A Direction Adaptation Evaluation 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 this paper, we develop a direction adaptation evolution strategy (`DAES`)—a new `MAES`-type method—for noisy derivative-free optimization, designed to reconcile the population-based search mechanisms of evolution strategies with rigorous complexity analysis. Unlike standard `MAES` schemes, `DAES` fixes the adaptation matrix to the identity and replaces matrix adaptation with a structured direction-generation mechanism based on symmetric sampling, joint sorting–selection of noisy function values, three-group recombination, and a new triangular search direction. Specifically, candidates are sampled along paired positive and negative directions, their inexact function values are jointly ranked, and the reordered directions are partitioned into three groups to construct three recombination points whose geometry defines the triangular direction. A signed sufficient-decrease search and extrapolation mechanism is then applied along this direction. This structure yields a population-based `MAES`-type algorithm that retains competitive practical behavior while being amenable to nonasymptotic analysis under noisy evaluations. We establish high-probability complexity bounds for nonconvex, convex, and strongly convex objective functions and derive corresponding guarantees at the noise-limited accuracy level. To the best of our knowledge, these results provide the first high-probability complexity guarantees for a noisy `MAES`-type derivative-free method with this direction-adaptation structure. Finally,

numerical experiments on the 655 `prince` test problems from the `BARON` collection compare `DAES` with the advanced `MAES`-type solver `MADFO` and show a favorable trade-off between evaluation efficiency and ultimate robustness.

Keywords: Noisy Derivative-Free Optimization, Evolution Strategy, Randomized Optimization, Direction Adaptation, Complexity Results, Heuristic Optimization

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].

1.1 Related work

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`) [23–27]. For advanced DFO methods that combine `DS` or `LS` with model-based techniques or with `MAES`, we refer the reader to [26–29].

`DS` methods explore the objective function along a finite set of polling directions and accept trial points that satisfy a forcing-function-based—typically proportional to the squared step-size—sufficient decrease condition. They require no derivative information and update step-sizes through simple increase–decrease rules. While `DS` methods are robust and effective for low-dimensional problems, their reliance on polling directions can lead to poor scalability as the problem dimension increases.

On the other hand, several `LS` methods also explore descent along prescribed directions, but they differ in the mechanism used to enforce sufficient decrease. Some approximate directional derivatives and apply Armijo or Wolfe conditions, while others rely on a forcing function similar to that of `DS` methods. This classification underscores the close connection between derivative-free `LS` strategies and `DS` schemes. Moreover, these `LS` variants employ an extrapolation step to move away from regions containing saddle points or maximizers and to rapidly approach an approximate minimizer.

MAES consists of three phases: mutation, selection, and recombination. In the mutation phase, candidate solutions are generated by sampling random directions and applying an adaptive affine transformation, which requires matrix construction and matrix–vector multiplications. In the selection phase, candidates are ranked according to their inexact objective function values, and the best mutation directions are retained. Finally, the recombination phase computes a weighted average of the selected directions to form a new point and update the step-sizes and transformation matrix.

In noisy regimes, an algorithm may mistakenly identify points with spuriously good function values. In particular, this discrepancy can undermine the reliability of descent-based strategies and can mislead the sorting and selection process in MAES. Failure in the sorting–selection phase of MAES can adversely affect the recombination phase, as the ordering of individuals may be inaccurate. Addressing this challenge by enhancements in different phases is the central objective of the present paper.

1.2 Our contribution

In this paper, we develop a new MAES-type method for noisy DFO, called the direction adaptation evolution strategy (DAES), by combining structured direction adaptation with an extrapolation step inspired by derivative-free LS methods. Unlike standard MAES, which rely on adaptive matrix updates and form recombination directions from selected mutation directions, DAES fixes the matrix to the identity and enables exploration over all mutation directions. It generates mutation points by sampling symmetric pairs of random directions (a direction and its negative), thereby promoting balanced exploration and mitigating the directional bias associated with one-sided sampling. The noisy function values are jointly sorted in ascending order, and the corresponding directions are partitioned into three ranked groups. Although this ranking can be unreliable under high noise, three recombination points—one from each group—are constructed, and their geometry defines a new triangular recombination direction designed to guide the search toward promising regions. This redesign replaces adaptive matrix updates with a structured population-based direction mechanism that retains practical effectiveness while being amenable to rigorous complexity analysis under noisy evaluations. From a theoretical perspective, fixing the adaptation matrix to the identity eliminates the need for additional assumptions on the spectrum or conditioning of the matrix adaptation, leading to a cleaner analysis of the proposed sampling and recombination scheme in noisy DFO.

In addition to the enhancements introduced in DAES, we provide a complexity analysis. To the best of our knowledge, these are the first high-probability complexity results for a noisy MAES-type method covering nonconvex, convex, and strongly convex objectives and attaining the corresponding complexity orders reported for related optimization frameworks in [7, 8, 17, 30]. To support the theoretical foundation of DAES, we derive high-probability complexity guarantees through a probabilistic convergence analysis. Under bounded evaluation noise, we show that the triangular recombination directions generated by the algorithm satisfy a probabilistic angle condition with respect to the

true gradient, based on sharp bounds that explicitly account for bounded evaluation noise. By controlling the associated per-iteration failure probabilities, we establish that the required directional and descent properties hold uniformly with high probability. Building on these results and standard tools from nonconvex, convex, and strongly convex optimization, we derive total-iteration and function-evaluation complexity bounds at the noise-limited accuracy level, with explicit dependence on the noise level ω and the sampling budget λ.

We compare the full version of `DAES`, incorporating symmetric sampling, grouped sorting–selection, triangular recombination, and extrapolation, with `MADFO` [27], an advanced `MAES`-type derivative-free solver, on 655 `prince` test problems from the `BARON` collection [31] with dimensions ranging from 2 to 100. The results show that `DAES` achieves the strongest overall behavior among its variants, with triangular recombination and extrapolation contributing most to efficiency and robustness, and is more efficient than `MADFO` over small and moderate budgets, whereas `MADFO` is ultimately more robust for substantially larger budgets.

1.3 Notation and Organization

In this paper, $[n] := \{1, 2, \ldots, n\}$ and $[n]_0 := \{0, 1, \ldots, n\}$. The notation $\mathcal{N}(0, I_n)$ denotes the standard multivariate normal distribution with mean $0 \in \mathbb{R}^n$ and identity covariance matrix $I_n \in \mathbb{R}^{n\times n}$, and $a \sim \mathcal{U}(0, 1)$ refers to a as a random variable drawn from the uniform distribution on $(0, 1)$. The abbreviation i.i.d. stands for independent and identically distributed random variables and $\|\cdot\|$ stands for the Euclidean norm. In the sorting–selection phase at iteration ℓ, the original direction $d^{j\ell}$ is represented by the ranked direction $d^{\pi(j)\ell}$ after sorting. Thus, subscript π denotes the labeling induced by the sorting–selection phase; otherwise, it refers to the number pi. We also adopt the conventional asymptotic notation $\mathcal{O}$, Ω, and Θ to characterize growth rates, and use sup and inf to denote the least upper bound and greatest lower bound, respectively.

The paper is organized as follows. Section 2 introduces `DAES` and details its mechanism. We establish complexity guarantees in Sections 3.1 and 3.2 for deterministic and randomized variants of `DAES`, respectively. Numerical experiments, demonstrating the effectiveness of `DAES` compared to existing approaches, are reported in Section 4. Finally, Section 5 concludes the paper and outlines directions for future research.

2 `DAES`: A New DFO Method

This section introduces `DAES`, a DFO method in the `MAES` framework. To solve (1), `DAES` starts from an initial guess x^0 and proceeds iteratively with iteration index $\ell \geq 1$ by three distinct enhanced phases: mutation, sorting–selection, and recombination.

In the following subsections, we detail the process involved in each phase. The complete procedure of `DAES` is outlined in Subsection 2.3.

2.1 Mutation Phase

The goal of the mutation phase is to generate a finite number (λ) of mutation points, which are randomly sampled. At iteration ℓ of `DAES`, this phase involves generating three components: the distribution directions, $\{z^{j\ell}\}$, the mutation step-sizes, $\{\alpha^{j\ell}\}$, and the mutation points, $\{y_{\pm}^{j\ell}\}$ for $j \in [\lambda]$. The mutation directions are defined as the product of the affine scaling matrix and the distribution directions. In contrast to `MAES`, where a scaling matrix is typically employed to generate the λ mutation directions by transforming the distribution directions, `DAES` adopts the identity matrix. This means that the mutation directions are directly derived from the sampled (distribution) directions without any covariance shaping. Obviously, this avoids the computational overhead associated with matrix-based adaptations in `DAES`. Given mutation directions and step-sizes, the goal of the mutation phase is achieved. In the following, we describe how each component of the mutation phase is constructed.

2.1.1 Mutation Directions

In some DFO approaches, such as `LS`-based methods, function evaluations are typically carried out along both symmetric directions to determine which one results in greater descent. Inspired by this idea, `DAES` also generates mutation directions in symmetric pairs, bypassing the need for an adaptation matrix.

Assuming that λ is an even multiple of three[1], `DAES` constructs $\lambda/2$ sampled vectors $\{z^{j\ell}\}$ for $j \in [\lambda/2]$, stored in $Z^{\ell} \in \mathbb{R}^{n\times\lambda/2}$, where each column satisfies $z^{j\ell} \sim \mathcal{N}(0, I_n)$. Although this isotropic distribution is unbiased in expectation, finite samples may exhibit directional asymmetries. To mitigate such sample-induced bias, `DAES` includes both $+z^{j\ell}$ and $-z^{j\ell}$, yielding a symmetric set with zero empirical mean. The corresponding raw symmetric Gaussian samples are stored as

$$Z_{\pm}^{\ell} := \begin{pmatrix} Z^{\ell} & -Z^{\ell} \end{pmatrix} \in \mathbb{R}^{n\times\lambda},$$

and are retained for the scaling step-size adaptation.

Since the affine scaling matrix is fixed to the identity, the mutation directions are obtained directly from these sampled vectors, followed by normalization to ensure consistent directional scaling. More precisely, let $\widehat{d}^{j\ell}$ denote the corresponding raw symmetric vector, so that

$$\widehat{d}^{j\ell} := \begin{cases} z^{j\ell}, & j \in [\lambda/2], \\ -z^{j-\lambda/2,\ell}, & j \in [\lambda] \setminus [\lambda/2]. \end{cases}$$

[1] Users can handle any λ value; this assumption is made for simplicity in the presentation.

Each mutation direction is then normalized as

$$d^{j\ell} := \frac{\widehat{d}^{j\ell}}{\|\widehat{d}^{j\ell}\|}, \qquad \text{for all} \quad j \in [\lambda]. \tag{2}$$

Hence, letting

$$U^{\ell} := \left(\frac{z^{1\ell}}{\|z^{1\ell}\|}, \ldots, \frac{z^{\lambda/2,\ell}}{\|z^{\lambda/2,\ell}\|} \right) \in \mathbb{R}^{n \times \lambda/2},$$

we store the normalized symmetric mutation directions as

$$D^{\ell} := \begin{pmatrix} U^{\ell} & -U^{\ell} \end{pmatrix} \in \mathbb{R}^{n \times \lambda}.$$

Thus, $Z^{\ell}_{\pm}$ contains the raw symmetric Gaussian samples used in the scaling step-size adaptation, whereas D^{ℓ} contains the normalized symmetric mutation directions used to generate mutation points and in the subsequent sorting–selection and recombination phases.

Empirical results also confirm that symmetric sampling improves stability and search performance compared to asymmetric directions. Besides improving the geometric coverage of the search space by avoiding directional bias, symmetric directions are also known to reduce the impact of noise when gradient approximations are constructed from function-value differences, a common strategy in DFO[2].

2.1.2 Mutation Points

For generating mutation points (candidates), the mutation step-sizes play a crucial role in ensuring adequate exploration of the search space. Let $[\alpha_{\min}, \alpha_{\max}]$ denote the admissible interval of step-sizes. Once the step-sizes $\alpha^{j\ell} > 0$, for all $j \in [\lambda]$, are selected from this interval at iteration ℓ, the λ mutation points are computed as

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

where $x^{\ell}_{\text{trec}} \in \mathbb{R}^n$ denotes the current iterate, which is updated in the recombination phase (Subsection 2.3.2), $x^{0}_{\text{trec}} = x^0$ is the given initial point, and $d^{j\ell} \in D^{\ell}$.

Using the normalized antipodal directions and the paired step-sizes $\alpha^{j+\lambda/2,\ell} = \alpha^{j\ell}$, the λ mutation points can equivalently be written as

$$y^{j\ell}_{+} := x^{\ell}_{\text{trec}} + \alpha^{j\ell} \frac{z^{j\ell}}{\|z^{j\ell}\|}, \qquad y^{j\ell}_{-} := x^{\ell}_{\text{trec}} - \alpha^{j\ell} \frac{z^{j\ell}}{\|z^{j\ell}\|}, \qquad j \in [\lambda/2]. \tag{3}$$

[2]This property motivates our follow-up study, which builds upon the present work.

2.1.3 Mutation Step-Sizes

We now explain how the mutation step-sizes $\alpha^{j\ell} \in [\alpha_{\min}, \alpha_{\max}]$ are determined and adaptively adjusted; see [27, 32]. Using element-wise division, let us define

$$b^{j\ell} := \left| x^{\ell}_{\text{trec}} / q_{\alpha} \right|, \qquad q_{\alpha} := z^{j\ell} \neq 0 \qquad \text{for} \quad j \in [\lambda/2]. \tag{4}$$

Then

$$\alpha^{j\ell} = \sqrt{\sigma^{\ell-1} \cdot \text{median}(b^{j\ell})}, \qquad \text{for} \quad j \in [\lambda/2], \tag{5}$$

where $\sigma^{\ell-1}$ is a scaling step-size (with $\sigma^0 = 1$), and the median is computed using only the valid entries of $b^{j\ell} \in \mathbb{R}^n$, excluding NaN, infinite, and zero values. Recalling $Z^{\ell}_{\pm} = (Z^{\ell} \ -Z^{\ell})$, $D^{\ell} = (U^{\ell} \ -U^{\ell})$, and (3), `DAES` sets $\alpha^{j+\lambda/2,\ell} := \alpha^{j\ell}$ for $j \in [\lambda/2]$. At iteration ℓ, the scaling step-size is updated as

$$\sigma^{\ell} = \sigma^{\ell-1} \exp\big(c_{\sigma} d_{\sigma}^{-1} \left(\|s^{\ell-1}_{\sigma}\|_2 / e_{\sigma} - 1 \right)\big) \in [\alpha_{\min}, \alpha_{\max}], \tag{6}$$

where $c_{\sigma} \leq 1$ is a learning rate, $d_{\sigma} \approx 1$ is a damping parameter, and e_{σ} approximates $\mathbb{E}(\|u\|_2)$ for $u \sim \mathcal{N}(0, I_n)$. In (6), $s^{\ell-1}_{\sigma}$ (with $s^0_{\sigma} = 0 \in \mathbb{R}^n$) is the evaluation path [25].

2.2 Sorting–Selection Phase

A proper sorting–selection process directly influences the quality of recombination. In classical `MAES`, the best mutation points are selected according to their function values, and their directions are averaged to define the recombination direction. Standard approaches [23–25, 27] split mutation points into a high-quality subset with the lowest function values, used to improve the current solution, and a discarded subset. However, selecting only the lowest noisy function values can be misleading.

To avoid misleading information from spuriously good points and bias the search direction, `DAES` ranks all symmetric mutation directions according to the noisy function values of their mutation points. For each mutation point $y^{j\ell}_{+}$ and $y^{j\ell}_{-}$ in (3), let $\tilde{f}^{j\ell}_{+}$ and $\tilde{f}^{j\ell}_{-}$ denote their corresponding noisy function values, respectively. Letting the vectors

$$\tilde{F}^{\ell}_{+} := (\tilde{f}^{1\ell}_{+}, \ldots, \tilde{f}^{\lambda/2,\ell}_{+}), \qquad \tilde{F}^{\ell}_{-} := (\tilde{f}^{1\ell}_{-}, \ldots, \tilde{f}^{\lambda/2,\ell}_{-}).$$

and $\tilde{F}^{\ell} = (\tilde{F}^{\ell}_{+} \ \ \tilde{F}^{\ell}_{-}) \in \mathbb{R}^{1\times\lambda}$, `DAES` forms a sorted vector $\tilde{F}^{\ell}_{\pi} = (\tilde{f}^{1\ell}_{\pi}, \ldots, \tilde{f}^{\lambda\ell}_{\pi})$ with $\tilde{f}^{1\ell}_{\pi} \leq \cdots \leq \tilde{f}^{\lambda\ell}_{\pi}$. The same sorting permutation is applied consistently to both the raw symmetric storage $Z^{\ell}_{\pm}$ and the normalized direction matrix D^{ℓ}, yielding $(Z^{\ell}_{\pm})_{\pi}$ and D^{ℓ}_{π}, respectively. Assuming $\lambda = 3\mu$, the reordered directions in $(Z^{\ell}_{\pm})_{\pi}$ and the reordered normalized directions in D^{ℓ}_{π} are partitioned into three groups of size μ, corresponding to the μ smallest, intermediate, and largest noisy function values, respectively.

2.3 Recombination Phase

This phase aims to guide the search toward reliable regions of the solution space by forming a robust direction. Such a heuristic search direction is constructed from information aggregated from the ranked groups in the sorting–selection phase. The direction p^{ℓ}_{trec} (Subsection 2.3.1) is next used to update the current iterate x^{ℓ}_{trec} (Subsection 2.3.2).

2.3.1 Triangular Recombination Direction

To improve the reliability of the search, let us consider weight parameters as below

$$w_1 \geq \ldots \geq w_{\mu} \geq \ldots \geq w_{2\mu} \geq \ldots \geq w_{\lambda} > 0, \qquad \text{with} \qquad \lambda = 3\mu. \tag{7}$$

Letting $\mathcal{G}_1 := [\mu]$, $\mathcal{G}_2 := [\mu] + \mu$, and $\mathcal{G}_3 := [\mu] + 2\mu$, and using three ranked groups partitioned from D^{ℓ}_{π}, we compute recombination directions d^{ℓ}_{rec1}, d^{ℓ}_{rec2}, and d^{ℓ}_{rec3} as

$$d^{\ell}_{\text{rec}\,k} = \sum_{j \in \mathcal{G}_k} \bar{w}_j d^{j\ell}_{\pi}, \qquad \sum_{j \in \mathcal{G}_k} \bar{w}_j = 1, \qquad \bar{w}_j = \frac{w_j}{\sum_{i \in \mathcal{G}_k} w_i}, \qquad k \in [3]. \tag{8}$$

Note that normalizing group weights ensures that each recombination direction reflects its internal ranking without being influenced by weight magnitudes across groups.

