1} = \gamma \alpha_k.$$

Proof. By the definitions of m_j^ϵ in (4.4) and (4.6), and of n_j^ϵ in (4.5), every x_k lies outside $S_\leq^\epsilon$ if $k \in [m_j^\epsilon, n_j^\epsilon - 1]$, so $\|\nabla f(x_k)\| > \epsilon$. Lemma 4.2 then gives $\alpha_{k+1} = \gamma \alpha_k$. □

itely_often⟩ **Lemma 4.4.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, if*

$$\limsup_{k \to \infty} \|\nabla f(x_k)\| \ > \ 0,$$

then

$$m_j^\epsilon \ < \ \infty \quad \textit{and} \quad n_j^\epsilon \ < \ \infty \quad \textit{for each } j \geq 0,$$

where $\epsilon = \limsup_k \|\nabla f(x_k)\|/3$.

Proof. The result follows readily from Theorem 3.1, but we give a self-contained proof by induction.

Base case. Since $\limsup_k \|\nabla f(x_k)\| = 3\epsilon > 2\epsilon$, infinitely many iterates lie in $S_>^\epsilon$, so $m_0^\epsilon < \infty$. If $n_0^\epsilon = \infty$, then Lemma 4.3 gives $\alpha_{k+1} = \gamma \alpha_k$ for all $k \geq m_0^\epsilon$, contradicting $\alpha_k \to 0$ (Lemma 3.1). Hence $n_0^\epsilon < \infty$.

Induction step. Assume that m_j^ϵ and n_j^ϵ are finite. Since infinitely many iterates lie in $S_>^\epsilon$ and $x_{n_j^\epsilon} \in S_\leq^\epsilon$, there exists a first index $k > n_j^\epsilon$ at which $x_k \in S_>^\epsilon$ and $x_{k-1} \notin S_>^\epsilon$, so $m_{j+1}^\epsilon < \infty$. If $n_{j+1}^\epsilon = \infty$, the same contradiction as in the base case arises. Hence $n_{j+1}^\epsilon < \infty$, completing the induction. □

For each $j \geq 0$, we define

$$\Delta f_j^\epsilon \;=\; f(x_{m_j^\epsilon}) - f(x_{n_j^\epsilon}), \tag{4.7}$$

eq:delta_f_j_e

and

$$D_j^\epsilon \;=\; \|x_{m_j^\epsilon} - x_{n_j^\epsilon}\|. \tag{4.8}$$

eq:D_j_epsilon

lower_bound⟩ **Lemma 4.5.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, if*

$$\limsup_{k\to\infty} \|\nabla f(x_k)\| \;>\; 0,$$

then

$$D_j^\epsilon \;\geq\; \epsilon/L, \quad \textit{for each } j \geq 0,$$

where $\epsilon = \limsup_k \|\nabla f(x_k)\|/3$.

Proof. Since $x_{m_j^\epsilon} \in S_>^\epsilon$ and $x_{n_j^\epsilon} \in S_\leq^\epsilon$, Lemma 4.1 gives

$$D_j^\epsilon \;=\; \|x_{m_j^\epsilon} - x_{n_j^\epsilon}\| \;\geq\; \mathrm{dist}(S_\leq^\epsilon, S_>^\epsilon) \;\geq\; \epsilon/L. \qquad \square$$

ta_f_j_to_0⟩ **Lemma 4.6.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, if*

$$\limsup_{k\to\infty} \|\nabla f(x_k)\| \;>\; 0,$$

then $\Delta f_j^\epsilon \to 0$ *as* $j \to \infty$, *where* $\epsilon = \limsup_k \|\nabla f(x_k)\|/3$.

Proof. The proof follows from Assumption 3.1 that f is lower bounded. □

unded_ratio⟩ **Lemma 4.7.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2, if*

$$\limsup_{k\to\infty} \|\nabla f(x_k)\| \;>\; 0,$$

then there exists a constant $C > 0$ *such that*

$$\frac{\Delta f_j^\epsilon}{(D_j^\epsilon)^p} \;\geq\; C \quad \textit{for all } j \geq 0,$$

where $\epsilon = \limsup_k \|\nabla f(x_k)\|/3$.

