A Unified Series Condition for the Convergence of Derivative-Free Trust Region and Direct Search

Cunxin Huang* Zaikun Zhang†

February 10, 2026

Abstract

Derivative-free trust-region and direct-search methods are two popular classes of derivative-free optimization algorithms. In this paper, we propose a unified perspective for their convergence analysis. Specifically, we show that the behavior of an algorithm-determined series governs asymptotic convergence, thereby generalizing existing results in both deterministic and randomized settings. Although our analysis of direct-search methods requires complete polling, we provide a counterexample showing that this requirement is essential for our convergence result.

Keywords: Derivative-free optimization, Trust region, Direct search, Sufficient decrease, Convergence analysis

1 Introduction

We consider the unconstrained optimization problem

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

eq:unconstrain

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. When gradient information for f in (1.1) is unavailable, derivative-free optimization (DFO) provides a powerful alternative [1, 7, 15]. DFO methods generally fall into two categories: model-based methods and direct-search methods.

*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).

In model-based methods, especially trust-region methods [5, 6, 7], one constructs a local (usually quadratic) model of the objective function within a trust region via interpolation or regression of function values, and updates the iterate by approximately minimizing this model. In contrast, direct-search methods [1, 7, 14] do not build such models. Instead, they explore the search space by evaluating the objective function along a finite set of directions and update the iterate based on comparisons of function values. In this paper, we focus on direct-search methods equipped with the "sufficient decrease" globalization strategy [9, 16]. Moreover, randomization techniques have been introduced in both categories; see [2, 3, 4, 12] for trust-region methods and [11] for direct-search methods.

Despite these algorithmic differences, natural questions arise: Is there a unified theoretical framework to analyze the convergence of these two classes of methods? Can we characterize their convergence behavior through a common condition?

In this paper, we provide an affirmative answer to these questions as a first step toward such a framework. We focus on a simplified first-order DFO trust-region method and a direct-search method based on sufficient decrease and complete polling. We show that the asymptotic convergence of these methods can be characterized by the behavior of an algorithm-determined series

$$H \;=\; \sum_{k=0}^{\infty}\prod_{\ell=0}^{k-1}\gamma^{y_\ell}\theta^{1-y_\ell}, \tag{1.2}$$?eq:series_intr

where $\gamma \in [1,\infty)$ and $\theta \in (0,1)$ are algorithmic parameters for step-size updates, and $\{y_k\}$ is a sequence of algorithm-determined indicators of whether the iteration is "good" (e.g., the model is sufficiently accurate or the direction set is well poised). In particular, we show that if the series H diverges, then the iterates generated by the algorithm admit a subsequence along which the gradients converge to zero.

Characterizing convergence via the behavior of a series is not new in optimization theory. For example, the well-known Zoutendijk condition [17, 18, 19]

$$\sum_{k=0}^{\infty}(\nabla f(x_k)^\mathsf{T} d_k)^2/\|d_k\|^2 \;<\; \infty$$

is central to the analysis of line-search methods. Similarly, for conjugate gradient methods, Dai et al. [8] proved a convergence result of the form

$$\sum_{k=0}^{\infty}\|\nabla f(x_k)\|^4/\|d_k\|^2 \;=\; \infty.$$

Our result echoes these classical theories by establishing a unified series condition for DFO methods.

The rest of the paper is organized as follows. In Section 2, we present the unified convergence theory based on the series condition. In Section 3, we apply this theory to a derivative-free trust-region method under both deterministic and randomized settings. In Section 4, we extend the analysis to direct-search methods with sufficient decrease and complete polling. A counterexample is provided in Subsection 4.3 to illustrate the necessity of complete polling for our convergence result. Finally, we discuss connections to the non-convergence results in [13] and open questions in Section 5, and we conclude in Section 6.

2 Unified series-based convergence theory

⟨sec:series⟩ This section presents a simple abstract framework that captures the step-size mechanisms shared by trust-region and direct-search methods. We then identify an algorithm-determined series whose divergence guarantees first-order convergence.

2.1 Deterministic framework

Algorithm 2.1 Deterministic general framework with adaptive step sizes

l-framework⟩ Select $x_0 \in \mathbb{R}^n$, $\theta \in (0, 1)$, $\gamma \in [1, \infty)$, and $\alpha_0 > 0$.

For $k = 0, 1, 2, \ldots,$ do the following.

1. Generate a step $s_k(\alpha_k) \in \mathbb{R}^n$ using a local model $m_k : \mathbb{R}^n \to \mathbb{R}$ deterministically.
2. If $s_k(\alpha_k)$ satisfies a sufficient decrease condition, set $x_{k+1} = x_k + s_k(\alpha_k)$; otherwise, set $x_{k+1} = x_k$.
3. If m_k satisfies a quality condition, set $\alpha_{k+1} = \gamma\alpha_k$; otherwise, set $\alpha_{k+1} = \theta\alpha_k$.

To distinguish iterations with "reliable" local information from those with "unreliable" local information, we introduce a binary sequence $\{y_k\}$. Intuitively, $y_k = 1$ indicates that the local model $m_k : \mathbb{R}^n \to \mathbb{R}$ (or, in a direct-search setting, the local direction set) is sufficiently good for the current step size, whereas $y_k = 0$ indicates the opposite. The precise definition of y_k depends on the algorithmic instance (e.g., fully linear models in trust-region methods, or a positive cosine measure in direct-search methods). We specify these choices in Sections 3 and 4, respectively. We assume that $y_k \in \{0, 1\}$ is determined by the local model m_k at iteration k.

The following assumption summarizes the key property we require of a good iteration.

⟨ass:yk_ass⟩ **Assumption 2.1.** *Consider Algorithm 2.1. For every* $\varepsilon > 0$*, there exists a constant* $\bar{\alpha} > 0$ *(possibly depending on* ε*) such that, for each* $k \geq 0$*, if* $y_k = 1$*,* $\|\nabla f(x_k)\| \geq \varepsilon$*, and* $\alpha_k \leq \bar{\alpha}$*, then*

- *the sufficient decrease condition and the quality condition are satisfied, so that*

$$x_{k+1} = x_k + s_k(\alpha_k) \quad \textit{and} \quad \alpha_{k+1} = \gamma \alpha_k;$$

- *the objective value decreases by at least* $\zeta \|\nabla f(x_k)\| \alpha_k$*, i.e.,*

$$f(x_k) - f(x_{k+1}) \;\geq\; \zeta \|\nabla f(x_k)\| \alpha_k,$$

for some constant $\zeta > 0$ *independent of* k *and* ε*.*

We also require that the step size converges to zero.

tep_to_zero⟩ **Assumption 2.2.** *For Algorithm 2.1, the step size* $\alpha_k \to 0$ *as* $k \to \infty$*.*

Finally, we impose standard smoothness and boundedness assumptions on the objective function for the remainder of the paper.

ss:function⟩ **Assumption 2.3.** *The objective function* f *is bounded from below and continuously differentiable in* $\mathbb{R}^n$*. The gradient* ∇f *is Lipschitz continuous in* $\mathbb{R}^n$ *with Lipschitz constant* L*.*

We now define the central series associated with the step-size update rule in Algorithm 2.1.