The use of ranked convex combinations in (8) introduces a potential *within-group cancellation* mechanism. In particular, even if one sampled direction has a sufficiently negative inner product with the true gradient and is therefore well aligned with local descent, its contribution may be partially offset by the remaining directions in the same ranked group. In the subsequent probabilistic analysis, the sampled direction with the most negative gradient inner product is referred to as the *latent descent direction*; it is "latent" because it is defined through the unavailable true gradient and is not directly identified by the algorithm.

The subsequent analysis therefore imposes a within-group dominance property on the ranked group containing this latent descent direction. More precisely, for some $\varepsilon_{\bar{w}} \in (0, 1)$, the normalized weight assigned to its best-ranked element is required to carry at least $1 - \varepsilon_{\bar{w}}$ of the group weight, while the remaining directions together carry at most $\varepsilon_{\bar{w}}$. This condition is quantified in Lemma 1, below. Its role is to prevent the other directions in the same group from completely canceling the alignment inherited from the latent descent direction.

Lemma 1 *Let $\{u_j\}_{j \in \mathcal{G}_k}$ be unit vectors, and let $\{\bar{w}_j\}_{j \in \mathcal{G}_k}$ be nonnegative weights. Suppose that, for some $\varepsilon_{\bar{w}} \in (0, 1)$, there exists $j^{\star} \in \mathcal{G}_k$ such that*

$$\bar{w}_{j^{\star}} \;\geq\; 1 - \varepsilon_{\bar{w}}, \qquad \text{and} \qquad \sum_{j \in \mathcal{G}_k \setminus \{j^{\star}\}} \bar{w}_j \;\leq\; \varepsilon_{\bar{w}}. \tag{9}$$

Then, for any unit vector $q \in \mathbb{R}^n$ *and* $v := \sum_{j \in \mathcal{G}_k} \bar{w}_j u_j$*, the following lower bound holds:*

$$|q^T v| \;\geq\; (1 - \varepsilon_{\bar{w}})|q^T u_{j^\star}| - \varepsilon_{\bar{w}}. \tag{10}$$

In particular, if, for some $t > 0$*,* $\varepsilon_{\bar{w}} < t/(1+t)$ *and* $|q^T u_{j^\star}| \geq t$*, then*

$$|q^T v| \;\geq\; (1 - \varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}} \;>\; 0. \tag{11}$$

Proof By the definition of v, $q^T v = \bar{w}_{j^\star} q^T u_{j^\star} + \sum_{j \in \mathcal{G}_k \setminus \{j^\star\}} \bar{w}_j q^T u_j$. Applying the reverse triangle inequality gives

$$|q^T v| \geq \bar{w}_{j^\star}|q^T u_{j^\star}| - \sum_{j \in \mathcal{G}_k \setminus \{j^\star\}} \bar{w}_j |q^T u_j| \geq \bar{w}_{j^\star}|q^T u_{j^\star}| - \sum_{j \in \mathcal{G}_k \setminus \{j^\star\}} \bar{w}_j, \tag{12}$$

where the second inequality follows from $|q^T u_j| \leq \|q\| \, \|u_j\| = 1$. Using (9) in (12) yields

$$|q^T v| \geq (1 - \varepsilon_{\bar{w}})|q^T u_{j^\star}| - \varepsilon_{\bar{w}},$$

which proves (10). Finally, if $|q^T u_{j^\star}| \geq t$, then $|q^T v| \geq (1 - \varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}}$. Moreover, $(1 - \varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}} > 0$ if and only if $\varepsilon_{\bar{w}} < t/(1+t)$. Therefore (11) follows. □

Given (8), the corresponding recombination points are

$$x^\ell_{\mathrm{rec}\,k} = x^\ell_{\mathrm{trec}} + \alpha^\ell_{\mathrm{rec}\,k} d^\ell_{\mathrm{rec}\,k}, \qquad k \in [3], \tag{13}$$

where $\alpha^\ell_{\mathrm{rec}\,1}$, $\alpha^\ell_{\mathrm{rec}\,2}$, and $\alpha^\ell_{\mathrm{rec}\,3} \in [\alpha_{\min}, \alpha_{\max}]$ are recombination step-sizes obtained via (4), (5), and (6), where $q_\alpha := d^\ell_{\mathrm{rec}\,k}$ for all $k \in [3]$, respectively.

Now, we introduce a new heuristic direction inspired by [27]. As shown in Fig. 1, let $\triangle^\ell_1 := \triangle(v^\ell_1, v^\ell_2, v^\ell_3)$ be a triangle with the vertices $v^\ell_k := x^\ell_{\mathrm{rec}\,k}$ for $k \in [3]$. We also build $\triangle^\ell_2 := \triangle(v^\ell_1, v^\ell_{12}, v^\ell_{13})$ and $\triangle^\ell_3 := \triangle(v^\ell_2, v^\ell_{21}, v^\ell_{23})$ near v^ℓ_1 and v^ℓ_2, where

$$v^\ell_{12} = v^\ell_{21} := \frac{v^\ell_1 + v^\ell_2}{2}, \qquad v^\ell_{13} := \frac{v^\ell_1 + v^\ell_3}{2}, \qquad \text{and} \qquad v^\ell_{23} := \frac{v^\ell_2 + v^\ell_3}{2}. \tag{14}$$

Now, given $a^\ell, \bar{a}^\ell \sim \mathcal{U}(0,1)$, and $b^\ell := \sqrt{1 - (a^\ell)^2}$, $\bar{b}^\ell := \sqrt{1 - (\bar{a}^\ell)^2}$, we generate

$$p^\ell_1 = a^\ell v^\ell_{12} + b^\ell v^\ell_{13}, \qquad p^\ell_2 = \bar{a}^\ell v^\ell_{21} + \bar{b}^\ell v^\ell_{23}. \tag{15}$$

Using (14) and (15), we introduce

$$\widehat{p}^\ell_{\mathrm{trec}} = \frac{1}{w^\ell_p}\big(\eta p^\ell_1 + (1-\eta) p^\ell_2\big) - x^\ell_{\mathrm{trec}}, \tag{16}$$

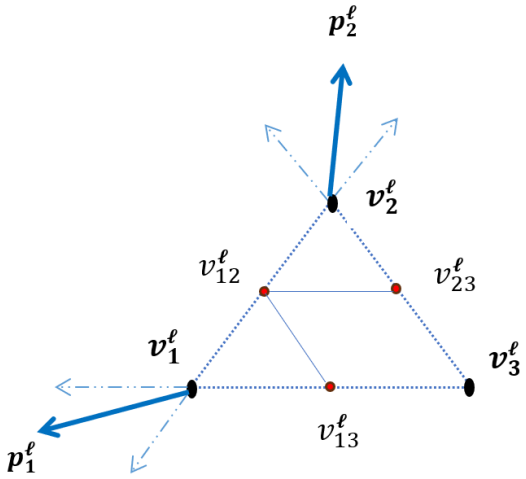

Fig. 1 Generating two components for the triangular recombination

where

$$w_p^\ell := \eta(a^\ell + b^\ell) + (1-\eta)(\bar{a}^\ell + \bar{b}^\ell), \qquad \text{and} \qquad \eta \in (0,1). \tag{17}$$

Using (15), (16), and (17), by some manipulation, this direction gets the final form as

$$\widehat{p}_{\text{trec}}^\ell = \frac{\sum_{k=1}^3 \theta_k^\ell \, \alpha_{\text{rec}\,k}^\ell \, d_{\text{rec}\,k}^\ell}{\sum_{k=1}^3 \theta_k^\ell}, \tag{18}$$

where

$$\begin{aligned} \theta_1^\ell &:= \frac{1}{2}\big(\eta(a^\ell + b^\ell) + (1-\eta)\bar{a}^\ell\big), \\ \theta_2^\ell &:= \frac{1}{2}\big(\eta a^\ell + (1-\eta)(\bar{a}^\ell + \bar{b}^\ell)\big), \qquad \theta_3^\ell := \frac{1}{2}\big(\eta b^\ell + (1-\eta)\bar{b}^\ell\big), \end{aligned} \tag{19}$$

and

$$\sum_{k=1}^3 \theta_k^\ell = w_p^\ell. \tag{20}$$

We term $\widehat{p}_{\text{trec}}^\ell$ as the ***triangular recombination direction***.

The representation (18) reveals a second possible cancellation mechanism, occurring *across ranked groups*. Indeed, even when the group containing a well-aligned latent descent direction preserves sufficient alignment through the within-group recombination in (8), its contribution to (18) may still be offset by the other two group-recombination directions. Positivity of the coefficients θ_k^ℓ alone does not exclude such cancellation.

Accordingly, the subsequent angle-transfer analysis treats the two mechanisms separately. The within-group dominance property controls cancellation inside the ranked

group containing the latent descent direction, whereas a quantitative cross-group margin property controls cancellation among the three terms in (18). The latter condition is expressed in terms of the effective coefficients $\theta_k^\ell \alpha_{\mathrm{rec}\,k}^\ell$ and ensures that, after the within-group alignment has been established, the contribution of the relevant group is not canceled by the remaining group-recombination contributions. This distinction is used explicitly in Proposition 4, below.

Since the magnitude of $\widehat{p}_{\mathrm{trec}}^\ell$ may lead to overly large or small displacements in noisy settings, it is controlled by some $\delta > 0$. Hence,

$$p_{\mathrm{trec}}^\ell = \delta \frac{\widehat{p}_{\mathrm{trec}}^\ell}{\|\widehat{p}_{\mathrm{trec}}^\ell\|}, \qquad \widehat{p}_{\mathrm{trec}}^\ell \neq 0. \tag{21}$$

The fixed-norm scaling in (21) controls the magnitude of the final search direction while preserving its angular relation with the gradient. Consequently, at every iteration for which the triangular recombination direction is nonzero, its norm is fixed at δ. This separation between the geometric construction in (18) and the magnitude control in (21) is used later in the probabilistic angle-transfer and complexity analyses.

Note that the subtraction of x_{trec}^ℓ in (16) is mathematically necessary to convert the trial position $\frac{1}{w_p^\ell}\big(\eta p_1^\ell + (1-\eta)p_2^\ell\big)$ into a proper search direction, as it is intrinsically tied to x_{trec}^ℓ. Removing this positional bias gives the pure direction $\widehat{p}_{\mathrm{trec}}^\ell$.

2.3.2 Triangular Recombination Step-Size and Point

Given the current iterate x_{trec}^ℓ and the constructed direction p_{trec}^ℓ, let us introduce the ***noisy derivative-free sufficient descent condition*** (`NDF-SDC`) as

$$\tilde{\mu}_s(\alpha) > \tilde{\beta}, \qquad \text{for some} \quad \tilde{\beta} \in (0,1), \tag{22}$$

for some $s \in \{-1, 1\}$, where

$$\tilde{\mu}_s(\alpha) := \frac{\tilde{f}(x_{\mathrm{trec}}^\ell + s\alpha p_{\mathrm{trec}}^\ell) - \tilde{f}(x_{\mathrm{trec}}^\ell)}{-\alpha^2}, \qquad \alpha > 0. \tag{23}$$

Using the `NDF-SDC` and an extrapolation mechanism, our method computes a step-size $\alpha_{\mathrm{trec}}^\ell > 0$ to update x_{trec}^ℓ. To describe the mechanism, we introduce an expansion factor $\gamma_e > 1$ and stress on (23) where both trial points $x_{\mathrm{trec}}^\ell \pm s\alpha p_{\mathrm{trec}}^\ell$ generated along the signed directions $\pm p_{\mathrm{trec}}^\ell$ are tested through the same criterion (22). Once the initial step-size $\alpha := \alpha^{\ell 0}$ satisfies (22), $\alpha^{\ell 0}$ is successively expanded by the factor γ_e along the same signed direction while the condition remains satisfied, i.e., $\alpha := \alpha^{\ell t} = \gamma_e \alpha^{\ell(t-1)}$

for $t \in [T]$ with a fixed budget T of iterations. The initial step-size $\alpha^{\ell 0} \in [\alpha_{\min}, \alpha_{\max}]$ is computed by the procedure in Subsection 2.1.3 via (4), (5) and (6) where $q_\alpha := p^\ell_{\text{trec}}$.

Let $\alpha^{\ell \bar{t}}$ denote the last successful step-size generated by the extrapolation mechanism. The extrapolation phase terminates either when the next expanded trial fails the `NDF-SDC` or when the prescribed budget T is exhausted. Once $\alpha^{\ell \bar{t}}$ is obtained along either $+p^\ell_{\text{trec}}$ or $-p^\ell_{\text{trec}}$, `DAES` updates the current iterate accordingly. If no successful initial step-size is found in one direction, the opposite direction is tested; if both signed initial trials fail, the iterate remains unchanged. Using $\alpha^{\ell \bar{t}}$, the ***triangular recombination step-size*** is defined as $\alpha^\ell_{\text{trec}} := \alpha^{\ell \bar{t}}$ and the new ***triangular recombination point*** is accepted as

$$x^{\ell+1}_{\text{trec}} = x^\ell_{\text{trec}} \pm \alpha^\ell_{\text{trec}} p^\ell_{\text{trec}}, \qquad x^0_{\text{trec}} = x^0. \tag{24}$$

2.4 The `DAES` Algorithm

Algorithm 1 outlines pseudocode of a single iteration of `DAES`. Each iteration starts by sampling symmetric mutation directions from a standard normal distribution to generate mutation points. These are grouped into three groups based on function values and sorted. Next, a recombination direction is built for each group and combined via a heuristic strategy into a single triangular direction. Then, a step-size is selected based on the `NDF-SDC` condition (22), with at most T extrapolation trials along a successful signed direction. The algorithm `DAES` distinguishes between successful and unsuccessful iterations. A successful iteration is one in which either p^ℓ_{trec} or $-p^\ell_{\text{trec}}$ satisfies the `NDF-SDC` and the iterate is updated. An unsuccessful iteration is one in which both signed initial trial directions fail the `NDF-SDC`, so the triangular recombination direction is rejected. The process repeats until a termination criterion is met.

Lines 30–34 in `DAES` introduce an additional control mechanism for unsuccessful iterations. This mechanism is required to ensure that unsuccessful iterations are properly regulated, thereby enabling us to derive complexity bounds on the total number of iterations, rather than only on the successful ones.

To make the control mechanism precise, unsuccessful recombination iterations contract the scaling step-size by the fixed factor ρ_u, while safeguarding it below by $\alpha_{\min}$; i.e.,

$$\sigma^{\ell+1} := \max\{\alpha_{\min}, \rho_u \sigma^\ell\}, \qquad \rho_u \in (0, 1), \tag{25}$$

for $\ell \in \mathcal{U}_T$, where $\mathcal{U}_T$ denotes the set of unsuccessful iterations. This update prevents the scaling parameter from remaining unchanged above the lower safeguard and provides an explicit mechanism for reducing the trial scale after an unsuccessful recombination iteration.

Algorithm 1 DAES

Tuning parameters: λ, $\ell_{\max} \in \mathbb{N}$, $\tilde{\beta} \in (0,1)$, $\gamma_e > 1$, $\delta > 0$, the scaling parameters in (6), $\rho_u \in (0,1)$, $T \in \mathbb{N}$, and $0 < \alpha_{\min} < \alpha_{\max} < \infty$;
Input: x^{ℓ}_{trec}, $\tilde{f}^{\ell}_{\mathrm{trec}}$, and σ^{ℓ};

---

for $i \in [\lambda/2]$ **do**
  sample $z^{i\ell} \sim \mathcal{N}(0, I_n)$ and store in Z^{ℓ};
  obtain $\alpha^{i\ell} \in [\alpha_{\min}, \alpha_{\max}]$ via (4) and (5) with $q_\alpha := z^{i\ell}$;
  generate $y^{i\ell}_{\pm}$ via (3), evaluate $\tilde{f}(y^{i\ell}_{\pm})$ and store in $F^{\ell}_{\pm}$;
end for
construct $Z^{\ell}_{\pm} = [Z^{\ell} \ -Z^{\ell}]$ and $D^{\ell} = [U^{\ell} \ -U^{\ell}]$, form $\tilde{F}^{\ell} = [F^{\ell}_{+} \ F^{\ell}_{-}]$, sort $\tilde{F}^{\ell}$, and apply the same permutation to $Z^{\ell}_{\pm}$ and D^{ℓ} to obtain $(Z^{\ell}_{\pm})_{\pi}$ and D^{ℓ}_{π};
compute $d^{\ell}_{\mathrm{rec}\,k}$ via (8) and $x^{\ell}_{\mathrm{rec}\,k}$ via (13) for $k \in [3]$;
obtain $\alpha^{\ell}_{\mathrm{rec}\,k} \in [\alpha_{\min}, \alpha_{\max}]$ via (4) and (5) with $q_\alpha := d^{\ell}_{\mathrm{rec}\,k}$ for $k \in [3]$;
construct $\widehat{p}^{\ell}_{\mathrm{trec}}$ via (16), and scale it via (21), resulting in p^{ℓ}_{trec};
obtain initial step-size $\alpha^{\ell 0} \in [\alpha_{\min}, \alpha_{\max}]$ via (4) and (5) with $q_\alpha := p^{\ell}_{\mathrm{trec}}$;
set $\mathsf{succ}_{\ell} := 0$;
for $s \in \{+1, -1\}$ **do** ▷ Search along signed directions
  if $\tilde{\mu}_s(\alpha^{\ell 0}) > \tilde{\beta}$ **then**
    set $\alpha^{\ell}_{\mathrm{trec}} := \alpha^{\ell 0}$ and $\mathsf{succ}_{\ell} := 1$; **break**;
  end if
end for
if succ_{ℓ} **then** ▷ Extrapolation along selected direction
  for $t \in [T]$ **do**
    set $\alpha_{\mathrm{temp}} := \alpha^{\ell}_{\mathrm{trec}}$ and $\alpha^{\ell}_{\mathrm{trec}} := \gamma_e \alpha^{\ell}_{\mathrm{trec}}$; compute $\tilde{f}(x^{\ell}_{\mathrm{trec}} + s\alpha^{\ell}_{\mathrm{trec}} p^{\ell}_{\mathrm{trec}})$;
    if $\tilde{\mu}_s(\alpha^{\ell}_{\mathrm{trec}}) \le \tilde{\beta}$ **then**
      set $\alpha^{\ell}_{\mathrm{trec}} := \alpha_{\mathrm{temp}}$;
      break;
    end if
  end for
  set $x^{\ell+1}_{\mathrm{trec}} := x^{\ell}_{\mathrm{trec}} + s\alpha^{\ell}_{\mathrm{trec}} p^{\ell}_{\mathrm{trec}}$ and $\tilde{f}^{\ell+1}_{\mathrm{trec}} := \tilde{f}(x^{\ell+1}_{\mathrm{trec}})$;
else
  set $x^{\ell+1}_{\mathrm{trec}} := x^{\ell}_{\mathrm{trec}}$ and $\tilde{f}^{\ell+1}_{\mathrm{trec}} := \tilde{f}^{\ell}_{\mathrm{trec}}$;
end if
if succ_{ℓ} **then**
  update $\sigma^{\ell+1}$ via (6);
else
  reduce $\sigma^{\ell+1}$ via (25);
end if
Output: $x^{\ell+1}_{\mathrm{trec}}$, $\tilde{f}^{\ell+1}_{\mathrm{trec}}$, and $\sigma^{\ell+1}$;

3 Analytical Study of DAES

In this section, we establish high-probability complexity bounds for the proposed DAES algorithm for noisy DFO. To the best of our knowledge, these results constitute the first complexity analysis for a MAES-type method under noisy settings covering nonconvex, convex, and strongly convex objectives, while matching the complexity orders reported in [7, 8, 17, 30].

