A Direction Adaptation Evolution Strategy for Noisy Derivative-Free Optimization

Morteza Kimiaei[1] and Mahsa Yousefi[2]

[1]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, Vienna, A-1090, Austria.

[2]Department of Industrial Engineering, University of Florence, Viale G.B. Morgagni 40, Florence, 50134, Italy.

Contributing authors: morteza.kimiaei@univie.ac.at; mahsa.yousefi@unifi.it;

Abstract

In noisy derivative-free optimization (DFO), methods based on the matrix adaptation evolution strategy (`MAES`) often mistakenly identify points with spuriously good function values, undermining selection and leading to unreliable search directions. To address this, we develop a novel direction adaptation evolution strategy (`DAES`) that targets the algorithmic phases prone to noise propagation, combining selected-population search with complexity analysis. Fixing the transformation matrix to the identity, `DAES` employs a structured direction-generation mechanism consisting of symmetric sampling, joint sorting–selection of noisy values, partitioning of all sampled directions into three ranked groups, and construction of a heuristic search direction. Along this direction, a noisy derivative-free sufficient descent condition with extrapolation is applied. Using a probabilistic angle condition preserved through our proposed framework, we obtain, to the best of our knowledge, the first high-probability evaluation-complexity bounds for a noisy `MAES`-type method across nonconvex, convex, and strongly convex objectives, under bounded noise. Finally, experiments on 230 `BARON BCP` problems across eight noise levels compare `DAES` with five DFO solvers, notably `MADFO`, showing greater robustness at intermediate-to-small budgets.

Keywords: Noisy DFO, Evolution Strategies, Randomized Optimization, Direction Adaptation, Complexity Analysis

1 Introduction

We consider the unconstrained derivative-free optimization (DFO) problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is smooth but not directly accessible. Instead, we assume access to a *noisy oracle* that, for a given point $x \in \mathbb{R}^n$, returns a noisy function value $\tilde{f}(x)$, contaminated by an unknown noise term $\tilde{f}(x) - f(x)$. We make no assumptions regarding the explicit form of $f(x)$, the availability of its gradient $g(x) := \nabla f(x)$, Lipschitz continuity, or the statistical properties of the noise. This setting is commonly referred to as the *noisy DFO problem*; see, e.g., [1, 2].

To solve the noisy DFO problem (1), numerous deterministic and randomized methods have been proposed, including line search (`LS`) solvers [3–11], trust-region (`TR`) methods [3, 12–15], direct search (`DS`) methods [3, 16–22], and matrix adaptation evolution strategies (`MAES`) [23–26]. For advanced DFO methods that combine `DS` or `LS` with model-based techniques or with `MAES`, we refer the reader to [25–27].

`DS` methods evaluate the objective function along a finite set of polling directions, accepting trial points that satisfy a forcing-function-based decrease condition (typically proportional to the squared step size). Requiring no derivative information, they adjust step sizes using simple increase–decrease rules. They are robust and effective for low-dimensional problems.

Conversely, several `LS` methods explore descent along prescribed directions, differing mainly in how they enforce sufficient decrease. While some variants approximate directional derivatives using Armijo/Wolfe conditions, others rely on forcing functions akin to `DS` methods—highlighting the close link between derivative-free `LS` and `DS` schemes. Additionally, these variants employ extrapolation steps to bypass saddle points or local maximizers and rapidly converge toward an approximate minimizer.

`MAES` consists of three phases: mutation, selection, and recombination. In the first phase, candidate solutions are generated by sampling random directions and applying a transformation matrix, which requires a specific construction and matrix–vector multiplications. In the second phase, mutation points (candidates) are ranked according to their noisy objective function values, and the corresponding mutation directions associated with the lowest values are retained. The last phase computes a weighted average of the selected directions to form a direction by which a new recombination point is generated. Finally, the transformation matrix and step-sizes are updated. In noisy regimes, an `MAES`-based algorithm may mistakenly identify points with spuriously good function values, undermining the reliability of the selection process and leading to misleading search directions. Failure in this phase can adversely affect the recombination phase. Furthermore, global and linear convergence are known for certain variants [28–30], but these results do not establish evaluation-complexity bounds for noisy `MAES`-type methods. We address these challenges in this paper.

1.1 Our Contribution

To address the limitation in MAES-type methods, the central objective of the present paper is twofold: first, to mitigate noise sensitivity through targeted algorithmic modifications across phases with the potential to control noise; and second, to establish complexity guarantees for the resulting method.

For noisy DFO, we propose the direction adaptation evolution strategy (DAES), a new MAES-type method. DAES promotes balanced exploration by sampling symmetric pairs of random directions. It selects a specialized population from multiple generated populations and applies structured within- and cross-group weighting schemes to jointly target instabilities arising from evaluation noise and directional cancellation. Specifically, all mutation directions of the selected population are partitioned into three ranked groups, yielding three directions whose geometric arrangement forms a novel search direction toward promising regions. This redesign replaces covariance matrix adaptation with a structured population-based direction mechanism; indeed, unlike standard MAES, DAES fixes the transformation matrix to the identity, simplifying its theoretical analysis while preserving efficiency. To guarantee sufficient decrease, DAES adopts a forcing function-based condition with an extrapolation strategy.

Under standard assumptions, we establish a probabilistic angle condition for the novel search direction and control per-iteration failure probabilities to guarantee uniform high-probability directional and descent properties. Across nonconvex, convex, and strongly convex settings, we derive complexity bounds matching deterministic rates at the noise-limited accuracy level—to our knowledge, the first for a noisy MAES-type method. Numerically, we evaluate the structurally simple DAES via ablation studies—which favor single-population and extrapolation settings for efficiency—and comparisons against five solvers across 230 BCP problems at eight noise levels, confirming superior robustness over MADFO [26] at intermediate-to-small budgets.

1.2 Notation and Organization

In this paper, we denote $[n] := \{1, \ldots, n\}$, $[n]_0 := \{0, \ldots, n\}$, $\mathcal{N}(0, I_n)$ as the standard n-dimensional normal distribution, and $a \sim \mathcal{U}(0, 1)$ as a uniform random variable on $(0, 1)$. For a σ-algebra $\mathcal{F}$, $\mathbb{P}(\cdot \mid \mathcal{F})$ denotes conditional probability given $\mathcal{F}$. We use $\mathbb{E}$, $\|\cdot\|$, $\lceil\cdot\rceil$, $\propto$ and $\emptyset$ to denote the mathematical expectation, Euclidean norm, ceiling function, proportionality and empty set, respectively. The abbreviation i.i.d. stands for independent and identically distributed random variables. Unless specified as the mathematical constant, π represents the permutation induced during the sorting–selection step. We also adopt the conventional asymptotic notation $\mathcal{O}$, Ω, and Θ to characterize growth rates, and sup for the least upper bound.

The paper is organized as follows. Section 2 introduces DAES and details its mechanism. We establish complexity guarantees in Sections 3.2 and 3.3 for deterministic and randomized variants of DAES, respectively. In Section 4, numerical experiments are reported. Section 5 concludes the paper.

2 DAES: A New DFO Method

This section details the DAES framework for solving (1), starting from an initial point x^0 at $\ell = 0$. We detail its three enhanced algorithmic phases and summarize the complete procedure in Subsection 2.4.

2.1 Mutation Phase

During iteration ℓ of DAES, assuming an even population size λ, the mutation phase generates three components to construct the offspring population as follows:

2.1.1 Mutation Directions

Although sampling from the standard normal distribution $\mathcal{N}(0, I_n)$ is unbiased in expectation, finite sample sets may exhibit directional asymmetries. Inspired by symmetric search strategies in some DFO methods (e.g., [8, 10]), DAES mitigates such sample-induced bias by drawing $\lambda/2$ independent random vectors $z^{j\ell} \sim \mathcal{N}(0, I_n)$ and forming symmetric pairs $(+z^{j\ell}, -z^{j\ell})$, for $j \in [\lambda/2]$, thereby guaranteeing a zero empirical mean. Moreover, unlike MAES, which uses a scaling matrix to transform distribution directions into mutation directions, DAES sets this transformation matrix to the identity. This eliminates the computational overhead of matrix adaptations. Thus, mutation directions are derived directly from the sampled directions. To maintain consistent directional scaling, they are normalized and then assembled into

$$D^{\ell}_{\pm} := \left(\widehat{Z}^{\ell} \; -\widehat{Z}^{\ell}\right) \in \mathbb{R}^{n\times\lambda}, \qquad \text{where} \qquad \widehat{Z}^{\ell} := \left(\frac{z^{1\ell}}{\|z^{1\ell}\|}, \ldots, \frac{z^{\lambda/2,\ell}}{\|z^{\lambda/2,\ell}\|}\right). \tag{2}$$

2.1.2 Mutation Step-Sizes

The mutation step-sizes play a crucial role in ensuring adequate exploration of the search space; see [26, 31]. Using the element-wise division $\oslash$, let us define

$$b^{j\ell} := \left|x^{\ell}_{\text{trec}} \oslash q_{\alpha}\right|, \qquad q_{\alpha} \neq 0 \in \mathbb{R}^n. \tag{3}$$

Here, $q_{\alpha} = z^{j\ell} \sim \mathcal{N}(0, I_n)$. The division is not well-defined in $\mathbb{R}^n$ whenever q_{α} has zero entries, making $b^{j\ell}$ undefined. Hence, to obtain the mutation step-sizes $\alpha^{j\ell}$, we first define an active index set $\mathcal{I} := \{i \in [n] : (q_{\alpha})_i \neq 0\}$, then evaluate the median of $b^{j\ell}_{\mathcal{I}} := (b^{j\ell}_i)_{i\in\mathcal{I}}$ and finally project the resulting value onto $[\alpha_{\min}, \alpha_{\max}]$; i.e.,

$$\alpha^{j\ell} = \min\left\{\alpha_{\max}, \max\left\{\alpha_{\min}, \sqrt{\sigma^{\ell} \cdot \text{median}(b^{j\ell}_{\mathcal{I}})}\right\}\right\} > 0, \qquad \text{for} \quad j \in [\lambda/2]. \tag{4}$$

In (4), σ^ℓ is the step-size scaling factor [24], updated and projected as

$$\sigma^{\ell+1} = \min\left\{\alpha_{\max}, \max\left\{\alpha_{\min}, \sigma^\ell \exp\left(c_\sigma d_\sigma^{-1}\left(\|s_\sigma^\ell\|_2/e_\sigma - 1\right)\right)\right\}\right\}, \tag{5}$$

where $\sigma^0 \in [\alpha_{\min}, \alpha_{\max}]$, $0 < c_\sigma \leq 1$ is a learning rate, $d_\sigma > 0$ is a damping parameter, typically $d_\sigma \approx 1$, and $e_\sigma > 0$ approximates $\mathbb{E}(\|u\|_2)$ for a vector $u \sim \mathcal{N}(0, I_n)$. In (5), s_σ^ℓ is the evaluation path, initialized by $s_\sigma^0 = 0 \in \mathbb{R}^n$ and updated as

$$s_\sigma^{\ell+1} = (1 - c_\sigma)s_\sigma^\ell + \sqrt{c_\sigma(2 - c_\sigma)} \cdot \frac{x_{\text{trec}}^{\ell+1} - x_{\text{trec}}^\ell}{\sigma^\ell}, \tag{6}$$

which accumulates past normalized displacements, serving as a filter to smooth out noise and adaptively guide the step-size scaling in (5).

2.1.3 Mutation Points

Once the mutation step-size $\alpha^{j\ell} \in [\alpha_{\min}, \alpha_{\max}]$ in (4) at iteration ℓ is determined, the λ mutation points $y^{j\ell}$ can be computed as $y^{j\ell} := (y_+^{j\ell} \;\; y_-^{j\ell})$ defined by

$$y_+^{j\ell} := x_{\text{trec}}^\ell + \alpha^{j\ell}\widehat{z}^{j\ell}, \qquad y_-^{j\ell} := x_{\text{trec}}^\ell - \alpha^{j\ell}\widehat{z}^{j\ell}, \qquad j \in [\lambda/2], \tag{7}$$

where $\widehat{z}^{j\ell}$ is the j-th column of $\widehat{Z}^\ell$ in (2), and $x_{\text{trec}}^\ell \in \mathbb{R}^n$ is the current iterate updated in the recombination phase (Subsection 2.3.2), with $x_{\text{trec}}^0 = x^0$. Alternatively, using a unified indexing notation, the λ mutation points can be written as

$$y^{j\ell} = x_{\text{trec}}^\ell + \alpha^{j\ell} d^{j\ell}, \qquad j \in [\lambda],$$

where $d^{j\ell}$ is the j-th column of $D_\pm^\ell \in \mathbb{R}^{n\times\lambda}$ in (2), with $\alpha^{j+\lambda/2,\ell} = \alpha^{j\ell}$ for all $j \in [\lambda/2]$.

2.2 Sorting–Selection Phase

To potentially mitigate noise-induced ranking errors that can arise in single-population classical `MAES` approaches [23, 24, 26], `DAES` is designed to generate N_{sel} independent populations, retain the most reliable one, and aggregate all its symmetric directions for recombination.

Assuming $\lambda = 3\mu$ with $\mu \geq 2$, let $\tilde{f}_\pm^{j\ell} := \tilde{f}(y^{i\ell})$ denote the noisy function evaluations at the j-th mutation point $y^{j\ell} := (y_+^{j\ell} \;\; y_-^{j\ell})$ defined in (7). We aggregate the evaluated objective values into the vector $\tilde{F}_\pm^\ell := (\tilde{F}_+^\ell \;\; \tilde{F}_-^\ell) \in \mathbb{R}^\lambda$

$$\tilde{F}_+^\ell := \left(\tilde{f}_+^{1\ell}, \ldots, \tilde{f}_+^{(\lambda/2)\ell}\right) \quad \text{and} \quad \tilde{F}_-^\ell := \left(\tilde{f}_-^{1\ell}, \ldots, \tilde{f}_-^{(\lambda/2)\ell}\right). \tag{8}$$

DAES sorts the entries of $\tilde{F}_{\pm}^{\ell}$ in ascending order, $\tilde{f}_{\pi}^{1\ell} \leq \cdots \leq \tilde{f}_{\pi}^{\lambda\ell}$, yielding $\tilde{F}_{\pi}^{\ell}$. The corresponding sorting permutation is then applied to the columns of $D_{\pm}^{\ell}$ in (2) to align each direction with its reordered function value, obtaining D_{π}^{ℓ}. Assuming $\lambda = 3\mu$, the columns of D_{π}^{ℓ} are partitioned into three ranked groups with index sets:

$$\mathcal{G}_{\pi}^{1} := \{1, \ldots, \mu\}, \qquad \mathcal{G}_{\pi}^{2} := \{\mu + 1, \ldots, 2\mu\}, \qquad \mathcal{G}_{\pi}^{3} := \{2\mu + 1, \ldots, 3\mu\},$$

corresponding to the μ best, intermediate, and worst objective values, respectively.

Let us define the best element associated with the q-th population and k-th group as

$$j_k^q \in \operatorname*{argmin}_{j \in \mathcal{G}_{\pi}^{k}} \tilde{f}_{\pi}^{j\ell}, \qquad k \in [3], \qquad q \in [N_{\mathrm{sel}}]. \tag{9}$$

Then, the ***minimum within-group noisy separation*** is defined by

$$\Delta_{\mathrm{sel}}^{\ell,q} := \min_{k \in [3]} \min_{r \in \mathcal{G}_{\pi}^{k} \setminus \{j_k^q\}} \left\{ \tilde{f}_{\pi}^{r\ell} - \tilde{f}_{\pi}^{j_k^q \ell} \right\}, \qquad q \in [N_{\mathrm{sel}}]. \tag{10}$$

After generating N_{sel} independent populations and observing $\Delta_{\mathrm{sel}}^{\ell,q}$ for $q \in [N_{\mathrm{sel}}]$, DAES selects the population having the largest observed within-group separation; i.e.,

$$q^{\star} \in \operatorname*{argmax}_{q \in [N_{\mathrm{sel}}]} \Delta_{\mathrm{sel}}^{\ell,q}. \tag{11}$$

This is intended to enhance ranking (sorting) reliability under unknown noise ω and mitigate potential rank inversion. Moreover, this criterion plays a key analytical role in Section 3.3.

Algorithm 1 Mutation and sorting–selection procedure in DAES

Input: x_{trec}^{ℓ}, σ^{ℓ}, s_{σ}^{ℓ}, N_{sel}, and $\lambda = 3\mu$ where $\mu \geq 2$;
for $q \in [N_{\mathrm{sel}}]$ **do** ▷ *Independent populations*
  for $j \in [\lambda/2]$ **do**
    ▶ sample $z^{j\ell} \sim \mathcal{N}(0, I_n)$;
    ▶ obtain $\alpha^{j\ell}$ via (4) with $q_{\alpha} := z^{j\ell}$, and set $\alpha^{j+\lambda/2,\ell} := \alpha^{j\ell}$;
    ▶ generate $y_{\pm}^{j\ell}$ via (7) and evaluate $\tilde{f}(y_{\pm}^{j\ell})$;
  end for
  ▶ form $D_{\pm}^{\ell}$ via (2), $\tilde{F}_{\pm}^{\ell} := \left(\tilde{F}_{+}^{\ell} \;\; \tilde{F}_{-}^{\ell}\right)$ via (8) and $\tilde{F}_{\pi}^{\ell}$ after ascending sorting;
  ▶ permute $D_{\pm}^{\ell}$ to get D_{π}^{ℓ} and partitioned groups $\mathcal{G}_{\pi}^{k}$, $k \in [3]$;
  ▶ compute $\Delta_{\mathrm{sel}}^{\ell,q}$ via (10);
end for
▶ set $q^{\star} = \operatorname{argmax}_{q \in [N_{\mathrm{sel}}]} \Delta_{\mathrm{sel}}^{\ell,q}$ as in (11);
Output: D_{π}^{ℓ} and $\mathcal{G}_{\pi}^{k}$, $k \in [3]$, associated with $q^{\star}$;

2.3 Recombination Phase

This phase converts the ranked candidates retained from the sorting–selection phase into a robust search direction p^{ℓ}_{trec}. Subsection 2.3.1 forms group-representative vectors via within-group weighting and combines them through a triangular construction to obtain p^{ℓ}_{trec}. Subsection 2.3.2 applies this direction in a noisy descent condition, triggers extrapolation, and performs success–failure updates on the iterate.