$$H \;=\; \sum_{k=0}^{\infty} \prod_{\ell=0}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell}. \tag{2.1}$$ eq:series

The next theorem shows that divergence of this series forces first-order stationarity along a subsequence.

convergence⟩ **Theorem 2.1.** *Consider Algorithm 2.1. Under Assumptions 2.1, 2.2, and 2.3, if the series* H *defined in* (2.1) *diverges, i.e.,* $H = \infty$*, then*

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

Proof. We prove the result by contradiction. Let $\zeta > 0$ be the constant in Assumption 2.1. Suppose that $\liminf_k \|\nabla f(x_k)\| > 0$. Then, since $\alpha_k \to 0$ by Assumption 2.2, there exist an integer $K \geq 0$ and constants $\varepsilon > 0$ and $\bar{\alpha} > 0$ such that, for each $k \geq K$,

$$\|\nabla f(x_k)\| \;\geq\; \varepsilon \quad \text{and} \quad \alpha_k \;\leq\; \bar{\alpha},$$

where $\bar{\alpha}$ is the positive constant in Assumption 2.1. Then, for each $k \geq K$, if $y_k = 1$, the sufficient decrease condition and the quality condition are both satisfied by Assumption 2.1, so the iterate is updated and the step size is expanded. Thus, for each $k \geq k' \geq K$, we obtain the following lower bound on the step size

$$\alpha_k \;\geq\; \alpha_{k'-1} \prod_{\ell=k'}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell}. \tag{2.2}$$

eq:lower_bound

Define $k^* = \inf\{k \geq K : y_k = 1\}$, $\mathcal{I}_0 = \{k \geq k^* : y_k = 0\}$, and $\mathcal{I}_1 = \{k \geq k^* : y_k = 1\}$. Since $H = \infty$, we have $k^* < \infty$ and $\operatorname{card}(\mathcal{I}_1) = \infty$. Suppose i_1 and i_2 are two consecutive indices in $\mathcal{I}_1$ with $i_1 < i_2$. Then we have

$$\sum_{\substack{k \in \mathcal{I}_0, \\ i_1 \leq k < i_2}} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell} \;\leq\; \frac{\gamma}{1-\theta} \prod_{\ell=k^*}^{i_1-1} \gamma^{y_\ell} \theta^{1-y_\ell},$$

meaning that the sum of the terms in $\mathcal{I}_0$ between two consecutive terms in $\mathcal{I}_1$ is bounded by a multiple of the earlier term, due to the convergence of the geometric series $\sum_{k=0}^{\infty} \theta^k$. By summing the above inequality over all pairs of consecutive indices in $\mathcal{I}_1$, we obtain

$$\sum_{k \in \mathcal{I}_0} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell} \;\leq\; \frac{\gamma}{1-\theta} \sum_{k \in \mathcal{I}_1} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell},$$

which implies

$$\sum_{k=k^*}^{\infty} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell} \;\leq\; \frac{1+\gamma-\theta}{1-\theta} \sum_{k \in \mathcal{I}_1} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell}. \tag{2.3}$$

eq:bound_I0_by

Let f^* be a lower bound of f. Then, by Assumption 2.1,

$$f_0 - f^* \;\geq\; \sum_{k \in \mathcal{I}_1} [f(x_k) - f(x_{k+1})] \;\geq\; \zeta\varepsilon \sum_{k \in \mathcal{I}_1} \alpha_k.$$

By applying the lower bound of α_k in (2.2) and the inequality in (2.3), we have

$$\begin{aligned} f_0 - f^* \;&\geq\; \zeta\varepsilon\, \alpha_{k^*-1} \sum_{k \in \mathcal{I}_1} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell} \\ &\geq\; \zeta\varepsilon\, \alpha_{k^*-1} \frac{1-\theta}{1+\gamma-\theta} \sum_{k=k^*}^{\infty} \prod_{\ell=k^*}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell}. \end{aligned} \tag{2.4}$$

eq:final_contr

Since $H = \infty$, the series on the right-hand side of (2.4) diverges, contradicting the assumption that f is bounded below. □

We obtain the following corollary from Theorem 2.1.

terministic⟩ **Corollary 2.1.** *Consider Algorithm 2.1. Under Assumptions 2.3, 2.1, and 2.2, if* $y_k = 1$ *for each* $k \geq 0$*, then*

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

Proof. If $y_k = 1$ for each $k \geq 0$, then $\prod_{\ell=0}^{k-1} \gamma^{y_\ell}\theta^{1-y_\ell} = \gamma^k$, so $H = \sum_{k=0}^{\infty} \gamma^k$ diverges because $\gamma \geq 1$. The conclusion follows from Theorem 2.1. □

2.2 Probabilistic framework

We next consider a randomized version of Algorithm 2.1, in which the local model (and hence the step) is generated randomly.

Algorithm 2.2 Probabilistic general framework with adaptive step sizes

l-framework⟩ Identical to Algorithm 2.1 except that the local model and the step in Step 1 are generated randomly.

For clarity, we summarize the notation for the random elements and their realizations in Table 1.

Table 1: Notation for random elements and their realizations

ion_general⟩

	Local model	Step	Iterate	Step size	Indicator	Series
Random element	M_k	S_k	X_k	A_k	Y_k	$\mathcal{H}$
Realization	m_k	s_k	x_k	α_k	y_k	H

We define the filtration $\{\mathcal{F}_k\}$, where for each $k \geq 0$,

$$\mathcal{F}_k \;=\; \sigma(M_0, A_0, X_1, \ldots, M_k, A_k, X_{k+1}), \tag{2.5}$$ eq:sigma-algeb

which is the σ-algebra generated by M_0, A_0, X_1, $\ldots$, M_k, A_k, and X_{k+1}. In addition, we define

$$\mathcal{F}_{-1} \;=\; \{\emptyset, \Omega\}.$$

istic_model⟩ **Definition 2.1.** Consider Algorithm 2.2. The sequence of random models $\{M_k\}$ is said to be p-probabilistically "good" if

$$\mathbb{P}(Y_k = 1 \mid \mathcal{F}_{k-1}) \;\geq\; p. \tag{2.6}$$?eq:probabilist

d_increment⟩ **Lemma 2.1** ([10, Theorem 5.3.1]). *Let $\{W_k\}$ be a martingale with $|W_{k+1} - W_k| \le M < \infty$. Let*

$$C \;=\; \{\lim_{k\to\infty} W_k \textit{ exists and is finite}\},$$
$$D \;=\; \{\limsup_{k\to\infty} W_k = +\infty \textit{ and } \liminf_{k\to\infty} W_k = -\infty\}.$$

Then we have

$$\mathbb{P}\,(C \cup D) \;=\; 1.$$

obabilistic⟩ **Lemma 2.2.** *Consider Algorithm 2.2. If the sequence of random models $\{M_k\}$ is p_0-probabilistically "good" with*

$$p_0 \;=\; \frac{\log\theta}{\log(\gamma^{-1}\theta)}, \tag{2.7}$$

?eq:def_p0?

then we have

$$\mathbb{P}\,(\mathcal{H} = \infty) \;=\; 1.$$

Proof. For each $k \ge 0$, we define

$$Z_k \;=\; \sum_{\ell=0}^{k-1} \left[Y_\ell \log\gamma + (1 - Y_\ell)\log\theta\right].$$