We first analyze an idealized deterministic setting, where the search directions are not subject to sampling randomness, to isolate the role of the angle condition in the convergence analysis. Since DFO has no access to the true gradient, only probabilistic analogues of the angle condition can be established. The deterministic analysis therefore serves solely as a theoretical foundation for the subsequent randomized analysis.

We begin by introducing a set of standard assumptions on compact level **S**et ($\mathbf{A_S}$), **L**ipschitz continuity of the gradient ($\mathbf{A_L}$), uniform **N**oise bound on the objective function ($\mathbf{A_N}$) which are commonly adopted in DFO studies, see, e.g., [7, 8, 10, 30, 33, 34], and another assumption by which all step-sizes in DAES satisfy a two-sided **B**ound ($\mathbf{A_B}$) which is also found in [7, 10, 33]. Under these assumptions, we first analyze the deterministic variant of DAES in Section 3.1, and then extend the analysis to its randomized variant in Section 3.2.

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

$$\mathcal{L}(x^0) := \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}, \tag{26}$$

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*

$$\|g(x) - g(y)\| \leq L \, \|x - y\|, \qquad x, y \in \mathbb{R}^n. \tag{27}$$

Letting $\underline{f} := f(\underline{x}) = \inf\{f(x) \mid x \in \mathcal{L}(x^0)\} > -\infty$, where $\underline{x} \in \mathcal{L}(x^0)$ is a global minimizer, these assumptions imply that $f(x)$ is bounded below by $\underline{f}$. Moreover, from (27),

$$|f(x + \hat{s}) - f(x) - \hat{s}^T g(x)| \leq \frac{L}{2}\|\hat{s}\|^2, \qquad \hat{s} \in \mathbb{R}^n. \tag{28}$$

Assumption 3 ($\mathbf{A_N}$). *For some noise level $0 < \omega < \infty$, the noisy function $\tilde{f}(x)$ satisfies*

$$|\tilde{f}(x) - f(x)| \leq \omega, \qquad x \in \mathbb{R}^n. \tag{29}$$

For solving a DFO problem such as (1) under Assumption ($\mathbf{A_N}$), we can only expect to find an ε_ω-*approximate stationary point* for some *limit accuracy* $\varepsilon_\omega > 0$. In other

words, in such a scenario, the *complexity bound* of an algorithm is defined as an upper bound on the number of function evaluations required to find a point satisfying the following condition, where x is called the ε_ω-approximate stationary point:

$$f(x) \leq \sup \left\{ f(y) \,\middle|\, y \in \mathcal{L}(x^0), \quad \|g(y)\| \leq \varepsilon_\omega \right\}. \tag{30}$$

Assumption 4 ($\mathbf{A_B}$). *All step-sizes generated within the three phases of* `DAES` *lie in a noise-calibrated range. In particular, for any step-size α produced by the algorithm, it holds*

$$\underline{\kappa}' \sqrt{\omega L^{-1}} \leq \alpha_{\min} \leq \alpha \leq \alpha_{\max} \leq \overline{\kappa}' \sqrt{\omega L^{-1}}, \tag{31}$$

where ω denotes the noise level, and $0 < \underline{\kappa}' \leq \overline{\kappa}'$ are tuning constants such that

$$\underline{\kappa}' > \sqrt{2L\tilde{\beta}^{-1}} > 0, \qquad \text{for some} \qquad \tilde{\beta} \in (0,1). \tag{32}$$

3.1 Complexity with Deterministic Directions

Let $x := x^\ell_{\text{trec}}$ and $p := p^\ell_{\text{trec}}$ for simplicity. We first favor a reproducible direction p, enabling a systematic exploration of the local geometry of $f(x)$ without additional sampling randomness. In DFO, a sufficient decrease along such a direction does not reliably reflect the norm of $g(x)$. Thus, an alignment (angle) condition is needed:

$$|g(x)^T p| \geq \Delta_{\text{a}} \|g(x)\| \, \|p\|, \qquad \text{for some} \quad \Delta_{\text{a}} > 0. \tag{33}$$

To quantify this geometric effect, we define

$$v(x,p) := \frac{\|g(x)\| \, \|p\|}{|g(x)^T p|}, \qquad |g(x)^T p| \neq 0. \tag{34}$$

Small values of $v(x,p)$ indicate an acute angle between p and $g(x)$. This quantity links directional decrease bounds (Proposition 1) to gradient-norm bounds (Theorem 2), and then to a posteriori accuracy guarantees under convexity (Theorem 3).

Proposition 1 (**Directional Derivative Bound**) *Under Assumptions* ($\mathbf{A_S}$), ($\mathbf{A_L}$), ($\mathbf{A_N}$), *and* ($\mathbf{A_B}$), *let $p \in \mathbb{R}^n$ be the search direction in* `DAES` *with $\|p\| = \delta$, and let $\alpha > 0$ be the corresponding step-size. Then*

$$|g(x)^T p| \leq \gamma_e \alpha \Big(\tilde{\beta} + \frac{L}{2} \delta^2 \Big) + \frac{2\omega}{\alpha}, \qquad \text{for some} \quad \tilde{\beta} \in (0,1), \quad \gamma_e > 1. \tag{35}$$

Proof We consider two possible cases for the ℓ-th iteration of the `DAES`.

Case 1. Let both trial points $x \pm \alpha p$ fail the `NDF-SDC`. In this case, we have $\tilde{\mu}_s(\alpha) \leq \tilde{\beta}$, i.e.,

$\tilde{f}(x+s\alpha p)-\tilde{f}(x)\geq-\tilde{\beta}\alpha^2$ for $s\in\{-1,1\}$. Using this and (28), (29), we obtain

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

Therefore

$$-s\,g(x)^Tp\leq\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}\leq\alpha\gamma_e\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}.$$

Applying the above inequality with both $s=+1$ and $s=-1$ yields (35).

Case 2. Let α denote the last successful step-size obtained along some $s\in\{-1,1\}$. By construction, $x+s\alpha p$ satisfies the `NDF-SDC`, whereas the next extrapolated trial point $x+s\gamma_e\alpha p$ does not; i.e., $\tilde{\mu}_s(\alpha)>\tilde{\beta}$ and $\tilde{\mu}_s(\gamma_e\alpha)\leq\tilde{\beta}$.

Case 2A. Let $\tilde{\mu}_1(\alpha)>\tilde{\beta}$, i.e., $\tilde{f}(x+\alpha p)-\tilde{f}(x)<-\tilde{\beta}\alpha^2$. By this and (28), (29), we have

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

Therefore

$$g(x)^Tp<-\tilde{\beta}\alpha+\frac{L}{2}\alpha\delta^2+\frac{2\omega}{\alpha}<\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}<\alpha\gamma_e\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}.\tag{36}$$

By construction, since $\tilde{\mu}_1(\gamma_e\alpha)\leq\tilde{\beta}$, i.e., $\tilde{f}(x+\gamma_e\alpha p)-\tilde{f}(x)\geq-\tilde{\beta}\gamma_e^2\alpha^2$, (28) and (29), we have

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

Therefore

$$-g(x)^Tp\leq\gamma_e\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\gamma_e\alpha}\leq\gamma_e\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}.\tag{37}$$

Combining one-sided bounds from (36) and (37) yields (35).

Case 2B. Let $\tilde{\mu}_{-1}(\alpha)>\tilde{\beta}$, i.e., $\tilde{f}(x-\alpha p)-\tilde{f}(x)<-\tilde{\beta}\alpha^2$. By this and (28), (29), we end up with

$$-g(x)^Tp<-\tilde{\beta}\alpha+\frac{L}{2}\alpha\delta^2+\frac{2\omega}{\alpha}<\gamma_e\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}.\tag{38}$$

By construction, since $\tilde{\mu}_{-1}(\gamma_e\alpha)\leq\tilde{\beta}$, i.e., $\tilde{f}(x-\gamma_e\alpha p)-\tilde{f}(x)\geq-\tilde{\beta}\gamma_e^2\alpha^2$, (28) and (29), we end up with

$$g(x)^Tp\leq\gamma_e\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\gamma_e\alpha}\leq\gamma_e\alpha\Big(\tilde{\beta}+\frac{L}{2}\delta^2\Big)+\frac{2\omega}{\alpha}.\tag{39}$$

Combining (38) and (39) gives (35), which completes the proof. □

Theorem 1 (Local bound on $\|g(x)\|$) *Under Assumptions* $(\mathbf{A_S})$, $(\mathbf{A_L})$, $(\mathbf{A_N})$, *and* $(\mathbf{A_B})$, *suppose that the conditions of Proposition 1 hold at all relevant iterations. Then*

$$\|g(x)\|\leq v(x,p)\,\Gamma(\alpha),\qquad\textit{where}\tag{40}$$

$$\Gamma(\alpha):=\frac{\alpha}{\delta}\Big(\gamma_e\big(\tilde{\beta}+\frac{L}{2}\delta^2\big)+\frac{2\omega}{\alpha^2}\Big).\tag{41}$$

Here, $v(x,p)$ is defined as in (34). Moreover, it holds $\|g(x)\|=\mathcal{O}\big(v(x,p)\sqrt{L\omega}\big)$. In addition, if the angle condition (33) holds, then

$$\|g(x)\|=\mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big).\tag{42}$$

Proof Using the geometric factor (34) and the upper bound in (35), we have

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

By $\delta := \|p\|$ and (41), the upper bound (40) is obtained. By $(\mathbf{A_B})$, $\alpha = \Theta\Big(\sqrt{\omega L^{-1}}\Big)$, which yields $\Gamma(\alpha) = \mathcal{O}(\sqrt{L\omega})$ and therefore $\|g(x)\| = \mathcal{O}\big(v(x,p)\sqrt{L\omega}\big)$. Invoking the angle condition (33), we have $v(x,p) \leq \Delta_{\mathrm{a}}^{-1}$, and the claim (42) follows. □

Theorem 2 (**Global bound on** $\|g(x)\|$) *Let ℓ_T denote the termination iteration of the* DAES *algorithm. Under Assumptions* $(\mathbf{A_S})$, $(\mathbf{A_L})$, $(\mathbf{A_N})$, *and* $(\mathbf{A_B})$, *suppose that the conditions of Proposition 1 hold for all iterations $\ell \in [\ell_T]_0$. Then*

$$\min_{0\leq\ell\leq\ell_T} \|g(x^\ell)\| \leq v_{\min} \max_{0\leq\ell\leq\ell_T} \Gamma(\alpha^\ell), \qquad \textit{where} \qquad v_{\min} := \min_{0\leq\ell\leq\ell_T} v(x^\ell, p^\ell), \tag{43}$$

with the ℓ-th geometric factor $v(x^\ell, p^\ell)$ defined in (34) and $\Gamma(\alpha^\ell)$ as in (41). Moreover,

$$\min_{0\leq\ell\leq\ell_T} \|g(x^\ell)\| = \mathcal{O}\Big(v_{\min}\sqrt{L\omega}\Big). \tag{44}$$

In particular, if the search direction p^ℓ satisfies the angle condition (33) for all $\ell \in [\ell_T]_0$, then

$$\min_{0\leq\ell\leq\ell_T} \|g(x^\ell)\| = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\sqrt{L\omega}\big). \tag{45}$$

Proof Let $v_{\min}$ in (43) be attained at iteration ℓ^*. Thus, $\|g(x^{\ell^*})\| \leq v_{\min}\Gamma(\alpha^{\ell^*})$, from (40). It follows immediately from the inequality below, which proves (43):

$$\min_{0\leq\ell\leq\ell_T} \|g(x^\ell)\| \leq v_{\min} \max_{0\leq\ell\leq\ell_T} \Gamma(\alpha^\ell). \tag{46}$$

By substituting $\alpha^\ell = \Theta\Big(\sqrt{\omega L^{-1}}\Big)$ from $(\mathbf{A_B})$ into $\Gamma(\alpha^\ell)$ in (41), since $\Gamma(\alpha^\ell) = \mathcal{O}\Big(\sqrt{L\omega}\Big)$ and thus $\max_{0\leq\ell\leq\ell_T}\Gamma(\alpha^\ell) = \mathcal{O}\Big(\sqrt{L\omega}\Big)$, the claim (44) is obtained. In particular, invoking (33) where we have $v_{\min} \leq \Delta_{\mathrm{a}}^{-1}$, the claim (45) follows. □

Theorem 3 (**Error Bounds**) *Under Assumptions* $(\mathbf{A_S})$, $(\mathbf{A_L})$, $(\mathbf{A_N})$, *and* $(\mathbf{A_B})$, *suppose that the conditions of Proposition 1 hold iterations. Let the minimum gradient norm (44) occur at iteration ℓ^{**} of the* DAES *algorithm. Then, the following guarantees hold with the limit accuracy*

$$\varepsilon_\omega := \mathcal{O}\big(\sqrt{L\omega}\big). \tag{47}$$

(i) If f is convex, then

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(v_{\min}\varepsilon_\omega\big). \tag{48}$$

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

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(v_{\min}^2\varepsilon_\omega^2/\mu_c\big), \qquad \textit{and} \qquad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}\big(v_{\min}\varepsilon_\omega/\mu_c\big). \tag{49}$$

In particular, if the deterministic angle condition (33) holds and

(iii) if f is convex, then

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\varepsilon_\omega\big). \tag{50}$$

(iv) if f is μ_c-strongly convex, then

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-2}\varepsilon_\omega^2/\mu_c\big), \qquad \textit{and} \qquad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}\big(\Delta_{\mathrm{a}}^{-1}\varepsilon_\omega/\mu_c\big). \tag{51}$$

Proof Let $\underline{x}$ be a global minimizer, i.e., $g(\underline{x}) = 0$, and $\{x^\ell\}_{\ell\in[\ell_T]_0}$ lies in a compact set $\mathcal{L}(x^0)$.

(i) By convexity of f, we have $f(x) - f(y) \leq g(x)^T(x-y)$. Applying this with $x = x^{\ell^{**}}$ and $y = \underline{x}$ and using the Cauchy–Schwarz inequality yields $f(x^{\ell^{**}}) - \underline{f} \leq \|x^{\ell^{**}} - \underline{x}\| \cdot \|g(x^{\ell^{**}})\|$. Here, $\|x^{\ell^{**}} - \underline{x}\|$ is bounded as $\mathcal{L}(x^0)$ is compact: thus, $\|x^{\ell^{**}} - \underline{x}\| \leq D$ for $D > 0$. Invoking (44), we obtain $\|g(x^{\ell^{**}})\| = \mathcal{O}(v_{\min}\varepsilon_\omega)$ which completes the proof of (48).

(ii) If f is μ_c-strongly convex, then $f(x) - \underline{f} \leq 1/2{\mu_c}^{-1}\|g(x)\|^2$, based on the Polyak–Łojasiewicz inequality. Substituting $x = x^{\ell^{**}}$ into this inequality, applying strong convexity which implies $\|x^{\ell^{**}} - \underline{x}\| \leq {\mu_c}^{-1}\|g(x^{\ell^{**}})\|$, and using (44), (47), the claim (49) follows.

In particular, if (33) holds where $v_{\min} \leq \Delta_{\mathrm{a}}^{-1}$, the results in (iii)–(iv) follow directly. □

We are now in a position to derive the overall complexity bounds for the deterministic `DAES` algorithm. To this end, we first present a practical lemma and a proposition to establish a sufficient decrease criterion.

Lemma 2 (Consistency of `NDF-SDC`) *Under Assumptions* $(\mathbf{A_S})$, $(\mathbf{A_L})$, $(\mathbf{A_N})$, *and* $(\mathbf{A_B})$, *let $p \in \mathbb{R}^n$ be the search direction in* `DAES` *and $\alpha > 0$ be the corresponding step-size which satisfies the noisy derivative-free sufficient descent condition* (22) *for some $s \in \{-1, 1\}$. Then, the corresponding noiseless sufficient decrease condition is guaranteed to hold, i.e.,*

$$\mu_s(\alpha) \geq \beta, \qquad \textit{for some} \quad \beta \in (0,1), \tag{52}$$

where

$$\mu_s(\alpha) := \frac{f(x + s\alpha p) - f(x)}{-\alpha^2}, \qquad \alpha > 0. \tag{53}$$

Proof By (23) and (53), we have

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

Using the noise bound (29), it follows that

$$|\tilde{\mu}_s(\alpha) - \mu_s(\alpha)| \leq \frac{2\omega}{\alpha^2}. \tag{54}$$

Hence, $\mu_s(\alpha) \geq \tilde{\mu}_s(\alpha) - 2\omega/\alpha^2$. Since the NDF-SDC (22) holds, we have $\tilde{\mu}_s(\alpha) > \tilde{\beta}$. Therefore, $\mu_s(\alpha) > \tilde{\beta} - 2\omega/\alpha^2$. Using (31), i.e., $\alpha \geq \underline{\kappa}'\sqrt{\omega L^{-1}}$, yields $\mu_s(\alpha) > \tilde{\beta} - 2L(\underline{\kappa}')^{-2}$. By (32) and letting $\beta := \tilde{\beta} - 2L(\underline{\kappa}')^{-2}$, we obtain $\beta > 0$ which completes the proof. □

Proposition 2 (**Efficiency Condition**) *Under Assumptions* ($\mathbf{A_S}$), ($\mathbf{A_L}$), ($\mathbf{A_N}$), *and* ($\mathbf{A_B}$), *suppose that the conditions of Lemma 2 hold and that the angle condition* (33) *holds. Then, with* $\delta := \|p\|$, *the following efficiency estimate holds:*

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

Proof Using the bound $|g(x)^T p|$ in (35) and the angle condition (33), we have

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

Reordering this inequality and the definition of K_1 in (55) yields

$$\|g(x)\| \leq K_1\alpha + K_2\frac{\omega}{\alpha}, \qquad \text{where} \qquad K_2 := 2(\Delta_a\delta)^{-1}. \tag{56}$$

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

$$K_1\alpha^2 - \|g(x)\|\alpha + K_2\omega = 0.$$

The roots are

$$\alpha_\pm = \left(\|g(x)\| \pm \sqrt{\Delta}\right)/2K_1, \qquad \text{where} \qquad \Delta = \|g(x)\|^2 - 4K_1K_2\omega. \tag{57}$$

By $\alpha \geq \underline{\kappa}'\sqrt{\omega L^{-1}} > \sqrt{2\tilde{\beta}^{-1}\omega}$ from (31) and (32), we consider the following cases for Δ.