2.3.1 Triangular Recombination Direction

Using D^{ℓ}_{π}, `DAES` constructs group-representative recombination directions, for $k \in [3]$,

$$d^{\ell}_{\mathrm{rec}\,k} = \sum_{j \in \mathcal{G}^{k}_{\pi}} \bar{w}_j d^{j\ell}_{\pi}, \qquad \text{such that} \sum_{j \in \mathcal{G}^{k}_{\pi}} \bar{w}_j = 1. \tag{12}$$

Here, $\bar{w}_j > 0$ is the ***normalized within-group weight*** assigned to the j-th direction. These weights are defined below, with rationale provided in Subsection 3.3.2.

Let $t_a > 0$ be a ***baseline alignment threshold*** from high-dimensional isotropic Gaussian sampling. Under noisy evaluations and adaptive step-sizes in $[\alpha_{\min}, \alpha_{\max}]$, for the best direction indexed as $j^{\star}_k \in \operatorname{argmin}_{j \in \mathcal{G}^{k}_{\pi}} \tilde{f}^{j\ell}_{\pi}$ in $\mathcal{G}^{k}_{\pi}$, we define a ***particular alignment threshold*** $t_{\mathrm{sel}} > 0$. Obviously, $j^{\star}_1 = 1$, $j^{\star}_2 = \mu + 1$ and $j^{\star}_3 = 2\mu + 1$. To ensure that the other elements within $\mathcal{G}^{k}_{\pi}$ do not cancel the contribution of the best direction, the ***within-group dominance parameter*** $\varepsilon_{\bar{w}} \in (0, 1)$ is defined such that

$$\varepsilon_{\bar{w}} < \frac{t_{\mathrm{sel}}}{1 + t_{\mathrm{sel}}}, \qquad \text{where} \qquad t_{\mathrm{sel}} := (\alpha_{\min}/2\alpha_{\max}) t_a, \quad t_a := \tau/\sqrt{n}, \tag{13}$$

with $n \geq 2$ and $0 < \tau < \sqrt{\pi/2}$. Given $\varepsilon_{\bar{w}} \in (0, 1)$, define

$$\bar{w}_{j^{\star}_k} := 1 - \varepsilon_{\bar{w}}, \qquad \bar{w}_j := \frac{\varepsilon_{\bar{w}}}{\mu - 1}, \qquad j \in \mathcal{G}^{k}_{\pi} \setminus \{j^{\star}_k\}, \qquad k \in [3]. \tag{14}$$

We now formulate the main search direction initiated from the following points:

$$x^{\ell}_{\mathrm{rec}\,k} := x^{\ell}_{\mathrm{trec}} + \alpha^{\ell}_{\mathrm{rec}\,k} d^{\ell}_{\mathrm{rec}\,k}, \qquad \alpha^{\ell}_{\mathrm{rec}\,k} \in [\alpha_{\min}, \alpha_{\max}], \qquad k \in [3], \tag{15}$$

where $d^{\ell}_{\mathrm{rec}\,k}$ is a direction given in (12), yielding a well-defined step-size $\alpha^{\ell}_{\mathrm{rec}\,k}$ via (4), with $q_{\alpha} := d^{\ell}_{\mathrm{rec}\,k}$. Note that any $d^{j\ell}_{\pi}$ is unit based on (2), thus

$$\|d^{\ell}_{\mathrm{rec}\,k}\| \geq \bar{w}_{j^{\star}_k} \|d^{j^{\star}_k \ell}_{\pi}\| - \sum_{j \neq j^{\star}_k} \bar{w}_j \|d^{j\ell}_{\pi}\| = 1 - 2\varepsilon_{\bar{w}} > 0. \tag{16}$$

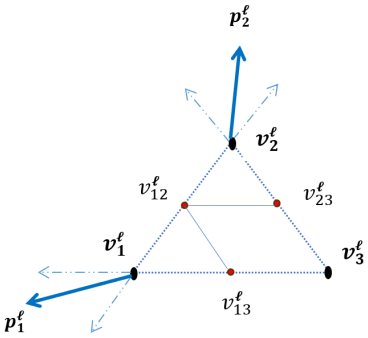


Fig. 1 Generating two components for the triangular recombination

Given (15), as shown in Fig. 1, let us consider the main triangle $\triangle(v_1^\ell, v_2^\ell, v_3^\ell)$ formed by the vertices $v_k^\ell := x_{\mathrm{rec}\,k}^\ell$, for $k \in [3]$. Near vertices v_1^ℓ and v_2^ℓ, we also define the sub-triangles $\triangle(v_1^\ell, v_{12}^\ell, v_{13}^\ell)$ and $\triangle(v_2^\ell, v_{21}^\ell, v_{23}^\ell)$, respectively, where the midpoints are

$$v_{12}^\ell = v_{21}^\ell := \frac{v_1^\ell + v_2^\ell}{2}, \qquad v_{13}^\ell := \frac{v_1^\ell + v_3^\ell}{2}, \qquad \text{and} \qquad v_{23}^\ell := \frac{v_2^\ell + v_3^\ell}{2}. \tag{17}$$

Letting $\Phi(\eta, a, b) := \eta a + (1-\eta)b$, we introduce a preliminary direction as

$$\tilde{p}_{\mathrm{trec}}^\ell = \frac{1}{w_p^\ell}\Phi(\eta, p_1^\ell, p_2^\ell) - x_{\mathrm{trec}}^\ell, \qquad \eta \in (0,1), \tag{18}$$

where

$$p_1^\ell := h^\ell v_{12}^\ell + b^\ell v_{13}^\ell, \qquad p_2^\ell := \bar{h}^\ell v_{21}^\ell + \bar{b}^\ell v_{23}^\ell, \qquad w_p^\ell := \Phi(\eta, h^\ell + b^\ell, \bar{h}^\ell + \bar{b}^\ell),$$

such that

$$h^\ell, \bar{h}^\ell \sim \mathcal{U}(0,1), \quad b^\ell := \sqrt{1-(h^\ell)^2}, \quad \bar{b}^\ell := \sqrt{1-(\bar{h}^\ell)^2}. \tag{19}$$

Rather than imposing a deterministic combination of vertices, $\tilde{p}_{\mathrm{trec}}^\ell$ samples a trial position inside $\triangle(v_1^\ell, v_2^\ell, v_3^\ell)$, biasing the search trajectory toward the superior vertices v_1^ℓ and v_2^ℓ. Subtracting x_{trec}^ℓ is mathematically necessary to convert the trial point into a centered search direction, eliminating positional bias. Algebraically, $\tilde{p}_{\mathrm{trec}}^\ell$ simplifies to a weighted combination of $v_1^\ell, v_2^\ell, v_3^\ell$, where the emerging coefficients correspond to:

$$\tilde{\theta}_1^\ell := \frac{1}{2}\Phi(\eta, h^\ell + b^\ell, \bar{h}^\ell), \qquad \tilde{\theta}_2^\ell := \frac{1}{2}\Phi(\eta, h^\ell, \bar{h}^\ell + \bar{b}^\ell), \qquad \tilde{\theta}_3^\ell := \frac{1}{2}\Phi(\eta, b^\ell, \bar{b}^\ell). \tag{20}$$

It is easy to verify that

$$\sum_{k=1}^{3} \tilde{\theta}_k^\ell = w_p^\ell, \qquad w_p^\ell > 0. \tag{21}$$

However, contributions from v_2^ℓ and v_3^ℓ may disrupt the trajectory of v_1^ℓ as the superior vertex–associated with the superior group ($\mathcal{G}_\pi^1$)–because raw weights (20) depend on local random geometry. To maintain the directional integrity of v_1^ℓ (aligned with $d_\pi^{1\ell}$), we define the **cross-group scaling factor**

$$s_\theta^\ell := \min\left\{1,\ \frac{\rho\bar{\varepsilon}\,\tilde{\theta}_1^\ell \alpha_{\mathrm{rec},1}^\ell}{2\sum_{k\neq 1}\tilde{\theta}_k^\ell \alpha_{\mathrm{rec},k}^\ell}\right\}, \qquad \rho\in(0,1) \tag{22}$$

and the **triangular weights** as

$$\theta_1^\ell := \tilde{\theta}_1^\ell, \qquad \theta_2^\ell := s_\theta^\ell \tilde{\theta}_2^\ell, \qquad \theta_3^\ell := s_\theta^\ell \tilde{\theta}_3^\ell. \tag{23}$$

Combining (22) and (23) directly yields $\sum_{k\neq 1}\theta_k^\ell \alpha_{\mathrm{rec},k}^\ell \leq \frac{1}{2}\rho\bar{\varepsilon}\,\theta_1^\ell \alpha_{\mathrm{rec},1}^\ell$, guaranteeing the intrinsic **cross-group margin property** of DAES:

$$\theta_1^\ell \alpha_{\mathrm{rec},1}^\ell \bar{\varepsilon}\rho - \sum_{k\neq 1}\theta_k^\ell \alpha_{\mathrm{rec}\,k}^\ell > 0. \tag{24}$$

Using (23), the **unscaled triangular recombination direction** is

$$\widehat{p}_{\mathrm{trec}}^\ell = \frac{\sum_{k=1}^3 \theta_k^\ell\, \alpha_{\mathrm{rec}\,k}^\ell\, d_{\mathrm{rec}\,k}^\ell}{\sum_{k=1}^3 \theta_k^\ell}. \tag{25}$$

Finally, normalizing $\widehat{p}_{\mathrm{trec}}^\ell$ yields the **triangular recombination direction** as

$$p_{\mathrm{trec}}^\ell = \delta\frac{\widehat{p}_{\mathrm{trec}}^\ell}{\|\widehat{p}_{\mathrm{trec}}^\ell\|}, \qquad \widehat{p}_{\mathrm{trec}}^\ell \neq 0, \qquad \text{for some constant} \quad \delta>0. \tag{26}$$

If, exceptionally, $\widehat{p}_{\mathrm{trec}}^\ell = 0$, setting $p_{\mathrm{trec}}^\ell := \delta d_\pi^{1\ell}$ preserves $\|p_{\mathrm{trec}}^\ell\| = \delta$ as $\|d_\pi^{1\ell}\| = 1$.

Algorithm 2 Triangular recombination direction in DAES

Input: σ^ℓ, s_σ^ℓ, x_{trec}^ℓ, $\tilde{f}_{\mathrm{trec}}^\ell$, the ranked matrix D_π^ℓ and groups $\mathcal{G}_\pi^k$, $k\in[3]$;
▶ generate the within-group weights $\bar{w}_j$ via (14);
▶ compute $d_{\mathrm{rec}\,k}^\ell$ via (12), for $k\in[3]$;
▶ obtain $\alpha_{\mathrm{rec}\,k}^\ell$ via (4) with $q_\alpha := d_{\mathrm{rec}\,k}^\ell$, for $k\in[3]$;
▶ compute $x_{\mathrm{rec}\,k}^\ell$ via (15), for $k\in[3]$;
▶ sample $h^\ell, \bar{h}^\ell \sim \mathcal{U}(0,1)$ and obtain b^ℓ and $\bar{b}^\ell$ via (19);
▶ generate θ_k^ℓ via (23), for $k\in[3]$;
▶ construct $\widehat{p}_{\mathrm{trec}}^\ell$ via (25);
▶ **if** $\widehat{p}_{\mathrm{trec}}^\ell \neq 0$ **then**, obtain p_{trec}^ℓ via (26); **else**, set $p_{\mathrm{trec}}^\ell := \delta d_\pi^{1\ell}$; **end if**
Output: p_{trec}^ℓ;

Algorithm 3 Triangular recombination point in `DAES`

Input: $\alpha^{\ell 0}$, p^{ℓ}_{trec}, x^{ℓ}_{trec} and $\tilde{f}^{\ell}_{\text{trec}}$;
▶ set $\mathsf{succ}_\ell := 0$;
for $s \in \{+1, -1\}$ **do** ▷ *Search along signed directions*
  ▶ evaluate $\tilde{f}_{\text{val}} := \tilde{f}(x^{\ell}_{\text{trec}} + s\alpha^{\ell 0} p^{\ell}_{\text{trec}})$;
  if $\tilde{\mu}_s(\alpha^{\ell 0}) > \tilde{\beta}$ **then** set $\alpha^{\ell}_{\text{trec}} := \alpha^{\ell 0}$, $\mathsf{succ}_\ell := 1$, and **break**; **end if**
end for
if succ_ℓ **then** ▷ *Extrapolation along selected direction*
  while 1 **do**
    ▶ set $\alpha_{\text{temp}} := \alpha^{\ell}_{\text{trec}}$ and $\tilde{f}_{\text{temp}} := \tilde{f}_{\text{val}}$;
    ▶ set $\alpha^{\ell}_{\text{trec}} := \gamma_e \alpha^{\ell}_{\text{trec}}$; evaluate $\tilde{f}_{\text{val}} := \tilde{f}(x^{\ell}_{\text{trec}} + s\alpha^{\ell}_{\text{trec}} p^{\ell}_{\text{trec}})$;
    if $\tilde{\mu}_s(\alpha^{\ell}_{\text{trec}}) \le \tilde{\beta}$ **then**
      ▶ set $\alpha^{\ell}_{\text{trec}} := \alpha_{\text{temp}}$ and $\tilde{f}_{\text{val}} := \tilde{f}_{\text{temp}}$; **break**;
    end if
  end while
  ▶ set $x^{\ell+1}_{\text{trec}} := x^{\ell}_{\text{trec}} + s\alpha^{\ell}_{\text{trec}} p^{\ell}_{\text{trec}}$ and $\tilde{f}^{\ell+1}_{\text{trec}} := \tilde{f}_{\text{val}}$;
else
  ▶ set $x^{\ell+1}_{\text{trec}} := x^{\ell}_{\text{trec}}$ and $\tilde{f}^{\ell+1}_{\text{trec}} := \tilde{f}^{\ell}_{\text{trec}}$;
end if
Output: $x^{\ell+1}_{\text{trec}}$ and succ_ℓ;

2.3.2 Triangular Recombination Point

Given p^{ℓ}_{trec}, the iterate x^{ℓ}_{trec} is robustly updated via the following steps:

1. ***Initial signed-search step-size.*** It is computed via (4) using $q_\alpha := p^{\ell}_{\text{trec}}$ as a candidate $\hat{\alpha}^{\ell 0}$, and safeguarded within $[L_\ell, a^\ell]$, i.e.,

$$\alpha^{\ell 0} := \min\left\{a^\ell, \max\left\{L_\ell, \hat{\alpha}^{\ell 0}\right\}\right\}, \quad \text{where } L_\ell := \max\{\alpha_{\min}, c_\alpha^{-1} a^\ell\}, \tag{27}$$

   and $c_\alpha > 1$ is a constant and a^ℓ is an initial auxiliary step-size scaling factor, satisfying

$$\max\left\{\alpha_{\min}, c_\alpha^{-1} a^\ell\right\} \le \alpha^{\ell 0} \le a^\ell \le \alpha_{\max}. \tag{28}$$

2. ***Signed-search.*** Given $\alpha := \alpha^{\ell 0}$, the candidate $x^{\ell}_{\text{trec}} + s\alpha p^{\ell}_{\text{trec}}$ is evaluated using the **noisy derivative-free sufficient descent condition** (`NDF-SDC`):

$$\tilde{\mu}_s(\alpha) > \tilde{\beta}, \qquad \text{for some} \quad \tilde{\beta} \in (0,1) \ \text{ and } \ s \in \{-1, 1\}, \tag{29}$$

   where

$$\tilde{\mu}_s(\alpha) := \frac{\tilde{f}(x^{\ell}_{\text{trec}} + s\alpha p^{\ell}_{\text{trec}}) - \tilde{f}(x^{\ell}_{\text{trec}})}{-\alpha^2}, \qquad \alpha > 0. \tag{30}$$

3. ***Extrapolation and scale refinements.*** The candidate (and thus iteration ℓ) is classified as successful if (29) holds for at least one $s \in \{+1, -1\}$; otherwise, it is deemed unsuccessful. Given a successful direction, an extrapolation expansion successively inflates the step-size $\alpha := \gamma_e \cdot \alpha$ until (29) fails, where $\gamma_e > 1$. Using the last successful step-size $\alpha^\ell_{\text{trec}} := \alpha$ as the ***triangular recombination step-size***, the iterate is updated to the ***triangular recombination point***

$$x^{\ell+1}_{\text{trec}} = x^\ell_{\text{trec}} \pm \alpha^\ell_{\text{trec}} p^\ell_{\text{trec}}, \qquad x^0_{\text{trec}} = x^0. \tag{31}$$

Given an unsuccessful direction, the iterate remains unchanged, while the step-size scaling factors are contracted by $\rho_u \in (0, 1)$ as:

$$a^{\ell+1} := \max\{\alpha_{\min}, \rho_u a^\ell\}, \qquad \sigma^{\ell+1} := \max\{\alpha_{\min}, \rho_u \sigma^\ell\}. \tag{32}$$

Note that $\sigma^\ell \in [\alpha_{\min}, \alpha_{\max}]$ scales all mutation and recombination step-sizes (including initial signed ones), while $a^\ell \in [\alpha_{\min}, \alpha_{\max}]$ is an auxiliary factor safeguarding only the initial signed step-size. In successful iterations, σ^ℓ updates via (5), whereas a^ℓ remains unchanged. The refinements (32) are crucial for convergence in unsuccessful iterations.