Then

$$\mathcal{H} \;=\; \sum_{k=1}^{\infty} \exp\left(Z_k\right).$$

To prove $\mathcal{H} = \infty$ a.s., it suffices to show that $\limsup_k Z_k > -\infty$ a.s.. By Definition 2.1 and the definition of p_0, the sequence $\{Z_k\}$ is a submartingale. By Doob's decomposition theorem ([10, Theorem 5.2.10]), $\{Z_k\}$ admits the unique decomposition $Z_k = W_k + P_k$, where $\{W_k\}$ is a martingale and $\{P_k\}$ is a predictable increasing process with $P_0 = 0$. Since $|Z_{k+1} - Z_k| \le \max\{\log\gamma, -\log\theta\} < \infty$, both $\{W_k\}$ and $\{P_k\}$ have bounded increments (see the formulae for W_k and P_k in [10, Theorem 5.2.10]). Applying Lemma 2.1 yields $\limsup_k Z_k > -\infty$ a.s., which completes the proof. □

obabilistic⟩ **Corollary 2.2.** *Consider Algorithm 2.2. Under Assumptions 2.3, 2.1, and 2.2, if the sequence of random models $\{M_k\}$ is p_0-probabilistically "good" with*

$$p_0 \;=\; \frac{\log\theta}{\log(\gamma^{-1}\theta)},$$

then we have

$$\mathbb{P}\left(\liminf_{k\to\infty} \|\nabla f(X_k)\| = 0\right) \;=\; 1.$$

Proof. It suffices to prove $\mathcal{H} = \infty$ when $\{M_k\}$ is p_0-probabilistically "good", which is guaranteed by Lemma 2.2. □

3 Derivative-free trust region

rust-region⟩ We begin with a simplified derivative-free trust-region method and explain how its convergence fits into the abstract framework of Section 2. The method alternates between computing a trial step from a local model, accepting the step when sufficient decrease is observed, and updating the trust-region radius based on a model-quality criterion.

Algorithm 3.1 A simplified first-order derivative-free trust-region method

rust-region⟩ Select $\eta_1, \eta_2 > 0$, $x_0 \in \mathbb{R}^n$, $\delta_0 \in (0, \infty)$, $\theta \in (0, 1)$, $\gamma \in [1, \infty)$. For $k = 0, 1, 2, \ldots$, do the following.

1. Build a quadratic model $m_k(s)$ of f and compute s_k by approximately minimizing m_k in $\mathcal{B}(x_k, \delta_k)$ so that s_k satisfies (3.1).
2. Compute the ratio
$$\varrho_k \;=\; \frac{f(x_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)}.$$
3. If $\varrho_k \geq \eta_1$, set $x_{k+1} = x_k + s_k$; otherwise, set $x_{k+1} = x_k$.
4. If $\varrho_k \geq \eta_1$ and $\|g_k\| \geq \eta_2 \delta_k$, set $\delta_{k+1} = \gamma \delta_k$; otherwise, set $\delta_{k+1} = \theta \delta_k$.

3.1 Basic definitions, assumptions, and existing results

Throughout this section, m_k denotes a $\mathcal{C}^1$ surrogate model of f on the trust region $\mathcal{B}(x_k, \delta_k)$. To quantify model quality in a scale-sensitive yet dimension-free way, we use the standard notion of fully linear models.

lly-linear⟩? **Definition 3.1.** Let f be a $\mathcal{C}^1$ function. A $\mathcal{C}^1$ model m is said to be $(\kappa_{\text{eg}}, \kappa_{\text{ef}})$-fully linear for f on $\mathcal{B}(x, \delta)$ if, for all $s \in \mathcal{B}(0, \delta)$,

$$\begin{aligned} |m(s) - f(x+s)| \;&\leq\; \kappa_{\text{ef}} \delta^2, \\ \|\nabla m(s) - \nabla f(x+s)\| \;&\leq\; \kappa_{\text{eg}} \delta. \end{aligned}$$

At iteration k of Algorithm 3.1, we consider a quadratic surrogate model

$$m_k(s) \;=\; f(x_k) + g_k^{\mathsf{T}} s + \frac{1}{2} s^{\mathsf{T}} B_k s, \qquad s \in \mathcal{B}(0, \delta_k),$$

where $g_k \in \mathbb{R}^n$ and $B_k \in \mathbb{R}^{n \times n}$ approximate $\nabla f(x_k)$ and $\nabla^2 f(x_k)$, respectively. The following assumptions are standard in the trust-region literature and will be used throughout.

ded-hessian⟩ **Assumption 3.1.** *There exists a positive constant $B_{\max}$ such that $\|B_k\| \leq B_{\max}$ for all $k \geq 0$.*

actional-CD⟩ **Assumption 3.2.** *There exists a constant* $m \in (0, 1]$ *such that, for all* $k \geq 0$*, we can compute a step* s_k *satisfying the fractional Cauchy decrease condition*

$$m_k(0) - m_k(s_k) \;\geq\; \frac{m}{2} \|g_k\| \min\{\|g_k\|/\|B_k\|, \delta_k\}, \tag{3.1}$$ eq:fcd

where by convention $\|g_k\|/\|B_k\| = \infty$ *if* $\|B_k\| = 0$.

We first recall classical global convergence results for Algorithm 3.1, which will later appear as corollaries of our series-based analysis.

rust-region⟩ **Theorem 3.1** ([7, Theorem 10.12])**.** *Consider Algorithm 3.1. Under Assumptions 2.3, 3.1, and 3.2, suppose there exist constants* $\kappa_{\text{eg}} > 0$ *and* $\kappa_{\text{ef}} > 0$ *such that* m_k *is* $(\kappa_{\text{eg}}, \kappa_{\text{ef}})$*-fully linear for every* $k \geq 0$*. Then*

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

We next consider a randomized variant of Algorithm 3.1, in which the model construction in Step 1 is randomized.

Algorithm 3.2 Probabilistic first-order derivative-free trust-region

---

ust-region⟩? Identical to Algorithm 3.1 except that the surrogate model in Step 1 is generated randomly.

---

lly-linear⟩? **Definition 3.2** ([2, Definition 3.2])**.** Consider Algorithm 3.1 with f being continuously differentiable. The sequence of surrogate models $\{M_k\}$ is said to be p-probabilistically $(\kappa_{\text{eg}}, \kappa_{\text{ef}})$-fully linear if it satisfies

$$\mathbb{P}\left(M_k \text{ is } (\kappa_{\text{eg}}, \kappa_{\text{ef}})\text{-fully linear} \mid \mathcal{F}_{k-1}\right) \;\geq\; p \quad \text{for each } k \geq 0,$$

where $\mathcal{F}_{k-1}$ is defined in (2.5).

rust-region⟩ **Theorem 3.2** ([2, Theorem 4.2])**.** *Consider Algorithm 3.1. Under Assumptions 2.3, 3.1, and 3.2, if* $\{M_k\}$ *is* p_0*-probabilistically* $(\kappa_{\text{eg}}, \kappa_{\text{ef}})$*-fully linear with* $p_0 = \log\theta/\log(\gamma^{-1}\theta)$ *and* κ_{eg} *and* κ_{ef} *being positive constants, then we have*

$$\mathbb{P}\left(\liminf_{k \to \infty} \|\nabla f(X_k)\| = 0\right) \;=\; 1.$$

3.2 Convergence analysis based on the series

We now connect Algorithm 3.1 to the abstract framework of Section 2. Fix constants $\kappa_{\rm eg} > 0$ and $\kappa_{\rm ef} > 0$, and define the indicator

$$y_k(\kappa_{\rm eg}, \kappa_{\rm ef}) \;=\; \mathbb{1}\big(m_k \text{ is } (\kappa_{\rm eg}, \kappa_{\rm ef})\text{-fully linear on } \mathcal{B}(x_k, \delta_k)\big). \tag{3.2}$$ eq:yk_tr

When $y_k = 1$ and δ_k is sufficiently small relative to $\|\nabla f(x_k)\|$, the trust-region mechanism ensures acceptance of the trial step and expansion of the radius. Our goal is to verify Assumption 2.1 and then invoke Theorem 2.1.