Proof. By Lemma 4.3, for each $k \in [m_j^\epsilon, n_j^\epsilon - 1]$, we have

$$\|x_{k+1} - x_k\| \;=\; \alpha_k, \quad \alpha_{k+1} \;=\; \gamma\alpha_k, \quad \text{and} \quad f(x_k) - f(x_{k+1}) \;\geq\; c(\alpha_k)^p,$$

where the first equality holds because all directions are unit vectors (Assumption 2.1). Hence, by the definition of D_j^ϵ and Δf_j^ϵ, we have

$$D_j^\epsilon \;\le\; \sum_{k=m_j^\epsilon}^{n_j^\epsilon-1} \alpha_k \;=\; \alpha_{m_j^\epsilon} \sum_{k=0}^{n_j^\epsilon-m_j^\epsilon-1} \gamma^k \;=\; \frac{\alpha_{m_j^\epsilon}}{\gamma-1}\left(\gamma^{n_j^\epsilon-m_j^\epsilon}-1\right), \tag{4.9}$$

?eq:D_bound?

and

$$\Delta f_j^\epsilon \;\ge\; \sum_{k=m_j^\epsilon}^{n_j^\epsilon-1} c(\alpha_k)^p \;=\; c\alpha_{m_j^\epsilon}^p \sum_{k=0}^{n_j^\epsilon-m_j^\epsilon-1} (\gamma^k)^p \;=\; \frac{c\alpha_{m_j^\epsilon}^p}{\gamma^p-1}\left(\gamma^{p(n_j^\epsilon-m_j^\epsilon)}-1\right). \tag{4.10}$$

?eq:delta_f_bou

Hence, we have

$$\begin{aligned}
\frac{\Delta f_j^\epsilon}{(D_j^\epsilon)^p} \;&\ge\; \frac{c(\gamma-1)^p}{\gamma^p-1}\,\frac{(\gamma^p)^{n_j^\epsilon-m_j^\epsilon}-1}{(\gamma^{n_j^\epsilon-m_j^\epsilon}-1)^p}\\
&\ge\; \frac{c(\gamma-1)^p}{\gamma^p-1}\left[1-\frac{1}{\gamma^{p(n_j^\epsilon-m_j^\epsilon)}}\right]\\
&\ge\; \frac{c(\gamma-1)^p}{\gamma^p-1}\left[1-\frac{1}{\gamma^p}\right] \;=\; \frac{c(\gamma-1)^p}{\gamma^p} \;>\; 0,
\end{aligned}$$

which completes the proof. □

4.3 Proof of Theorem 4.1

:proof_main⟩ We are now ready to prove Theorem 4.1 by combining the lemmas established above.

Proof. By Theorem 3.1, it suffices to show that

$$\limsup_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{4.11}$$

eq:convergence_

We prove by contradiction. Suppose (4.11) is false. Then we define

$$\epsilon \;=\; \frac{1}{3}\limsup_{k\to\infty} \|\nabla f(x_k)\| \;>\; 0.$$

By Lemma 4.7, there exists a constant $C>0$ such that $\Delta f_j^\epsilon \ge C(D_j^\epsilon)^p$ for all $j\ge 0$. On the other hand, Lemma 4.6 gives $\Delta f_j^\epsilon \to 0$ as $j\to\infty$, which forces $D_j^\epsilon \to 0$. This contradicts $D_j^\epsilon \ge \epsilon/L > 0$ (Lemma 4.5), which completes the proof. □

5 Extensions to locally Lipschitz continuous gradients

:extensions⟩ The proof of Theorem 4.1 uses the global Lipschitz continuity of ∇f (Assumption 3.2) in two places: the descent lemma (Lemma 3.2) and the separation of the gradient level sets (Lemma 4.1). In this section, we show that the same proof technique applies under the weaker assumption that ∇f is only locally Lipschitz continuous, yielding a stationarity result for the limit points of the iterates.

y_lipschitz⟩ **Assumption 5.1.** *The gradient of the objective function ∇f is locally Lipschitz continuous.*

Note that Assumption 5.1 is weaker than Assumption 3.2. The conclusion of Lemma 3.1 ($\alpha_k \to 0$) remains valid since it does not require Assumption 3.2. We now state the main result of this section.

_stationary⟩ **Theorem 5.1.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, 2.3, 3.1, and 5.1, every limit point of $\{x_k\}$ is a stationary point of f.*

The remainder of this section proves Theorem 5.1. Let x^* be an arbitrary limit point of $\{x_k\}$; we fix it for the rest of this section. The argument parallels that of Theorem 4.1, with the gradient level sets $S^{\epsilon}_{\leq}$ and $S^{\epsilon}_{>}$ replaced by balls centered at x^*.