If $\Delta > 0$, then

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

and the quadratic inequality (56) holds only if $\alpha \leq \alpha_-$ or $\alpha \geq \alpha_+$. Rewriting the smaller root α_- and substituting the definition of K_1 yields

$$\alpha_- = \frac{4K_1K_2\omega}{2K_1\left(\|g(x)\| + \sqrt{\|g(x)\|^2 - 4K_1K_2\omega}\right)} < \sqrt{K_2{K_1}^{-1}\omega} \leq \sqrt{2\tilde{\beta}^{-1}\omega}. \tag{58}$$

From (58), we obtain $\alpha > \alpha_-$. Therefore, the branch $\alpha \leq \alpha_-$ is impossible, and when $\Delta > 0$, (56) can hold only on the second feasible branch $\alpha \geq \alpha_+$; i.e.,

$$\alpha \geq \alpha_+ = \frac{\|g(x)\| + \sqrt{\|g(x)\|^2 - 4K_1K_2\omega}}{2K_1} \geq \frac{\|g(x)\|}{2K_1}.$$

If $\Delta \leq 0$, then $\|g(x)\| \leq 2\sqrt{K_1K_2\omega}$, which, after substituting the definition of K_1, yields

$$\frac{\|g(x)\|}{2K_1} \leq \sqrt{K_2{K_1}^{-1}\omega} \leq \sqrt{2\tilde{\beta}^{-1}\omega} < \alpha.$$

Hence, in both cases,

$$\alpha^2 \geq \frac{1}{(2K_1)^2}\|g(x)\|^2. \tag{59}$$

Since the accepted signed trial point satisfies the `NDF-SDC`, Lemma 2 implies that $\mu_s(\alpha) \geq \beta$. By the definition of $\mu_s(\alpha)$, this is equivalent to $f(x) - f(x+s\alpha p) \geq \beta\alpha^2$. Using (59), we obtain

$$f(x) - f(x + s\alpha p) \geq \beta\alpha^2 \geq \frac{\beta}{4K_1^2}\|g(x)\|^2,$$

which proves (55). □

Theorem 4 (**Complexity Bounds**) *Under Assumptions* ($\mathbf{A_S}$), ($\mathbf{A_L}$), ($\mathbf{A_N}$), *and* ($\mathbf{A_B}$), *suppose that the conditions of Proposition 1 hold. Let* ℓ_T *denote the total number of iterations required to first attain the prescribed noise-limited stationarity level. Then, to achieve the desired accuracy, where* ε_ω *is from* (47), ℓ_{T_s} *is bounded as follows:*

(i) $\ell_{T_s} = \mathcal{O}\left(\varepsilon_\omega^{-2}\right)$ *for* ***nonconvex*** *functions;*

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

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

Moreover, the same orders hold for the total number of iterations and function evaluations.

Proof Let $f^\ell := f(x^\ell)$, $g^\ell := g(x^\ell)$, $\eta := \beta/(2K_1)^2$ and $\|g^\ell\| > \varepsilon_\omega$, see (30). Let $\mathcal{S}_T$ and $\mathcal{U}_T$ denote, respectively, the sets of successful and unsuccessful recombination iterations before termination. Hence, $|\mathcal{S}_T| = \ell_{T_s}$ and $\ell_T = |\mathcal{S}_T| + |\mathcal{U}_T|$ which count the successful and total number of recombination iterations, respectively.

(i) The condition (55) holds for all $\ell \in \mathcal{S}_T$. Therefore, $f^\ell - f^{\ell+1} \geq \eta\varepsilon_\omega^2$ for every $\ell \in \mathcal{S}_T$. Summing over $\ell < \ell_T$ yields summing over the successful steps as follows

$$f^0 - f^{\ell_T} = \sum_{\ell=0}^{\ell_T-1}(f^\ell - f^{\ell+1}) = \sum_{\ell\in\mathcal{S}_T}(f^\ell - f^{\ell+1}) \geq \eta\varepsilon_\omega^2\ell_{T_s}.$$

Since f is bounded below, i.e., $f^{\ell_T} \geq \underline{f} > -\infty$, we obtain $\ell_{T_s} \leq \big(f^0 - \underline{f}\big)\big(\eta\varepsilon_\omega^2\big)^{-1}$. Therefore, the number of successful iterations is $\ell_{T_s} = \mathcal{O}(\varepsilon_\omega^{-2})$. Note that for an unsuccessful step at iteration ℓ, we have $f^\ell = f^{\ell+1}$.

For parts (ii) and (iii), let $\{x^j\}_{j=0}^{\ell_{T_s}}$ denote the subsequence indexed by successful recombination iterations, and define $\psi_j := f(x^j) - \underline{f}$.

(ii) Let $d_{\max} := \sup_{x\in\mathcal{L}(x^0)}\|x - \underline{x}\|$, where $\underline{x}$ minimizes f with the value of $\underline{f}$. Convexity and the Cauchy-Schwarz inequality yield

$$f^\ell - \underline{f} \leq (x^\ell - \underline{x})^T g^\ell \leq \|x^\ell - \underline{x}\|\|g^\ell\| \leq d_{\max}\|g^\ell\|.$$

Rewriting this inequality gives $\|g^\ell\| \geq (f^\ell - \underline{f})d_{\max}^{-1}$. Substituting this into (55) results in

$$f^{\ell+1} \leq f^\ell - \eta\|g^\ell\|^2 \leq f^\ell - \eta d_{\max}^{-2}(f^\ell - \underline{f})^2.$$

This inequality gets the form $\psi^{\ell+1} \leq \psi^\ell - \eta d_{\max}^{-2}(\psi^\ell)^2$, where $\psi^\ell := f^\ell - \underline{f}$. By standard arguments (see Lemma 2.1 in [35]), we obtain $\psi^\ell \leq (\eta d_{\max}^{-2}\ell + (\psi^0)^{-1})^{-1}$. Thus, the error ψ^ℓ decreases at a sublinear rate of $\mathcal{O}(1/\ell)$. To achieve $\psi^\ell \leq \varepsilon_\omega$, it requires $\ell_{T_s} = \mathcal{O}(\varepsilon_\omega^{-1})$.

(iii) By the μ_c-strong convexity of f, we have $f^\ell - \underline{f} \leq (1/2\mu_c)\|g^\ell\|^2$. Rewriting this inequality for $\|g^\ell\|^2$, substituting it into (55) and unrolling the recursion ends up with

$$\psi^\ell \leq (1 - 2\eta\mu_c)^\ell \psi^0, \qquad \text{where} \qquad \psi^\ell := f^\ell - \underline{f}.$$

Thus, $\psi^\ell = \mathcal{O}\left((1 - 2\eta\mu_c)^\ell\right)$. To achieve $\psi^\ell \leq \varepsilon_\omega^2/(2\mu_c)$, it suffices $\ell_{T_s} = \mathcal{O}\left(\log(\varepsilon_\omega^{-1})\right)$.

Now we prove that the total number of iterations is of the same order as the number of successful iterations. The key observation is that an unsuccessful recombination iteration already implies attainment of the noise-limited stationarity scale.

Consider an unsuccessful recombination iteration ℓ. Then both signed trial points fail the `NDF-SDC`. By Proposition 1,

$$|g(x^\ell)^T p_{\text{trec}}^\ell| \leq \gamma_e \alpha^\ell \left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha^\ell}, \qquad \|p_{\text{trec}}^\ell\| = \delta,$$

where $\delta > 0$ is fixed. Combining this estimate with the angle condition

$$|g(x^\ell)^T p_{\text{trec}}^\ell| \geq \Delta_{\text{a}}\|g(x^\ell)\|\,\delta$$

gives

$$\|g(x^\ell)\| \leq \frac{1}{\Delta_{\text{a}}\delta}\left[\gamma_e \alpha^\ell \left(\tilde{\beta} + \frac{L}{2}\delta^2\right) + \frac{2\omega}{\alpha^\ell}\right].$$

Using the step-size bounds in Assumption ($\mathbf{A_B}$),

$$\underline{\kappa}'\sqrt{\frac{\omega}{L}} \leq \alpha^\ell \leq \overline{\kappa}'\sqrt{\frac{\omega}{L}},$$

we obtain $\|g(x^\ell)\| \leq C_u\sqrt{L\omega}$, where

$$C_u := \frac{1}{\Delta_{\text{a}}\delta}\left[\gamma_e\overline{\kappa}'\left(\frac{\tilde{\beta}}{L} + \frac{\delta^2}{2}\right) + \frac{2}{\underline{\kappa}'}\right].$$

Hence, every unsuccessful recombination iteration satisfies $\|g(x^\ell)\| = \mathcal{O}\left(\Delta_{\text{a}}^{-1}\sqrt{L\omega}\right)$. Thus, an unsuccessful recombination iteration already attains the noise-limited stationarity scale. To make the counting argument precise, define

$$\ell_\star := \inf\left\{\ell \geq 0 : \|g(x^\ell)\| \leq C_u\sqrt{L\omega}\right\}.$$

If the initial iterate already satisfies this bound, the conclusion is immediate. Otherwise, for every $\ell < \ell_\star$, $\|g(x^\ell)\| > C_u\sqrt{L\omega}$. Consequently, no iteration $\ell < \ell_\star$ can be unsuccessful, since an unsuccessful recombination iteration would imply $\|g(x^\ell)\| \leq C_u\sqrt{L\omega}$, which contradicts

the definition of $\ell_\star$. Therefore, all iterations strictly preceding the first attainment of the noise-limited stationarity scale are successful. The attainment iteration itself may be unsuccessful, so the iteration prefix counted up to first attainment contains at most one unsuccessful iteration. Hence,

$$N_u \le 1, \qquad N_{\text{total}} := N_s + N_u \le N_s + 1.$$

It follows that the total number of iterations required to attain the noise-limited stationarity scale has the same asymptotic order as the number of successful iterations.

Combining this observation with (i)–(iii), we obtain the total iteration complexity bounds $N_{\text{total}} = \mathcal{O}(\varepsilon_\omega^{-2})$ for nonconvex functions, $N_{\text{total}} = \mathcal{O}(\varepsilon_\omega^{-1})$ for convex functions, and $N_{\text{total}} = \mathcal{O}\Big(\log(\varepsilon_\omega^{-1})\Big)$ for μ_c-strongly convex functions.

Each unsuccessful iteration uses two function evaluations along $\pm p_{\text{trec}}$. For a successful iteration, the signed initial search uses at most two function evaluations, while the extrapolation phase uses at most T additional function evaluations, where $T \in \mathbb{N}$ is the prescribed extrapolation budget. Indeed, after a signed trial point satisfies the `NDF-SDC`, the step-size is expanded geometrically by the factor $\gamma_e > 1$ along the selected signed direction, with at most T extrapolation trials. The extrapolation phase terminates earlier if an expanded trial fails the `NDF-SDC`; otherwise, it stops when the prescribed budget T is exhausted. Hence, including the λ mutation-point evaluations, each iteration uses at most $\lambda + 2 + T$ function evaluations. Therefore, for fixed λ and fixed T, the total function-evaluation complexity has the same asymptotic order as the total iteration complexity. □

Corollary 1 *Under Assumptions* $(\mathbf{A_S})$, $(\mathbf{A_L})$, $(\mathbf{A_N})$, *and* $(\mathbf{A_B})$, *suppose that the conditions of Theorems 2 and 4 hold. Then, the effective gradient accuracy under arbitrary directions satisfies* (44). *Moreover, under the angle condition* (33), *we obtain* (45); *i.e.,*

$$\min_{0\le \ell \le \ell_T} \|g(x^\ell)\| = \mathcal{O}\big(\Delta_{\text{a}}^{-1}\sqrt{L\omega}\big).$$

This specializes to a coordinate polling set, for which at least one coordinate direction satisfies the angle bound with $\Delta_{\text{a}} \ge n^{-1/2}$.

The complexity bounds established in Theorem 4 are consistent with the well-known complexity estimates for derivative-free and gradient-based optimization methods in [7, 8, 17, 30, 36].

3.2 Complexity with Randomized Directions

We extend the deterministic analysis of `DAES` to the randomized setting, where the search directions are constructed from Gaussian samples. We study the conditions under which the required angle property holds with high probability (Proposition 4), leading to bounds on the gradient norm (Theorem 5), convergence accuracy (Theorem 6), and finally the overall complexity (Theorem 7). For these purposes, we introduce two additional assumptions and distinguish two failure mechanisms. The

first is a *sampling failure*, occurring when none of the independent Gaussian samples satisfies a prescribed angle condition. The second is a *sorting–selection failure*, occurring when noisy ranking prevents the latent well-aligned direction from satisfying the separation property required for its correct within-group identification. We control these mechanisms separately. A prescribed global sampling-failure tolerance $\gamma_0 \in (0, 1/2)$ is used to calibrate the population size λ, in the same spirit as sample-size calibration in randomized DFO methods, while a second global tolerance $\gamma_{\mathrm{sel}} \in (0, 1/2)$ controls the accumulated sorting–selection failures. The two mechanisms are combined only after their individual bounds have been established.

3.2.1 Probabilistic Assumptions and Failure Events

We now introduce the additional probabilistic assumptions required for the randomized analysis of `DAES`. The analysis distinguishes two failure mechanisms. The first is a *sampling failure*, which occurs when none of the freshly generated Gaussian directions satisfies a prescribed angle condition with the current gradient. The second is a *sorting–selection failure*, which occurs when the noisy ranking does not provide the separation property required to identify the latent well-aligned direction as the best-ranked element within its ranked group.

These two mechanisms are controlled separately. The sampling failure is bounded through the Gaussian sampling model, antipodal symmetry, and a population-size calibration based on a spherical-cap probability estimate. The sorting–selection failure is controlled through an accumulated probability budget. The two mechanisms are combined only after the corresponding failure events have been defined and bounded.

Assumption 5 ($\mathbf{A_F}$). *Let $n \geq 2$, let $\gamma_0 \in (0, 1/2)$ be a prescribed global sampling-failure tolerance, and let $\ell_{\max} \in \mathbb{N}$ be a prescribed maximum number of outer iterations of* `DAES`*, fixed before the algorithm is run and before the Gaussian samples used in the randomized analysis are generated. Set $\bar{\ell}_T := \ell_{\max}$. The maximum-iteration safeguard is included in the termination criterion of* `DAES`*; hence, if ℓ_T denotes the realized termination iteration, then*

$$\ell_T \leq \bar{\ell}_T$$

holds by construction. In particular, $\bar{\ell}_T$ is a deterministic analysis horizon fixed independently of the realized Gaussian samples, noisy sorting outcomes, and termination iteration ℓ_T. Let

$$0 < \tau < \sqrt{\pi/2}, \qquad t := \frac{\tau}{\sqrt{n}}, \quad r_\tau := \tau\sqrt{2/\pi} \in (0, 1), \tag{60}$$

so that the prescribed angle threshold has the natural dimension-dependent scale $t = \Theta(n^{-1/2})$. The number λ of mutation points, taken as an even multiple of three, is chosen as

$$\lambda = 6 \left\lceil \frac{\log\big(\bar{\ell}_T/\gamma_0\big)}{3\log(1/r_\tau)} \right\rceil. \tag{61}$$

The calibration (61) differs from a lower bound based directly on the exact spherical density. It exploits the dimension-scaled threshold $t = \tau/\sqrt{n}$ together with a spherical-cap estimate and the antipodal sampling structure of `DAES`. Since $\bar{\ell}_T = \ell_{\max}$ is a prescribed deterministic maximum-iteration horizon fixed before the run, the population size λ is also fixed before the Gaussian samples are generated. In particular, for fixed τ, γ_0, and $\bar{\ell}_T$, $\lambda = \mathcal{O}\big(\log(\bar{\ell}_T/\gamma_0)\big)$ with no additional explicit dependence on n. For each iteration ℓ, let $\{z^{j\ell}\}_{j\in[\lambda/2]}$ be i.i.d. random vectors drawn from $\mathcal{N}(0, I_n)$ and define their normalized orientations by

$$u^{j\ell} := \frac{z^{j\ell}}{\|z^{j\ell}\|}, \qquad d^{j\ell} := u^{j\ell}, \qquad d^{j+\lambda/2,\ell} := -u^{j\ell}, \qquad j \in [\lambda/2]. \tag{62}$$

For notational convenience, extend the unit-direction notation to the antipodal half of the population by setting $u^{j+\lambda/2,\ell} := -u^{j\ell}$ for $j \in [\lambda/2]$. Thus, $d^{j\ell} = u^{j\ell}$, $\|d^{j\ell}\| = \|u^{j\ell}\| = 1$ for $j \in [\lambda]$, where only the first $\lambda/2$ directions are independently sampled and the remaining $\lambda/2$ directions are their antipodal counterparts. In particular, the symmetric construction does not create λ independent samples.

After sorting, let $u^{\pi(j)\ell}$ denote the corresponding reordered unit direction $u^{j\ell}$. Since the mutation directions are normalized, the sorting operation changes only their labels, and hence $u^{\pi(j)\ell} = d^{\pi(j)\ell}$ and $\|u^{\pi(j)\ell}\| = 1$ for $j \in [\lambda]$. Let $\mathcal{F}_{\ell-1}$ denote the σ-algebra generated by the complete history of the algorithm prior to sampling the fresh Gaussian directions at iteration ℓ. In particular, $\mathcal{F}_{\ell-1}$ contains the previous Gaussian samples, noisy function evaluations, auxiliary random variables, sorting outcomes, and algorithmic decisions generated through iteration $\ell - 1$. Consequently, x^ℓ and $g(x^\ell)$ are $\mathcal{F}_{\ell-1}$-measurable, whereas the fresh samples $\{z^{j\ell}\}_{j\in[\lambda/2]}$ are independent of $\mathcal{F}_{\ell-1}$.

For $g(x^\ell) \neq 0$, we define the sampling-failure event

$$E_{\text{samp},\ell} := \left\{ \max_{j\in[\lambda]} \frac{|g(x^\ell)^T d^{j\ell}|}{\|g(x^\ell)\|\,\|d^{j\ell}\|} < t \right\}. \tag{63}$$

Recalling from (60) that $t = \tau/\sqrt{n}$, and using the antipodal construction in (62), the event (63) is equivalently written as

$$E_{\text{samp},\ell} = \left\{ \max_{j\in[\lambda/2]} \left| \frac{g(x^\ell)^T u^{j\ell}}{\|g(x^\ell)\|} \right| < \frac{\tau}{\sqrt{n}} \right\},$$

which makes explicit that only $\lambda/2$ independent directions contribute to the sampling probability. When $g(x^\ell) = 0$, we set $E_{\text{samp},\ell} := \emptyset$, since stationarity has already been attained and no angle condition is then required. For notational convenience, if the algorithm terminates before $\bar{\ell}_T = \ell_{\max}$, we set $E_{\text{samp},\ell} := \emptyset$ for every $\ell > \ell_T$.