-successful⟩ **Lemma 3.1** ([12, Lemma 2.7]). *Consider Algorithm 3.1 and fix constants* $\kappa_{\rm eg} > 0$ *and* $\kappa_{\rm ef} > 0$. *Under Assumptions 2.3, 3.1, and 3.2, if* $y_k(\kappa_{\rm eg}, \kappa_{\rm ef}) = 1$ *and* $\delta_k \le c_1\|\nabla f(x_k)\|$ *with*

$$c_1 \;=\; \left(\kappa_{\rm eg} + \max\left\{\eta_2, B_{\max}, \frac{4\kappa_{\rm ef}}{m(1-\eta_1)}\right\}\right)^{-1}, \tag{3.3}$$?eq:def_c1?

then $x_{k+1} = x_k + s_k$ *and* $\delta_{k+1} = \gamma\delta_k$, *meaning that both the sufficient decrease condition and the quality condition are satisfied.*

ecrease_tr⟩? **Lemma 3.2.** *Consider Algorithm 3.1 and fix constants* $\kappa_{\rm eg} > 0$ *and* $\kappa_{\rm ef} > 0$. *Under Assumptions 2.3, 3.1, and 3.2, if* $y_k(\kappa_{\rm eg}, \kappa_{\rm ef}) = 1$ *and* $\delta_k \le c_2\|\nabla f(x_k)\|$ *with*

$$c_2 \;=\; \left(2\kappa_{\rm eg} + \max\left\{\eta_2, B_{\max}, \frac{4\kappa_{\rm ef}}{m(1-\eta_1)}\right\}\right)^{-1}, \tag{3.4}$$ eq:def_c2

then we have

$$f(x_k) - f(x_{k+1}) \;\ge\; \frac{m\eta_1}{4}\|\nabla f(x_k)\|\delta_k.$$

Proof. By Lemma 3.1, we have

$$\begin{aligned} f(x_k) - f(x_{k+1}) \;&\ge\; \eta_1(m_k(0) - m_k(s_k)) \\ &\ge\; \frac{m\eta_1}{2}\|g_k\| \min\{\|g_k\|/\|B_k\|, \delta_k\}, \end{aligned}$$

where the last inequality is due to Assumption 3.2. Since $\delta_k \le c_2\|\nabla f(x_k)\|$, we have $\delta_k \le \|\nabla f(x_k)\|/(B_{\max} + \kappa_{\rm eg})$, which implies $\|g_k\|/\|B_k\| \ge \delta_k$. Thus, we have

$$f(x_k) - f(x_{k+1}) \;\ge\; \frac{m\eta_1}{2}\|g_k\|\delta_k \;\ge\; \frac{m\eta_1}{2}(\|\nabla f(x_k)\| - \kappa_{\rm eg}\delta_k)\delta_k,$$

where the last inequality is due to the definition of $y_k(\kappa_{\rm eg}, \kappa_{\rm ef})$. We complete the proof by noting that

$$\|\nabla f(x_k)\| - \kappa_{\rm eg}\delta_k \;\ge\; \frac{1}{2}\|\nabla f(x_k)\|$$

due to the definition of c_2. □

We next make the connection to Algorithm 2.1 explicit. In Algorithm 3.1, we set $\alpha_k = \delta_k$ and view a successful trust-region iteration as a "good" iteration in the sense of Section 2. More precisely, Algorithm 3.1 can be regarded as an instance of Algorithm 2.1 via the following identifications.

- The step is defined as $s_k(\alpha_k) = s_k$ with $\alpha_k = \delta_k$.
- The local model is the trust-region model m_k.
- The sufficient decrease condition is

$$f(x_k) - f(x_k + s_k) \;\geq\; \eta_1\big(m_k(0) - m_k(s_k)\big).$$

- The quality condition is

$$f(x_k) - f(x_k + s_k) \;\geq\; \eta_1\big(m_k(0) - m_k(s_k)\big) \quad \text{and} \quad \|g_k\| \;\geq\; \eta_2\delta_k.$$

With the above identifications, Assumptions 2.2 and 2.1 are verified by the following lemmas.

The next lemma records the standard fact that the trust-region radius converges to zero.

to_zero_tr⟩? **Lemma 3.3.** *Consider Algorithm 3.1. Under Assumptions 2.3, 3.1, and 3.2, for any realization, we have $\delta_k \to 0$ as $k \to \infty$.*

The next lemma shows that, on good iterations with sufficiently small δ_k, the method achieves a decrease proportional to $\|\nabla f(x_k)\|\delta_k$.

trategy_tr⟩? **Lemma 3.4.** *Consider Algorithm 3.1. Under Assumptions 2.3, 3.1, and 3.2, for any $\kappa_{\mathrm{eg}} > 0$ and $\kappa_{\mathrm{ef}} > 0$, the definition of $y_k(\kappa_{\mathrm{eg}}, \kappa_{\mathrm{ef}})$ satisfies Assumption 2.1 with $\bar{\alpha} = c_2\varepsilon$ and $\zeta = m\eta_1/4$, where c_2 is defined in (3.4).*

We can now state the series-based convergence result for the trust-region method.

rust_region⟩ **Theorem 3.3.** *Consider Algorithm 3.1 and the configuration of $\{y_k\}$ as (3.2). Under Assumptions 2.3, 3.1, and 3.2, if there exist constants $\kappa_{\mathrm{eg}} > 0$ and $\kappa_{\mathrm{ef}} > 0$ such that $H = \infty$, then we have*

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

Theorems 3.1 and 3.2 follow from Theorem 3.3 combined with Corollaries 2.1 and 2.2, respectively.

4 Direct search based on sufficient decrease

rect-search⟩ We next consider a simplified direct-search method based on the sufficient decrease condition. The presentation emphasizes the adaptive step-size update and prepares for the connection with the series-based framework in Section 2.

Algorithm 4.1 A simplified direct-search method with 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

$$f(x_k) - f(x_k + \alpha_k d) \ > \ \rho(\alpha_k). \tag{4.1}$$

eq:sufficient_

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

4.1 Basic definitions, assumptions, and existing results

In Algorithm 4.1, a function $\rho : (0,\infty) \to (0,\infty)$ is called a forcing function if it is nondecreasing and satisfies $\rho(\alpha) = o(\alpha)$ as $\alpha \to 0^+$. The inequality (4.1) is the sufficient decrease condition.