2.4 The DAES Algorithm

To complete the algorithmic specification of DAES, we detail the explicit calibration of the population size λ and outline the complete procedure.

Population size λ. Let $\ell_{\max}$ be the prescribed maximum outer iteration limit, distinct from the stationarity index $\ell_T \le \ell_{\max}$ used in the complexity analysis. Given an ***overall confidence parameter*** $\gamma_0 \in (0, 1/2)$, the ***per-population confidence budget*** γ_{pop} is defined, from which the population size λ is explicitly calibrated as:

$$\lambda = 6 \left\lceil \frac{\log(\gamma_{\text{pop}}^{-1})}{3 \log(1/(\tau\sqrt{2/\pi}))} \right\rceil, \qquad \text{where} \qquad \gamma_{\text{pop}} := \frac{\gamma_0}{\ell_{\max} N_{\text{sel}}}, \tag{33}$$

and $0 < \tau < \sqrt{\pi/2}$. The multiplier 6 allows the $\lambda/2$ symmetric direction pairs to be partitioned into three ranked groups. In the randomized analysis (Section 3.3), this logarithmic calibration ensures that the sampling-failure probability of each population remains bounded by γ_{pop}, securing the high-probability convergence guarantees.

Algorithmic Outline. Algorithm 4 integrates three subroutines into the complete pseudo-code for a single outer iteration. First, it generates some populations and executes the sorting–selection mechanism (Algorithm 1) to isolate the most noise-resilient population. Next, the triangular recombination procedure (Algorithm 2) synthesizes the search direction from that population. Finally, the signed-search and extrapolation scheme determines the step-size and updates the iterate (Algorithm 3).

Algorithm 4 DAES: the $(\ell+1)$-th iteration

Tuning parameters: $\ell_{\max}, N_{\text{sel}} \in \mathbb{N}$, $\gamma_0 \in (0, 1/2)$, $0 < \tau < \sqrt{\pi/2}$, $\tilde{\beta} \in (0,1)$, $\eta, \rho \in (0,1)$, $\delta > 0$, $\gamma_e > 1$, weight budget $\varepsilon_{\bar{w}}$ in (13), parameters required in (5), $\rho_u \in (0,1)$, $c_\alpha > 1$, $0 < \alpha_{\min} < \alpha_{\max} < \infty$, and $a^0 \in [\alpha_{\min}, \alpha_{\max}]$;
Input: σ^ℓ, s_σ^ℓ, a^ℓ, x_{trec}^ℓ, succ_ℓ, and $\tilde{f}_{\text{trec}}^\ell := \tilde{f}(x_{\text{trec}}^\ell)$;
set the population size λ according to (33);

▶ apply Algorithm 1 to have D_π^ℓ and $\mathcal{G}_\pi^k$, $k \in [3]$, associated with $q^\star$;
▶ apply Algorithm 2 to have p_{trec}^ℓ;
▶ compute $\hat{\alpha}^{\ell 0}$ via (4) with $q_\alpha := p_{\text{trec}}^\ell$ and obtain $\alpha^{\ell 0}$ via (27);
▶ apply Algorithm 3 to have succ_ℓ and $x_{\text{trec}}^{\ell+1}$;
▶ update $s_\sigma^{\ell+1}$ via (6);
if succ_ℓ **then**
    ▶ update $\sigma^{\ell+1}$ via (5), set $a^{\ell+1} := a^\ell$;
else
    ▶ reduce $\sigma^{\ell+1}$ and $a^{\ell+1}$ via (32);
end if
Output: $x_{\text{trec}}^{\ell+1}$, $\tilde{f}_{\text{trec}}^{\ell+1}$, $s_\sigma^{\ell+1}$, $\sigma^{\ell+1}$, and $a^{\ell+1}$;

3 Analytical Study of DAES

Adopting standard assumptions, this section establishes deterministic (Subsection 3.2) and high-probability (Subsection 3.3) complexity bounds for DAES. To our knowledge, this is the first complexity analysis for a noisy MAES-type method across nonconvex, convex, and strongly convex objectives, matching the rates in [7, 8, 17, 32].

3.1 Assumptions

Assumption 1 ($\mathbf{A_S}$). *Given the starting point $x^0 \in \mathbb{R}^n$, the following level set is compact:*

$$\mathcal{L}(x^0) := \{x \in \mathbb{R}^n \mid f(x) \le f(x^0)\}. \tag{34}$$

Assumption 2 ($\mathbf{A_L}$). *The objective function $f(x)$ is continuously differentiable on $\mathbb{R}^n$, and its gradient $g(x)$ is Lipschitz continuous; i.e., there exists a constant $L > 0$ such that*

$$\|g(x) - g(y)\| \le L\ \|x - y\|, \qquad \text{for all } x, y \in \mathbb{R}^n. \tag{35}$$

These assumptions ensure that f attains its minimum $\underline{f} := f(\underline{x}) > -\infty$ on $\mathcal{L}(x^0)$ and satisfies the Taylor-error bound

$$|f(x + \hat{s}) - f(x) - \hat{s}^T g(x)| \le \frac{L}{2}|\hat{s}|^2, \qquad \text{for all } x, \hat{s} \in \mathbb{R}^n. \tag{36}$$

Assumption 3 ($\mathbf{A_N}$). *For some noise level*

$$0 < \omega < \frac{\tilde{\beta}}{2}\alpha_{\min}^2, \qquad \tilde{\beta} \in (0,1), \tag{37}$$

the noisy function $\tilde{f}(x)$ *satisfies*

$$|\tilde{f}(x) - f(x)| \leq \omega, \qquad \text{for all } x \in \mathbb{R}^n. \tag{38}$$

Condition (37) is purely analytical and not computed by `DAES`. Using the aforementioned assumptions, for the theoretical analysis, we also define two problem-dependent constants

$$\beta := \tilde{\beta} - \frac{2\omega}{\alpha_{\min}^2} \in (0,1), \qquad \text{and} \qquad A_{\max} := \sqrt{\beta^{-1}\big(f(x^0) - \underline{f}\big)}. \tag{39}$$

Intuitively, β is the effective decrease factor on the true objective f, while $A_{\max}$ bounds the length of any successful step under this decrease.

Approximate Stationarity. Solving (1) generally yields at best an ε_ω-approximate stationary point for some limiting accuracy $\varepsilon_\omega > 0$. Accordingly, algorithm complexity is measured by the function evaluations required to find $x \in \mathcal{L}(x^0)$ satisfying

$$\|g(x)\| \leq \varepsilon_\omega, \qquad \text{and} \qquad f(x) \leq \sup\big\{f(y) \mid y \in \mathcal{L}(x^0),\ \|g(y)\| \leq \varepsilon_\omega\big\}. \tag{40}$$

While mathematically redundant, this second requirement provides a meaningful worst-case objective-value benchmark.

3.2 Complexity with Deterministic Directions

The deterministic results proceed as follows, laying the foundation for the randomized analysis in Subsection 3.3. Step-sizes satisfying `NDF-SDC` yield true sufficient decrease and finite extrapolation termination (Lemma 1, Proposition 1), while directional-derivative bounds (Proposition 2) yield local gradient bounds (Theorem 1). Crucially, Lemma 2 proves that successive unsuccessful iterations drive the search scale into the noise-balanced regime, establishing the worst-case bound (Corollary 1) and yielding a global gradient bound (Theorem 2). Next, Theorem 3 bounds optimality gaps and iterate errors, establishing sublinear (convex) and linear (strongly convex) accuracy. Finally, Proposition 3 bounds decrease efficiency, enabling Theorem 4 to establish evaluation complexity across nonconvex, convex, and strongly convex functions. All proofs are provided in Appendix A.

Remark 1 *Although fixed, the alignment threshold* Δ_{a} *and the Lipschitz constant* $L > 0$ *are explicitly retained in Theorems 2–4 to isolate the distinct contributions of directional geometry and objective-governed accuracy, while facilitating a direct comparison with other methods in Table 1.*

Let us define

$$v(x^\ell, p^\ell) := \begin{cases} \dfrac{\|g(x^\ell)\|\|p^\ell\|}{|g(x^\ell)^T p^\ell|}, & \text{if } |g(x^\ell)^T p^\ell| > 0, \\ +\infty, & \text{if } g(x^\ell) \neq 0 \text{ and } g(x^\ell)^T p^\ell = 0, \\ 1, & \text{if } g(x^\ell) = 0, \end{cases} \tag{41}$$

to quantify the geometric effect of the alignment (angle) condition

$$|g(x^\ell)^T p^\ell| \geq \Delta_{\mathrm{a}} \, \|g(x^\ell)\| \, \|p^\ell\|, \qquad \text{for some} \quad \Delta_{\mathrm{a}} \in (0, 1]. \tag{42}$$

Lemma 1 (**Consistency of** `NDF-SDC`) *Under Assumption* $\mathbf{A_N}$, *let* $p^\ell \in \mathbb{R}^n$ *be a deterministic search direction in* `DAES` *and let* $\alpha \in [\alpha_{\min}, \infty)$ *be a corresponding step-size at iteration* ℓ *which satisfies* `NDF-SDC` (29) *for some* $s \in \{-1, 1\}$. *Then, for the constant* $\beta \in (0, 1)$ *defined in* (39), *the corresponding noiseless sufficient decrease condition is guaranteed to hold, i.e.,*

$$\mu_s(\alpha) > \beta, \qquad \textit{where} \quad \mu_s(\alpha) := \frac{f(x^\ell + s\alpha p^\ell) - f(x^\ell)}{-\alpha^2}. \tag{43}$$

Proof: *See Appendix A.1.*

Proposition 1 (**Termination of extrapolation**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *let* $\alpha_{\min}$ *be the lower step-size limit, while* $A_{\max}$ *is the bound defined in* (39). *If* $A_{\max} > \alpha_{\min}$, *then the total number of expansion trials in* `DAES` *(including the final unsuccessful one) is at most*

$$N_e := \left\lceil \log_{\gamma_e} (A_{\max}/\alpha_{\min}) \right\rceil, \qquad \gamma_e > 1. \tag{44}$$

If $A_{\max} \leq \alpha_{\min}$, *no successful outer iteration initiates extrapolation trials; i.e.,* $N_e = 0$.

Proof: *See Appendix A.2.*

Proposition 2 (**Directional derivative bound**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *and the analytical compatibility condition* (39), *let* $p^\ell \in \mathbb{R}^n$ *be a deterministic search direction in* `DAES` *with* $\delta = \|p^\ell\|$. *Then, for both unsuccessful and successful iterations,*

$$|g(x^\ell)^T p^\ell| \leq \gamma_e \alpha \left(\tilde{\beta} + \frac{L}{2}\delta^2 \right) + \frac{2\omega}{\alpha}, \qquad \textit{for some constants} \quad \tilde{\beta} \in (0, 1), \ \gamma_e > 1, \tag{45}$$

where $\alpha := \alpha^{\ell 0}$ *for unsuccessful iterations and* $\alpha := \alpha^\ell_{\mathrm{trec}}$ *for successful iterations.*

Proof: *See Appendix A.3.*

Theorem 1 (**Local bound on** $\|\mathbf{g}(\mathbf{x}^\ell)\|$) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *suppose that the conditions of Proposition 2 hold and* $v(x^\ell, p^\ell)$ *is defined as in* (41). *Then*

$$\|g(x^\ell)\| \leq v(x^\ell, p^\ell) \, \Gamma(\alpha), \qquad \textit{where} \qquad \Gamma(\alpha) := \frac{\alpha}{\delta}\left(\gamma_e \left(\tilde{\beta} + \frac{L}{2}\delta^2 \right) + \frac{2\omega}{\alpha^2} \right), \tag{46}$$

Proof: *See Appendix A.4.*

Lemma 2 (**Noise-balanced scale crossing**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *suppose that the conditions of Proposition 2 hold. Define the noise-balanced scale* a_ω *by*

$$a_\omega := \kappa_\star \sqrt{\omega/L}, \qquad \text{where } \kappa_\star := \sqrt{2c_\alpha/(\tilde{\beta}/L + \delta^2/2)}, \tag{47}$$

with $c_\alpha > 1$ *and initial auxiliary factor* a^0 *satisfying* (28) *and*

$$a^0 \geq a_\omega \geq \alpha_{\min}. \tag{48}$$

If DAES *does not terminate, it eventually reaches the first unsuccessful iteration* ℓ_ω *at scale* a^{ℓ_ω} *such that*

$$\rho_u a_\omega < a^{\ell_\omega} \leq a_\omega, \qquad \rho_u \in (0,1). \tag{49}$$

Proof: *See Appendix A.5.*

Corollary 1 (**Worst-case stationarity bound**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *let the conditions of Theorem 1 hold for the unsuccessful iteration* ℓ_ω *in Lemma 2, and*

$$C_d^0 := \delta^{-1}\left[\gamma_e \overline{\kappa}'\left(\tilde{\beta}/L + \delta^2/2\right) + 2/\underline{\kappa}'\right], \qquad \underline{\kappa}' := \rho_u \kappa_\star / c_\alpha, \quad \overline{\kappa}' := \kappa_\star. \tag{50}$$

If, in addition, the angle condition (42) *holds, then*

$$\|g(x^{\ell_\omega})\| \leq \Delta_{\mathrm{a}}^{-1} C_d^0 \sqrt{L\omega}. \tag{51}$$

Proof: *See Appendix A.6.*

Theorem 2 (**Global bound on** $\|\mathbf{g}(\mathbf{x}^\ell)\|$) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *let the worst-case iteration* ℓ_ω *be defined as in Lemma 2, and the conditions of Proposition 2 hold. If the angle condition* (42) *holds, then*

$$\min_{\ell \in \mathcal{K}_\omega} \|g(x^\ell)\| = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big), \qquad \mathcal{K}_\omega := \{\ell \in \mathbb{N}_0 : 0 \leq \ell \leq \ell_\omega < \ell_{\max}\}, \tag{52}$$

where $\ell_{\max}$ *denotes the prescribed maximum number of iterations.*

Proof: *See Appendix A.7.*

Theorem 3 (**Error Bounds**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *let the conditions of Theorem 2 hold,* $\underline{f} := f(\underline{x})$ *at the minimizer* $\underline{x}$ *and* $\ell^{**} \in \operatorname{argmin}_{\ell \in \mathcal{K}_\omega} \|g(x^\ell)\|$. *If the angle condition* (42) *holds, then*

(i) if f *is convex*

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big); \tag{53}$$

(ii) if f *is* μ_c*-strongly convex,*

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-2} L\omega\big), \qquad \text{and} \qquad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big). \tag{54}$$

Proof: *See Appendix A.8.*

Proposition 3 (**Efficiency**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, *and* $\mathbf{A_N}$, *suppose that the conditions of Proposition 2 and the angle condition (42) hold at a successful iteration. Then, for the constant* $\beta \in (0,1)$ *defined in (39), it holds:*

$$f(x) - f(x + s\alpha p) \geq \frac{\beta}{4K_1^2}\|g(x^\ell)\|^2, \qquad \text{where} \qquad K_1 := \frac{\gamma_e\left(\tilde{\beta} + L\delta^2/2\right)}{\Delta_a \delta}. \tag{55}$$

Proof: *See Appendix A.9.*

Defining the noise-limited accuracy as ε_ω and C_d^0 as in (50), we specify the ***termination iteration*** ℓ_T as the first iteration satisfying

$$\|g(x^\ell)\| \leq \varepsilon_\omega, \qquad \varepsilon_\omega := C_d\sqrt{L\omega} \geq \Delta_a^{-1} C_d^0 \sqrt{L\omega}, \tag{56}$$

where $C_d > 0$ is fixed independently of ω. This formally justifies the noise-scale unsuccessful iteration ℓ_ω in Lemma 2 as the worst-case termination bound ($\ell_T \leq \ell_\omega$).

