Revisiting Implicit and Explicit Averaging for Noisy Optimization

Ali Ahrari*−, *Member, IEEE* Saber Elsayed*, *Senior Member, IEEE,* Ruhul Sarker*, *Member, IEEE,* Daryl Essam*, *Member, IEEE,* and Carlos A. Coello Coello+, *Fellow, IEEE*

***Abstract*—Explicit and implicit averaging are two well-known strategies for noisy optimization. Both strategies can counteract the disruptive effect of noise; however, a critical question remains: which one is more efficient? This question has been raised in many studies, with conflicting preferences and, in some cases, findings. Nevertheless, theoretical findings on the noisy sphere problem with additive Gaussian noise supports the superiority of implicit averaging, which may have had a strong impact on the preference of implicit averaging in more recent evolutionary methods for noisy optimization. This study speculates that the analytically supported superiority of implicit averaging relies on specific features of the noisy sphere problem with additive noise, which cannot be generalized to other problems. It enumerates these features and designs controlled numerical experiments to investigate this potential reliance. Each experiment gradually suppresses one specific feature, and the progress rate is numerically calculated for different values of the sample size given a fixed evaluation budget. Our empirical results indicate that for a wide range of noise strength and evaluation budget per iteration, the more these specific features are suppressed, the more the optimal averaging strategy deviates from implicit towards explicit averaging, which confirms our speculations. Consequently, the optimal sample size, which is regarded as the trade-off between implicit and explicit averaging, depends on the problem characteristics and should be learned during optimization for maximum efficiency.**

***Index Terms*—Continuous optimization, uncertainty, evolutionary algorithm, noisy problem**

I. Introduction

Many real-world optimization problems are subject to uncertainty, meaning that the quality or feasibility of a candidate solution cannot be calculated with certainty. This uncertainty may affect the decision parameters; for example, the fabrication precision for a product has limited accuracy [1]. This means the decision parameters in the fabricated products from a unique solution will follow random distributions. A good solution should be robust to such uncertainties, and perform well if the implemented solution (slightly) deviates from the computationally selected one, or when the working conditions are variable. *Robust optimization* [2], [3] and *reliability-based optimization* [4] can find such solutions for these problems, which frequently arise in different fields such as engineering design [5] and operations research [3].

In some other problems, uncertainty affects the assessment of a solution, which means it cannot be accurately evaluated. Finding the optimal solution(s) to these problems requires solving a *noisy optimization* problem [6]. Such problems may arise in different applications, such as machine learning [7], [8], chemistry and material sciences [9], and online optimization of feedback controllers [10].

Robust optimization and noisy optimization share many similar features and challenges; however, there is one significant difference between their goals. In robust optimization, the goal is to find the *global robust optimizer*, the one with the best expected or average fitness with respect to the given uncertainties. The robust optimizer may slightly or substantially depend on the distribution of uncertainties. In contrast, in noisy optimization, the goal is to find the best solution for the noiseless problem; however, the presence of uncertainties in the calculated fitness makes this task more difficult.

Rakshit et al. [6] classified strategies for handling noise into five groups: (i) explicit averaging, (ii) implicit averaging, (iii) alternative fitness estimation methods, (iv) specialized search strategies, and (v) robust selection. Among them, averaging techniques are robust, simple and can be easily integrated with any population-based search method.

In explicit averaging, also known as *resampling*, a candidate solution is evaluated κ times independently, in which κ is known as the sample size [6]. The true fitness is estimated using some statistical indicators. The mean of these κ values is a commonly used statistic for this purpose [6], although the median is a more robust statistic, especially when the distribution of the evaluated fitness has a heavy-tail or includes outliers [11], [12]. Given a population size of λ, the required number of function evaluations per iteration is $FEpI = \kappa\lambda$. In contrast, implicit averaging does not use resampling but uses the largest possible population size for a given *FEpI*, which is $\lambda = FEpI$. Implicit averaging mitigates the disruptive effect of noise by assessing more sample solutions instead of lowering the noise. This can be particularly beneficial when recombination is used since it moderates the erroneous effect of noise in selection [10].