Proposition 3 establishes the conditional per-iteration sampling-failure bound

$$\mathbb{P}(E_{\mathrm{samp},\ell} \,|\, \mathcal{F}_{\ell-1}) \le \frac{\gamma_0}{\bar{\ell}_T}, \tag{64}$$

which, by the union bound over the deterministic horizon $\bar{\ell}_T$, yields the global sampling-success guarantee

$$\mathbb{P}\left(\bigcap_{\ell=1}^{\ell_T} E^c_{\mathrm{samp},\ell}\right) \ge 1 - \gamma_0. \tag{65}$$

The second failure mechanism concerns the noisy sorting–selection phase. On the sampling-success event $E^c_{\mathrm{samp},\ell}$, the symmetric direction set contains at least one direction whose absolute alignment with the gradient is at least $t = \tau/\sqrt{n}$. Since the antipodal counterpart of every sampled direction is also present, there exists a descent-oriented direction satisfying

$$g(x^\ell)^T d^{i\ell} \le -t\|g(x^\ell)\|.$$

We select the most descending direction through the following latent index.

Assumption 6 ($\mathbf{A_R}$). *On the sampling-success event $E^c_{\mathrm{samp},\ell}$, let*

$$i^\star_\ell \in \operatorname*{argmin}_{i\in[\lambda]} g(x^\ell)^T d^{i\ell}, \tag{66}$$

where a fixed deterministic tie-breaking rule is used whenever the minimizer is not unique. By the symmetric sampling construction and $E^c_{\mathrm{samp},\ell}$,

$$g(x^\ell)^T d^{i^\star_\ell} \le -t\|g(x^\ell)\|.$$

Let $r^\star_\ell = \pi(i^\star_\ell)$ denote the ranked position of this same latent direction after sorting, so that

$$d^{r^\star_\ell} = d^{i^\star_\ell}. \tag{67}$$

If equal noisy function values occur, the sorting permutation is defined using a fixed deterministic tie-breaking rule. Let $\mathcal{G}_{k^\star}$ be the unique ranked group containing $r^\star_\ell$, where $k^\star = k^\star_\ell$ may depend on the iteration. Define the sorting–selection failure event by

$$\begin{aligned} E_{\mathrm{sel},\ell} := E^c_{\mathrm{samp},\ell} \cap \Bigg\{ \exists r \in \mathcal{G}_{k^\star} \setminus \{r^\star_\ell\} : \alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r^\star_\ell} g(x^\ell)^T d^{r^\star_\ell} \\ \le \frac{L}{2}\left((\alpha^{r\ell})^2 + (\alpha^{r^\star_\ell})^2\right) + 2\omega \Bigg\}. \end{aligned} \tag{68}$$

The accumulated sorting–selection failure probability satisfies

$$\sum_{\ell=1}^{\bar{\ell}_T} \mathbb{P}(E_{\mathrm{sel},\ell}) \le \gamma_{\mathrm{sel}}, \qquad \gamma_{\mathrm{sel}} \in (0, 1/2). \tag{69}$$

Assumption ($\mathbf{A_R}$) is an accumulated global failure-budget assumption for the noisy sorting–selection mechanism. In contrast to the sampling bound (64), the estimate (69) is not derived solely from Assumptions ($\mathbf{A_L}$), ($\mathbf{A_N}$), and ($\mathbf{A_B}$); rather, it explicitly controls the probability that noisy ranking fails to preserve the separation property required in the subsequent within-group identification argument.

To clarify the role of (68), its complement on the sampling-success event implies that, for every $r \in \mathcal{G}_{k^\star} \setminus \{r_\ell^\star\}$,

$$\alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} > \frac{L}{2}\left((\alpha^{r\ell})^2 + (\alpha^{r_\ell^\star})^2\right) + 2\omega.$$

Thus, the first-order advantage of the latent descent direction dominates the worst-case curvature and bounded-noise perturbations appearing in the comparison of the corresponding noisy mutation values.

For intuition, if the two compared directions use a common step-size $\alpha > 0$, the strict separation condition reduces to

$$g(x^\ell)^T d^{r\ell} - g(x^\ell)^T d^{r_\ell^\star} > L\alpha + \frac{2\omega}{\alpha}.$$

Hence, the separation threshold explicitly balances local curvature through $L\alpha$ and bounded evaluation noise through $2\omega/\alpha$. Assumption ($\mathbf{A_B}$) places the relevant step-sizes on the noise-calibrated scale $\alpha = \Theta(\sqrt{\omega L^{-1}})$, under which these two terms are of the same characteristic order.

We now combine the two failure mechanisms. Define the global good event

$$\mathcal{G}_T := \bigcap_{\ell=1}^{\ell_T} \left(E_{\mathrm{samp},\ell}^c \cap E_{\mathrm{sel},\ell}^c\right). \tag{70}$$

Since $\ell_T \le \bar{\ell}_T$, its complement satisfies the deterministic event containment

$$\mathcal{G}_T^c \subseteq \left(\bigcup_{\ell=1}^{\bar{\ell}_T} E_{\mathrm{samp},\ell}\right) \cup \left(\bigcup_{\ell=1}^{\bar{\ell}_T} E_{\mathrm{sel},\ell}\right).$$

Therefore, by Proposition 3, below, Assumption ($\mathbf{A_R}$), and the union bound,

$$\mathbb{P}(\mathcal{G}_T^c) \le \sum_{\ell=1}^{\bar{\ell}_T} \mathbb{P}(E_{\mathrm{samp},\ell}) + \sum_{\ell=1}^{\bar{\ell}_T} \mathbb{P}(E_{\mathrm{sel},\ell}) \le \gamma_0 + \gamma_{\mathrm{sel}}. \tag{71}$$

Consequently,

$$\mathbb{P}(\mathcal{G}_T) \ge 1 - (\gamma_0 + \gamma_{\mathrm{sel}}). \tag{72}$$

No independence between the sampling and sorting–selection failure events is required.

Remark 1 *The dimension dependence in the randomized angle mechanism is now carried by the natural threshold $t = \tau/\sqrt{n}$, rather than by the number of independently sampled Gaussian directions. For fixed τ, γ_0, and $\bar{\ell}_T$, the population calibration (61) introduces no additional explicit dependence on n. This distinction is important when converting iteration complexity into function-evaluation complexity.*

Remark 2 *The probability model used here should be distinguished from a stochastic noise model. Assumption ($\mathbf{A_N}$) imposes a uniform deterministic bound on the evaluation error and does not assign a probability distribution to that error. The high-probability statements above arise from the randomized Gaussian sampling mechanism together with the accumulated sorting–selection failure budget in Assumption ($\mathbf{A_R}$).*

Remark 3 *Assumption ($\mathbf{A_R}$) is stated as an accumulated failure budget because the subsequent randomized analysis proceeds by conditioning on the uniform global good event $\mathcal{G}_T$. This is distinct from analyses based on counting the frequency of favorable iterations through per-iteration conditional success probabilities. The present formulation is therefore tailored to the global-event transfer argument used below.*

Remark 4 *The deterministic analysis in Subsection 3.1 suppresses the sampling and sorting–selection failure mechanisms and provides the deterministic, pathwise component used after conditioning on $\mathcal{G}_T$. The randomized analysis developed below supplements that deterministic foundation by proving that the required directional properties hold jointly on $\mathcal{G}_T$ with the probability bound (72).*

3.2.2 Probabilistic Angle Conditions

In the probabilistic analysis of randomized DFO methods, a key challenge is to ensure that the sampled directions are sufficiently aligned with the true gradient. To this end, we first establish Proposition 3, which yields a uniform high-probability sampling-angle property over the prescribed deterministic horizon. We then analyze how this property is preserved through the noisy sorting, within-group recombination, and triangular recombination mechanisms of `DAES`, leading to the angle-transfer result in Proposition 4.

Unlike the analysis in [8, Proposition 3], which relies on uniformly sampled box directions, `DAES` generates Gaussian samples and uses their normalized orientations. By rotational invariance, these orientations are uniformly distributed on the unit sphere. Under Assumption ($\mathbf{A_F}$), the angle threshold is $t = \tau/\sqrt{n}$, $0 < \tau < \sqrt{\pi/2}$, and the antipodal sampling construction permits a population-size calibration with no additional explicit dependence on n.

Proposition 3 *Under Assumption* ($\mathbf{A_F}$), *let* $\{z^{j\ell}\}_{j\in[\lambda/2]}$ *be i.i.d. random vectors drawn from* $\mathcal{N}(0, I_n)$ *and let the symmetric normalized directions* $\{d^{j\ell}\}_{j\in[\lambda]}$ *be constructed as in* (62)*; i.e.,*

$$d^{j\ell} = \frac{z^{j\ell}}{\|z^{j\ell}\|}, \qquad d^{j+\lambda/2,\ell} = -d^{j\ell}, \qquad j \in [\lambda/2].$$

Let x^ℓ *be the current iterate of* `DAES`*, and let* $\mathcal{F}_{\ell-1}$ *denote the history of the algorithm prior to sampling the fresh Gaussian directions at iteration* ℓ*. Then*

$$\mathbb{P}\big(E_{\text{samp},\ell} \,\big|\, \mathcal{F}_{\ell-1}\big) \le \frac{\gamma_0}{\bar{\ell}_T}, \qquad \ell \in [\bar{\ell}_T]. \tag{73}$$

Consequently, with probability at least $1-\gamma_0$*, for every* $\ell \in [\ell_T]$ *such that* $g(x^\ell) \neq 0$*, there exists an index* $j \in [\lambda]$ *satisfying*

$$\frac{|g(x^\ell)^T d^{j\ell}|}{\|g(x^\ell)\|\,\|d^{j\ell}\|} \ge t = \frac{\tau}{\sqrt{n}}. \tag{74}$$

If $g(x^\ell) = 0$*, stationarity has already been attained and no angle condition is required.*

Proof Fix an iteration $\ell \in [\bar{\ell}_T]$ and condition on $\mathcal{F}_{\ell-1}$. By construction, x^ℓ and $g(x^\ell)$ are $\mathcal{F}_{\ell-1}$-measurable, whereas the Gaussian samples $\{z^{j\ell}\}_{j\in[\lambda/2]}$ are fresh and independent of $\mathcal{F}_{\ell-1}$. Hence, conditionally on $\mathcal{F}_{\ell-1}$, x^ℓ and $g(x^\ell)$ are fixed with respect to the new Gaussian samples.

If $g(x^\ell) = 0$, then $E_{\text{samp},\ell} = \emptyset$ by definition, and therefore $\mathbb{P}\big(E_{\text{samp},\ell} \,\big|\, \mathcal{F}_{\ell-1}\big) = 0$. It remains to consider the case $g(x^\ell) \neq 0$. Define $q^\ell := g(x^\ell)/\|g(x^\ell)\|$.

For each $j \in [\lambda/2]$, let $u^{j\ell} := z^{j\ell}/\|z^{j\ell}\|$. By rotational invariance of the standard Gaussian distribution, $\{u^{j\ell}\}_{j=1}^{\lambda/2}$ are conditionally i.i.d. and uniformly distributed on the unit sphere $\mathbb{S}^{n-1}$. Moreover, by (62), $d^{j\ell} = u^{j\ell}$ and $d^{j+\lambda/2,\ell} = -u^{j\ell}$.

For the dimension-scaled threshold $t = \tau/\sqrt{n}$, a spherical-cap estimate gives, for every fixed unit vector q^ℓ,

$$\mathbb{P}\left((q^\ell)^T u^{j\ell} \ge \frac{\tau}{\sqrt{n}} \,\middle|\, \mathcal{F}_{\ell-1}\right) \ge \frac{1}{2} - \frac{\tau}{\sqrt{2\pi}}. \tag{75}$$

This estimate is the single-direction spherical-cap bound established in the proof of [17, Lemma B.2, Eq. (52)].

By symmetry of the uniform distribution on $\mathbb{S}^{n-1}$,

$$\mathbb{P}\left((q^\ell)^T u^{j\ell} \le -\frac{\tau}{\sqrt{n}} \,\middle|\, \mathcal{F}_{\ell-1}\right) = \mathbb{P}\left((q^\ell)^T u^{j\ell} \ge \frac{\tau}{\sqrt{n}} \,\middle|\, \mathcal{F}_{\ell-1}\right).$$

Since the two events are disjoint for $\tau > 0$, it follows from (75) that

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

Recalling $r_\tau := \tau\sqrt{2/\pi} \in (0,1)$, we therefore obtain

$$\mathbb{P}\left(\left|(q^\ell)^T u^{j\ell}\right| < \frac{\tau}{\sqrt{n}} \,\middle|\, \mathcal{F}_{\ell-1}\right) \le r_\tau. \tag{76}$$

The antipodal counterparts do not provide additional independent samples. Indeed, for every $j \in [\lambda/2]$,

$$\frac{|g(x^\ell)^T d^{j\ell}|}{\|g(x^\ell)\|\,\|d^{j\ell}\|} = \frac{|g(x^\ell)^T d^{j+\lambda/2,\ell}|}{\|g(x^\ell)\|\,\|d^{j+\lambda/2,\ell}\|} = \left|(q^\ell)^T u^{j\ell}\right|.$$

Consequently, the sampling-failure event (63) can be written as

$$E_{\mathrm{samp},\ell} = \left\{ \max_{j\in[\lambda/2]} \left|(q^\ell)^T u^{j\ell}\right| < \frac{\tau}{\sqrt{n}} \right\}.$$

Using the conditional independence of the $\lambda/2$ fresh normalized Gaussian directions and (76), we obtain

$$\mathbb{P}\big(E_{\mathrm{samp},\ell} \,\big|\, \mathcal{F}_{\ell-1}\big) \le r_\tau^{\lambda/2}.$$

From (61), $\frac{\lambda}{2}\log(1/r_\tau) \ge \log\big(\bar{\ell}_T/\gamma_0\big)$ is obtained, so that

$$r_\tau^{\lambda/2} = \exp\left(-\frac{\lambda}{2}\log(1/r_\tau)\right) \le \exp\big(-\log\big(\bar{\ell}_T/\gamma_0\big)\big) = \gamma_0/\bar{\ell}_T.$$

Thus, $\mathbb{P}\big(E_{\mathrm{samp},\ell} \,\big|\, \mathcal{F}_{\ell-1}\big) \le \gamma_0/\bar{\ell}_T$, which proves (73) and is consistent with (64).

Taking expectations with respect to $\mathcal{F}_{\ell-1}$ and using the tower property gives

$$\mathbb{P}(E_{\mathrm{samp},\ell}) = \mathbb{E}\big[\mathbb{P}\big(E_{\mathrm{samp},\ell} \,\big|\, \mathcal{F}_{\ell-1}\big)\big] \le \frac{\gamma_0}{\bar{\ell}_T}.$$

Hence, by the union bound over the fixed deterministic horizon,

$$\mathbb{P}\left(\bigcup_{\ell=1}^{\bar{\ell}_T} E_{\mathrm{samp},\ell}\right) \le \sum_{\ell=1}^{\bar{\ell}_T} \mathbb{P}(E_{\mathrm{samp},\ell}) \le \gamma_0.$$

Since $\ell_T \le \bar{\ell}_T$, $\big\{\exists\, \ell \in [\ell_T] : E_{\mathrm{samp},\ell}\big\} \subseteq \bigcup_{\ell=1}^{\bar{\ell}_T} E_{\mathrm{samp},\ell}$. Therefore,

$$\mathbb{P}\left(\bigcap_{\ell=1}^{\ell_T} E^c_{\mathrm{samp},\ell}\right) \ge 1-\gamma_0,$$

which proves (65). Finally, by the definition of $E_{\mathrm{samp},\ell}$, on this global sampling-success event every nonstationary iteration $\ell \in [\ell_T]$ admits at least one $j \in [\lambda]$ satisfying (74). This completes the proof. □

Remark 5 *With the geometric factor $v(x,p)$ defined in (34), Proposition 3 implies that, with probability at least $1-\gamma_0$, for every nonstationary $\ell \in [\ell_T]$ there exists $j \in [\lambda]$ such that*

$$v(x^\ell, d^{j\ell}) = \frac{\|g(x^\ell)\|\,\|d^{j\ell}\|}{|g(x^\ell)^T d^{j\ell}|} \le \frac{1}{t} = \frac{\sqrt{n}}{\tau}.$$

Thus, the geometric scaling is of order $\sqrt{n}$, as is natural for random spherical directions at an angle threshold of order $n^{-1/2}$. The important distinction is that the required population size does not incur an additional explicit dimension factor. Indeed, for fixed τ, (61) gives $\lambda = \mathcal{O}\big(\log(\bar{\ell}_T/\gamma_0)\big)$. Hence, the dimension dependence is carried by the geometric threshold $t = \tau/\sqrt{n}$, rather than by an additional growth of the population size.

Remark 6 *The parameter τ controls the trade-off between alignment and population size. Since $t = \tau/\sqrt{n}$, larger values of τ yield a stronger angle threshold. At the same time, $r_\tau = \tau\sqrt{2/\pi}$ increases with τ, and therefore the population size required by (61) also increases. The restriction $0 < \tau < \sqrt{\pi/2}$ ensures that $r_\tau \in (0,1)$ and hence that the geometric failure probability $r_\tau^{\lambda/2}$ decays exponentially with the number of independently sampled antipodal pairs.*

Proposition 3 guarantees that, on the global sampling-success event (65), every nonstationary iteration admits at least one sampled direction whose absolute alignment with the gradient is at least $t = \tau/\sqrt{n}$. However, `DAES` does not directly use an individual sampled direction as its final search direction. Instead, the sampled directions are first ranked according to noisy objective values, aggregated within three ranked groups, and then combined into the triangular recombination direction p^ℓ_{trec}. Consequently, the sampled angle property does not automatically transfer to p^ℓ_{trec}.

To establish such a transfer, we separate two possible cancellation mechanisms discussed in Subsection 2.3.1. The first occurs *within a ranked group*, where the contribution of a well-aligned direction may be offset by the remaining directions in the same convex combination. The second occurs *across ranked groups*, where the contribution of the group containing the latent well-aligned direction may be offset by the other group-recombination directions. The analysis below controls these mechanisms separately through a within-group dominance property and a cross-group margin property.

We now make precise the role of the sampling angle event (74). Fix a nonstationary iteration ℓ and suppose that $E^c_{\mathrm{samp},\ell}$ occurs. Then (74) implies the existence of at least one $j \in [\lambda]$ such that

$$|g(x^\ell)^T d^{j\ell}| \geq t\|g(x^\ell)\|,$$

because $\|d^{j\ell}\| = 1$. Since the sampled direction set is antipodally symmetric, whenever $d^{j\ell}$ is present, so is $-d^{j\ell}$. Hence at least one member of the corresponding antipodal pair satisfies

$$g(x^\ell)^T d^{i\ell} \leq -t\|g(x^\ell)\|.$$

Recalling the latent index introduced in (66), we therefore obtain

$$i^\star_\ell \in \operatorname*{argmin}_{i\in[\lambda]} g(x^\ell)^T d^{i\ell}, \qquad \text{and} \qquad g(x^\ell)^T d^{i^\star_\ell} \leq -t\|g(x^\ell)\|. \tag{77}$$

The second inequality in (77) follows because the minimum over the complete symmetric direction set cannot exceed the inner product of the descent-oriented member of the well-aligned antipodal pair.