Theorem 4 (**Complexity Bounds**) *Under Assumptions* $\mathbf{A_S}$, $\mathbf{A_L}$, $\mathbf{A_N}$, *and the condition (48), let the hypotheses of Proposition 2 and the angle condition (42) hold before termination with the noise-limited accuracy* ε_ω *defined in (56). Then, given* N_{sel} *selection populations and maximum* N_e *extrapolation steps, the number of function evaluations is bounded by*

$$N_f \leq 1 + (\lambda N_{\text{sel}} + 2 + N_e)\,\ell_T, \qquad \text{and} \tag{57}$$

(i) $\ell_T = \mathcal{O}\left(L^2\varepsilon_\omega^{-2}\right)$ *and* $N_f = \mathcal{O}\left(\varepsilon_\omega^{-2}\right)$ *for nonconvex functions;*

(ii) $\ell_T = \mathcal{O}\left(\varepsilon_\omega^{-1}\right)$ *and* $N_f = \mathcal{O}\left(\varepsilon_\omega^{-1}\right)$ *for convex functions;*

(iii) $\ell_T = \mathcal{O}\left(\log(\varepsilon_\omega^{-1})\right)$ *and* $N_f = \mathcal{O}\left(\log(\varepsilon_\omega^{-1})\right)$ *for* μ_c*-strongly convex functions.*

Proof: *See Appendix A.10.*

3.3 Complexity with Randomized Directions

We extend the analysis in Subsection 3.2 to the randomized setting, where the search direction is a triangular recombination direction. To establish high-probability convergence, we first construct the formal probabilistic framework. Next, we prove that Gaussian sampling yields a probabilistic angle property (Proposition 4). Following that, we establish the directional alignment of p_{trec}^ℓ via a three-stage hierarchy, leveraging two weighting mechanisms described in Subsection 2.3.1. Finally, we ensure that `DAES` drives the gradient norm down to the noise-accuracy threshold with high probability (Theorem 5), provide objective and iterate accuracy guarantees across all convexity regimes (Theorem 6) and obtain the overall high-probability evaluation complexity for `DAES` (Theorem 7).

3.3.1 Probabilistic Sampling and Ranking Stability

At each iteration ℓ, DAES generates N_{sel} independent Gaussian populations indexed by $q \in [N_{\text{sel}}]$. In the sorting–selection phase, it retains the one maximizing the observable within-group separation $\Delta_{\text{sel}}^{\ell,q}$ among those populations. For theoretical purposes, we define the ***analytical certificate-success event*** $\mathcal{C}_{\ell,q}$ and its complementary ***analytical certificate-failure event*** $E_{\text{sel},\ell,q}$ as

$$\mathcal{C}_{\ell,q} := \left\{ \Delta_{\text{sel}}^{\ell,q} > 2\omega \right\}, \qquad \text{and} \qquad E_{\text{sel},\ell,q} := \mathcal{C}_{\ell,q}^{c} = \left\{ \Delta_{\text{sel}}^{\ell,q} \leq 2\omega \right\}. \tag{58}$$

While $\mathcal{C}_{\ell,q}$ is an event with respect to Gaussian sampling, it yields a deterministic ranking-stability certificate under bounded noise ω as

$$f_{\pi}^{r\ell} - f_{\pi}^{j_k^\star \ell} \geq \tilde{f}_{\pi}^{r\ell} - \tilde{f}_{\pi}^{j_k^\star \ell} - 2\omega > 0, \tag{59}$$

for every $r \in \mathcal{G}_{\pi}^{k} \setminus \{j_k^\star\}$, where $j_k^\star \in \mathcal{G}_{\pi}^{k}$ for $k \in [3]$. This guarantees that the noisy best-ranked element of each group is strictly superior in terms of the true objective f.

By (11), $\Delta_{\text{sel}}^{\ell,q^\star} \geq \Delta_{\text{sel}}^{\ell,q}$ for all $q \in [N_{\text{sel}}]$. If at least one of the N_{sel} populations satisfies $\mathcal{C}_{\ell,q}$, then the selected population $q^\star$ also satisfies $\mathcal{C}_{\ell,q^\star}$. Accordingly, the ***selected certificate-failure event*** $E_{\text{sel},\ell}$ occurs when all populations fail on $\mathcal{C}_{\ell,q}$, i.e.,

$$E_{\text{sel},\ell} := \left\{ \Delta_{\text{sel}}^{\ell,q^\star} \leq 2\omega \right\} = \left(\bigcup_{q=1}^{N_{\text{sel}}} \mathcal{C}_{\ell,q} \right)^{c} = \bigcap_{q=1}^{N_{\text{sel}}} E_{\text{sel},\ell,q}. \tag{60}$$

Serving as an analytical tool, $E_{\text{sel},\ell}$ merely indicates that (59) is unavailable, not an incorrect noisy ranking. It is neither evaluated by DAES nor assigned a probability bound. Crucially, the probabilistic guarantees stem strictly from Gaussian sampling, while Assumption $\mathbf{A_N}$ serves solely as an analytical tool under a bounded deterministic noise level. Hence, conditional on sampling a well-aligned direction, all subsequent steps (sorting–selection and weighting) proceed deterministically.

For any candidate population indexed by $q \in [N_{\text{sel}}]$ at iteration ℓ, let $d^{j\ell}$ with $\|d^{j\ell}\| = 1$ denote the generated direction corresponding to the j-th column of $D_{\pm}^{\ell}$. For $g(x^\ell) \neq 0$, the ***sampling-failure event*** for the q-th population is defined as

$$E_{\text{samp},\ell,q} := \left\{ \max_{j \in [\lambda/2]} \frac{|g(x^\ell)^T d^{j\ell}|}{\|g(x^\ell)\|} < t_a \right\}, \tag{61}$$

where $t_a > 0$ is the baseline alignment threshold. This event represents the realization in which no sampled direction in the q-th population achieves the required alignment with the true gradient. If $g(x^\ell) = 0$, we set $E_{\text{samp},\ell,q} := \emptyset$.

Conversely, $E^c_{\text{samp},\ell,q}$ defines the ***sampling-success event***, on which there exists at least one index $j \in [\lambda/2]$ in population q satisfying

$$|g(x^\ell)^T d^{j\ell}| \geq t_a \|g(x^\ell)\|. \tag{62}$$

To guarantee that sampling success holds uniformly across the entire iteration horizon (covering all $\ell_{\max} \times N_{\text{sel}}$ sampled populations), the ***global good event*** E_G as the joint occurrence of sampling success is defined as

$$E_G := \bigcap_{\ell=0}^{\ell_{\max}-1} \bigcap_{q=1}^{N_{\text{sel}}} E^c_{\text{samp},\ell,q}. \tag{63}$$

To rigorously analyze the probability of E_G under sequential population generation, let $\mathcal{F}_{\ell,q-1}$ denote the σ-algebra generated by the algorithmic history prior to generating the q-th population at iteration ℓ, with $\mathcal{F}_{\ell,0} := \mathcal{F}_{\ell-1}$, where the initial σ-algebra is generated by the initial point x^0. It ensures that x^ℓ and $g(x^\ell)$ are $\mathcal{F}_{\ell,q-1}$-measurable, while the fresh Gaussian samples are independent of $\mathcal{F}_{\ell,q-1}$. It provides the filtration necessary to apply the tower property of conditional expectation when evaluating the joint failure probability ($\mathbb{P}(E_G^c)$) in the subsequent proofs.

3.3.2 Probabilistic Angle Conditions

In our randomized analysis, we first bound the conditional failure probability of the Gaussian sampling mechanism for any individual population. As a consequence, we establish that on the global good event E_G, every candidate population generated across the entire iteration horizon contains at least one direction sufficiently aligned with the negative true gradient with high probability.

Proposition 4 *Let the symmetric normalized directions $\{d^{j\ell}\}_{j\in[\lambda]}$ be generated by* `DAES` *for $0 \leq \ell < \ell_{\max}$, where the population size λ is calibrated as in (33). Then, for each of the N_{sel} generated populations with confidence budget γ_{pop}, the conditional sampling-failure probability is at most γ_{pop} given the σ-algebra $\mathcal{F}_{\ell,q-1}$; i.e.,*

$$\mathbb{P}\big(E_{\text{samp},\ell,q} \,\big|\, \mathcal{F}_{\ell,q-1}\big) \leq \gamma_{\text{pop}}, \qquad \gamma_{\text{pop}} := \frac{\gamma_0}{\ell_{\max} N_{\text{sel}}}, \tag{64}$$

where $\gamma_0 \in (0, 1/2)$ is given as the overall confidence parameter and used for selecting λ. Consequently, on the global good event E_G, there exists a sampled direction for which the angle alignment property

$$\frac{|g(x^\ell)^T d^{j\ell}|}{\|g(x^\ell)\|} \geq t_a, \quad j \in [\lambda] \tag{65}$$

holds with probability at least $1 - \gamma_0$, where the baseline alignment threshold $t_a := \tau/\sqrt{n}$ is defined as in (13) with $n \geq 2$ and $0 < \tau < \sqrt{\pi/2}$.

Proof Let $\mathcal{F}_{\ell,q-1}$ be the sigma-algebra generated by the algorithmic history prior to generating the q-th population at iteration ℓ. Let us fix $0 \le \ell < \ell_{\max}$ and $q \in [N_{\mathrm{sel}}]$ and denote the n-dimensional unit sphere by $\mathbb{S}^{n-1}$. By the construction, x^ℓ and $g(x^\ell)$ are $\mathcal{F}_{\ell,q-1}$-measurable, whereas the fresh Gaussian vectors $\{z^{j\ell}\}_{j\in[\lambda/2]}$ are independent of $\mathcal{F}_{\ell,q-1}$. They are conditionally i.i.d. and uniformly distributed on $\mathbb{S}^{n-1}$. Let $u^\ell := g(x^\ell)/\|g(x^\ell)\|$. The following finite-dimensional spherical-cap estimate follows the arguments in [17], letting $T := (u^\ell)^T \widehat{z}^{j\ell}$.

First, suppose that $n \ge 3$. Then T has density

$$f_n(t) = c_n(1-t^2)^{(n-3)/2}, \qquad c_n := \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)}, \qquad -1 \le t \le 1.$$

Since $(n-3)/2 \ge 0$, the density f_n is nonincreasing on $[0,1]$. Therefore, with $t_a = \tau/\sqrt{n}$,

$$\mathbb{P}\big(0 \le T < t_a \,\big|\, \mathcal{F}_{\ell,q-1}\big) = \int_0^{\tau/\sqrt{n}} c_n(1-t^2)^{(n-3)/2}\,dt \le c_n \frac{\tau}{\sqrt{n}}.$$

By the log-convexity of the Gamma function,

$$\Gamma\left(x+\frac{1}{2}\right)^2 \le \Gamma(x)\Gamma(x+1) = x\Gamma(x)^2, \qquad x > 0.$$

Taking $x = (n-1)/2$ gives

$$\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \le \sqrt{\frac{n-1}{2}}, \quad \text{and hence} \quad \frac{c_n}{\sqrt{n}} \le \sqrt{\frac{n-1}{2n\pi}} < \frac{1}{\sqrt{2\pi}}.$$

Using the symmetry of T, we obtain

$$\mathbb{P}\big(T \ge t_a \,\big|\, \mathcal{F}_{\ell,q-1}\big) = \frac{1}{2} - \mathbb{P}\big(0 \le T < t_a \,\big|\, \mathcal{F}_{\ell,q-1}\big) \ge \frac{1}{2} - \frac{\tau}{\sqrt{2\pi}}.$$

Now, suppose that $n = 2$. In this case, T has density

$$f_2(t) = \frac{1}{\pi\sqrt{1-t^2}}, \qquad -1 < t < 1.$$

Since $0 < \tau < \sqrt{\pi/2}$, we have $0 < \tau/\sqrt{2} < 1$, and therefore

$$\mathbb{P}\left(T \ge \frac{\tau}{\sqrt{2}} \,\middle|\, \mathcal{F}_{\ell,q-1}\right) = \frac{1}{2} - \frac{1}{\pi}\arcsin\left(\frac{\tau}{\sqrt{2}}\right).$$

Since $\arcsin(t)$ is convex on $[0,1]$ and joins the points $(0,0)$ and $(1,\pi/2)$,

$$\arcsin(t) \le \frac{\pi}{2}t, \qquad 0 \le t \le 1, \quad \text{so that} \quad \frac{1}{\pi}\arcsin\left(\frac{\tau}{\sqrt{2}}\right) \le \frac{\tau}{2\sqrt{2}} \le \frac{\tau}{\sqrt{2\pi}},$$

where the last inequality follows from $\pi \le 4$. Since $t_a = \tau/\sqrt{n} \le \tau/\sqrt{2}$ for all $n \ge 2$,

$$\mathbb{P}\big(T \ge t_a \,\big|\, \mathcal{F}_{\ell,q-1}\big) \ge \frac{1}{2} - \frac{\tau}{\sqrt{2\pi}}.$$

Using symmetry, complementation and the conditional independence of $\{\widehat{z}^{j\ell}\}_{j\in[\lambda/2]}$, we have

$$\mathbb{P}\left(E_{\mathrm{samp},\ell,q} \,\middle|\, \mathcal{F}_{\ell,q-1}\right) \le r_\tau^{\lambda/2}, \qquad r_\tau := \tau\sqrt{2/\pi} \in (0,1).$$

From (33), we have $\lambda \log(1/r_\tau)/2 \geq \log(1/\gamma_{\text{pop}})$, thereby proving (64). On $E^c_{\text{samp},\ell,q}$, by the antipodal construction of $D^\ell_\pm$ in (2), there exists at least one index $j \in [\lambda]$ for which (65) holds with conditional probability at least $1-\gamma_{\text{pop}}$ given $\mathcal{F}_{\ell,q-1}$. To prove that the probability guarantee holds simultaneously for all nonstationary iterations, consider the global good event defined in (63). By the union bound and tower property of conditional expectation,

$$\mathbb{P}(E_G^c) \leq \sum_{\ell=0}^{\ell_{\max}-1} \sum_{q=1}^{N_{\text{sel}}} \mathbb{E}\left[\mathbb{P}\left(E_{\text{samp},\ell,q} \,\middle|\, \mathcal{F}_{\ell,q-1}\right)\right] \leq \ell_{\max} N_{\text{sel}} \gamma_{\text{pop}} = \gamma_0,$$

which establishes that $\mathbb{P}(E_G) \geq 1-\gamma_0$ and completes the proof. □

On $E^c_{\text{samp},\ell,q}$ for any fixed candidate population, the antipodal construction of $D^\ell_\pm$ ensures that some direction satisfies $g(x^\ell)^T d^{j\ell} \leq -t_a \|g(x^\ell)\|$. For the retained population selected as in (11), if it achieves sampling success, its ***latent descent direction*** $d^{j^\star \ell}$ satisfies

$$g(x^\ell)^T d^{j^\star \ell} \leq -t_a \|g(x^\ell)\|, \qquad \text{where} \quad j^\star \in \arg\min_{j\in[\lambda]} g(x^\ell)^T d^{j\ell}. \tag{66}$$

This is an ***analytical descent-angle inequality***, distinct from the algorithmic test `NDF-SDC` in (29).

We establish the directional alignment of p^ℓ_{trec} via a three-stage hierarchy, leveraging two weighting mechanisms designed according to the algorithmic conditions (13) and (24): first, bounded-noise ordering transfers a sufficient descent angle from the latent (unobservable) direction $d^{j^\star \ell}$ to $d^{1\ell}_\pi$ (Proposition 5); second, the within-group dominance parameter $\varepsilon_{\bar{w}}$ (13) preserves alignment inside $\mathcal{G}^1_\pi$, transferring it to $d^\ell_{\text{rec}\,1}$ (Lemma 3); third, the cross-group margin property (24) prevents cancellation from secondary groups, securing the overall alignment of p^ℓ_{trec} (Proposition 6).

Proposition 5 *Suppose that Assumptions* $\mathbf{A_L}$ *and* $\mathbf{A_N}$ *hold, and define*

$$G_{\text{sel}} := \frac{2\alpha_{\max}}{\alpha_{\min} t_a} B_{\text{sel}}, \qquad B_{\text{sel}} := \frac{L}{2}\left(\alpha_{\max} + \frac{\alpha^2_{\max}}{\alpha_{\min}}\right) + \frac{2\omega}{\alpha_{\min}}, \tag{67}$$

where t_a is given as in (65), and consider the retained population selected as in (11). If this population achieves sampling success, then for any iteration $0 \leq \ell < \ell_{\max}$ satisfying the nonstationarity condition

$$\|g(x^\ell)\| > G_{\text{sel}}, \tag{68}$$

the best noisy-ranked direction $d^{1\ell}_\pi$ achieves the alignment guarantee

$$g(x^\ell)^T d^{1\ell}_\pi < -t_{\text{sel}} \|g(x^\ell)\|, \tag{69}$$

with the particular alignment threshold $t_{\text{sel}} > 0$, which directly yields

$$\frac{|g(x^\ell)^T d^{1\ell}_\pi|}{\|g(x^\ell)\|} \geq t_{\text{sel}}, \qquad \text{where} \quad t_{\text{sel}} := \frac{\alpha_{\min}}{2\alpha_{\max}} t_a. \tag{70}$$

Consequently, on the global good event E_G, which occurs with probability at least $1-\gamma_0$, the angle alignment property (70) holds simultaneously for all iterations satisfying (68).