Although there is general agreement that both averaging techniques can improve performance in noisy problems, there remains a crucial question that has been raised in many studies on noisy optimization: *which averaging strategy is more beneficial* [13], [14], [15], [16]? This means that the given *FEpI* (which may change after each iteration) should

*Authors are with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT, Australia (E-mail: aliahrari1983@gmail.com, {s.elsayed; r.sarker; d.essam}@adfa.edu.au)

−The author is now with the School of Aerospace, Mechanical, and Mechatronics Engineering, University of Sydney, Sydney, NSW, Australia (E-mail: ali.ahrari@sydney.edu.au)

+The author is with is with CINVESTAV-IPN, Departamento de Computación, Mexico City, Mexico; Basque Center for Applied Mathematics (BCAM) & Ikerbasque, Spain (e-mail: ccoello@cs.cinvestav.mx)

be allocated to provide rough fitness estimations of many solutions (implicit averaging) or more accurate fitness estimations of fewer solutions (explicit averaging). There have been contradicting preferences, and in some cases, answers, to this question [17], [8].

Theoretical studies on simple noisy problems, such as the sphere function, generally favor implicit averaging for population-based methods with a reasonable parent to offspring size ratio (e.g., between 0.25 and 0.5) [13], [18], [19]. A similar conclusion has been drawn for the more general case of quadratic problems [1]. For robust optimization, however, the theoretical findings in [20] revealed the benefits of explicit averaging with moderate values of κ for a specific test function. More recent sophisticated methods for noisy optimization generally employ implicit averaging [14], [21], which is potentially motivated by existing theoretical findings that favor implicit averaging for noisy quadratic problems and with different noise models [1], [19].

This study provides empirical evidence that the superiority of implicit averaging for noisy sphere problems with additive noise, which is supported by theoretical methods [19], [1], depends on specific properties of this problem, some of which might be absent in many other problems. It discriminates these properties and designs and performs controlled experiments which reveal that by gradually suppressing *any* of these properties, the optimal sample size is likely to gradually increase for a fixed *FEpI*.

The rest of this article is organized as follows: Section II reviews related theoretical and empirical findings in noisy optimization. Section III provides preliminaries related to this work. Section IV discusses our empirical methodology. Section V performs controlled experiments to explore the dependency of the optimal sample size on specific properties of the sphere problem. Finally, Section VI summarizes the empirical findings and draws conclusions.

II. Related Studies

Explicit averaging is a popular and general technique that is commonly used in the literature for handling evaluation noise by reducing noise strength [22], [10]. Assuming that the noise follows a normal distribution with standard deviation of σ_ϵ, mean of κ independent evaluations of the solution $\boldsymbol{x}$ has a standard deviation of $\sigma_\epsilon/\sqrt{\kappa}$. Moreover, resampling can provide an estimate of the noise strength of each solution. It can also reveal the impact of noise, e.g., by analyzing the change in the rank of solutions after resampling [10].

Another advantage of explicit averaging is that it is not necessary to reevaluate all sampled solutions κ times. Dynamic sampling strategies [6] allocate different sample sizes to different solutions. For example, Branke [23] investigated different averaging options and suggested either reevaluation of the best solutions κ times or estimating a weighted fitness based on a large archive of evaluated solutions. Adaptive sampling strategies make a more efficient use of *FEpI* by gradually increasing κ for certain solutions until a sufficiently reliable selection can be made [24], [17], [16]. A comprehensive review of dynamic sampling strategies can be found in [6], [25].

When the population has only one parent, theoretical studies on optimizing simple noisy problems with evolution strategies [19] and genetic algorithms [22] support the advantages of explicit over implicit averaging. In contrast, theoretical analyses in [19], [18], [1] demonstrate the superiority of implicit averaging for multi-parent evolution strategies on quadratic functions and different noise distributions, as long as the ratio of parents to offspring is selected properly. The analysis in [13] for binary genetic algorithms suggested that increasing the population size is always a better choice than increasing κ, assuming that the computational cost of the optimization mainly originates from solutions' evaluations.

A few studies performed a theoretical analysis of the optimal κ for robust optimization of functions with noise induced multimodality [20], [2]. In these functions, the global optimum is not affected by the noise parameter if it is below a certain threshold. For a noise parameter greater than this threshold, local optima emerge, and the robust global and local optima become dependent on the value of the noise parameter. Four different functions in this class, denoted by f_1, f_2, f_3, and f_4 were introduced and analyzed in different studies [20], [2], [26]. These analyses employed steady state behavior, the state in which both control and decision parameters fluctuate around their expected values, as a measure for the success of optimization. For f_4, both averaging strategies could provide an arbitrarily exact approximation of the robust global optimizer given a sufficiently large *FEpI*; however, implicit averaging emerged as the most efficient choice, and thus, resampling was discouraged [20].