We refer to (77) as a *descent-angle inequality* for the latent direction. It should be distinguished from the `NDF-SDC` (22): the former is a geometric relation involving the unavailable true gradient, whereas the latter is an algorithmic acceptance condition expressed through noisy function values.

Sorting changes only the position of the latent direction, not the direction itself. By (67), $d^{r_\ell^\star} = d^{i_\ell^\star}$. Therefore, the descent-angle inequality is preserved after relabeling:

$$g(x^\ell)^T d^{r_\ell^\star} \le -t\|g(x^\ell)\|. \tag{78}$$

Let $\mathcal{G}_{k^\star}$, with $k^\star = k_\ell^\star$, denote the unique ranked group containing $r_\ell^\star$. We next show that, on the joint sampling and sorting–selection success event, the latent direction is the unique best-ranked element of this group (Lemma 3, below).

Lemma 3 *Suppose that Assumptions* $(\mathbf{A_L})$, $(\mathbf{A_N})$, $(\mathbf{A_B})$, *and* $(\mathbf{A_R})$ *hold. On the event*

$$E_{\mathrm{samp},\ell}^c \cap E_{\mathrm{sel},\ell}^c,$$

let $i_\ell^\star$ *denote the latent descent index defined in* (66), *let* $r_\ell^\star$ *be its ranked position after sorting, i.e.,* $d^{r_\ell^\star} = d^{i_\ell^\star}$, *and let* $\mathcal{G}_{k^\star}$, *with* $k^\star = k_\ell^\star$, *be the unique ranked group containing* $r_\ell^\star$. *Then, for every* $r \in \mathcal{G}_{k^\star} \setminus \{r_\ell^\star\}$,

$$\tilde{f}\left(x^\ell + \alpha^{r_\ell^\star} d^{r_\ell^\star}\right) < \tilde{f}\left(x^\ell + \alpha^{r\ell} d^{r\ell}\right). \tag{79}$$

Consequently, the latent descent direction is the unique best-ranked element within $\mathcal{G}_{k^\star}$.

Proof Fix $r \in \mathcal{G}_{k^\star} \setminus \{r_\ell^\star\}$. By Assumption $(\mathbf{A_B})$, $\alpha^{r\ell}, \alpha^{r_\ell^\star} \in [\alpha_{\min}, \alpha_{\max}]$. Define the corresponding mutation points by $y_r := x^\ell + \alpha^{r\ell} d^{r\ell}$ and $y_{r_\ell^\star} := x^\ell + \alpha^{r_\ell^\star} d^{r_\ell^\star}$. Since the gradient of f is L-Lipschitz continuous by Assumption $(\mathbf{A_L})$, the first-order expansions at x^ℓ can be written as

$$f(y_r) = f(x^\ell) + \alpha^{r\ell} g(x^\ell)^T d^{r\ell} + R_r,$$

and

$$f(y_{r_\ell^\star}) = f(x^\ell) + \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} + R_{r_\ell^\star},$$

where, using the unit normalization of the ranked directions,

$$|R_r| \le \frac{L}{2}(\alpha^{r\ell})^2, \qquad |R_{r_\ell^\star}| \le \frac{L}{2}(\alpha^{r_\ell^\star})^2, \qquad \text{with} \qquad \|d^{r\ell}\| = \|d^{r_\ell^\star}\| = 1. \tag{80}$$

Subtracting the two expansions gives

$$f(y_r) - f(y_{r_\ell^\star}) = \alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} + R_r - R_{r_\ell^\star}.$$

Using (80), we obtain

$$f(y_r) - f(y_{r_\ell^\star}) \geq \alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} - \frac{L}{2}\left((\alpha^{r\ell})^2 + (\alpha^{r_\ell^\star})^2\right). \tag{81}$$

By Assumption ($\mathbf{A_N}$), each noisy evaluation differs from its exact counterpart by at most ω. Hence

$$\tilde{f}(y_r) - \tilde{f}(y_{r_\ell^\star}) \geq f(y_r) - f(y_{r_\ell^\star}) - 2\omega. \tag{82}$$

Combining (81) and (82) yields

$$\tilde{f}(y_r) - \tilde{f}(y_{r_\ell^\star}) \geq \alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} - \frac{L}{2}\left((\alpha^{r\ell})^2 + (\alpha^{r_\ell^\star})^2\right) - 2\omega. \tag{83}$$

By (68), on $E^c_{\text{samp},\ell} \cap E^c_{\text{sel},\ell}$, no index $r \in \mathcal{G}_{k^\star} \setminus \{r_\ell^\star\}$ satisfies the separation violation. Consequently, for every such r,

$$\alpha^{r\ell} g(x^\ell)^T d^{r\ell} - \alpha^{r_\ell^\star} g(x^\ell)^T d^{r_\ell^\star} > \frac{L}{2}\left((\alpha^{r\ell})^2 + (\alpha^{r_\ell^\star})^2\right) + 2\omega.$$

Substituting this strict inequality into (83) gives $\tilde{f}(y_r) - \tilde{f}(y_{r_\ell^\star}) > 0$. Hence, $\tilde{f}(y_{r_\ell^\star}) < \tilde{f}(y_r)$ for every $r \in \mathcal{G}_{k^\star} \setminus \{r_\ell^\star\}$, which proves (79). Thus, on $E^c_{\text{samp},\ell} \cap E^c_{\text{sel},\ell}$, the latent descent direction is the unique minimizer of the noisy mutation values within $\mathcal{G}_{k^\star}$ and therefore the unique best-ranked element of that group. □

We now apply Lemma 1 to the ranked group $\mathcal{G}_{k^\star}$ containing the latent descent direction. On $E^c_{\text{samp},\ell} \cap E^c_{\text{sel},\ell}$, Lemma 3 implies that $r_\ell^\star$ is the unique best-ranked element of $\mathcal{G}_{k^\star}$. Hence, if (9) holds with $j^\star = r_\ell^\star$, then, using (8) and (78),

$$\left| \frac{g(x^\ell)^T}{\|g(x^\ell)\|} d^\ell_{\text{rec}\,k^\star} \right| \geq (1 - \varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}}.$$

In particular, if $\varepsilon_{\bar{w}} < t/(1+t)$, then the strict positivity conclusion (11) holds and the sampled alignment is preserved after within-group recombination.

Remark 7 *Since $t = \tau/\sqrt{n}$, the positivity requirement becomes $\varepsilon_{\bar{w}} < \tau/(\sqrt{n}+\tau)$. Thus, for a dimension-uniform asymptotic interpretation, the within-group dominance parameter must scale as $\mathcal{O}(n^{-1/2})$, with a sufficiently small coefficient to preserve strict positivity. This requirement is separate from the population-size calibration in (61), which introduces no additional explicit polynomial dependence on n.*

In addition to the probabilistic assumptions above, the transfer from the within-group recombination direction to the final triangular direction requires a deterministic cross-group margin condition. This condition is an analytical hypothesis used to control cancellation among the three ranked groups; it is not imposed or checked by the numerical implementation since in a genuine DFO setting, k^* depends on the latent descent direction defined through the unavailable true gradient.

Proposition 4 *Under Assumptions* $(\mathbf{A_L})$, $(\mathbf{A_N})$, $(\mathbf{A_B})$, $(\mathbf{A_F})$, *and* $(\mathbf{A_R})$, *suppose that the conditions of Lemmas 3 and 1 hold. Let*

$$\bar{\varepsilon} := (1 - \varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}} > 0, \qquad \varepsilon_{\bar{w}} < \frac{t}{1+t}, \tag{84}$$

where $t = \tau/\sqrt{n}$, *and suppose that, for some constant* $c_\theta > 0$, *the following cross-group margin condition holds at every relevant iteration:*

$$\theta^\ell_{k^\star} \alpha^\ell_{\mathrm{rec},k^\star} \bar{\varepsilon} - \sum_{k \neq k^\star} \theta^\ell_k \alpha^\ell_{\mathrm{rec},k} \geq c_\theta \bar{\varepsilon} \sum_{k \in [3]} \theta^\ell_k \alpha^\ell_{\mathrm{rec},k}, \tag{85}$$

where $k^\star = k^\star_\ell$ *denotes the unique ranked group containing* $r^\star_\ell$. *Then, on the global good event* $\mathcal{G}_T$, *for every nonstationary iteration* $\ell \in [\ell_T]$,

$$\frac{|g(x^\ell)^T d^{r^\star_\ell}|}{\|g(x^\ell)\| \, \|d^{r^\star_\ell}\|} \geq t, \tag{86}$$

and

$$\frac{|g(x^\ell)^T p^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\| \, \|p^\ell_{\mathrm{trec}}\|} \geq c_\theta \bar{\varepsilon}. \tag{87}$$

In particular, (85) *implies that the unscaled triangular vector* $\widehat{p}^\ell_{\mathrm{trec}}$ *is nonzero, and hence the fixed-norm scaled direction* p^ℓ_{trec} *is well defined. Consequently, the conclusions* (86) *and* (87) *hold simultaneously for all nonstationary iterations* $\ell \in [\ell_T]$ *with probability at least* $1 - (\gamma_0 + \gamma_{\mathrm{sel}})$. *If* $g(x^\ell) = 0$ *at some iteration, stationarity has already been attained and no angle condition is required.*

Proof Fix a nonstationary iteration $\ell \in [\ell_T]$ and condition on the global good event $\mathcal{G}_T$. By its definition in (70), $\mathcal{G}_T \subseteq E^c_{\mathrm{samp},\ell} \cap E^c_{\mathrm{sel},\ell}$. Hence, both the sampling-success and sorting–selection-success properties hold at iteration ℓ.

Define $q^\ell := g(x^\ell)/\|g(x^\ell)\|$. Since $E^c_{\mathrm{samp},\ell}$ occurs, Proposition 3, together with the antipodal construction of the sampled direction set, yields the latent descent index $i^\star_\ell$ defined in (66). By (77), $g(x^\ell)^T d^{i^\star_\ell} \leq -t\|g(x^\ell)\|$. After sorting, the same direction appears at ranked position $r^\star_\ell$, namely, $d^{r^\star_\ell} = d^{i^\star_\ell}$. Therefore, by (78), $(q^\ell)^T d^{r^\star_\ell} \leq -t$. Since the ranked mutation directions are unit vectors, $\|d^{r^\star_\ell}\| = 1$, and consequently

$$\left|(q^\ell)^T d^{r^\star_\ell}\right| \geq t, \qquad \|d^{r^\star_\ell}\| = 1, \qquad \|q^\ell\| = 1. \tag{88}$$

This proves (86).

We next transfer the alignment from the latent ranked direction to its group-recombination direction. Since $E^c_{\mathrm{sel},\ell}$ also occurs, Lemma 3 implies that $r^\star_\ell$ is the unique best-ranked element of the group $\mathcal{G}_{k^\star}$ containing it, where $k^\star = k^\star_\ell$.

Suppose that the within-group weights satisfy (9) with dominant index $j^\star = r^\star_\ell$. Applying Lemma 1 with

$$u_j = d^{\pi(j)\ell}, \qquad v = d^\ell_{\mathrm{rec}\,k^\star} = \sum_{j \in \mathcal{G}_{k^\star}} \bar{w}_j d^{\pi(j)\ell}, \qquad q = q^\ell,$$

and using (88), we obtain

$$\left|(q^\ell)^T d^\ell_{\mathrm{rec}\,k^\star}\right| \geq (1-\varepsilon_{\bar{w}})t - \varepsilon_{\bar{w}} = \bar{\varepsilon}. \tag{89}$$

Consider now the unscaled triangular recombination vector

$$\widehat{p}^\ell_{\mathrm{trec}} = \frac{1}{w^\ell_p}\sum_{k=1}^{3}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k} d^\ell_{\mathrm{rec}\,k}, \qquad w^\ell_p = \sum_{k=1}^{3}\theta^\ell_k.$$

Taking the inner product with q^ℓ and separating the contribution of the group $\mathcal{G}_{k^\star}$ yields

$$\left|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}\right| = \frac{1}{w^\ell_p}\left|\theta^\ell_{k^\star}\alpha^\ell_{\mathrm{rec}\,k^\star}(q^\ell)^T d^\ell_{\mathrm{rec}\,k^\star} + \sum_{k\neq k^\star}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}(q^\ell)^T d^\ell_{\mathrm{rec}\,k}\right|.$$

Applying the reverse triangle inequality gives

$$\left|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}\right| \geq \frac{1}{w^\ell_p}\left[\theta^\ell_{k^\star}\alpha^\ell_{\mathrm{rec}\,k^\star}\left|(q^\ell)^T d^\ell_{\mathrm{rec}\,k^\star}\right| - \sum_{k\neq k^\star}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}\left|(q^\ell)^T d^\ell_{\mathrm{rec}\,k}\right|\right].$$

For every $k \in [3]$, the vector $d^\ell_{\mathrm{rec}\,k}$ is a convex combination of unit vectors. Hence

$$\|d^\ell_{\mathrm{rec}\,k}\| \leq \sum_{j\in\mathcal{G}_k}\bar{w}_j\|d^{\pi(j)\ell}\| = 1.$$

Since $\|q^\ell\| = 1$, the Cauchy–Schwarz inequality gives $\left|(q^\ell)^T d^\ell_{\mathrm{rec}\,k}\right| \leq 1$. Using this estimate together with (89), we obtain

$$\left|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}\right| \geq \frac{1}{w^\ell_p}\left[\bar{\varepsilon}\theta^\ell_{k^\star}\alpha^\ell_{\mathrm{rec}\,k^\star} - \sum_{k\neq k^\star}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}\right]. \tag{90}$$

Applying the cross-group margin condition (85) to (90) yields

$$\left|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}\right| \geq \frac{c_\theta}{w^\ell_p}\bar{\varepsilon}\sum_{k=1}^{3}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}. \tag{91}$$

Since the right-hand side is strictly positive, $\widehat{p}^\ell_{\mathrm{trec}} \neq 0$. On the other hand, using the definition of $\widehat{p}^\ell_{\mathrm{trec}}$, $\|d^\ell_{\mathrm{rec}\,k}\| \leq 1$, and the triangle inequality, we have

$$\|\widehat{p}^\ell_{\mathrm{trec}}\| \leq \frac{1}{w^\ell_p}\sum_{k=1}^{3}\theta^\ell_k \alpha^\ell_{\mathrm{rec}\,k}, \qquad \text{where} \qquad \alpha^\ell_{\mathrm{rec}\,k} \in [\alpha_{\min}, \alpha_{\max}]. \tag{92}$$

Combining (91) and (92) gives $\left|(q^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}\right| \geq c_\theta\bar{\varepsilon}\|\widehat{p}^\ell_{\mathrm{trec}}\|$. Hence,

$$\frac{|g(x^\ell)^T \widehat{p}^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\|\,\|\widehat{p}^\ell_{\mathrm{trec}}\|} \geq c_\theta\bar{\varepsilon}.$$

The actual search direction is obtained by the fixed positive scaling $p^{\ell}_{\mathrm{trec}} = \delta \widehat{p}^{\ell}_{\mathrm{trec}} / \|\widehat{p}^{\ell}_{\mathrm{trec}}\|$. Positive scaling preserves the angle ratio, and therefore

$$\frac{|g(x^{\ell})^T p^{\ell}_{\mathrm{trec}}|}{\|g(x^{\ell})\| \, \|p^{\ell}_{\mathrm{trec}}\|} = \frac{|g(x^{\ell})^T \widehat{p}^{\ell}_{\mathrm{trec}}|}{\|g(x^{\ell})\| \, \|\widehat{p}^{\ell}_{\mathrm{trec}}\|} \geq c_{\theta} \bar{\varepsilon},$$

which proves (87). The argument above holds for every nonstationary $\ell \in [\ell_T]$ on the single global event $\mathcal{G}_T$. Therefore, (86) and (87) hold simultaneously over all such iterations on $\mathcal{G}_T$. Finally, by (72), $\mathbb{P}(\mathcal{G}_T) \geq 1 - (\gamma_0 + \gamma_{\mathrm{sel}})$, which proves the stated high-probability conclusion. □

Remark 8 *The within-group dominance condition (9) and the cross-group margin condition (85) play distinct roles in Proposition 4. The former controls cancellation within the ranked group containing the latent descent direction, whereas the latter controls cancellation between the three group-recombination contributions. In particular, the cross-group margin (85) is a quantitative analytical transfer condition and does not follow merely from*

$$\theta^{\ell}_k > 0 \qquad \text{and} \qquad \sum_{k \in [3]} \theta^{\ell}_k = w^{\ell}_p.$$

Accordingly, it must hold uniformly over the iterations to which Proposition 4 is applied. No independence assumption is required for this deterministic geometric transfer once the global good event $\mathcal{G}_T$ has occurred.

Remark 9 *The distinction between the unscaled triangular vector and the final search direction is essential. The vector*

$$\widehat{p}^{\ell}_{\mathrm{trec}} = \frac{1}{w^{\ell}_p} \sum_{k=1}^{3} \theta^{\ell}_k \, \alpha^{\ell}_{\mathrm{rec},k} \, d^{\ell}_{\mathrm{rec}\,k}$$

satisfies

$$\|\widehat{p}^{\ell}_{\mathrm{trec}}\| \leq \frac{1}{w^{\ell}_p} \sum_{k=1}^{3} \theta^{\ell}_k \, \alpha^{\ell}_{\mathrm{rec},k} \leq \max_{k \in [3]} \alpha^{\ell}_{\mathrm{rec}\,k}.$$

The actual search direction is then defined by $p^{\ell}_{\mathrm{trec}} = \delta \widehat{p}^{\ell}_{\mathrm{trec}} / \|\widehat{p}^{\ell}_{\mathrm{trec}}\|$ with $\delta > 0$, so that $\|p^{\ell}_{\mathrm{trec}}\| = \delta$ by construction. Moreover,

$$\frac{|g(x^{\ell})^T p^{\ell}_{\mathrm{trec}}|}{\|g(x^{\ell})\| \, \|p^{\ell}_{\mathrm{trec}}\|} = \frac{|g(x^{\ell})^T \widehat{p}^{\ell}_{\mathrm{trec}}|}{\|g(x^{\ell})\| \, \|\widehat{p}^{\ell}_{\mathrm{trec}}\|}.$$

Thus, the angle-transfer conclusion of Proposition 4 is invariant under the fixed-norm positive scaling used by Algorithm `DAES`.