Convergence analysis for variants of Algorithm 4.1 can be found in [9, 11, 14, 16]. We briefly recall the key geometric notion needed in these results.

ine_measure⟩ **Definition 4.1** (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, denoted by $\mathrm{cm}(\mathcal{D}, v)$, is defined as

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

The cosine measure of $\mathcal{D}$, denoted by $\mathrm{cm}(\mathcal{D})$, is defined by

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

We impose the following standard normalization on the direction sets.

unit_length⟩ **Assumption 4.1.** *For each $k \geq 0$, the direction set $\mathcal{D}_k$ is a finite set of unit vectors in $\mathbb{R}^n$.*

Remark 4.1. Assumption 4.1 can be replaced by assuming uniform lower and upper bounds on the lengths of vectors in all direction sets. Without loss of generality, we normalize all directions to unit length.

With these assumptions in place, global convergence for deterministic direction sets follows from classical analysis.

ministic_ds⟩ **Theorem 4.1** ([14, Theorem 3.11]). *Consider Algorithm 4.1. Under Assumptions 2.3 and 4.1, if there exists a constant* $\kappa > 0$ *such that* $\mathrm{cm}(\mathcal{D}_k) \geq \kappa$ *for each* $k \geq 0$*, then we have*

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

We also consider a randomized variant, in which the direction set is generated randomly.

Algorithm 4.2 Probabilistic direct search

rect-search⟩ Identical to Algorithm 4.1 except that the direction set in Step 1 is generated randomly.

We denote the random direction set by $\mathfrak{D}_k$, with realization $\mathcal{D}_k$. The convergence analysis of Algorithm 4.2 relies on the notion of "p-probabilistically κ-descent".

-k_descent⟩? **Definition 4.2** ([11, Definition 3.1]). Consider Algorithm 4.2 with f being differentiable. The direction set sequence $\{\mathfrak{D}_k\}$ is said to be p-probabilistically κ-descent if it satisfies

$$\mathbb{P}\left(\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa \mid \mathcal{F}^{\mathfrak{D}}_{k-1}\right) \;\geq\; p \quad \text{for each } k \geq 0,$$

where $\mathcal{F}^{\mathfrak{D}}_{k-1} = \sigma(\mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1})$ and $\mathcal{F}^{\mathfrak{D}}_{-1}$ is the trivial σ-algebra.

Using this notion, [11] established the global convergence of Algorithm 4.1 under random direction sets.

bilistic_ds⟩ **Theorem 4.2** ([11, Theorem 3.4]). *Consider Algorithm 4.1. Under Assumptions 2.3 and 4.1, if* $\{\mathcal{D}_k\}$ *is p-probabilistically* κ*-descent with* $p = \log\theta/\log(\gamma^{-1}\theta)$ *and a positive constant* κ*, then we have*

$$\mathbb{P}\left(\liminf_{k\to\infty} \|\nabla f(X_k)\| = 0\right) \;=\; 1.$$

4.2 Convergence analysis based on the series

c:ds_series⟩ We now connect Algorithm 4.1 to the abstract framework of Section 2. Fix $\kappa > 0$, and define the indicator

$$y_k(\kappa) \;=\; \mathbb{1}\big(\mathrm{cm}(\mathcal{D}_k, -\nabla f(x_k)) \geq \kappa\big). \tag{4.2}$$ eq:yk_ds

Intuitively, $y_k(\kappa) = 1$ means that $\mathcal{D}_k$ contains a direction making an angle uniformly bounded away from $\pi/2$ with the steepest-descent direction.

_decreasing⟩ **Lemma 4.1.** *Consider Algorithm 4.1 and fix a constant $\kappa > 0$. Under Assumptions 2.3 and 4.1, if $y_k(\kappa) = 1$ and $\alpha_k \le \kappa\|\nabla f(x_k)\|/L$, then we have*

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

Proof. By Definition 4.1, when $y_k(\kappa) = 1$, there exists $d_k^* \in \mathcal{D}_k$ such that

$$-\nabla f(x_k)^\mathsf{T} d_k^* \;\ge\; \kappa\|\nabla f(x_k)\|\|d_k^*\| \;=\; \kappa\|\nabla f(x_k)\|,$$

where the last equality comes from Assumption 4.1. Then we have

$$\begin{aligned}\max_{d\in\mathcal{D}_k}\{f(x_k) - f(x_k + \alpha_k d)\} &\;\ge\; f(x_k) - f(x_k + \alpha_k d_k^*)\\ &\;\ge\; -\nabla f(x_k)^\mathsf{T} d_k^* \alpha_k - \frac{L}{2}\alpha_k^2\\ &\;\ge\; \kappa\|\nabla f(x_k)\|\alpha_k - \frac{L}{2}\alpha_k^2,\end{aligned}$$

where the second inequality is due to the Lipschitz continuity of ∇f by Assumption 2.3. The result follows from the fact that $\kappa\|\nabla f(x_k)\|\alpha_k - L\alpha_k^2/2 \ge \kappa\|\nabla f(x_k)\|\alpha_k/2$ when $\alpha_k \le \kappa\|\nabla f(x_k)\|/L$. □

For the series-based argument below, we also impose a complete polling rule: whenever an iterate is updated, the best trial point among all polling directions is selected.

ete_polling⟩ **Assumption 4.2.** *Consider Algorithm 4.1. If $x_{k+1} \ne x_k$, then we have*

$$f(x_{k+1}) \;=\; \min_{d\in\mathcal{D}_k} f(x_k + \alpha_k d).$$

We next make the connection to Algorithm 2.1 explicit. Algorithm 4.1 can be viewed as an instance of Algorithm 2.1 via the following choices.

- The step is defined as $s_k(\alpha_k) = \alpha_k d_k$ with

$$d_k \;=\; \operatorname*{argmin}_{d\in\mathcal{D}_k} f(x_k + \alpha_k d).$$

- The local model is defined as

$$m_k(s) \;=\; \begin{cases} f(x_k + \alpha_k s), & \text{if } s \in \mathcal{D}_k \cup \{0\},\\ \infty, & \text{otherwise.}\end{cases}$$

- The sufficient decrease condition and the quality condition are defined as

$$f(x_k) - f(x_k + s_k(\alpha_k)) \;\geq\; \rho(\alpha_k).$$

With the above identifications, Assumptions 2.2 and 2.1 are verified by the following lemmas.

epsize_to_0⟩ **Lemma 4.2** ([14, Theorem 3.4])**.** *Consider Algorithm 4.1. Under Assumption 2.3, we have* $\alpha_k \to 0$.

trategy_ds⟩? **Lemma 4.3.** *Consider Algorithm 4.1. Under Assumptions 2.3, 4.1, and 4.2, for any* $\kappa > 0$*, the definition of* $y_k(\kappa)$ *satisfies Assumption 2.1 with some* $\bar{\alpha}$ *depending on* ρ *and* ε *and* $\zeta = \kappa/2$.