Proof Let $d^{j^\star \ell}$ be the (unobservable) direction in the retained population that achieves the theoretical alignment guarantee (66), and let $i^\star$ denote its ranked position. However, `DAES` identifies the search direction $d_\pi^{1\ell}$ via noisy function evaluations $\tilde{f}$. Let $\tilde{f}_\pi^{1\ell}$ be the observed value at the globally best noisy-ranked direction $d_\pi^{1\ell}$. Thus, $\tilde{f}(x^\ell + \alpha_1 d_\pi^{1\ell}) \leq \tilde{f}(x^\ell + \alpha_\star d_\pi^{i^\star \ell})$, where α_1 and $\alpha_\star$ are the mutation step-sizes along $d_\pi^{1\ell}$ and $d_\pi^{i^\star \ell}$, respectively. By $\mathbf{A_N}$, $f(x^\ell + \alpha_1 d_\pi^{1\ell}) \leq f(x^\ell + \alpha_\star d_\pi^{i^\star \ell}) + 2\omega$. Using (36) with $\hat{s} = d_\pi^{1\ell}$ and $\hat{s} = d_\pi^{i^\star \ell}$, respectively,

$$f\Big(x^\ell + \alpha_1 d_\pi^{1\ell}\Big) \geq f(x^\ell) + \alpha_1 g(x^\ell)^T d_\pi^{1\ell} - \frac{L}{2}\alpha_1^2, \qquad \text{and}$$

$$f\Big(x^\ell + \alpha_\star d_\pi^{i^\star \ell}\Big) \leq f(x^\ell) + \alpha_\star g(x^\ell)^T d_\pi^{i^\star \ell} + \frac{L}{2}\alpha_\star^2.$$

Combining these last three inequalities, and using (66), $\|d_\pi^{i^\star \ell}\| = 1$ and $\alpha_1, \alpha_\star \in [\alpha_{\min}, \alpha_{\max}]$, yields $g(x^\ell)^T d_\pi^{1\ell} \leq -2t_{\text{sel}}\|g(x^\ell)\| + B_{\text{sel}}$ followed by (69) because $B_{\text{sel}} < t_{\text{sel}}\|g(x^\ell)\|$ if $\|g(x^\ell)\| > G_{\text{sel}}$. Hence (70) follows immediately. On E_G, the sampling-success property holds for every candidate population and therefore for the selected population $q^\star$. The deterministic argument above then establishes (70) for every iteration satisfying (68). On E_G, consequently, by the union bound and tower property, (70) holds with overall probability at least $1-\gamma_0$. □

As formalized in Lemma 3 below, each group-representative recombination direction retains a positive alignment margin with the normalized gradient whenever its best direction is well-aligned. Since, by construction of `DAES`, the best directions of ranked groups are associated with $j_1^\star = 1$, $j_2^\star = \mu+1$, and $j_3^\star = 2\mu+1$, this property applies in particular to the representative recombination direction of the first group, i.e., $d_{\text{rec}\,1}$. Because its associated best direction $d_\pi^{1\ell}$ is well-aligned from Proposition 5, it provides the condition required for the final result in Proposition 6.

Lemma 3 *For each group $\mathcal{G}_\pi^k$, let $\{d_\pi^{j\ell}\}_{j\in\mathcal{G}_\pi^k}$ be the set of sorted unit directions from the retained population selected as in (11), and let $\{\bar{w}_j\}_{j\in\mathcal{G}_\pi^k}$ be their corresponding within-group recombination weights generated according to (14), given a within-group dominance parameter $\varepsilon_{\bar{w}}$ satisfying (13). If the best direction of each group achieves the alignment guarantee*

$$|g(x^\ell)^T d_\pi^{j_k^\star \ell}| \geq t_{\text{sel}}\|g(x^\ell)\|, \qquad k \in [3],$$

with the particular alignment threshold $t_{\text{sel}} > 0$, then each group-representative recombination direction $d_{\text{rec}\,k}^\ell$ with the within-group alignment margin $\bar{\varepsilon} > 0$ satisfies

$$\frac{|g(x^\ell)^T d_{\text{rec}\,k}^\ell|}{\|g(x^\ell)\|} \geq \bar{\varepsilon}, \quad k \in [3], \qquad \textit{where} \quad \bar{\varepsilon} := (1-\varepsilon_{\bar{w}})t_{\text{sel}} - \varepsilon_{\bar{w}}. \tag{71}$$

Proof Let $q^\ell := g(x^\ell)/\|g(x^\ell)\|$ and apply the reverse triangle inequality to $(q^\ell)^T d_{\text{rec}\,k}^\ell$. Using (12) and the fact that $\|q^\ell\|\,\|d_\pi^{j\ell}\| = 1$, we have $|(q^\ell)^T v_k^\ell| \geq (1-\varepsilon_{\bar{w}})|(q^\ell)^T d_\pi^{j_k^\star \ell}| - \varepsilon_{\bar{w}}$ where $\sum_j \bar{w}_j \leq \varepsilon_{\bar{w}}$ with $j \in \mathcal{G}_\pi^k \setminus \{j_k^\star\}$. In particular, if $|(q^\ell)^T d_\pi^{j_k^\star \ell}| \geq t_{\text{sel}}$, applying the definition of $\bar{\varepsilon}$ in (71) completes the proof. □

Proposition 6 *Under the Assumptions* $\mathbf{A_L}$ *and* $\mathbf{A_N}$*, consider the retained population selected as in* (11)*. Let the normalized within-group weights and triangular weights be generated according to* (14) *and* (23)*, respectively, with the cross-group margin property* (24)*. Then, on the global good event* E_G*, which occurs with probability at least* $1-\gamma_0$ *given the overall confidence budget* γ_0 *used to define the population-size* λ *in* (33)*, the triangular recombination direction* p^ℓ_{trec} *achieves the alignment guarantee*

$$\frac{|g(x^\ell)^T p^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\|\,\|p^\ell_{\mathrm{trec}}\|} \geq \Delta_r, \qquad \textit{where} \quad \Delta_r := \frac{(1-\rho)\bar{\varepsilon}}{1+\rho\bar{\varepsilon}} > 0, \tag{72}$$

with $\rho \in (0,1)$ *and* $\bar{\varepsilon}$ *defined in* (71)*, simultaneously for all iterations satisfying* (68)*. Consequently, the probability statement holds for the reciprocal alignment ratio*

$$v(x^\ell, p^\ell_{\mathrm{trec}}) \leq \Delta_r^{-1}. \tag{73}$$

Proof Given γ_0 and the retained population, the global good event E_G with $\mathbb{P}(E_G) \geq 1-\gamma_0$ guarantees sampling success across all nonstationary iterations; let $q^\ell := g(x^\ell)/\|g(x^\ell)\|$. On E_G, applying Proposition 5, $|(q^\ell)^T d^{1\ell}_\pi| \geq t_{\mathrm{sel}}$ and applying Lemma 3, $|(q^\ell)^T d^\ell_{\mathrm{rec}\,1}| \geq \bar{\varepsilon}$. Now, recall the unscaled triangular recombination direction (25). Applying the reverse triangle inequality to $(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}$, then using $|(q^\ell)^T d^\ell_{\mathrm{rec}\,k}| \leq \|q^\ell\|\|d^\ell_{\mathrm{rec}\,k}\| \leq 1$ and $\|d^\ell_{\mathrm{rec}\,k}\| \leq 1$ for all $k \in [3]$, and taking the algorithmic cross-group margin property (24) yields

$$|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}| \geq \frac{\left(\bar{\varepsilon}\theta^\ell_1 \alpha^\ell_{\mathrm{rec}\,1} - \sum_{k\neq 1} \theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}\right)}{\Theta^\ell} > \frac{(1-\rho)\bar{\varepsilon}\theta^\ell_1 \alpha^\ell_{\mathrm{rec}\,1}}{\Theta^\ell} > 0, \qquad \Theta^\ell := \sum_{k=1}^{3} \theta^\ell_k. \tag{74}$$

This strict positivity implies $\widehat{p}^\ell_{\mathrm{trec}} \neq 0$. Combining $\|\widehat{p}^\ell_{\mathrm{trec}}\| \leq \frac{1}{\Theta^\ell}\sum_{k=1}^3 \theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}$ with (74),

$$\frac{|g(x^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\|\|\widehat{p}^\ell_{\mathrm{trec}}\|} \geq \frac{(1-\rho)\bar{\varepsilon}\theta^\ell_1 \alpha^\ell_{\mathrm{rec}\,1}}{\sum_{k=1}^3 \theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}} \geq \frac{(1-\rho)\bar{\varepsilon}}{1+r^\ell}, \qquad \text{where} \quad r^\ell := \frac{\sum_{k\neq 1} \theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}}{\theta^\ell_1 \alpha^\ell_{\mathrm{rec}\,1}}.$$

Since $r^\ell < \rho\bar{\varepsilon}$ by (24), the alignment ratio is bounded below by Δ_r. Since directional alignment is invariant under unit-normalization in p^ℓ_{trec}, (72) follows immediately. To complete the proof, note that the strict positivity (74) ensures $|(q^\ell)^T p^\ell_{\mathrm{trec}}| > 0$. Using (41) and taking the reciprocal of the angular ratio in (72) gives (73), and the probability statement follows. □

Lemma 4 *Under Assumptions* $\mathbf{A_L}$ *and* $\mathbf{A_N}$*, suppose that the condition in Proposition 5 holds and let the noise-scale feasibility condition* (48) *hold. For all iterations satisfying* (68)*, the stationarity transfer threshold is tightly controlled by*

$$G_{\mathrm{sel}} \leq C_{\mathrm{sel}}\sqrt{L\omega}, \qquad \textit{where} \quad C_{\mathrm{sel}} := 2R_\alpha t_a^{-1} C^0_{\mathrm{sel}}, \tag{75}$$

with

$$C^0_{\mathrm{sel}} = \left\lceil \frac{\kappa_\star}{2}\left(R_\alpha + (R_\alpha)^2\right) + \sqrt{\frac{2\tilde{\beta}}{L}} \right\rceil, \qquad R_\alpha := \frac{\alpha_{\max}}{\alpha_{\min}}. \tag{76}$$

Proof From (48) in Lemma 2, $\alpha_{\min} \leq a_\omega = \kappa_\star\sqrt{\omega/L}$. Moreover, from $\mathbf{A_N}$, $\alpha^2_{\min} > 2\omega/\tilde{\beta}$. Thus, $B_{\mathrm{sel}} \leq C^0_{\mathrm{sel}}\sqrt{L\omega}$. Dividing by t_{sel} (70) completes the proof. □

3.3.3 Probabilistic Bounds and Complexity

In this subsection, we first establish the high-probability global gradient bound (Theorem 5) and subsequently translate it into noise-limited accuracy (Theorem 6). Next, we present a high-probability counterpart of Theorem 4 (Theorem 7).

To obtain the stationarity level in the randomized setting under (48), we utilize Proposition 6 and Lemma 4. If $\|g(x^\ell)\| \le G_{\text{sel}}$ at any iteration ℓ, the stationarity bound is directly satisfied. Otherwise, the randomized angle condition holds on the event E_G with parameter Δ_r via Proposition 6. Extending the deterministic worst-case bound (51) to the randomized setting yields $\|g(x^\ell)\| \le \Delta_r^{-1} C_d^0 \sqrt{L\omega}$. Thus, the overall worst-case stationarity level is governed by the maximum of both bounds with an ω-independent constant $\widehat{C}_r$ (and thus any C_r), leading to

$$\varepsilon_{\omega,r} := C_r \sqrt{L\omega}, \qquad \text{where} \quad C_r \ge \widehat{C}_r := \max\left\{C_{\text{sel}},\, \Delta_r^{-1} C_d^0\right\}. \tag{77}$$

Theorem 5 *Under Assumptions* $\mathbf{A_L}$, $\mathbf{A_N}$, *and* $\mathbf{A_S}$, *let the conditions of Proposition 6 be satisfied, and the hypotheses of Theorem 2 hold for the search directions* p_{trec}^ℓ *under* (48) *before termination. For the randomized stationarity level* $\varepsilon_{\omega,r}$ *in* (77), *it holds on* E_G *that*

$$\min_{\ell \in \mathcal{K}_\omega} \|g(x^\ell)\| = \mathcal{O}(\varepsilon_{\omega,r}), \qquad \textit{with probability at least } 1-\gamma_0\ . \tag{78}$$

Proof By definition of $\varepsilon_{\omega,r}$, whether $\|g(x^{\ell_\omega})\| \le G_{\text{sel}}$ or $\|g(x^{\ell_\omega})\| > G_{\text{sel}}$, it holds that $\|g(x^{\ell_\omega})\| \le \varepsilon_{\omega,r}$ on E_G. Since $\ell_\omega \in \mathcal{K}_\omega$, $\min_{\ell \in \mathcal{K}_\omega} \|g(x^\ell)\| \le \|g(x^{\ell_\omega})\| \le \varepsilon_{\omega,r}$, which yields (78). The probabilistic statement then follows from $\mathbb{P}(E_G) \ge 1-\gamma_0$, completing the proof. □

Theorem 6 *Suppose that Assumptions* $\mathbf{A_L}$, $\mathbf{A_N}$, *and* $\mathbf{A_S}$ *hold, the conditions of Theorem 5 are satisfied, and the hypotheses of Theorem 3 hold for the triangular recombination search directions. Then, the following conclusions hold on* E_G *with probability at least* $1-\gamma_0$*:*

(i) If f *is convex, then*

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}(\varepsilon_{\omega,r}). \tag{79}$$

(ii) If f *is* μ_c*-strongly convex, then*

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\left(\varepsilon_{\omega,r}^2\right), \qquad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}(\varepsilon_{\omega,r}). \tag{80}$$

Proof Theorem 5 guarantees $\min_{\ell \in \mathcal{K}_\omega} \|g(x^\ell)\| \le \varepsilon_{\omega,r}$, which implies $\|g(x^{\ell^{**}})\| \le \varepsilon_{\omega,r}$ by the definition of ℓ^{**}. The bounds in terms of $\varepsilon_{\omega,r}$ then follow directly from the properties of f as established in Theorem 3 (see Appendix A.8). Since $\mathbb{P}(E_G) \ge 1-\gamma_0$, the high-probability conclusions follow immediately, completing the proof. □

Theorem 7 *Under Assumptions* $\mathbf{A_L}$, $\mathbf{A_N}$, $\mathbf{A_S}$, *and the noise-scale feasibility condition* (48), *let the conditions of Theorem 5 be satisfied, and the hypotheses of Theorem 4 hold before termination. Then, the following conclusions hold on* E_G *with probability at least* $1 - \gamma_0$*:*

(i) $\ell_T = \mathcal{O}\left(L^2 \varepsilon_{\omega,r}^{-2}\right)$ *and* $N_f = \mathcal{O}\left(L^2 \varepsilon_{\omega,r}^{-2}\right)$ *for nonconvex functions;*

(ii) $\ell_T = \mathcal{O}\left(L^2 \varepsilon_{\omega,r}^{-1}\right)$ *and* $N_f = \mathcal{O}\left(L^2 \varepsilon_{\omega,r}^{-1}\right)$ *for convex functions;*

(iii) $\ell_T = \mathcal{O}\left(L^2 \log(\varepsilon_{\omega,r}^{-1})\right)$ *and* $N_f = \mathcal{O}\left(L^2 \log(\varepsilon_{\omega,r}^{-1})\right)$ *for* μ_c*-strongly convex functions.*

Proof The proof follows the same line of reasoning as the deterministic complexity result (Theorem 4; see Appendix A.10), replacing the deterministic parameters with their randomized counterparts.

Specifically, on E_G, for any pre-termination iteration $0 \leq \ell < \ell_T \leq \ell_\omega < \ell_{\max}$, it holds that $\|g(x^\ell)\| > \varepsilon_{\omega,r}$, where $\varepsilon_{\omega,r}$ is defined in (77). By definition of $\varepsilon_{\omega,r}$, two crucial properties are simultaneously guaranteed: first, $\|g(x^\ell)\| > \varepsilon_{\omega,r} \geq C_{\text{sel}}\sqrt{L\omega}$, ensuring the randomized angle condition holds with parameter Δ_r via Proposition 6, and second, $\|g(x^\ell)\| > \varepsilon_{\omega,r} \geq \Delta_r^{-1} C_d^0 \sqrt{L\omega}$. Thus, for any successful iteration $\ell \in S$, Proposition 3 applies with Δ_r, guaranteeing the sufficient decrease $f(x^\ell) - f(x^{\ell+1}) \geq \tilde{\eta}\|g(x^\ell)\|^2 > \tilde{\eta}\varepsilon_{\omega,r}^2$, where $\tilde{\eta} \propto L^{-2}\Delta_r^2$. Summing these decreases over S yields the bound on the number $|S|$ of successful iterations for nonconvex functions. For convex and strongly convex functions, the proof follows in an analogous manner to the deterministic analysis, replacing ε_ω with $\varepsilon_{\omega,r}$. Similarly, for unsuccessful iterations, the analysis follows directly from Theorem 4, where $a^\ell > a_\omega$ holds for all $\ell < \ell_T$, which bounds the unsuccessful iterations by $|U|$. Combining $\ell_T = |S| + |U|$ on E_G with $\mathbb{P}(E_G) \geq 1 - \gamma_0$ completes the proof. □

Table 1 summarizes the principal nonconvex function-evaluation complexity results for representative randomized DFO methods. Based on this table:

- While noiseless methods (e.g., `DSPFD`, `STP`, and `VRBBO`) target an arbitrary tolerance $\varepsilon > 0$, noise-aware algorithms like `VRDFON` (deterministic) and `DAES` (randomized) adopt accuracy limits $\varepsilon_{\omega,d} = \sqrt{L\omega}$ and $\varepsilon_{\omega,r} := C_r\sqrt{L\omega}$ defined in (77), respectively. Under condition (48), the randomized threshold $\varepsilon_{\omega,r}$ in our `DAES` matches the deterministic noise-order bound in `VRDFON`, up to the randomized scaling constant C_r.
- Sampling strategies vary across the baseline methods (see the respective references in Table 1). Specifically, `DAES` evaluates N_{sel} independent populations of λ directions (formed as $\lambda/2$ antipodal pairs), using an evaluation budget of λN_{sel} per outer iteration, where λ is calibrated via (33).