Corollary 2 *Suppose that the conditions of Proposition 4 hold, and define*

$$\bar{c} := c_{\theta} \bar{\varepsilon} = c_{\theta} \big((1 - \varepsilon_{\bar{w}}) t - \varepsilon_{\bar{w}} \big) > 0.$$

Then, on the global good event $\mathcal{G}_T$, for every nonstationary iteration $\ell \in [\ell_T]$,

$$v(x^{\ell}, p^{\ell}_{\mathrm{trec}}) := \frac{\|g(x^{\ell})\| \, \|p^{\ell}_{\mathrm{trec}}\|}{|g(x^{\ell})^T p^{\ell}_{\mathrm{trec}}|} \leq \frac{1}{\bar{c}}. \tag{93}$$

Consequently, the bound (93) *holds simultaneously for all nonstationary iterations* $\ell \in [\ell_T]$ *with probability at least* $1-(\gamma_0+\gamma_{\text{sel}})$. *If* $g(x^\ell)=0$ *at some iteration, stationarity has already been attained and the geometric factor* $v(x^\ell, p^\ell_{\text{trec}})$ *need not be invoked.*

Proof On $\mathcal{G}_T$, Proposition 4 gives, for every nonstationary $\ell \in [\ell_T]$,

$$\frac{|g(x^\ell)^T p^\ell_{\text{trec}}|}{\|g(x^\ell)\|\,\|p^\ell_{\text{trec}}\|} \geq c_\theta \bar{\varepsilon} = \bar{c} > 0.$$

In particular, Proposition 4 ensures that the unscaled triangular vector is nonzero and hence that the fixed-norm direction p^ℓ_{trec} is well defined. Taking reciprocals yields

$$\frac{\|g(x^\ell)\|\,\|p^\ell_{\text{trec}}\|}{|g(x^\ell)^T p^\ell_{\text{trec}}|} \leq \frac{1}{\bar{c}},$$

which is exactly (93). Since the above argument holds simultaneously over all nonstationary iterations on the single global event $\mathcal{G}_T$, and since (72) gives $\mathbb{P}(\mathcal{G}_T) \geq 1-(\gamma_0+\gamma_{\text{sel}})$, the stated high-probability conclusion follows. □

3.2.3 Probabilistic Bound on Gradient Norm

Using Corollary 2, we now derive a high-probability global bound on the minimum gradient norm generated by `DAES`. The argument combines the uniform geometric-factor bound established on the global good event $\mathcal{G}_T$ with the deterministic gradient-norm estimate of Theorem 2.

Theorem 5 *Suppose that Assumption* ($\mathbf{A_S}$) *holds, that the conditions of Proposition 4 hold, and that the hypotheses of Theorem 2 are satisfied for the triangular recombination directions* p^ℓ_{trec} *at the relevant iterations. Let* $\bar{c} := c_\theta \bar{\varepsilon} > 0$. *Then*

$$\min_{0\leq \ell \leq \ell_T} \|g(x^\ell)\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right), \qquad \textit{with probability at least} \qquad 1-(\gamma_0+\gamma_{\text{sel}}). \tag{94}$$

Proof Condition on the global good event $\mathcal{G}_T$. By (72), $\mathbb{P}(\mathcal{G}_T) \geq 1-(\gamma_0+\gamma_{\text{sel}})$. If there exists an iteration $\ell \in [\ell_T]$ such that $g(x^\ell)=0$, then $\min_{0\leq j\leq \ell_T}\|g(x^j)\|=0$, and (94) follows immediately. We may therefore restrict attention to the case in which all relevant iterations are nonstationary. On $\mathcal{G}_T$, Corollary 2 gives, simultaneously for every nonstationary iteration $\ell \in [\ell_T]$, $v(x^\ell, p^\ell_{\text{trec}}) \leq 1/\bar{c}$. Hence, the geometric factor associated with the triangular recombination direction is uniformly bounded over the relevant iterations. Applying Theorem 2 with $p^\ell = p^\ell_{\text{trec}}$ therefore yields

$$\min_{1\leq \ell \leq \ell_T} \|g(x^\ell)\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right).$$

Since $\min_{0\le\ell\le\ell_T} \|g(x^\ell)\| \le \min_{1\le\ell\le\ell_T} \|g(x^\ell)\|$, we obtain

$$\min_{0\le\ell\le\ell_T} \|g(x^\ell)\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right),$$

which is (94). The preceding argument holds on the single global event $\mathcal{G}_T$. Therefore, using (72), the conclusion holds with probability at least $1-(\gamma_0+\gamma_{\text{sel}})$. □

3.2.4 Probabilistic Bounds on Convergence Accuracy

We now translate the high-probability gradient bound of Theorem 5 into corresponding bounds on the objective accuracy and, under strong convexity, on the distance to the unique minimizer. Thus, the noise-limited stationarity guarantee obtained in the previous subsection yields explicit high-probability convergence guarantees for DAES without requiring exact gradient information.

Theorem 6 *Suppose that the hypotheses of Theorem 5 hold, and let $\bar{c} = c_\theta \bar{\varepsilon} > 0$ be as defined therein. Let $\ell^{**} \in \operatorname{argmin}_{0\le\ell\le\ell_T} \|g(x^\ell)\|$, and let $x^{\ell^{**}}$ denote the corresponding iterate of Algorithm* DAES*. Then, with probability at least $1-(\gamma_0+\gamma_{\text{sel}})$, the following conclusions hold:*

(i) If f is convex, then

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right). \tag{95}$$

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

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\left(\frac{L\omega}{\mu_c\,\bar{c}^{\,2}}\right), \qquad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\mu_c\,\bar{c}}\right). \tag{96}$$

Proof Condition on the global good event $\mathcal{G}_T$. By (72), $\mathbb{P}(\mathcal{G}_T) \ge 1-(\gamma_0+\gamma_{\text{sel}})$. On this event, Theorem 5 gives

$$\min_{0\le\ell\le\ell_T} \|g(x^\ell)\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right).$$

By the definition of ℓ^{**},

$$\|g(x^{\ell^{**}})\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right). \tag{97}$$

We first consider the convex case. By the same argument used in Theorem 3, convexity gives

$$f(x) - \underline{f} \le g(x)^T(x-\underline{x}) \le \|g(x)\|\,\|x-\underline{x}\|.$$

Under Assumption ($\mathbf{A_S}$), the relevant level set is compact; hence there exists a finite constant $d_{\max} > 0$, independent of ω, such that $\|x^\ell - \underline{x}\| \le d_{\max}$ for all iterates under consideration. Therefore,

$$f(x^\ell) - \underline{f} \le d_{\max}\|g(x^\ell)\|.$$

Evaluating this inequality at $\ell = \ell^{**}$ and using (97) yields

$$f(x^{\ell^{**}}) - \underline{f} \le d_{\max}\|g(x^{\ell^{**}})\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\bar{c}}\right),$$

which proves (95). We next consider the μ_c-strongly convex case. Strong convexity implies the standard gradient-dominance estimate

$$f(x) - \underline{f} \le \frac{1}{2\mu_c}\|g(x)\|^2,$$

and also $\|x - \underline{x}\| \le (1/\mu_c)\|g(x)\|$. Applying these inequalities at $x = x^{\ell^{**}}$ gives

$$f(x^{\ell^{**}}) - \underline{f} \le \frac{1}{2\mu_c}\|g(x^{\ell^{**}})\|^2,$$

and $\|x^{\ell^{**}} - \underline{x}\| \le (1/\mu_c)\|g(x^{\ell^{**}})\|$. Substituting (97) yields

$$f(x^{\ell^{**}}) - \underline{f} = \mathcal{O}\left(\frac{L\omega}{\mu_c\,\bar{c}^{\,2}}\right) \quad \text{and} \quad \|x^{\ell^{**}} - \underline{x}\| = \mathcal{O}\left(\frac{\sqrt{L\omega}}{\mu_c\,\bar{c}}\right),$$

which proves (96). Since the entire argument holds on the single global event $\mathcal{G}_T$, whose probability is bounded below by $1 - (\gamma_0 + \gamma_{\mathrm{sel}})$, the stated high-probability conclusions follow. □

3.2.5 Probabilistic Complexity of DAES

We now derive a high-probability counterpart of the deterministic complexity analysis of DAES. On the global good event $\mathcal{G}_T$, Corollary 2 provides a uniform bound on the geometric factor associated with the triangular recombination direction. Consequently, the deterministic efficiency mechanism can be applied pathwise on $\mathcal{G}_T$, with the deterministic angle constant replaced by the probabilistic transfer constant $\bar{c} = c_\theta \bar{\varepsilon} > 0$.

Consider a successful recombination iteration ℓ of DAES, and write

$$x^{\ell+1} = x^\ell + s^\ell \alpha^\ell p^\ell_{\mathrm{trec}}, \qquad s^\ell \in \{-1, 1\},$$

where the accepted step satisfies the NDF-SDC (22). By the scaling rule of Algorithm DAES, the triangular recombination direction has the fixed norm

$$\|p^\ell_{\mathrm{trec}}\| = \delta, \qquad \delta > 0.$$

On $\mathcal{G}_T$, Corollary 2 gives

$$\frac{|g(x^\ell)^T p^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\|\,\|p^\ell_{\mathrm{trec}}\|} \ge \bar{c}.$$

Hence Proposition 2 applies with

$$\Delta_{\mathrm{a}} := \bar{c}, \qquad p := p^\ell_{\mathrm{trec}}, \qquad \|p^\ell_{\mathrm{trec}}\| = \delta.$$

Define

$$\tilde{K}_1 := \frac{\gamma_e \left(\tilde{\beta} + \frac{L}{2}\delta^2 \right)}{\bar{c}\,\delta}.$$

Recalling from Lemma 2 that $\beta := \tilde{\beta} - 2L(\underline{\kappa}')^{-2} > 0$, the corresponding high-probability efficiency estimate is

$$f(x^\ell) - f(x^{\ell+1}) \;\geq\; \tilde{\eta}\,\|g(x^\ell)\|^2, \qquad \tilde{\eta} := \frac{\beta}{4(\tilde{K}_1)^2}. \tag{98}$$

Thus, on $\mathcal{G}_T$, every successful recombination iteration satisfying the hypotheses of Proposition 2 enjoys the same quadratic gradient-decrease structure as in the deterministic analysis, with the deterministic efficiency constant replaced by $\tilde{\eta}$.

Theorem 7 *Suppose that the hypotheses of Theorem 4 hold and that the conditions of Proposition 4 are satisfied. Let $\bar{c} = c_\theta \bar{\varepsilon} > 0$, and define*

$$\tilde{K}_1 := \frac{\gamma_e \left(\tilde{\beta} + \frac{L}{2}\delta^2 \right)}{\bar{c}\,\delta}, \qquad \tilde{\eta} := \frac{\beta}{4(\tilde{K}_1)^2},$$

where $\beta = \tilde{\beta} - 2L(\underline{\kappa}')^{-2} > 0$ and $\delta > 0$ is the fixed norm of the scaled triangular recombination direction. Then, with probability at least $1-(\gamma_0+\gamma_{\mathrm{sel}})$, the successful-iteration complexity guarantees established in Theorem 4 carry over to the randomized setting, with the deterministic efficiency constant replaced by $\tilde{\eta}$. Moreover, if the unsuccessful-iteration counting hypotheses used in Theorem 4 also hold, then the corresponding total iteration-complexity guarantees carry over with the same asymptotic orders.

Proof Condition on the global good event $\mathcal{G}_T$. By (72), $\mathbb{P}(\mathcal{G}_T) \geq 1-(\gamma_0+\gamma_{\mathrm{sel}})$. On this event, Corollary 2 gives $v(x^\ell, p^\ell_{\mathrm{trec}}) \leq 1/\bar{c}$ simultaneously over all relevant nonstationary iterations. Equivalently,

$$\frac{|g(x^\ell)^T p^\ell_{\mathrm{trec}}|}{\|g(x^\ell)\|\,\|p^\ell_{\mathrm{trec}}\|} \geq \bar{c}.$$

Consider any successful recombination iteration. By the scaling rule of Algorithm `DAES`,

$$\|p^\ell_{\mathrm{trec}}\| = \delta,$$

where $\delta > 0$ is fixed and independent of ℓ. The accepted iterate is of the form

$$x^{\ell+1} = x^\ell + s^\ell \alpha^\ell p^\ell_{\mathrm{trec}}, \qquad s^\ell \in \{-1, 1\},$$

and, by hypothesis, the accepted step satisfies the conditions required by Proposition 2.

Applying Proposition 2, with $\Delta_{\mathrm{a}} = \bar{c}$, $p = p^\ell_{\mathrm{trec}}$, and $\|p^\ell_{\mathrm{trec}}\| = \delta$, gives

$$f(x^\ell) - f(x^{\ell+1}) \geq \frac{\beta}{4(\tilde{K}_1)^2}\|g(x^\ell)\|^2,$$

where $\tilde{K}_1 = \gamma_e\left(\tilde{\beta} + L\delta^2/2\right)/(\bar{c}\,\delta)$. Since both δ and $\bar{c}$ are independent of ℓ, the constant $\tilde{K}_1$ is uniform over all relevant successful recombination iterations. Therefore,

$$f(x^\ell) - f(x^{\ell+1}) \geq \tilde{\eta}\|g(x^\ell)\|^2, \qquad \tilde{\eta} = \frac{\beta}{4(\tilde{K}_1)^2},$$

which is precisely (98).

Thus, conditioned on $\mathcal{G}_T$, the successful iterations satisfy the same deterministic decrease recursion used in Theorem 4, with the deterministic efficiency constant replaced by $\tilde{\eta}$. Hence the successful-iteration complexity arguments of Theorem 4 apply pathwise on $\mathcal{G}_T$, and the corresponding asymptotic orders remain unchanged, with constants rescaled according to $\tilde{\eta}^{-1}$. If, in addition, the unsuccessful-iteration counting hypotheses invoked in Theorem 4 hold, then the same deterministic relation between the numbers of unsuccessful and successful iterations applies on $\mathcal{G}_T$. Consequently, the corresponding total iteration-complexity orders also carry over. Finally, since $\mathbb{P}(\mathcal{G}_T) \geq 1 - (\gamma_0 + \gamma_{\text{sel}})$, all stated conclusions hold with at least this probability. □

The result above is a uniform-good-event transfer of the deterministic complexity analysis. On $\mathcal{G}_T$, the probabilistic angle estimate provides a uniform deterministic geometric bound, while the fixed scaling

$$\|p^\ell_{\text{trec}}\| = \delta$$

ensures that the efficiency constant is independent of the iteration index. To place the resulting bound in context, Table 1 compares the dominant nonconvex function-evaluation dependence with representative randomized DFO frameworks. The comparison distinguishes the oracle models, the random-direction budgets, and, for the noisy methods, the different definitions of the noise-limited accuracy scale. Sampling budgets that are fixed once the prescribed confidence parameters are chosen are absorbed into the principal asymptotic order; their explicit dependence is discussed below.

As shown in Table 1, the noiseless methods of Gratton et al. [17], `STP` [30], and `VRBBO` [8] target an arbitrary stationarity tolerance $\varepsilon > 0$, whereas `VRDFON` and `DAES` operate under bounded evaluation noise. This distinction is essential when comparing the complexity orders.

The random-direction budgets play analogous roles but are calibrated differently. Gratton et al. [17] derive the explicit function-evaluation bound $\mathcal{O}(mnL^2\varepsilon^{-2})$, where m is the number of random polling directions. Their theory permits the fixed choice $m = 2$ for admissible algorithmic parameters, for example $\gamma = 2$ and $\theta = 0.5$, and hence the corresponding principal order reduces to $\mathcal{O}(nL^2\varepsilon^{-2})$. The `STP` method samples one random direction at each iteration and tests the corresponding two signed trial points, leading directly to the same nonconvex dimension–accuracy order $\mathcal{O}(nL\varepsilon^{-2})$.

In `VRBBO`, the direction budget is chosen according to the requested confidence level as

$$R = \left\lceil \log_2(\eta^{-1}) \right\rceil.$$

Table 1 Dominant nonconvex function-evaluation dependence for representative randomized DFO frameworks. Problem- and algorithm-dependent constants not involving the displayed parameters are suppressed. For `DAES`, the displayed principal order uses $\bar{c} = \Theta(n^{-1/2})$, fixed δ, and the dominant L-dependence inherited from the efficiency factor $\tilde{\eta}^{-1}$. Fixed random-direction budgets are absorbed into the hidden constant.

Method	Oracle / guarantee	Direction budget	Principal nonconvex FE complexity
Gratton et al. [17]	noiseless / high probability	$m = 2$ admissible	$\mathcal{O}(nL^2\varepsilon^{-2})$
`STP` [30]	noiseless / expectation	one sampled direction	$\mathcal{O}(nL\varepsilon^{-2})$
`VRBBO` [8]	noiseless / high probability	$R = \lceil \log_2(\eta^{-1}) \rceil$	$\mathcal{O}(nL^2\varepsilon^{-2})$
`VRDFON` [7]	bounded noise / high probability	$R = \lceil T_0^{-1} \log_2(\eta^{-1}) \rceil$	$\mathcal{O}(nL^2\varepsilon_\omega^{-2})$
`DAES`	bounded noise / high probability	$\lambda \geq 6$ via (61)	$\mathcal{O}(nL^2\varepsilon_\omega^{-2})$

Thus, for a fixed confidence parameter η, R is a fixed constant and the parameter-explicit bound $\mathcal{O}(nRL^2\varepsilon^{-2})$ is conventionally reported as $\mathcal{O}(nL^2\varepsilon^{-2})$. The same fixed-confidence convention applies to `VRDFON`, where

$$R = \left\lceil T_0^{-1} \log_2(\eta^{-1}) \right\rceil .$$

In particular, $R = 2$ is possible only for confidence parameters satisfying

$$1 < T_0^{-1} \log_2(\eta^{-1}) \leq 2,$$

or equivalently $2^{-2T_0} \leq \eta < 2^{-T_0}$; it is not a universal choice. Nevertheless, once T_0 and η are fixed, R is fixed and is absorbed into the hidden constant of the reported function-evaluation complexity.

The corresponding situation for `DAES` is similar, with one important structural distinction. Because the present construction uses antipodal sampling together with three ranked groups, λ is taken as an even multiple of three. Hence, the smallest admissible population size is $\lambda = 6$, rather than 2. Under the calibration (61),

$$\lambda = 6 \left\lceil \frac{\log(\bar{\ell}_T/\gamma_0)}{3\log(1/r_\tau)} \right\rceil, \qquad r_\tau = \tau\sqrt{\frac{2}{\pi}},$$

and therefore $\lambda = 6$ is sufficient whenever $r_\tau^3 \leq \gamma_0/\bar{\ell}_T$. For more stringent confidence–horizon requirements, a larger multiple of six is required. However, once τ, γ_0, and

$\bar{\ell}_T$ are prescribed, λ is a fixed constant. Consequently, it is absorbed into the hidden constant of the principal asymptotic function-evaluation bound, in the same sense that fixed m or R is absorbed in the related frameworks.

More explicitly, since the mutation phase of `DAES` evaluates λ mutation points at each iteration, the budget-explicit function-evaluation bound contains the factor $\lambda + \mathcal{O}(1)$. Thus, under the dominant dependence displayed in Table 1, one may write $\mathcal{O}\big(\lambda n L^2 \varepsilon_\omega^{-2}\big)$ when the sampling budget is shown explicitly, whereas for fixed λ the principal order is $\mathcal{O}\big(nL^2\varepsilon_\omega^{-2}\big)$. The latter reporting convention is directly analogous to suppressing a fixed m in the Gratton et al. bound or a fixed R in the `VRBBO` and `VRDFON` bounds.