Proof. By Lemma 4.1, it remains to show that, for every $\varepsilon > 0$, there is a constant $\bar{\alpha} \in (0, \kappa/L]$ such that, if $y_k(\kappa) = 1$, $\|\nabla f(x_k)\| \geq \varepsilon$, and $\alpha_k \leq \bar{\alpha}$, then

$$\max_{d \in \mathcal{D}_k} \{f(x_k) - f(x_k + \alpha_k d)\} \;\geq\; \frac{\kappa}{2}\|\nabla f(x_k)\|\alpha_k \;\geq\; \frac{\kappa\varepsilon}{2}\alpha_k \;\geq\; \rho(\alpha_k), \qquad (4.3)$$ eq:ds_satisfy_

where the first inequality comes from Lemma 4.1 and $\bar{\alpha} \leq \kappa/L$. The existence of $\bar{\alpha}$ fulfilling the last inequality in (4.3) comes from the fact that $\rho(\alpha) = o(\alpha)$. □

We can now state the main convergence result for direct search.

e_series_ds⟩ **Theorem 4.3.** *Consider Algorithm 4.1 and the configuration of* $\{y_k\}$ *as* (4.2)*. Under Assumptions 2.3, 4.1, and 4.2, if there exists a constant* $\kappa > 0$ *such that* $H(\kappa) = \infty$*, then*

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

Theorems 4.1 and 4.2 follow from Theorem 4.3 combined with Corollaries 2.1 and 2.2, respectively.

4.3 Necessity of the complete polling assumption

nterexample⟩ We now show by a counterexample that the complete polling assumption (Assumption 4.2) cannot be dropped from Theorem 4.3. Specifically, we construct an instance of Algorithm 4.1 without complete polling for which $H(\kappa) = \infty$ for some $\kappa > 0$, yet $\liminf_k \|\nabla f(x_k)\| > 0$.

The construction exploits the contrast between the divergence of the harmonic series $\sum_{k=1}^{\infty} 1/k = \infty$ and the convergence of the sum of squared reciprocals $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$.

Setup

Consider the quadratic objective $f(x) = \frac{1}{2}x^\mathsf{T}x$ on $\mathbb{R}^2$, so that $\nabla f(x) = x$ and the unique minimizer is the origin. Define a decreasing sequence of radii

$$r_\ell \;=\; 1 + \frac{\pi^2}{6} - \sum_{i=1}^{\ell} \frac{1}{i^2}, \quad \ell \geq 1, \tag{4.4}$$

?eq:def_radii?

together with $r_0 = 1 + \pi^2/6$ and $r_\infty = \lim_{\ell\to\infty} r_\ell = 1$. These radii define a family of concentric circles centered at the origin, with

$$r_{\ell-1} - r_\ell \;=\; \frac{1}{\ell^2}, \qquad \ell \geq 1. \tag{4.5}$$

?eq:gap_radii?

We set the algorithmic parameters to $\gamma = 1/\theta = 2$ and the forcing function to $\rho(\alpha) = \frac{1}{2}\alpha^2$.

Construction of the sequences

We now define the sequences $\{x_k\}$, $\{\alpha_k\}$, $\{\mathcal{D}_k\}$, and an auxiliary index sequence $\{\ell_k\}$ that records which circle the current iterate lies on. Set $x_0 = (r_0, 0)$, $\alpha_0 = 1$, and $\ell_0 = 0$, with the convention $1/0 = \infty$. For each $k \geq 0$, define

$$\mathcal{D}_k \;=\; \begin{cases} \left\{-\frac{x_k}{\|x_k\|}\right\} \cup \left\{d \in \mathbb{R}^2 : \|d\| = 1,\ \|x_k + \alpha_k d\| = r_{\ell_k+1}\right\} & \text{if } \alpha_k \leq \frac{1}{\ell_k}, \\ \left\{\frac{x_k}{\|x_k\|}\right\} & \text{if } \alpha_k > \frac{1}{\ell_k}, \end{cases} \tag{4.6}$$

eq:def_Dk

$$\ell_{k+1} \;=\; \begin{cases} \ell_k + 1 & \text{if } \alpha_k \leq \frac{1}{\ell_k}, \\ \ell_k & \text{if } \alpha_k > \frac{1}{\ell_k}, \end{cases} \tag{4.7}$$

?eq:def_ellk?

$$x_{k+1} \;=\; \begin{cases} x_k + \alpha_k d_k, \text{ where } d_k \in \mathcal{D}_k \text{ satisfies } \|x_k + \alpha_k d_k\| = r_{\ell_{k+1}} & \text{if } \alpha_k \leq \frac{1}{\ell_k}, \\ x_k & \text{if } \alpha_k > \frac{1}{\ell_k}, \end{cases} \tag{4.8}$$

eq:def_xk

$$\alpha_{k+1} \;=\; \begin{cases} 2\alpha_k & \text{if } \alpha_k \leq \frac{1}{\ell_k}, \\ \frac{1}{2}\alpha_k & \text{if } \alpha_k > \frac{1}{\ell_k}. \end{cases} \tag{4.9}$$

eq:def_alphak

te_polling⟩? **Remark 4.2** (Role of incomplete polling)**.** The key mechanism of this counterexample is the choice of x_{k+1} in (4.8) when $\alpha_k \leq 1/\ell_k$. In this case, the direction set $\mathcal{D}_k$ contains two directions: the steepest-descent direction $-x_k/\|x_k\|$ and a nearly tangential direction d_k satisfying $\|x_k + \alpha_k d_k\| = r_{\ell_k+1}$. Among these, the steepest-descent direction produces a larger decrease in f because it moves x_k toward the origin:

$$f(x_k) - f(x_k - \alpha_k x_k/\|x_k\|) \;>\; f(x_k) - f(x_k + \alpha_k d_k).$$

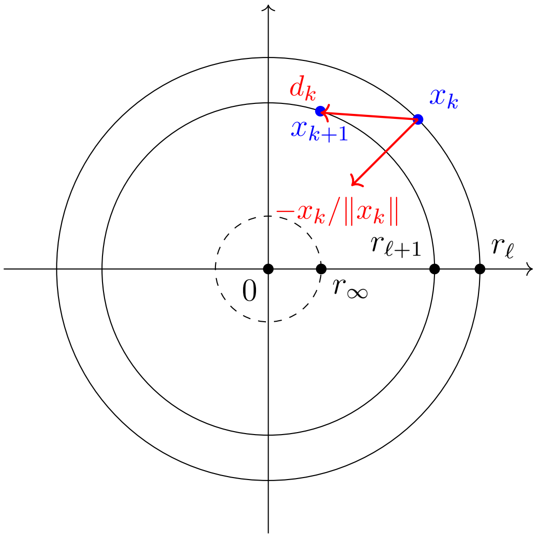

Figure 1: Illustration of the counterexample when $\alpha_k \le 1/\ell_k$. The direction set $\mathcal{D}_k$ contains two directions (red arrows): the steepest-descent direction $-x_k/\|x_k\|$ and a nearly tangential direction d_k that moves x_k to the next inner circle. The algorithm selects d_k and sets $x_{k+1} = x_k + \alpha_k d_k$ on the circle of radius $r_{\ell+1}$, instead of using $-x_k/\|x_k\|$, which would yield a larger decrease in f. This is precisely where complete polling is violated.

illustrate⟩?