A similar analysis of f_2 [2], [26] resulted in completely different conclusions. For this function, neither the simple evolution strategy nor the canonical genetic algorithm could reach the global optimum, even when $\lambda \to \infty$ [2]. For the evolution strategy, an ad hoc setting of $\mu/\lambda = 0.61$, could approximate the global optimum; however, it is far from the recommended setting for noiseless problems ($\mu/\lambda = 0.25$) and general noisy problems ($0.25 \leq \mu/\lambda \leq 0.5$) [20]. A more general alternative was to use a greater but moderate value for the sample size ($\kappa \leq 10$) [2], [27]; Nevertheless, increasing κ emerged as an inefficient alternative which should be used only when implicit averaging may not reach the global optimizer when $\lambda \to \infty$ [27].

Bosman et al. [14] reported little benefit from explicit averaging when using a relatively sophisticated optimization method on complex noisy problems of BBOB'2009 [28], even when the cost of reevaluation was not considered. More importantly, for noise models that do not follow the central limit theorem, e.g., the Cauchy noise, explicit averaging turned out to be detrimental since it increased the noise strength. Considering the findings in [11] and [12], one potential alternative to this challenge is to use the median of the calculated fitness values, instead of the mean.

Implicit averaging has thus been preferred in more recent evolutionary methods for noisy optimization. For example, explicit averaging was avoided in the BI-Population Covariance Matrix Adaption Evolution Strategy [21] while an increasing population size mechanism was kept active. Implicit averaging can also employ heuristics to adapt (generally increase) the population size whenever it is required or beneficial. For example, the population adapting rule proposed in [29] increases the population size when there is no improvement in the average fitness of the parent population. The population control mechanism in [15] increases the population size when there is no statistically significant improvement in the parental centroid fitness. Li et al. [30] utilized the noise quantification strategy introduced in [10] and proposed a population sizing rule which increases the population size when the normalized quantified noise is greater than zero and vice versa.

III. Preliminaries

This section briefly reviews preliminary notions and definitions which are used in this study.

A. Problem Formulation

For a predefined search space $\mathbf{S}$, the goal of noisy optimization is to find the global minimum $\boldsymbol{x}^* \in \mathbf{S}$, which optimizes the true objective function ($f(\boldsymbol{x})$). The minimization case is considered in this study without loss of generality:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbf{S}} f(\boldsymbol{x}). \tag{1}$$

In contrast to deterministic optimization, $f(\boldsymbol{x})$ cannot be calculated for an arbitrary $\boldsymbol{x}$. Instead, evaluation of $\boldsymbol{x}$ returns $\hat{f}(\boldsymbol{x}) = f(\boldsymbol{x}; \epsilon)$, a noisy estimate for $f(\boldsymbol{x})$ in which ϵ is the noise parameter. Each evaluation of $\boldsymbol{x}$ will return a different value for $\hat{f}(\boldsymbol{x})$, the standard deviation of which is referred to as the *noise strength* and denoted by σ_ϵ [31]. Since the goal of noisy optimization is to find the global minimum of the true objective function, $\boldsymbol{x}^*$ is independent of σ_ϵ. This is in contrast to robust optimization in which the robust global minimizer can be affected by the noise type and strength.

B. Noise Types

Two common noise types are considered: additive noise and fitness proportionate. Both noise types employ a Gaussian model. Additive Gaussian noise is the most commonly used noise model in the literature [32], according to which the noise strength is independent of the solution:

$$\hat{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon_{\text{A}}\,\mathcal{N}(0,1), \tag{2}$$

in which $\mathcal{N}(0,1)$ is a random number sampled from a standard normal distribution and $f(\boldsymbol{x})$ is the true fitness of solution $\boldsymbol{x}$. ϵ_{A} controls the noise strengths in which the the subscript "A" indicates that the noise model is additive. For additive noise, $\sigma_\epsilon = \sigma_{\text{A}}$.

Fitness proportionate noise [32], also known as multiplicative noise [15], may happen in certain applications, such as measuring devices, in which the accuracy is roughly a percentage of the measurement [32]:

$$\hat{f}(\boldsymbol{x}) = f(\boldsymbol{x})\left(1 + \epsilon_{\text{P}}\,\mathcal{N}(0,1)\right), \tag{3}$$

in which ϵ_{P} controls the strength of the fitness-proportionate noise.