The two noisy frameworks should not, however, be assigned the same definition of ε_ω. In `VRDFON`, the noise-limited scale is $\varepsilon_\omega = \mathcal{O}(\sqrt{nL\omega})$, and therefore

$$\mathcal{O}\big(nL^2\varepsilon_\omega^{-2}\big) = \mathcal{O}(L\omega^{-1}).$$

In contrast, the intrinsic noise scale of `DAES` is defined in (47) as $\varepsilon_\omega = \mathcal{O}(\sqrt{L\omega})$. The dimension dependence enters separately through the probabilistic geometric factor. Indeed, Theorem 5 gives

$$\min_{0\le\ell\le\ell_T} \|g(x^\ell)\| = \mathcal{O}\Big(\frac{\varepsilon_\omega}{\bar{c}}\Big).$$

Hence, under the dimension-scaled angle regime $\bar{c} = \Theta(n^{-1/2})$, the randomized stationarity level becomes $\mathcal{O}(\varepsilon_\omega/\bar{c}) = \mathcal{O}(\sqrt{nL\omega})$, which is comparable to the noise-limited stationarity scale used in `VRDFON`, although the underlying definition of ε_ω is different.

Overall, the comparison shows that the direction budgets m, R, and λ should be treated under a common asymptotic convention. A fixed budget may be absorbed into the hidden constant, whereas a confidence-explicit analysis should display its dependence. Under the present calibration, `DAES` cannot reduce the population to two directions because of its antipodal three-group construction, but it admits the fixed minimum $\lambda = 6$ whenever the prescribed confidence–horizon condition above is satisfied, and its population calibration introduces no additional explicit polynomial dependence on the dimension n.

4 Numerical Results

DAES vs. MADFO

The comparison with `MADFO` [27] evaluates the proposed selection and recombination scheme in `DAES` against a state-of-the-art matrix adaptation evolution strategy for noisy derivative-free optimization.

MADFO employs the standard mutation–selection–recombination framework with controlled mutation step-sizes and a randomized nonmonotone line search in the recombination phase. It relies on classical selection, where mutation points are ranked by their noisy objective values and only a subset with the smallest values is used in recombination, which may reduce robustness in noisy settings due to potentially misleading evaluations and the loss of information from the remaining points. In contrast, DAES retains all sorted mutation directions, partitions them into three ranked groups, and combines the resulting recombination directions and points through triangular recombination. The comparison with MADFO, which uses the adaptation matrix M, evaluates this grouped sorting–selection and recombination strategy within the MAES framework for noisy DFO.

We set $M = I$ and do not perform affine matrix updates in DAES. This avoids anisotropy effects that may arise in noisy settings, where the matrix M can become highly ill-conditioned and lead to nearly collinear mutation directions after the transformation $d = Mz$, thereby reducing directional diversity and weakening the effect of symmetric sampling and triangular recombination. This design is consistent with the goal of DAES, which is to study the effect of enhanced mutation, sorting–selection, and triangular recombination mechanisms rather than improving MAES via matrix adaptation. Fixing $M = I$ isolates these components and avoids confounding effects from affine transformations.

Test Problems

For the numerical comparison between DAES variants and MADFO, we consider 655 test problems with $2 \leq n \leq 100$ from the prince collection of the BARON software package [31], available at https://minlp.com/optimization-test-problems. Each deterministic test problem is combined with absolute and relative Gaussian noise as well as uniform noise (cf. [27, Section 3.4]), and evaluated under seven noise levels $\omega \in \{10^i\}_{i=0}^{2} \cup \{4^i\}_{i=1}^{4}$. This results in a total benchmark of $4 \times 7 \times 655 = 18340$

Data and Performance Profiles

To evaluate the *robustness* and *efficiency* of the solvers under comparison, we use data profiles, introduced by Moré and Wild [37], and performance profiles, proposed by Dolan and Moré [38]. Let $\mathcal{S}$ denote the set of solvers and let $\mathcal{P}$ denote the set of test problems.

For a solver $s \in \mathcal{S}$, the data profile measures the fraction of problems solved within a budget of at most κ groups of $n_p + 1$ function evaluations, where n_p is the dimension of problem p. It is defined by

$$\delta_s(\kappa) := \frac{1}{|\mathcal{P}|} \left| \left\{ p \in \mathcal{P} \;\middle|\; cr_{p,s} := \frac{c_{p,s}}{n_p + 1} \leq \kappa \right\} \right|. \tag{99}$$

Here, $c_{p,s}$ is the cost of solving problem p by solver s, and $cr_{p,s}$ is the corresponding normalized cost ratio. The performance profile compares the cost of each solver with the best cost achieved on each problem by any solver in the test set. It is defined as

$$\rho_s(\tau) := \frac{1}{|\mathcal{P}|} \left| \left\{ p \in \mathcal{P} \;\middle|\; pr_{p,s} := \frac{c_{p,s}}{\min\{c_{p,\bar{s}} : \bar{s} \in \mathcal{S}\}} \leq \tau \right\} \right|. \tag{100}$$

Thus, $\rho_s(\tau)$ is the fraction of problems with performance ratio $pr_{p,s} \leq \tau$. In particular, $\rho_s(1)$ is the fraction where solver s achieves the best observed performance. For large τ and κ, $\rho_s(\tau)$ and $\delta_s(\kappa)$ approximate the overall fraction of problems solved by s.

Stopping Criteria

For each solver $s \in \mathcal{S}$, performance is measured by the normalized residual

$$q_s = (f_s - f_{\text{opt}})/(f_0 - f_{\text{opt}}).$$

Here, f_s is the best value returned by solver s, f_0 is the starting value, and f_{opt} is a reference value obtained from global or high-quality local solutions via DFO methods. The ratio q_s is used only for benchmarking since f_{opt} is generally unavailable in black-box optimization. Solver s solves a problem if $q_s \leq \varepsilon_q$ before reaching `nfmax` or `secmax`; otherwise, it is deemed unsuccessful. Parameters are chosen so that the best-performing solver solves at least half the test problems, except under high noise, where larger budgets provide no benefit. We use `secmax` $= 600$, `nfmax` $= 12000$, and $\varepsilon_q = 10^{-4}$.

Initial Point

Following [7, 27], we do not use the origin directly as the starting point. Instead, the nominal initial point is shifted by a vector $\xi \in \mathbb{R}^n$ whose components are given by

$$\xi_i := (-1)^{i-1} \frac{2}{i+2}, \qquad \forall i \in [n].$$

This shift is introduced to avoid artificial advantages on simple test problems from the `prince` collection, for which the solution structure may be too easily inferred from the unshifted starting point. The use of a shifted initial point therefore makes the initialization less trivial and provides a more meaningful assessment of solver behavior.

Accordingly, we set $y^0 := 0$ and define the initial noisy function value as $\widetilde{f}_0 := \widetilde{f}(y^0)$. For all subsequent iterates, the noisy objective values are evaluated using

$$\widetilde{f}_\ell := \widetilde{f}(y^\ell + \xi), \qquad \ell \geq 0.$$

Tuning Parameters

For `MADFO`, we use the default tuning parameters recommended in [27]. For `DAES`, the sample size (i.e., number of mutation points) is set to

$$\lambda := \max\{6,\, 4 + \lfloor 3\log(n) \rfloor\}.$$

For the numerical experiments, this practical dimension-dependent population rule is used rather than the conservative confidence calibration (61). The choice keeps the `DAES` population size close to that used by `MADFO` for most tested dimensions, thereby supporting a fair comparison in terms of the number of sampled mutation points. When triangular recombination is used, the sorted mutation directions are partitioned into three ranked groups with

$$\mu := \lfloor \lambda/3 \rfloor,$$

where the first two groups have length μ and the third group absorbs any remaining ranked indices. The triangular recombination parameter is fixed at $\eta = 0.9$, see (16), with normalization radius $\delta = 10^5$, see (21). The scaling parameters required in (6) are chosen as in [27]. Other step-size parameters required in (25) and Subsection 2.3.2 are $\rho_u = 1/2$, $\tilde{\beta} = 10^{-12}$, $\gamma_e = 4$, $T = 10$, $\alpha_{\min} = 10^{-2}$ and $\alpha_{\max} = 1/2$. In addition, for `DAES`, we set $\ell_{\max} = 12000$.

For the numerical implementation, the raw ranking weights are computed as

$$w_j = \max\{\log(\lambda + 0.5) - \log j,\, \epsilon_{\mathrm{mach}}\}\,, \qquad j \in [\lambda],$$

and are normalized separately within each ranked group according to (8); when λ is not divisible by three, the third group absorbs the remaining ranked indices. Since the raw weights decrease with the rank, let j_k^{best} denote the best-ranked element of $\mathcal{G}_k$. We record, without modifying the algorithm, the induced within-group dominance quantities

$$\varepsilon_{\bar{w},k} := 1 - \bar{w}_{j_k^{\mathrm{best}}}, \qquad \varepsilon_{\bar{w}}^{\mathrm{num}} := \max_{k\in[3]} \varepsilon_{\bar{w},k}.$$

By the group normalization in (8), $\varepsilon_{\bar{w},k} = \sum_{j\in\mathcal{G}_k\setminus\{j_k^{\mathrm{best}}\}} \bar{w}_j$, so $\varepsilon_{\bar{w}}^{\mathrm{num}}$ is the smallest common value satisfying (9) for the best-ranked elements of all three groups. This a posteriori diagnostic leaves the weights, search directions, and numerical iterates unchanged. The additional condition $\varepsilon_{\bar{w}}^{\mathrm{num}} < t/(1+t)$ in Lemma 1 is a sufficient condition for a strictly positive within-group angle margin and is not imposed in the numerical implementation.

Comparison of `DAES` Variants

We denote the full method, which includes all proposed components, by `DAES`, and consider two ablation variants: `DAES-ptrec` (without triangular recombination) and `DAES-extrap` (without extrapolation). The ablation without symmetric sampling is

omitted since its impact on the profiles is negligible. Figure 2 shows the corresponding performance and data profiles, where higher curves indicate better performance.

The performance profile shows that the full `DAES` consistently outperforms both ablation variants. In particular, at $\tau = 1$, it solves a substantially larger fraction of problems, indicating improved efficiency. The `DAES-ptrec` variant performs uniformly worse, showing that removing triangular recombination degrades both efficiency and robustness. The deterioration is more pronounced for `DAES-extrap`, whose curve remains lowest across almost all τ, indicating that extrapolation is crucial for overall performance.

The data profile confirms the same conclusion. Although all three variants behave similarly under very small computational budgets, the full `DAES` method becomes clearly superior as the budget increases and eventually solves a substantially larger fraction of the test problems. The `DAES-ptrec` variant shows intermediate performance, whereas `DAES-extrap` achieves the lowest final solved fraction. Overall, these results indicate that removing either of the two main components of the method, triangular recombination or extrapolation, can significantly degrade both efficiency and robustness. The best and most balanced performance is therefore obtained when these two mechanisms are used together in the full `DAES` algorithm.

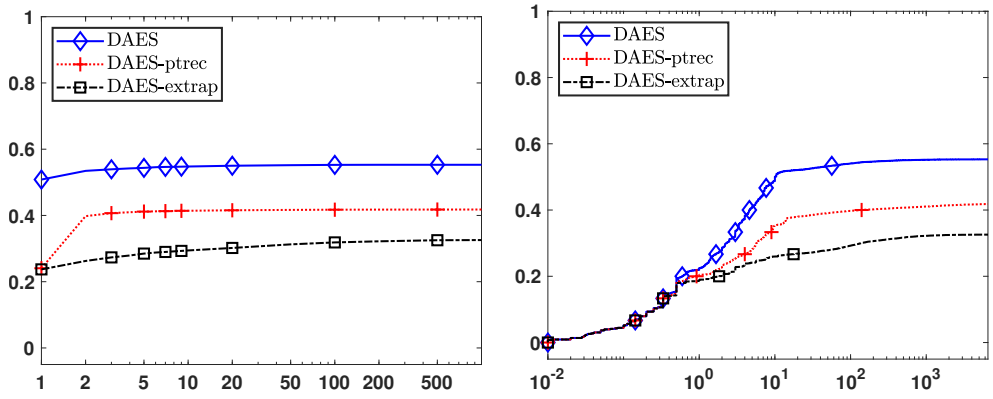

Fig. 2 Performance (left) and data (right) profiles of `DAES` variants in terms of `nf`, using the objective-quality criterion $q_{\mathrm{sol}} \leq \varepsilon_q = 10^{-4}$. The performance profile plots $\rho(\tau)$ versus the performance ratio τ, whereas the data profile plots $\delta(\kappa)$ versus the cost ratio κ.

A comparison between `DAES` *and* `MADFO`

Figure 3 compares the full `DAES` method with `MADFO`. The performance profile shows that `DAES` has a substantially larger value at $\tau = 1$, meaning that it attains the best observed cost on a larger fraction of test instances. This indicates that `DAES` is more efficient than `MADFO` on the problems it solves fastest. As τ increases, however, `MADFO` gradually catches up with and eventually overtakes `DAES`, reaching a higher

final fraction of solved problems. This shows that MADFO is more robust overall when larger performance ratios are allowed. The data profile gives a consistent qualitative conclusion. After the smallest-budget regime, DAES solves a larger fraction of problems over a broad range of small and moderate values of κ, whereas MADFO eventually overtakes it and reaches a higher final solved fraction for sufficiently large budgets.

Overall, the comparison suggests that DAES achieves the intended trade-off. It is deliberately simpler and more suitable for complexity analysis, yet it remains competitive with the more heuristic MADFO solver. In particular, DAES exhibits a clear efficiency advantage at $\tau = 1$ and over a broad range of practically relevant evaluation budgets, while MADFO achieves greater ultimate robustness when substantially larger budgets are allowed.

The differences can be attributed to their structural design. MADFO is a heuristic MAES-type method with adaptive matrix and noise-robust mechanisms. In contrast, DAES removes matrix adaptation and relies on symmetric sampling, grouped selection, triangular recombination, and extrapolation, targeting a simpler and more analyzable structure with competitive performance.

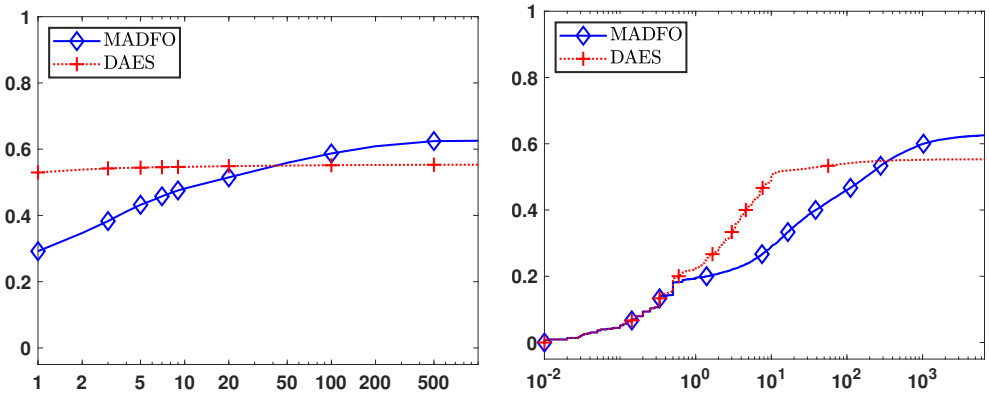

Fig. 3 Performance (left) and data (right) profiles of the full DAES method and MADFO in terms of nf, using the objective-quality criterion $q_{\text{sol}} \leq \varepsilon_q = 10^{-4}$. The performance profile plots $\rho(\tau)$ versus the performance ratio τ, whereas the data profile plots $\delta(\kappa)$ versus the cost ratio κ.

5 Conclusion

We proposed DAES, a simplified MAES for noisy derivative-free optimization that fixes the adaptation matrix to the identity and uses all sampled mutation directions, reducing computational cost while preserving robustness to noise. The main algorithmic contribution of DAES lies in its mutation, sorting–selection, and recombination mechanisms. Symmetric sampling is used to improve directional balance under noise. In the sorting–selection phase, all sampled points are ranked and partitioned into three

groups instead of discarding nonselected ones. From these groups, three recombination directions are constructed and combined via a triangular recombination scheme, yielding a structured search direction that is more informative than classical `MAES` selection rules.

We also develop a high-probability complexity analysis under bounded noise. In particular, we show that the triangular recombination direction satisfies a probabilistic angle condition with the true gradient, which enables sufficient descent arguments with high probability. Based on this, we derive iteration and function-evaluation complexity bounds for nonconvex, convex, and strongly convex problems.

The numerical experiments on `prince` problems from the `BARON` collection support the proposed design. The ablation study shows that the best performance is obtained when triangular recombination and extrapolation are used together, while removing either component degrades performance. Compared with `MADFO`, `DAES` is more efficient on problems solved with small budgets, whereas `MADFO` is more robust under larger budgets, reflecting their different design goals: heuristic robustness versus simplicity and theoretical tractability.

Funding 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.

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)

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

[4] Berahas, A.S., Byrd, R.H., Nocedal, J.: Derivative-free optimization of noisy functions via quasi-newton methods. SIAM Journal on Optimization **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)

[12] Bandeira, A.S., Scheinberg, K., Vicente, L.N.: Convergence of trust-region methods based on probabilistic models. SIAM Journal on Optimization **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 Journal on Optimization **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 Journal on Optimization **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.: Design principles for matrix adaptation evolution strategies. In: Genetic and Evolutionary Computation Conference Companion, vol. 2, pp. 682–700. ACM, Cancún Mexico (2020). https://doi.org/10.1145/3377929.3389870

[25] Beyer, H.-G., Sendhoff, B.: Simplify your covariance matrix adaptation evolution strategy. IEEE Transactions on Evolutionary Computation **21**(5), 746–759 (2017)

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

[27] 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

[28] 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

[29] Ma, K., Rios, L.M., Zheng, H., Sahinidis, N.V., Rajagopalan, S.: Model-and-search: a derivative-free local optimization algorithm. Computational Optimization and Applications **92**(3), 889–921 (2025) https://doi.org/10.1007/s10589-025-00686-9

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

[31] 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

[32] Hansen, N.: The CMA evolution strategy: A tutorial. preprint arXiv:1604.00772 (2016)

[33] C. Elster, A.N.: A grid algorithm for bound constrained optimization of noisy functions. IMA Journal of Numerical Analysis **15**(4), 585–608 (1995) https://doi.org/10.1093/imanum/15.4.585

[34] Gratton, S., Royer, C.W., Vicente, L.N., Zhang, Z.: Direct search based on probabilistic descent. SIAM Journal on Optimization **25**(3), 1515–1541 (2015)

[35] Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization, vol. 87. Springer, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-1-4419-8853-9

[36] Cartis, C., Gould, N.I.M., Toint, P.L.: Evaluation Complexity of Algorithms for Nonconvex Optimization: Theory, Computation and Perspectives. Society for Industrial and Applied Mathematics, Philadelphia (2022). https://doi.org/10.1137/1.9781611976991

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

[38] 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