Setup. Suppose, for contradiction, that $\nabla f(x^*) \neq 0$. By Assumption 5.1 and the continuity of ∇f, there exist $\delta > 0$ and $L_* > 0$ such that ∇f is L_*-Lipschitz continuous on $\mathcal{B}(x^*, 2\delta)$ and

$$\|\nabla f(x)\| \;\geq\; \|\nabla f(x^*)\|/2 \quad \text{for all } x \in \mathcal{B}(x^*, 2\delta). \tag{5.1}$$ eq:grad_lower_

Set $\epsilon = \|\nabla f(x^*)\|/2 > 0$ and define

$$A \;=\; \mathcal{B}(x^*, \delta/2), \qquad B \;=\; \mathcal{B}(x^*, \delta)^{\complement}. \tag{5.2}$$?eq:AB_sets?

Note that $\operatorname{dist}(A, B) = \delta/2 > 0$.

The following lemma is the counterpart of Lemma 4.2.

egion_local⟩ **Lemma 5.1.** *Under the setup above, there exists $K \geq 0$ such that, for all $k \geq K$, if $x_k \in \mathcal{B}(x^*, \delta)$, then the sufficient decrease condition* (2.1) *is satisfied, and hence $\alpha_{k+1} = \gamma\alpha_k$.*

Proof. The proof follows that of Lemma 4.2, with L replaced by L_* and "$x_k \notin S^{\epsilon}_{\leq}$" replaced by "$x_k \in \mathcal{B}(x^*, \delta)$". By (5.1), every $x_k \in \mathcal{B}(x^*, \delta)$ satisfies $\|\nabla f(x_k)\| \geq \epsilon$. Since $\alpha_k \to 0$ (Lemma 3.1), we choose K large enough that $\alpha_k \leq \min\{\kappa\epsilon/L_*,\ \delta,\ (\kappa\epsilon/(2c))^{1/(p-1)}\}$ for all $k \geq K$. The condition $\alpha_k \leq \delta$ ensures $x_k + \alpha_k d \in \mathcal{B}(x^*, 2\delta)$ whenever $x_k \in \mathcal{B}(x^*, \delta)$ and $\|d\| = 1$, so the Lipschitz bound applies to the descent inequality. □

We define the stopping times as in (4.4)–(4.6), with A and B playing the roles of $S^{\epsilon}_{>}$ and $S^{\epsilon}_{\leq}$:

$$\begin{aligned} m_0 &= \inf\{k \geq K : x_k \in A\}, && (5.3) \\ n_j &= \inf\{k > m_j : x_k \in B\}, && (5.4) \\ m_{j+1} &= \inf\{k > n_j : x_k \in A\}. && (5.5) \end{aligned}$$

?eq:m0_local? ?eq:nj_local? ?eq:mj_local?

The next lemma shows that, between entering A and exiting $\mathcal{B}(x^*, \delta)$, the iterates behave exactly as in the crossings of Section 4.2.

amics_local⟩ **Lemma 5.2.** *Under the setup above, for each $j \geq 0$ and each $k \in [m_j, n_j - 1]$, we have*

$$x_k \in \mathcal{B}(x^*, \delta), \quad \|\nabla f(x_k)\| \geq \epsilon, \quad \textit{and} \quad \alpha_{k+1} = \gamma \alpha_k.$$

Proof. Identical to the proof of Lemma 4.3, with $S^{\epsilon}_{\leq}$ replaced by $B = \mathcal{B}(x^*, \delta)^{\complement}$ and Lemma 4.2 replaced by Lemma 5.1. □

The next lemma is the counterpart of Lemma 4.4: the iterates cross between A and B infinitely often.

sing_local⟩? **Lemma 5.3.** *Under the setup above, $m_j < \infty$ and $n_j < \infty$ for each $j \geq 0$.*

Proof. Since x^* is a limit point, infinitely many iterates lie in $A = \mathcal{B}(x^*, \delta/2)$, so $m_j < \infty$ for each j. If $n_j = \infty$ for some j, then Lemma 5.2 gives $\alpha_{k+1} = \gamma \alpha_k$ for all $k \geq m_j$, contradicting $\alpha_k \to 0$ (Lemma 3.1). Hence $n_j < \infty$. □

For each $j \geq 0$, define $\Delta f_j = f(x_{m_j}) - f(x_{n_j})$ and $D_j = \|x_{m_j} - x_{n_j}\|$ as in (4.7)–(4.8). The final lemma combines the counterparts of Lemmas 4.5, 4.6, and 4.7.

ounds_local⟩ **Lemma 5.4.** *Under the setup above:*

(a) *$D_j \geq \delta/2$ for each $j \geq 0$.*

(b) *$\Delta f_j \to 0$ as $j \to \infty$.*

(c) *There exists a constant $C > 0$ such that $\Delta f_j \geq C\,(D_j)^p$ for all $j \geq 0$.*

Proof. (a) Since $x_{m_j} \in A$ and $x_{n_j} \in B$, we have $D_j \geq \operatorname{dist}(A, B) = \delta/2$.