However, our construction deliberately selects the direction d_k that moves the iterate to the next inner circle while avoiding the steepest-descent direction. Under complete polling (Assumption 4.2), the algorithm would be required to choose the direction that minimizes f, namely $-x_k/\|x_k\|$, and the resulting decrease in f would be much larger. It is this "wasteful" choice of direction — permitted only when complete polling is not enforced — that allows the total decrease in f to remain finite even though the series H diverges.

Validity of the construction

We first verify that the above sequences form a valid realization of Algorithm 4.1 (without complete polling).

mple_valid⟩? **Lemma 4.4.** *The sequences* $\{x_k\}$, $\{\alpha_k\}$, *and* $\{\mathcal{D}_k\}$ *defined in* (4.6)–(4.9) *form a valid realization of Algorithm 4.1 with the parameters* $\gamma = 2$, $\theta = 1/2$, $\rho(\alpha) = \frac{1}{2}\alpha^2$, *and* $f(x) = \frac{1}{2}x^\top x$. *In particular, the direction sets consist of unit vectors, and the step-size update rule in* (4.9) *agrees with Step* 3 *of Algorithm 4.1 for all* k *with* $\ell_k \ge 3$.

Proof. We verify each step of Algorithm 4.1.

Step 1 (Direction set). By construction, $\mathcal{D}_k$ in (4.6) is a finite set of unit vectors for each $k \geq 0$, satisfying Assumption 4.1.

Step 2 (Sufficient decrease check). We need to show that the step-size update rule matches the sufficient decrease condition (4.1). Consider the case $\alpha_k \leq 1/\ell_k$ with $\ell_k \geq 3$. In this case, $x_{k+1} = x_k + \alpha_k d_k$ with $\|x_{k+1}\| = r_{\ell_k+1}$, and

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= \frac{1}{2}\big(r_{\ell_k}^2 - r_{\ell_k+1}^2\big) = \frac{1}{2}(r_{\ell_k} + r_{\ell_k+1})(r_{\ell_k} - r_{\ell_k+1}) \\ &> r_{\ell_k} - r_{\ell_k+1} = \frac{1}{(\ell_k+1)^2} \geq \frac{1}{2}\alpha_k^2 = \rho(\alpha_k), \end{aligned} \tag{4.10}$$

?eq:counterexam

where the first inequality uses $r_{\ell_k} + r_{\ell_k+1} > 2$ (since both radii exceed 1) and the second inequality holds because $\alpha_k \leq 1/\ell_k$ and $\ell_k \geq 3$ imply $\frac{1}{2}\alpha_k^2 \leq \frac{1}{2\ell_k^2} \leq \frac{1}{(\ell_k+1)^2}$. Hence the direction $d_k \in \mathcal{D}_k$ achieves sufficient decrease, so Step 3 of Algorithm 4.1 expands the step size: $\alpha_{k+1} = \gamma\alpha_k = 2\alpha_k$, consistent with (4.9).

Consider now the case $\alpha_k > 1/\ell_k$ with $\ell_k \geq 3$. Here $\mathcal{D}_k = \{x_k/\|x_k\|\}$, and the only trial point is $x_k + \alpha_k x_k/\|x_k\|$, which has $\|x_k + \alpha_k x_k/\|x_k\|\| = r_{\ell_k} + \alpha_k > r_{\ell_k}$, so $f(x_k + \alpha_k x_k/\|x_k\|) > f(x_k)$. No direction achieves sufficient decrease, so $x_{k+1} = x_k$ and $\alpha_{k+1} = \theta\alpha_k = \alpha_k/2$, again consistent with (4.9).

Step 3 (Iterate and step-size update). By the above analysis, the iterate update in (4.8) and the step-size update in (4.9) agree with Step 3 of Algorithm 4.1 for all k with $\ell_k \geq 3$. For the finitely many initial iterations with $\ell_k < 3$, the sequences can be checked directly. □

Divergence of the series

We now show that $H(\kappa) = \infty$ for $\kappa = 1/2$.

_H_diverges⟩ **Lemma 4.5.** *For the sequences defined in (4.6)–(4.9) with $\kappa = 1/2$ and $y_k(\kappa)$ defined as in (4.2), we have $H(1/2) = \infty$.*

Proof. We first determine $y_k(1/2)$. When $\alpha_k \leq 1/\ell_k$, the direction $-x_k/\|x_k\|$ belongs to $\mathcal{D}_k$, and $\mathrm{cm}(\mathcal{D}_k, -\nabla f(x_k)) = \mathrm{cm}(\mathcal{D}_k, -x_k) = 1 \geq 1/2$, so $y_k(1/2) = 1$. When $\alpha_k > 1/\ell_k$, we have $\mathcal{D}_k = \{x_k/\|x_k\|\}$, so $\mathrm{cm}(\mathcal{D}_k, -x_k) = -1 < 1/2$ and hence $y_k(1/2) = 0$. Therefore, for $\ell_k \geq 3$, the indicator $y_k(1/2) = 1$ if and only if $\alpha_k \leq 1/\ell_k$, and the step-size update satisfies $\alpha_{k+1} = \gamma^{y_k(1/2)}\theta^{1-y_k(1/2)}\alpha_k$. Consequently, there exists a constant $c > 0$ such that

$$H(1/2) = c\sum_{k=1}^{\infty} \alpha_k. \tag{4.11}$$

eq:H_equals_c_

We claim that $\alpha_k \geq 1/(2\ell_k)$ for each $k \geq 1$. The base case holds since $\alpha_1 = 2 \geq 1/2 = 1/(2\ell_1)$ (as $\ell_1 = 1$). For the inductive step, suppose $\alpha_n \geq 1/(2\ell_n)$. We consider two cases.

- If $\alpha_n > 1/\ell_n$, then $\ell_{n+1} = \ell_n$ and $\alpha_{n+1} = \alpha_n/2 > 1/(2\ell_n) = 1/(2\ell_{n+1})$.
- If $\alpha_n \le 1/\ell_n$, then $\ell_{n+1} = \ell_n + 1$ and $\alpha_{n+1} = 2\alpha_n \ge 1/\ell_n \ge 1/(2\ell_{n+1})$.

In both cases $\alpha_{n+1} \ge 1/(2\ell_{n+1})$, completing the induction. Since $\alpha_k \to 0$ by Lemma 4.2, we deduce $\ell_k \to \infty$. Moreover, each value $\ell \ge 1$ is attained by ℓ_k for at least one index k (because ℓ_k increases by at most 1 at each step), so

$$\sum_{k=1}^{\infty} \alpha_k \;\ge\; \sum_{k=1}^{\infty} \frac{1}{2\ell_k} \;\ge\; \frac{1}{2} \sum_{\ell=1}^{\infty} \frac{1}{\ell} \;=\; \infty.$$

Together with (4.11), this gives $H(1/2) = \infty$. □

Non-convergence of the iterates

onvergence⟩? **Lemma 4.6.** *For the sequences defined in* (4.6)–(4.9)*, we have* $\lim_{k\to\infty} \|\nabla f(x_k)\| = 1$*. In particular,* $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 1 \ne 0$.