Table 1 Nonconvex function-evaluation complexity of randomized DFO methods.

Method	Guarantee in	Direction budget	Complexity
`DSPFD` [17]	high-probability	$m = 2$ admissible	$\mathcal{O}(nL^2\varepsilon^{-2})$
`STP` [32]	expectation	One sampled direction	$\mathcal{O}(nL\varepsilon^{-2})$
`VRBBO` [8]	high-probability	$R = \lceil \log_2(\eta_0^{-1}) \rceil$	$\mathcal{O}(nL^2\varepsilon^{-2})$
`VRDFON` [7]	high-probability	$R = \lceil T_0^{-1} \log_2(\eta_0^{-1}) \rceil$	$\mathcal{O}(L^2\varepsilon_{\omega,d}^{-2})$
`DAES`	high-probability	λN_{sel}	$\mathcal{O}(L^2\varepsilon_{\omega,r}^{-2})$

Remark 2 *In `DAES`, the dimension factor $\Delta_r^{-2} = \Theta(n)$ stems directly from the alignment threshold $t_a = \tau/\sqrt{n}$, which governs t_{sel}, $\bar{\varepsilon}$, and consequently Δ_r. Since the stationarity level satisfies $\varepsilon_{\omega,r} \leq \max\{C_{\mathrm{sel}}, C_r\}\sqrt{L\omega}$, with both $C_{\mathrm{sel}} \propto t_a^{-1} = \mathcal{O}(\sqrt{n})$ and $C_r \propto \Delta_r^{-1} = \mathcal{O}(\sqrt{n})$, this yields the explicit factor n in the evaluation complexity $\mathcal{O}\big(nL^2(\sqrt{L\omega})^{-2}\big)$, which is equivalent to the bound reported in Table 1. By comparison, `VRDFON` [7] defines its stationarity parameter directly as $\varepsilon_{\omega,d} := \sqrt{nL\omega}$, leading to a complexity of $\mathcal{O}\big(nL^2(\sqrt{nL\omega})^{-2}\big)$, matching the entry displayed in Table 1.*

4 Numerical Experiments

This section presents an ablation study of `DAES` and compares it with `VRDFON` [7], `VRBBO` [8], `NMSMAX` [18], `MADFO` [26], and `DFNDFL` [33]. These experiments do not aim to claim overall superiority over advanced solvers like `MADFO`, which integrate empirical heuristics. Instead, they measure how effectively a design built for provable complexity bounds performs in practice. The experimental setup is structured as follows.

Test Problems. We consider 230 test problems with the dimensions $2 \leq n \leq 30$ from the `BCP` collection of the `BARON` software package [34], available at https://minlp.com/optimization-test-problems. Each deterministic test problem is perturbed only by absolute uniform noise (see [26]) and evaluated at eight noise levels from $\{10^i\}_{i=-5}^{2}$.

Comparing Metrics. We use *data profiles* [35] and *performance profiles* [36]. Data profiles show how many problems a solver can solve as the available computational budget increases, while performance profiles show how often a solver performs close to the best solver on each problem. Let $\mathcal{S}$ be the set of solvers, and $\mathcal{P}$ and Ω denote the 230 test problems and the eight noise levels, respectively, and $\mathcal{I} = \mathcal{P} \times \Omega$ with $|\mathcal{I}| = 1840$. For $i \in \mathcal{I}$, let $c_{i,s}$ be the function-evaluation cost for solver s, with unsuccessful runs assigned infinite cost, and let n_i denote the dimension of problem p. The data profile $(\delta_s(\kappa))$ and the performance profile $(\rho_s(\bar{\tau}))$ are defined as

$$\delta_s(\kappa) := \frac{1}{|\mathcal{I}|}\left|\left\{i \in \mathcal{I} : \frac{c_{i,s}}{n_i} \leq \kappa\right\}\right|, \qquad \rho_s(\bar{\tau}) := \frac{1}{|\mathcal{I}|}\left|\left\{i \in \mathcal{I} : \frac{c_{i,s}}{\min_{r\in\mathcal{S}} c_{i,r}} \leq \bar{\tau}\right\}\right|.$$

Table 2 Tuning parameters used in `DAES`.

Parameter	Value	Role
η	0.99	triangular recombination parameter
δ	2	norm of the triangular search direction
γ_e	4	extrapolation expansion factor
σ^0	1	initial step-size scale
$\widetilde{\beta}$	10^{-12}	`NDF-SDC` forcing parameter
$\alpha_{\min}$	10^{-5}	minimum step-size scale
$\alpha_{\max}$	10^{2}	maximum step-size scale
ρ_u	0.99	unsuccessful-iteration contraction factor
c_α	2	initial signed-search safeguard
τ	0.25	within-group dominance parameter
ξ	0.5	within-group weight scaling parameter
ρ	0.5	parameter in the cross-group margin property
a^0	1	initial auxiliary signed-search scale

Stopping Criteria. For each solver $s \in \mathcal{S}$, the ratio $q_s = (f_s - f_{\text{tar}})/(f_0 - f_{\text{tar}})$ measures the *performance*, where f_s is the best value returned by the solver, f_0 is a starting value, and f_{tar} is a target evaluated at a high-quality local minimizer, or potentially an approximate global minimizer, identified by sequentially applying a set of DFO solvers. The ratio is used only for benchmarking since f_{tar} is generally unavailable in real-world problems. Solver s solves a problem if $q_s \leq \varepsilon_q$ before reaching `nfmax` or `secmax`; otherwise, it is deemed unsuccessful. We set `secmax` $= 600$ sec, `nfmax` $= 20000$, and $\varepsilon_q = 10^{-4}$.

Initial Point. Following [7, 26], the nominal starting point is shifted by $\bar{\xi} \in \mathbb{R}^n$ with components $\bar{\xi}_i := (-1)^{i-1}\frac{2}{i+2}$ for $i \in [n]$. This shift avoids trivial initialization advantages on `BCP` benchmark problems, ensuring a more realistic performance evaluation. Setting $y^0 := 0$, all noisy function evaluations are computed at $y^\ell + \bar{\xi}$ for $\ell \geq 0$.

Tuning Parameters. For `DAES`, we use the parameter values summarized in Table 2. We set $N_{\text{sel}} = 1$ and $\ell_{\max} = 2 \times 10^4$. Accordingly, $\lambda = 18$, where the population size λ is obtained by (33). In our recombination phase, the normalized within-group weights of each group $\mathcal{G}_k$, $k \in [3]$ are defined as in (14), whose within-group dominance parameter $\varepsilon_{\bar{w}}$ satisfying (13) is set as $\varepsilon_{\bar{w}} = \xi t_{\text{sel}}/(1 + t_{\text{sel}})$ with $\xi > 0$. Normalizing the group weights in (12) ensures that each recombination direction reflects its internal ranking independently of the weight magnitudes across groups. All other solvers are run with their default parameter settings; see [26].

4.1 Noise-Sensitivity and Ablation Study

To evaluate solver performance under noise, we report two complementary metrics: *robustness*, defined as the total number of problem instances solved prior to reaching the evaluation limit, and *efficiency*, measured by the normalized n_f-efficiency statistic representing the computational evaluation cost per solved instance. The evaluated algorithm variants are structured as follows.

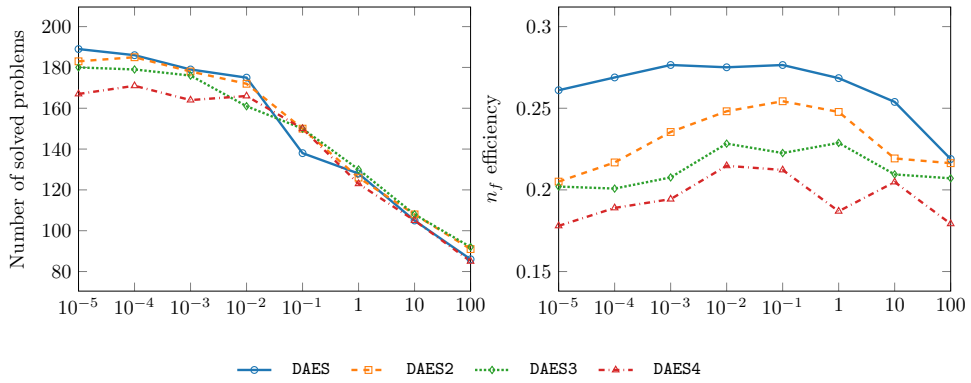


Fig. 2 Sensitivity of the complete `DAES` method to N_{sel} as the noise magnitude increases.

The variants `DAES` and `DAES2`–`DAES4` retain the complete algorithm while systematically varying the population parameter $N_{\text{sel}} \in \{1, 2, 3, 4\}$, respectively, with `DAES` ($N_{\text{sel}} = 1$) serving as the baseline implementation. `DAES5`–`DAES7` isolate the contributions of key algorithmic mechanisms: `DAES5` removes symmetric sampling to assess the role of antipodal directions; `DAES6` replaces the proposed triangular recombination scheme with a traditional weighted recombination mechanism over the top $\mu = \lfloor \lambda/2 \rfloor$ directions (see, e.g., [26]); `DAES7` deactivates the extrapolation component to assess its effect on convergence acceleration.

In Figure 2, the left panel measures robustness, and the right panel gives the reported normalized n_f-efficiency. At the smallest noise level 10^{-5}, the baseline `DAES` solves 189 problems, compared with 183, 180, and 167 for `DAES2`–`DAES4`, respectively. At the largest level 10^2, the counts are 86, 91, 92, and 85. Thus, higher N_{sel} values do not yield uniform or monotonic robustness improvements across noise levels. The right panel reveals a clear trend: the baseline achieves the highest n_f-efficiency across all tested noise levels. For instance, at a noise level of 10^{-5}, the baseline reaches an efficiency of approximately 0.261, compared to 0.205, 0.202, and 0.178 for `DAES2`–`DAES4`, respectively. A similar advantage persists under severe noise.

In Figure 3, the left panel shows that `DAES` and `DAES5`–`DAES7` retain comparable solved counts, although their rankings change with noise. Hence, higher noise levels decrease the number of solved problems across all variants. The right panel highlights the efficiency trade-offs among the structural variants. Notably, the complete `DAES` framework maintains an almost stable n_f-efficiency profile across (almost) all tested noise levels, demonstrating strong efficiency robustness against increasing noise. While removing triangular recombination (`DAES6`) or symmetric sampling (`DAES5`) slightly increases efficiency under specific low-to-moderate noise levels, removing extrapolation (`DAES7`) systematically reduces efficiency across all noise regimes. Overall, the baseline design provides a balanced, noise-insensitive trade-off between robustness and computational efficiency, rather than strict pointwise dominance in every scenario.

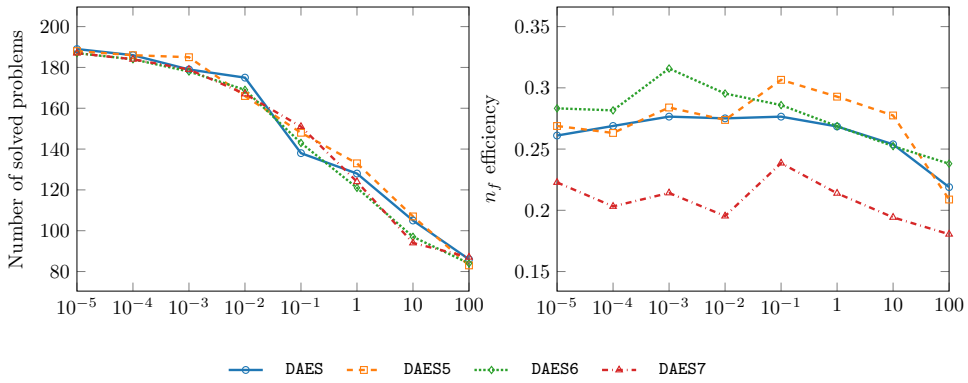


Fig. 3 Structural ablation of DAES across increasing noise.

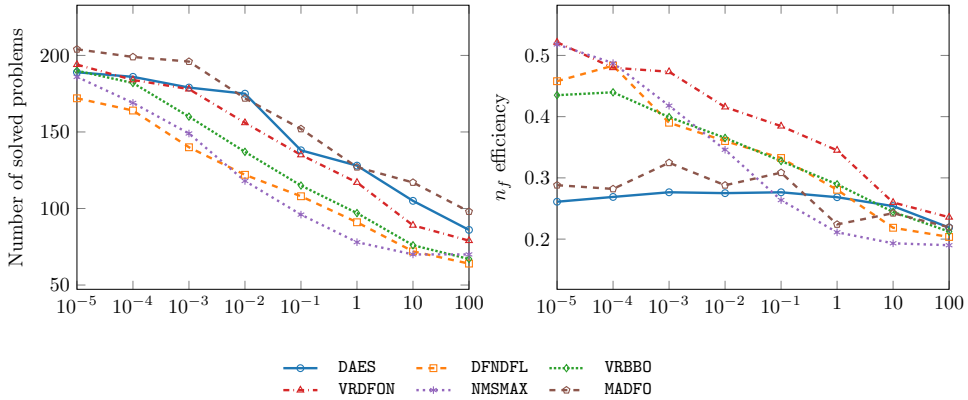


Fig. 4 A comparison between DAES and the external solvers as a function of noise magnitude.

4.2 Comparison with Other DFO Solvers

Figure 4 compares DAES with five external derivative-free solvers. On the left, DAES solves 189 and 86 problems at noise levels 10^{-5} and 10^{2}, respectively, whereas MADFO solves 204 and 98. MADFO achieves the highest solved count across most noise levels. Nevertheless, DAES remains competitive with the other external methods. Based on the right panel, under high noise regimes, DAES is comparable with all the solvers. While MADFO exhibits greater overall robustness—consistent with its emphasis on practical robustness and more elaborate implementation—these curves alone do not identify which individual enhancements cause the performance gap.

With $\mathcal{S} = \{\texttt{DAES}, \texttt{MADFO}\}$, Figure 5 aggregates all 1840 instances and compares the distributions of evaluation costs. The left panel shows a meaningful, but limited, evaluation-budget advantage for DAES. At $\kappa = 5$, its data-profile fraction is approximately 0.258, compared with 0.171 for MADFO. It overtakes DAES at approximately $\kappa = 33.4$. Hence, in terms of robustness, DAES is advantageous in an intermediate,

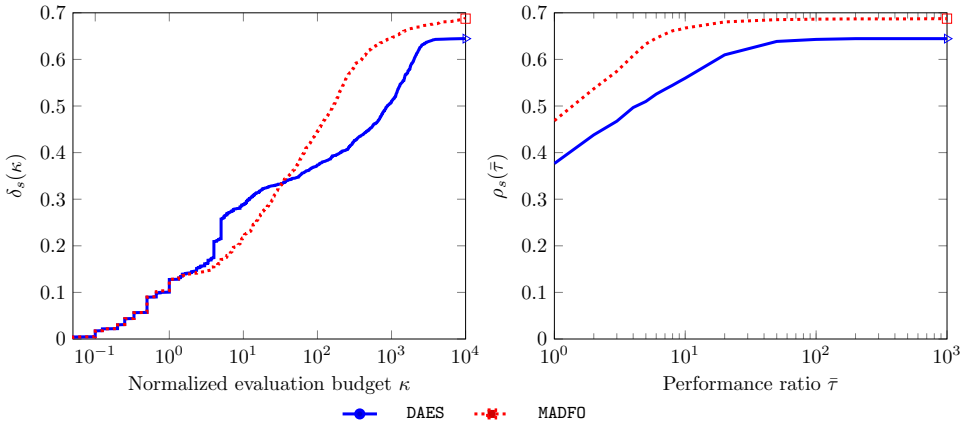


Fig. 5 Data profile (left) and performance profile (right) for `DAES` and `MADFO`.

small-budget regime. The right panel shows that `MADFO` has greater relative efficiency, with a fastest-solver frequency approximately 9.2 percentage points higher than `DAES`, while its overall robustness is approximately 4.3 percentage points higher. These differences may reflect the heuristic techniques and transformation matrices used in `MADFO`, although their individual effects are not isolated by these experiments.

5 Conclusion

For noisy DFO, we proposed `DAES`, an `MAES`-type method that fixes the transformation matrix to the identity. By integrating symmetric sampling, grouped sorting–selection of all samples, and a novel triangular recombination mechanism, `DAES` constructs search directions that systematically mitigate two distinct sources of instability: noise-induced ranking distortion and directional cancellation. The resulting angle guarantee yields high-probability iteration and evaluation-complexity bounds for nonconvex, convex, and strongly convex problems under bounded noise, matching the asymptotic orders of various noisy or noiseless methods referenced in Table 1. To the best of our knowledge, this represents the first complexity guarantee of its kind within this algorithmic class.

Our empirical analysis highlights three main findings for `DAES`. Increasing N_{sel} yields no systematic efficiency gain, while a single population provides a simple baseline with analytical tractability under noise. Although $N_{\text{sel}} > 1$ under (9)–(11) yields no performance gain, exploring alternative selection criteria could unlock its potential in future work. Ablating core components—even without guaranteed theoretical convergence—reveals a trade-off between noise sensitivity and efficiency, whereas the complete method exhibits relatively stable efficiency across the tested noise levels. Finally, compared with `MADFO`, `DAES` remains competitive in robustness under intermediate, small-budget regimes. Overall, with its simple structure, `DAES` bridges complexity guarantees and competitive empirical performance relative to `MADFO`.