(b) Since f is nonincreasing along the iterates and bounded below (Assumption 3.1), the intervals $[m_j, n_j]$ are disjoint, and $\sum_j \Delta f_j \leq f(x_{m_0}) - \inf f < \infty$. Hence $\Delta f_j \to 0$.

(c) By Lemma 5.2, between m_j and $n_j - 1$ the step sizes grow geometrically and each step achieves sufficient decrease. The computation is then identical to that of Lemma 4.7, yielding $\Delta f_j/(D_j)^p \geq c(\gamma - 1)^p/\gamma^p > 0$. □

Proof of Theorem 5.1. By Lemma 5.4(c), $\Delta f_j \geq C\,(D_j)^p$ for all j. By Lemma 5.4(b), $\Delta f_j \to 0$, which forces $D_j \to 0$. This contradicts $D_j \geq \delta/2 > 0$ from Lemma 5.4(a). Hence, every limit point of $\{x_k\}$ is stationary. □

Remark 5.1. As noted in the remark following Assumption 2.3, the proofs of both Theorem 4.1 and Theorem 5.1 remain valid if, on unsuccessful iterations, the algorithm sets x_{k+1} to any point satisfying $f(x_{k+1}) \leq f(x_k)$.

6 Concluding remarks

We have shown that a direct-search method based on sufficient decrease achieves the full-limit gradient convergence $\lim_k \|\nabla f(x_k)\| = 0$ under standard assumptions (Theorem 4.1). The key idea is to analyze the repeated crossings of the iterates between small-gradient and large-gradient regions: the Lipschitz continuity of ∇f forces each crossing to cover a positive distance, while the step-size dynamics ensure that each crossing costs a function decrease bounded below relative to the distance traveled. Since the total function decrease is finite, these crossings cannot occur infinitely often, giving a contradiction. We have also extended this result to the locally Lipschitz setting, showing that every limit point of the iterates is stationary (Theorem 5.1).

We close with two open questions.

- Our analysis requires the step-size expansion factor $\gamma > 1$ in Algorithm 2.1. The classical theory allows $\gamma = 1$, i.e., no expansion on successful steps. Can the full-limit result $\lim_k \|\nabla f(x_k)\| = 0$ still be established when $\gamma = 1$?

- Can these results be extended to probabilistic direct-search methods [8]? In particular, does the gradient norm converge to zero almost surely under appropriate probabilistic assumptions on the direction sets?

References

[1] C. Audet and J. E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, 17:188–217, 2006.

[2] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Springer, Cham, 2017.

[3] E. H. Bergou, Y. Diouane, V. Kungurtsev, and C. W. Royer. Direct-search methods for decentralized blackbox optimization. *arXiv:2504.04269*, 2025.

ver_Yger_2025 [4] B. Cavarretta, F. Goyens, C. W. Royer, and F. Yger. Complexity guarantees and polling strategies for Riemannian direct-search methods. *arXiv:2511.15360*, 2025.

Vicente_2009b [5] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*, volume 8 of *MOS-SIAM Ser. Optim.* SIAM, Philadelphia, 2009.

ce_Zhang_2016 [6] M. Dodangeh, L. N. Vicente, and Z. Zhang. On the optimal order of worst case complexity of direct search. *Optim. Lett.*, 10:699–708, 2016.

_Zeffiro_2025 [7] K. J. Dzahini, F. Rinaldi, C. W. Royer, and D. Zeffiro. Direct-search methods in the year 2025: Theoretical guarantees and algorithmic paradigms. *EURO J. Comput. Optim.*, 13:1–24, 2025.

ce_Zhang_2015 [8] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search based on probabilistic descent. *SIAM J. Optim.*, 25:1515–1541, 2015.

_Torczon_2003 [9] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45:385–482, 2003.

y_Wild_2019 [10] J. Larson, M. Menickelly, and S. M. Wild. Derivative-free optimization methods. *Acta Numer.*, 28:287–404, 2019.

ndrone_2002 [11] S. Lucidi and M. Sciandrone. On the global convergence of derivative-free methods for unconstrained optimization. *SIAM J. Optim.*, 13:97–116, 2002.

Torczon_1997 [12] V. Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim.*, 7:1–25, 1997.

Vicente_2013 [13] L. N. Vicente. Worst case complexity of direct search. *EURO J. Comput. Optim.*, 1:143–153, 2013.