Proof. By construction, $\|x_k\| = r_{\ell_k}$ for each $k \ge 0$. Since $\ell_k \to \infty$ (as shown in the proof of Lemma 4.5), we have $\|\nabla f(x_k)\| = \|x_k\| = r_{\ell_k} \to r_\infty = 1 > 0$. □

5 Relation to the non-convergence result and open questions

sec:remarks⟩ Recall that, in Section 4.2, for direct search with sufficient decrease and complete polling, we defined the indicator

$$y_k(\kappa) \;=\; \mathbb{1}\big(\mathrm{cm}(\mathcal{D}_k, -\nabla f(x_k)) \ge \kappa\big)$$

for a fixed $\kappa > 0$, and the series

$$H(\kappa) \;=\; \sum_{k=0}^{\infty} \prod_{\ell=0}^{k-1} \gamma^{y_\ell(\kappa)} \theta^{1-y_\ell(\kappa)}.$$

Theorem 4.3 shows that first-order convergence of direct search follows once there exists $\kappa > 0$ such that $H(\kappa) = \infty$.

In contrast, the paper [13] studies the opposite question: under what conditions can (probabilistic) direct search fail to converge? There, the indicator is defined as

$$z_k \;=\; \mathbb{1}\big(\mathrm{cm}(\mathcal{D}_k, -\nabla f(x_k)) > 0\big).$$

Based on $\{z_k\}$, they define another series

$$\hat{H} = \sum_{k=0}^{\infty} z_k \prod_{\ell=0}^{k-1} \gamma^{z_\ell} \theta^{1-z_\ell}. \tag{5.1}$$

?eq:hat_H?

Assuming that f is smooth and convex, [13] shows that when $\hat{H} < \infty$, direct search with sufficient decrease can fail to converge (without requiring complete polling).

To compare these series, define

$$\tilde{H}(\kappa) = \sum_{k=0}^{\infty} y_k(\kappa) \prod_{\ell=0}^{k-1} \gamma^{y_\ell(\kappa)} \theta^{1-y_\ell(\kappa)}.$$

Then we always have $H(\kappa) \geq \tilde{H}(\kappa)$. Moreover, since $y_k(\kappa) = 1$ implies $z_k = 1$ for any $\kappa > 0$, we have $\tilde{H}(\kappa) \geq \hat{H}$ for any $\kappa > 0$. Hence, the convergence condition $H(\kappa) = \infty$ in Theorem 4.3 is (in general) stronger than requiring $\hat{H} = \infty$, while the non-convergence regime $\hat{H} < \infty$ studied in [13] does not contradict our result.

Several questions remain open. First, for direct search, the borderline case $\kappa = 0$ in the definition of $y_k(\kappa)$ is not covered by our analysis. Second, although Subsection 4.3 shows that complete polling is necessary for our deterministic series-based convergence result, it is unclear whether complete polling is also essential in the probabilistic setting. Third, it would be interesting to develop an analogous non-convergence theory for derivative-free trust-region methods, complementing the series-based convergence results in Section 3. Fourth, we only study a simplified DFO trust-region framework. In contrast, the classical derivative-free trust-region algorithm (see, e.g., Algorithm 10.1 in [7]) includes a criticality step that explicitly addresses the regime of small model gradients by reducing the trust-region radius to be commensurate with the model gradient. It would be interesting to understand whether this classical mechanism implies our series assumption automatically, i.e., whether it necessarily generates sufficiently many fully linear models so that $H = \infty$.

6 Conclusion

:conclusion⟩ We proposed a unified series-based perspective for the convergence analysis of derivative-free trust-region and direct-search methods. In particular, we identified an algorithm-determined series of the form

$$H = \sum_{k=0}^{\infty} \prod_{\ell=0}^{k-1} \gamma^{y_\ell} \theta^{1-y_\ell},$$

and showed that its divergence provides a simple sufficient condition for first-order convergence, namely $\liminf_k \|\nabla f(x_k)\| = 0$.

We illustrated this framework by recovering convergence guarantees for a simplified derivative-free trust-region method (including the deterministic and probabilistic settings) and by establishing a parallel series condition for direct search based on sufficient decrease. For direct search, our analysis requires complete polling; we also provided a counterexample (Subsection 4.3) showing that complete polling cannot be removed from our series-based convergence theorem in general.

Finally, we discussed the connection between our convergence condition and the complementary non-convergence analysis in [13], highlighting how different series conditions reflect different algorithmic regimes. Together, these results suggest that series conditions offer a useful lens for understanding both convergence and potential failure modes of randomized and deterministic derivative-free methods.

References

let_Hare_2017 [1] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Springer, Cham, 2017.

_Vicente_2014 [2] A. S. Bandeira, K. Scheinberg, and L. N. Vicente. Convergence of trust-region methods based on probabilistic models. *SIAM J. Optim.*, 24:1238–1264, 2014.

het_etal_2019 [3] J. Blanchet, C. Cartis, M. Menickelly, and K. Scheinberg. Convergence rate analysis of a stochastic trust region method via supermartingales. *INFORMS J. Optim.*, 1:92–119, 2019.

_Roberts_2022 [4] C. Cartis and L. Roberts. Scalable subspace methods for derivative-free nonlinear least-squares optimization. *Math. Program.*, 199:461–524, 2023.

d_Toint_2000 [5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods.* SIAM, Philadelphia, 2000.

Vicente_2009a [6] A. R. Conn, K. Scheinberg, and L. N. Vicente. Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points. *SIAM J. Optim.*, 20:387–415, 2009.

Vicente_2009b [7] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*, volume 8 of *MOS-SIAM Ser. Optim.* SIAM, Philadelphia, 2009.

Yin_Yuan_2000 [8] Y. H. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y. Yuan. Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.*, 10:345–358, 2000.

_Vicente_2016 [9] M. Dodangeh and L. N. Vicente. Worst case complexity of direct search under convexity. *Math. Program.*, 155:307–332, 2016.

Durrett_2010 [10] R. Durrett. *Probability: Theory and Examples.* Camb. Ser. Stat. Probab. Math. Cambridge University Press, Cambridge, fourth edition, 2010.

e_Zhang_2015 [11] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search based on probabilistic descent. *SIAM J. Optim.*, 25:1515–1541, 2015.

e_Zhang_2018 [12] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Complexity and global rates of trust-region methods based on probabilistic models. *IMA J. Numer. Anal.*, 38:1579–1597, 2018.

Zhang_2026a [13] C. Huang and Z. Zhang. Non-convergence analysis of probabilistic direct search, 2026.

Torczon_2003 [14] 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 [15] J. Larson, M. Menickelly, and S. M. Wild. Derivative-free optimization methods. *Acta Numer.*, 28:287–404, 2019.

Vicente_2013 [16] L. N. Vicente. Worst case complexity of direct search. *EURO J. Comput. Optim.*, 1:143–153, 2013.

Wolfe_1969 [17] Ph. Wolfe. Convergence conditions for ascent methods. *SIAM Rev.*, 11:226–235, 1969.

Wolfe_1971 [18] Ph. Wolfe. Convergence conditions for ascent methods. ii: some corrections. *SIAM Rev.*, 13:185–188, 1971.

tendijk_1970 [19] G. Zoutendijk. Nonlinear programming, computational methods. In J. Abadie, editor, *Integer and Nonlinear Programming*, pages 37–86. North-Holland, Amsterdam, 1970.