Acknowledgments. The authors kindly acknowledge Arnold Neumaier for his constructive comments and suggestions. Morteza Kimiaei acknowledges financial support from the Austrian Science Foundation under https://doi.org/10.55776/PAT2747625. Mahsa Yousefi is a member of the INdAM Research Group GNCS; her work was partially supported by INDAM-GNCS through Progetti di Ricerca 2026, whose support is gratefully acknowledged.

A Proofs in Deterministic Analysis

For simplicity of notation throughout most of the proofs, let $p := p^\ell$ and $x := x^\ell$.

A.1 Proof of Lemma 1

Using (30) and $\mu_s(\alpha)$ in (43),

$$\tilde{\mu}_s(\alpha) - \mu_s(\alpha) = -\alpha^{-2}\big(\tilde{f}(x + s\alpha p) - f(x + s\alpha p) - \big(\tilde{f}(x) - f(x)\big)\big).$$

Using (38), we end up with $|\tilde{\mu}_s(\alpha) - \mu_s(\alpha)| \leq 2\omega/\alpha^2$. Now, since $\tilde{\mu}_s(\alpha) > \tilde{\beta}$, we obtain $\mu_s(\alpha) > \beta$ and $\beta < \tilde{\beta} < 1$ from (39). Moreover, (37) yields $\beta > 0$, which completes the proof.

A.2 Proof of Proposition 1

Let $\alpha := \gamma_e^t \alpha^{\ell 0}$ (for $t \geq 0$) denote any successful step-size during the extrapolation phase at any successful outer iteration ℓ, i.e., satisfying $\tilde{\mu}_s(\alpha) > \tilde{\beta}$ for some $s \in \{-1, 1\}$. By Lemma 1, any such step satisfies the noiseless sufficient decrease $\mu_s(\alpha) > \beta$ with β defined in (39), which implies $f(x^\ell_{\text{trec}} + s\alpha p^\ell_{\text{trec}}) < f(x^\ell_{\text{trec}}) - \beta\alpha^2$. By induction on the successful iterates, we have $f(x^\ell_{\text{trec}}) \leq f(x^0)$. Combined with the lower bound $\underline{f} > -\infty$, this yields $\underline{f} \leq f(x^\ell_{\text{trec}} + s\alpha p^\ell_{\text{trec}}) < f(x^0) - \beta\alpha^2$, which implies

$$\alpha < \sqrt{\beta^{-1}(f(x^0) - \underline{f})} =: A_{\max}.$$

Thus, any successful step-size must strictly satisfy $\alpha < A_{\max}$. Combining the following two cases for $A_{\max}$ relative to $\alpha_{\min}$ confirms the definition of N_e, completing the proof:

1. Let $A_{\max} \leq \alpha_{\min}$: Since the initial signed search enforces $\alpha^{\ell 0} \geq \alpha_{\min}$ by (28), any successful trial step would require $\alpha_{\min} \leq \alpha^{\ell 0} < A_{\max} \leq \alpha_{\min}$, which is a contradiction. Consequently, no successful initial signed step can trigger extrapolation in this branch, yielding $N_e = 0$.

2. Let $A_{\max} > \alpha_{\min}$: Since $\alpha^{\ell 0} \geq \alpha_{\min}$, we have $\gamma_e^t \alpha^{\ell 0} \geq \gamma_e^t \alpha_{\min}$. An extrapolation trial step must be unsuccessful as soon as $\gamma_e^t \alpha_{\min} \geq A_{\max}$. Solving $\gamma_e^t \geq A_{\max}/\alpha_{\min}$ for the integer index t shows that an unsuccessful trial step occurs at or before
$$t = \left\lceil \log_{\gamma_e} (A_{\max}/\alpha_{\min}) \right\rceil .$$
Including this final unsuccessful trial step, the total number of extrapolation expansion trials per successful outer iteration is bounded above by N_e in (44).

A.3 Proof of Proposition 2

Case 1. Let $\alpha = \alpha^{\ell 0}$. If both $x + s\alpha p$, $s \in \{-1, 1\}$, fail `NDF-SDC`, then since $\tilde{\mu}_s(\alpha) \leq \tilde{\beta}$,

$$-\tilde{\beta}\alpha^2 \leq \tilde{f}(x + s\alpha p) - \tilde{f}(x) \overset{(38)}{\leq} f(x + s\alpha p) - f(x) + 2\omega \overset{(36)}{\leq} s\alpha g(x)^T p + \frac{L}{2}\alpha^2\delta^2 + 2\omega.$$

Hence,

$$-s\, g(x)^T p \leq \alpha(\tilde{\beta} + \frac{L}{2}\delta^2) + \frac{2\omega}{\alpha}, \qquad \text{and} \qquad |g(x)^T p| \leq \alpha(\tilde{\beta} + \frac{L}{2}\delta^2) + \frac{2\omega}{\alpha}.$$

Here, applying the first inequality for both $s = +1$ and $s = -1$ gives the second one, followed immediately by (45) for $\gamma_e > 1$.

Case 2. Let $\alpha := \alpha_{\text{trec}}^{\ell}$ be the last accepted successful step-size along sp, for some $s \in \{-1, 1\}$. Hence, by construction and Proposition 1, the accepted step is followed by the first unsuccessful expanded trial, so $\tilde{\mu}_s(\alpha) > \tilde{\beta}$ and $\tilde{\mu}_s(\gamma_e\alpha) \leq \tilde{\beta}$.

Case 2A. From $\tilde{\mu}_1(\alpha) > \tilde{\beta}$, $\tilde{f}(x + \alpha p) - \tilde{f}(x) < -\tilde{\beta}\alpha^2$. Using (36) and (38) implies

$$\alpha g(x)^T p - \frac{L}{2}\alpha^2\delta^2 \leq f(x + \alpha p) - f(x) \leq \tilde{f}(x + \alpha p) - \tilde{f}(x) + 2\omega < -\tilde{\beta}\alpha^2 + 2\omega.$$

Rearranging terms, we arrive at the following bound

$$g(x)^T p < -\tilde{\beta}\alpha + \frac{L}{2}\alpha\delta^2 + \frac{2\omega}{\alpha} < \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha}. \tag{81}$$

Similarly, from the condition $\tilde{\mu}_1(\gamma_e\alpha) \leq \tilde{\beta}$, using (36) and (38) yields an inequality. Dividing by $\gamma_e\alpha$ and rearranging its terms yields (82), which when combined with (81) directly gives (45):

$$-g(x)^T p \leq \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\gamma_e\alpha} \leq \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha}. \tag{82}$$

CASE 2B. Similarly, from $\tilde{\mu}_{-1}(\alpha) > \tilde{\beta}$ and $\tilde{\mu}_{-1}(\gamma_e\alpha) \leq \tilde{\beta}$, we end up with the following bounds, respectively, which, when combined, give (45):

$$-g(x)^T p < -\tilde{\beta}\alpha + \frac{L}{2}\alpha\delta^2 + \frac{2\omega}{\alpha} < \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha}.$$

$$g(x)^T p \leq \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\gamma_e\alpha} \leq \gamma_e\alpha\left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha}.$$

A.4 Proof of Theorem 1

From (41), if $g(x) = 0$, then $v(x,p) = 1$, and (46) follows immediately. Otherwise, if $g(x)^T p = 0$, then $v(x,p) = +\infty$, and (46) holds trivially. Now, let $|g(x)^T p| > 0$. Using the upper bound in (45), we obtain the upper bound in (46) as:

$$\|g(x)\| = v(x,p)\frac{|g(x)^T p|}{\|p\|} \leq v(x,p)\Big(\frac{\gamma_e\alpha}{\|p\|}\big(\tilde{\beta} + \frac{L}{2}\|p\|^2\big) + \frac{2\omega}{\alpha\|p\|}\Big), \qquad \delta = \|p\|.$$

A.5 Proof of Lemma 2

Let any fixed scale $a \in [\alpha_{\min}, \alpha_{\max}]$. At $a^\ell = a$, the initial step-size in (27) satisfies $\alpha^{\ell 0} \geq a/c_\alpha$. By Lemma 1, any successful iteration at scale a yields a strictly positive decrease $f(x^\ell) - f(x^{\ell+1}) \geq \beta(a/c_\alpha)^2 > 0$. Since f is bounded below by $\underline{f}$, infinitely many successful iterations at scale a are impossible. Therefore, at any scale a, only finitely many successful iterations can occur before an unsuccessful trial takes place.

If $a^0 = a_\omega$, the first unsuccessful iteration trivially occurs at scale $a^{\ell_\omega} = a_\omega \in (\rho_u a_\omega, a_\omega]$. Now, consider $a^0 > a_\omega \geq \alpha_{\min}$. Since a remains unchanged during successful iterations and strictly contracts via $a^{\ell+1} = \max\{\alpha_{\min}, \rho_u a^\ell\} < a^\ell$ at unsuccessful iterations whenever $a^\ell > a_\omega \geq \alpha_{\min}$, the sequence $\{a^\ell\}$ must eventually cross a_ω. Let a^- denote the last scale strictly greater than a_ω prior to the first contraction that brings the scale to or below a_ω. Let ℓ_ω be the specific unsuccessful iteration that triggers this contraction, yielding the updated scale $a^{\ell_\omega} := \max\{\alpha_{\min}, \rho_u a^-\}$. By definition of a^-, we have $a^{\ell_\omega} \leq a_\omega$. Furthermore, using $a^- > a_\omega$ and $\rho_u > 0$, we obtain $a^{\ell_\omega} = \max\{\alpha_{\min}, \rho_u a^-\} \geq \rho_u a^- > \rho_u a_\omega$. Combining these bounds yields $\rho_u a_\omega < a^{\ell_\omega} \leq a_\omega$. Thus, ℓ_ω is the first unsuccessful iteration at which $a^{\ell_\omega} \in (\rho_u a_\omega, a_\omega]$, completing the proof.

A.6 Proof of Corollary 1

Let $\ell = \ell_\omega$. From (28), since $\rho_u \in (0,1)$, $c_\alpha^{-1}\rho_u a_\omega < \alpha^{\ell 0} \leq a^\ell$. Using $\underline{\kappa}', \overline{\kappa}'$ in (50) and $a_\omega = \kappa_\star\sqrt{\omega/L}$ in (47), $\underline{\kappa}'\sqrt{\omega/L} < \alpha^{\ell 0} \leq \overline{\kappa}'\sqrt{\omega/L}$. Applying this to $\Gamma(\alpha)$ in (46), where $\alpha = \alpha^{\ell 0}$, gives the bound $\|g(x^{\ell_\omega})\| \leq v_m\,\Gamma(\alpha)$, where $\Gamma(\alpha) \leq C_d^0\sqrt{L\omega}$ and $v_m := v(x^{\ell_\omega}, p^{\ell_\omega}) \leq \Delta_{\mathrm{a}}^{-1}$. If, in addition, (42) holds, then (51) follows immediately.

A.7 Proof of Theorem 2

Over $\mathcal{K}_\omega$, $\min_{\ell\in\mathcal{K}_\omega}\|g(x^\ell)\| \leq \|g(x^{\ell_\omega})\| \overset{(51)}{\leq} \Delta_{\mathrm{a}}^{-1}C_d^0\sqrt{L\omega}$. Since $\underline{\kappa}'$ and $\overline{\kappa}'$ are constants independent of the noise level ω for fixed L, δ, $\tilde{\beta}$ and γ_e in C_d^0, we obtain (52); see Remark 1. This concludes the proof.

A.8 Proof of Theorem 3

By the true sufficient decrease guaranteed on successful iterations and the fact that unsuccessful iterations leave the iterate unchanged, $\{x^\ell\}_{\ell\in\mathcal{K}_\omega} \subseteq \mathcal{L}(x^0)$. Moreover, the global minimizer $\underline{x} \in \mathcal{L}(x^0)$. Since $\mathcal{L}(x^0)$ is compact, there exists a constant $\overline{r} > 0$ such that $\|x^\ell - \underline{x}\| \leq \overline{r}$ for $\ell \in \mathcal{K}_\omega$.

Since $\|g(x^{\ell^{**}})\| \leq \|g(x^{\ell_\omega})\| \overset{(51)}{\leq} \Delta_{\mathrm{a}}^{-1}C_d^0\sqrt{L\omega}$ and C_d^0 is independent of ω, we obtain $\|g(x^{\ell^{**}})\| = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big)$; see Remark 1. Applying this bound to the following arguments yields (53) and (54), which completes the proof:

(i) By convexity of f, $f(x^{\ell^{**}}) - f(\underline{x}) \leq g(x^{\ell^{**}})^T(x^{\ell^{**}} - \underline{x})$. Applying the Cauchy–Schwarz inequality and using compactness of $\mathcal{L}(x^0)$ gives

$$f(x^{\ell^{**}}) - \underline{f} \leq \|g(x^{\ell^{**}})\|\|x^{\ell^{**}} - \underline{x}\| \leq \overline{r}\,\|g(x^{\ell^{**}})\|.$$

(ii) By μ_c-strong convexity,

$$f(x^{\ell^{**}}) - \underline{f} \leq \frac{1}{2\mu_c}\|g(x^{\ell^{**}})\|^2 \qquad \text{and} \qquad \|x^{\ell^{**}} - \underline{x}\| \leq \frac{1}{\mu_c}\|g(x^{\ell^{**}})\|.$$

A.9 Proof of Proposition 3

Let $\alpha = \alpha^{\ell}_{\text{trec}}$. Using the bound $|g(x)^T p|$ in (45) and the angle condition (42), we obtain

$$\Delta_{\text{a}} \|g(x)\| \|p\| \leq |g(x)^T p| \leq \gamma_e \alpha \Big(\tilde{\beta} + \frac{L}{2} \delta^2 \Big) + \frac{2\omega}{\alpha}.$$

Reordering this inequality yields

$$\|g(x)\| \leq K_1 \alpha + K_2 \frac{\omega}{\alpha}, \qquad \text{where} \qquad K_2 := \frac{2}{\Delta_{\text{a}} \delta}. \tag{83}$$

Equivalently, $K_1 \alpha^2 - \|g(x)\| \alpha + K_2 \omega \geq 0$. For this, α must not lie strictly between the roots of the corresponding equation. The roots of $K_1 \alpha^2 - \|g(x)\| \alpha + K_2 \omega = 0$ are

$$\alpha_{\pm} = \frac{\|g(x)\| \pm \sqrt{\Delta}}{2K_1}, \qquad \text{where} \qquad \Delta = \|g(x)\|^2 - 4K_1 K_2 \omega. \tag{84}$$

Now, we consider the following cases for Δ in (84). If $\Delta > 0$, then

$$\alpha_- = \frac{\|g(x)\| - \sqrt{\Delta}}{2K_1} < \alpha_+ = \frac{\|g(x)\| + \sqrt{\Delta}}{2K_1}.$$

The quadratic inequality (83) holds only if $\alpha \leq \alpha_-$ or $\alpha \geq \alpha_+$. Rewriting the smaller root α_- and using

$$\frac{K_2}{K_1} = \frac{2\gamma_e^{-1}}{(\tilde{\beta} + L\delta^2/2)} \leq \frac{2}{\tilde{\beta}},$$

gives

$$\alpha_- = \frac{4K_1 K_2 \omega}{2K_1 \left(\|g(x)\| + \sqrt{\|g(x)\|^2 - 4K_1 K_2 \omega} \right)} < \sqrt{\frac{K_2}{K_1}} \omega \leq \sqrt{\frac{2\omega}{\tilde{\beta}}}. \tag{85}$$

Since $\alpha = \alpha^{\ell}_{\text{trec}}$ is obtained from extrapolation, $\alpha \geq \alpha^{\ell 0} \geq \alpha_{\min}$. Together with (37),

$$\alpha \geq \alpha_{\min} > \sqrt{2\omega/\tilde{\beta}}. \tag{86}$$

Combining (85) with (86) gives $\alpha > \alpha_-$. Therefore, the branch $\alpha \leq \alpha_-$ is impossible, and (83) can hold only on the second feasible branch $\alpha \geq \alpha_+$. Hence,

$$\alpha \geq \alpha_+ = \frac{\|g(x)\| + \sqrt{\|g(x)\|^2 - 4K_1 K_2 \omega}}{2K_1} \geq \frac{\|g(x)\|}{2K_1}.$$

If $\Delta \leq 0$, then $\|g(x)\| \leq 2\sqrt{K_1 K_2 \omega}$ and therefore

$$\frac{\|g(x)\|}{2K_1} \leq \sqrt{\frac{K_2}{K_1}\omega} \leq \sqrt{\frac{2\omega}{\tilde{\beta}}} < \alpha,$$

where the last inequality follows from (86). Hence, in both cases,

$$\alpha^2 \geq \frac{1}{(2K_1)^2}\|g(x)\|^2. \tag{87}$$

Combining (87) with Lemma 1 (under `NDF-SDC`) yields (55).

A.10 Proof of Theorem 4

Let $f^\ell := f(x^\ell)$ and $g^\ell := g(x^\ell)$ for simplicity. Moreover, let S and U be the sets of successful and unsuccessful iterations from the set of pre-termination iterations $\mathcal{K}_T^- := \{\ell \in \mathbb{N}_0 : 0 \leq \ell < \ell_T\}$, respectively, i.e., $\mathcal{K}_T^- = S \cup U$, $S \cap U = \emptyset$ and $\ell_T = |\mathcal{K}_T^-| = |S| + |U|$. Let us also define $\eta := \beta/(2K_1)^2$ in (55). We include the Lipschitz constant $L > 0$ in our conclusions; see Remark 1.

We obtain the number of unsuccessful iterations before termination. If $\ell_\omega < \ell_T$, then $\ell_\omega \in \mathcal{K}_T^-$. Thus, at iteration $\ell = \ell_\omega$, equations (49), (51), and (56) imply $\|g^{\ell_\omega}\| \leq \Delta_{\mathrm{a}}^{-1} C_d^0 \sqrt{L\omega} \leq C_d\sqrt{L\omega} = \varepsilon_\omega$, which contradicts the pre-termination condition $\|g^{\ell_\omega}\| > \varepsilon_\omega$. Therefore, we must have $\ell_T \leq \ell_\omega$. Since the step-size scale a^ℓ is reduced from its initial value a^0 strictly during unsuccessful iterations via (32) as $a^{\ell+1} = \rho_u a^\ell$, and since $a^\ell > a_\omega$ for all $\ell < \ell_T$, we obtain the following bound

$$|U| \leq N_u := \left\lceil \log_{1/\rho_u}\left(\frac{a_0}{a_\omega}\right)\right\rceil, \qquad |U| = \mathcal{O}\left(L\log(\varepsilon_\omega^{-1})\right). \tag{88}$$

The second result follows from a_ω and ε_ω in (47) and (56), $a_\omega = (\kappa_\star/(C_d L))\varepsilon_\omega$.

Since $\ell_T = |S| + |U|$, it remains only to bound $|S|$. Let $S := \{\ell_0, \ldots, \ell_{|S|-1}\}$, with $\ell_0 < \cdots < \ell_{|S|-1}$, as well as $x^j := x^{\ell_j}$, $f^j := f(x^j)$ and $g^j := g(x^j)$ for $j = 0, \ldots, |S|-1$, and $x^{|S|} := x^{\ell_T}$, $f^{|S|} := f(x^{\ell_T})$, and $g^{|S|} := g(x^{\ell_T})$.

(i) For $\ell \in S$ before termination, $\|g^\ell\| > \varepsilon_\omega$. Hence, $f^\ell - f^{\ell+1} \geq \eta\|g^\ell\|^2 > \eta\varepsilon_\omega^2$ by (55). Now, summing over $\ell < \ell_T$ gives

$$f^0 - f^{\ell_T} = \sum_{\ell=0}^{\ell_T-1}(f^\ell - f^{\ell+1}) = \sum_{\ell\in S}(f^\ell - f^{\ell+1}) \geq \eta\varepsilon_\omega^2|S|, \qquad \eta \propto L^{-2}\Delta_{\mathrm{a}}^2.$$

Since $f \geq \underline{f}$, we obtain $|S| \leq \big(f^0 - \underline{f}\big)\big(\eta\varepsilon_\omega^2\big)^{-1}$, yielding $|S| = \mathcal{O}(L^2\varepsilon_\omega^{-2})$. Finally,

$$\ell_T = |S| + |U| = \mathcal{O}(L^2\varepsilon_\omega^{-2}) + \mathcal{O}(L\log(\varepsilon_\omega^{-1})) = \mathcal{O}(L^2\varepsilon_\omega^{-2}).$$

(ii) Let $d_{\max} := \sup_{x\in\mathcal{L}(x^0)} \|x - \underline{x}\|$. By convexity and the Cauchy–Schwarz inequality,

$$f^j - \underline{f} \leq d_{\max}\|g^j\|, \qquad \text{for} \quad j = 0, \ldots, |S|. \tag{89}$$

Combining (89) with (55) for $j = 0, \ldots, |S|-1$ yields $f^{j+1} \leq f^j - \eta d_{\max}^{-2}(f^j - \underline{f})^2$ which is equivalent to $\psi_{j+1}^{-1} - \psi_j^{-1} \geq \eta d_{\max}^{-2}$, where $\psi_j := f^j - \underline{f}$ and $\psi_{j+1} \leq \psi_j$. Summing this inequality from $j = 0$ to $k-1$, simplifying the telescoping sum and rearranging terms yields the optimality gap of the k-th successful iteration as

$$\psi_k \leq \big(\eta d_{\max}^{-2} k + \psi_0^{-1}\big)^{-1}, \qquad k = 1, \ldots, |S|. \tag{90}$$

In particular, setting $k = |S|$ in (90) bounds the optimality gap in terms of the total number of successful iterations. On the other hand, at the terminal iterate $x^{|S|} \in \mathcal{L}(x^0)$, where $\|g^{|S|}\| \leq \varepsilon_\omega$, (89) yields $\psi_{|S|} \leq d_{\max}\varepsilon_\omega$. Let $j_1 \in \{0, \ldots, |S|\}$ be the first index satisfying $\psi_{j_1} \leq d_{\max}\varepsilon_\omega$. Achieving this requires $1/k = \mathcal{O}(\varepsilon_\omega)$, by (90), which yields $j_1 = \mathcal{O}(\varepsilon_\omega^{-1})$. Let us now discuss j_1.

If $j_1 = |S|$, there are no successful iterations after j_1; i.e., $|S| = \mathcal{O}(\varepsilon_\omega^{-1})$. Otherwise, the algorithm requires additional iterations until $\|g^{|S|}\| \leq \varepsilon_\omega$. If $j_1 < |S|$, then $\|g^j\| > \varepsilon_\omega$ for $j = j_1, \ldots, |S| - 1$. For each such j, by (55), $f^j - f^{j+1} \geq \eta\varepsilon_\omega^2$. However, $f^{j_1} - f^{|S|} \leq f^{j_1} - \underline{f} \leq d_{\max}\varepsilon_\omega$ specifies the total decrease after j_1. Hence, the number of successful iterations after j_1 is

$$\frac{\psi_{j_1}}{\psi_j - \psi_{j+1}} \leq \frac{d_{\max}\varepsilon_\omega}{\eta\varepsilon_\omega^2} = \mathcal{O}(L^2\varepsilon_\omega^{-1}), \qquad \eta \propto L^{-2}\Delta_{\mathrm{a}}^2.$$

Therefore, combining $j_1 = \mathcal{O}(\varepsilon_\omega^{-1})$ and $|S| - j_1 = \mathcal{O}(\varepsilon_\omega^{-1})$ yields $|S| = \mathcal{O}(L^2\varepsilon_\omega^{-1})$. Finally, in both cases $j_1 = |S|$ and $j_1 < |S|$,

$$\ell_T = |S| + |U| = \mathcal{O}(L^2\varepsilon_\omega^{-1}) + \mathcal{O}(L\log(\varepsilon_\omega^{-1})) = \mathcal{O}(L^2\varepsilon_\omega^{-1}).$$

(iii) The μ_c-strong convexity bound $f^j - \underline{f} \leq (2\mu_c)^{-1}\|g^j\|^2$ yields $\|g^j\|^2 \geq 2\mu_c\psi_j$ where $\psi_j := f^j - \underline{f}$ is the optimality gap for every successful index $j = 0, \ldots, |S| - 1$. Combining this inequality with (55) yields $\psi_{j+1} \leq (1 - 2\eta\mu_c)\psi_j$. To ensure a contraction factor $q(\eta) := 1 - 2\eta\mu_c \in (0, 1)$, we define $\widehat{\eta} := \min\{\eta, (4\mu_c)^{-1}\}$ so that $q_0 := q(\widehat{\eta}) \in [1/2, 1)$. Since $\widehat{\eta} \leq \eta$, we obtain $\psi_{j+1} \leq q_0\psi_j$, and hence $\psi_j \leq q_0^j\psi_0$. On the other hand, setting $x = x^j$ and $\hat{s} = -L^{-1}g^j$ in (36), together

with $f \geq \underline{f}$, yields $\|g^j\|^2 \leq 2L\psi_j$. Hence, the stationarity criterion $\|g^j\| \leq \varepsilon_\omega$ is guaranteed whenever $\psi_j \leq \varepsilon_\omega^2/(2L)$. Consequently, it suffices that

$$|S| \leq \left\lceil \Big(2\log(\varepsilon_\omega^{-1}) + \log(2L\psi_0)\Big) / \Big(-\log(q_0)\Big) \right\rceil, \qquad -\log(q_0) \propto L^{-2}\Delta_{\mathrm{a}}^2.$$

Therefore, $|S| = \mathcal{O}\left(L^2 \log(\varepsilon_\omega^{-1})\right)$. Finally,

$$\ell_T = |S| + |U| = \mathcal{O}\left(L^2 \log(\varepsilon_\omega^{-1})\right) + \mathcal{O}(L \log(\varepsilon_\omega^{-1})) = \mathcal{O}(L^2 \log(\varepsilon_\omega^{-1})).$$

Now, we establish a uniform bound on evaluations per outer iteration of `DAES`. By Proposition 1, the extrapolation phase needs at most $N_e \leq \lceil \log_{\gamma_e}(\alpha_{\max}/\alpha_{\min}) \rceil$ evaluations (independent of ε_ω). Each iteration ℓ requires at most $\lambda N_{sel} + 2 + N_e$ constant evaluations (covering mutation–selection, initial signed search, and extrapolation), yielding $N_f = \mathcal{O}(\ell_T)$ and completing the proof.

References

[1] Audet, C., Hare, W.: Derivative-Free and Blackbox Optimization. Springer, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-68913-5

[2] Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to derivative-free optimization. In: MPS-SIAM Series on Optimization (2009). https://doi.org/10.1137/1.9780898718768

[3] Larson, J., Menickelly, M., Wild, S.M.: Derivative-free optimization methods. Acta Numerica **28**, 287–404 (2019) https://doi.org/10.1017/S0962492919000060

[4] Berahas, A.S., Byrd, R.H., Nocedal, J.: Derivative-free optimization of noisy functions via quasi-newton methods. SIAM J. Optim. **29**(2), 965–993 (2019) https://doi.org/10.1137/18m1177718

[5] Brilli, A., Kimiaei, M., Liuzzi, G., Lucidi, S.: Worst case complexity bounds for linesearch-type derivative-free algorithms. Journal of Optimization Theory and Applications **203**(1), 419–454 (2024) https://doi.org/10.1007/s10957-024-02519-x

[6] Giovannelli, T., Liuzzi, G., Lucidi, S., Rinaldi, F.: Derivative-free methods for mixed-integer nonsmooth constrained optimization. Computational Optimization and Applications **82**(2), 293–327 (2022) https://doi.org/10.1007/s10589-022-00363-1

[7] Kimiaei, M.: An improved randomized algorithm with noise level tuning for large-scale noisy unconstrained dfo problems. Numerical Algorithms (2025) https://doi.org/10.1007/s11075-025-02016-w

[8] Kimiaei, M., Neumaier, A.: Efficient unconstrained black box optimization. Mathematical Programming Computation **14**(2), 365–414 (2022) https://doi.org/10.1007/s12532-021-00215-9

[9] Kimiaei, M., Neumaier, A., Faramarzi, P.: New subspace method for unconstrained derivative-free optimization. ACM Transactions on Mathematical Software **49**(4), 1–28 (2023) https://doi.org/10.1145/3618297

[10] Lucidi, S., Sciandrone, M.: A derivative-free algorithm for bound constrained optimization. Computational Optimization and Applications **21**(2), 119–142 (2002) https://doi.org/10.1023/a:1013735414984

[11] De Santis, A., Liuzzi, G., Lucidi, S.: A linesearch-based derivative-free method for noisy black-box problems. preprint arXiv:2508.00495 (2025) https://doi.org/10.48550/arXiv.2508.00495

[12] Bandeira, A.S., Scheinberg, K., Vicente, L.N.: Convergence of trust-region methods based on probabilistic models. SIAM J. Optim. **24**(3), 1238–1264 (2014) https://doi.org/10.1137/130915984

[13] Gratton, S., Toint, P.L., Tröltzsch, A.: An active-set trust-region method for derivative-free nonlinear bound-constrained optimization. Optimization Methods and Software **26**(4–5), 873–894 (2011) https://doi.org/10.1080/10556788.2010.549231

[14] Powell, M.J.: UOBYQA: unconstrained optimization by quadratic approximation. Mathematical Programming **92**(3), 555–582 (2002) https://doi.org/10.1007/s101070100290

[15] Liuzzi, G., Lucidi, S., Rinaldi, F., Vicente, L.N.: Trust-region methods for the derivative-free optimization of nonsmooth black-box functions. SIAM J. Optim. **29**(4), 3012–3035 (2019) https://doi.org/10.1137/19m125772x

[16] Dzahini, K.J., Rinaldi, F., Royer, C.W., Zeffiro, D.: Direct-search methods in the year 2025: Theoretical guarantees and algorithmic paradigms. EURO Journal on Computational Optimization **13**, 100110 (2025) https://doi.org/10.1016/j.ejco.2025.100110

[17] Gratton, S., Royer, C.W., Vicente, L.N., Zhang, Z.: Direct search based on probabilistic descent. SIAM J. Optim. **25**(3), 1515–1541 (2015) https://doi.org/10.1137/140961602

[18] Higham, N.J.: Optimization by direct search in matrix computations. SIAM Journal on Matrix Analysis and Applications **14**(2), 317–333 (1993) https://doi.org/10.1137/0614023

[19] Porcelli, M., Toint, P.L.: Exploiting problem structure in derivative free optimization. ACM Transactions on Mathematical Software **48**(1), 1–25 (2022) https://doi.org/10.1145/3474054

[20] Huyer, W., Neumaier, A.: Global optimization by multilevel coordinate search. Journal of Global Optimization **14**(4), 331–355 (1999) https://doi.org/10.1023/a:1008382309369

[21] Custódio, A.L., Madeira, J.F.A.: Glods: Global and local optimization using direct search. Journal of Global Optimization **62**(1), 1–28 (2014) https://doi.org/10.1007/s10898-014-0224-9

[22] Custódio, A.L., Madeira, J.F.A.: Multiglods: global and local multiobjective optimization using direct search. Journal of Global Optimization **72**(2), 323–345

(2018) https://doi.org/10.1007/s10898-018-0618-1

[23] Auger, A., Hansen, N.: A restart CMA evolution strategy with increasing population size. In: Congress on Evolutionary Computation, vol. 2, pp. 1769–1776. IEEE, Edinburgh, UK (2005). https://doi.org/10.1109/CEC.2005.1554902

[24] Beyer, H.-G., Sendhoff, B.: Simplify your covariance matrix adaptation evolution strategy. IEEE Transactions on Evolutionary Computation **21**(5), 746–759 (2017) https://doi.org/10.1109/TEVC.2017.2680320

[25] Kimiaei, M.: GS1400/MATRS: MATRSv3.0. Zenodo (2024). https://doi.org/10.5281/zenodo.13759724

[26] Kimiaei, M., Neumaier, A.: Effective matrix adaptation strategy for noisy derivative-free optimization. Mathematical Programming Computation **16**(3), 459–501 (2024) https://doi.org/10.1007/s12532-024-00261-z

[27] Audet, C., Le Digabel, S., Montplaisir, V.R., Tribes, C.: Algorithm 1027: Nomad version 4: Nonlinear optimization with the mads algorithm. ACM Transactions on Mathematical Software **48**(3), 1–22 (2022) https://doi.org/10.1145/3544489

[28] Diouane, Y., Gratton, S., Vicente, L.N.: Globally convergent evolution strategies. Mathematical Programming **152**, 467–490 (2015) https://doi.org/10.1007/s10107-014-0793-x

[29] Akimoto, Y., Auger, A., Glasmachers, T., Morinaga, D.: Global linear convergence of evolution strategies on more than smooth strongly convex functions. SIAM Journal on Optimization **32**(2), 1402–1429 (2022) https://doi.org/10.1137/20M1373815

[30] Touré, C., Auger, A., Hansen, N.: Global linear convergence of evolution strategies with recombination on scaling-invariant functions. Journal of Global Optimization **86**(1), 163–203 (2023) https://doi.org/10.1007/s10898-022-01249-6

[31] Hansen, N.: The CMA evolution strategy: A tutorial. preprint arXiv:1604.00772 (2016) https://doi.org/10.48550/arXiv.1604.00772

[32] Bergou, E.H., Gorbunov, E., Richtárik, P.: Stochastic three points method for unconstrained smooth minimization. SIAM J. Optim. **30**(4), 2726–2749 (2020) https://doi.org/10.1137/19m1244378

[33] Giovannelli, T., Liuzzi, G., Lucidi, S., Rinaldi, F.: Derivative-free methods for mixed-integer nonsmooth constrained optimization. Comput. Optim. Appl. **82**, 293–327 (2022) https://doi.org/10.1007/s10589-022-00363-1

[34] Zhang, Y., Sahinidis, N.V.: Solving continuous and discrete nonlinear programs with BARON. Computational Optimization and Applications **92**(3), 1123–1161 (2024) https://doi.org/10.1007/s10589-024-00633-0

[35] Moré, J.J., Wild, S.M.: Benchmarking derivative-free optimization algorithms. SIAM J. Optim. **20**(1), 172–191 (2009) https://doi.org/10.1137/080724083

[36] Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Mathematical Programming **91**(2), 201–213 (2002) https://doi.org/10.1007/s101070100263