C. Signal-to-Noise Ratio

It should be noted that noise only affects the selection step in the evolution process. The impact of the noise depends on its strength in comparison with the true difference between the solution values (signal). This is referred to as *signal-to-noise-ratio* (SNR) [10]. If SNR is small, even if the noise strength itself is great, the impact of noise on the selection operator will be small. This notion provides a useful means to analyze the detrimental impact of noise.

D. Evolution of the Population

Evolution strategies have been shown to be more robust to noise than other direct search methods [33]. This study employs $(\mu/\mu_{\text{I}}, \lambda)$-ES with isotropic mutation to evolve the population. At each iteration, λ solutions are generated by mutation of the population center ($\boldsymbol{x}_{\mathbf{C}}$) using a normal distribution with standard deviation of σ:

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_{\mathbf{C}} + \sigma\,\mathcal{N}_D(\mathbf{0},\mathbf{1}), i = 1, 2, \ldots, \lambda, \tag{4}$$

in which D is the problem dimensionality and $\mathcal{N}_D(\mathbf{0},\mathbf{1})$ is a vector of D independent numbers sampled from a standard normal distribution. σ is the mutation strength, also known as the *step size*. These λ solutions are evaluated and sorted according to their realized fitness. Then, the μ-best solutions are selected and recombined to update $\boldsymbol{x}_{\mathbf{C}}$ for the next generation:

$$\boldsymbol{x}_{\mathbf{C}} \leftarrow \left(\frac{1}{\mu}\right)\sum_{i=1}^{\mu} \boldsymbol{x}_i \tag{5}$$

This type of recombination in known as *global intermediate recombination*. The non-elitist selection of parents is particularly useful for noisy optimization since the error of overvalued solutions may not propagate for more than one iteration [10]. In this study, the truncation ratio (μ/λ) is set to 0.25. This choice is supported by theoretical results in [19] and the default setting for the (effective) number of parents in the most successful evolution strategies for noiseless problems [34], [35]. For the noisy sphere problem, the optimal truncation ratio increases with noise strength until it reaches $\mu/\lambda = 0.5$ [19]; however, the optimal value for μ/λ is not known beforehand and any $0.25 \leq \mu/\lambda \leq 0.5$ is thus a reasonable choice.

E. Improvement Measure

Two common improvement measures used in the literature are *progress rate* and *quality gain* [31], [36]. Progress rate measures the improvement in the solution space (i.e., how much closer the population is to the global minimum after one iteration) whereas quality gain measures the improvement in the objective space. The progress rate is the performance indicator used in this study, which is the difference between the distance of the population center ($\boldsymbol{x}_{\mathbf{C}}$) to the global minimum, before and after one iteration. Starting with the population center at $\boldsymbol{x}_{\mathbf{C}}^{\text{ini}}$ and the given σ (step size), λ solutions are sampled according to (4). Then, the center of the population is updated according to (5). This new center is denoted by $\boldsymbol{x}_{\mathbf{C}}^{\text{fin}}$. The progress rate is then calculated as follows:

$$\phi = \|\boldsymbol{x}_{\mathbf{C}}^{\text{ini}} - \boldsymbol{x}^*\| - \|\boldsymbol{x}_{\mathbf{C}}^{\text{fin}} - \boldsymbol{x}^*\|. \tag{6}$$

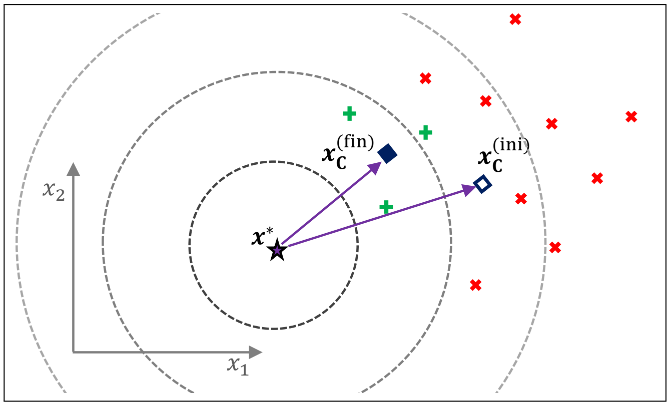


Fig. 1. Schematic illustration of progress rate ϕ. Cross and plus marks show the sampled solutions generated by the mutation of the initial population center ($\boldsymbol{x}_{\mathbf{C}}^{\text{ini}}$). Plus marks represent the selected parents used to calculate the new population center ($\boldsymbol{x}_{\mathbf{C}}^{\text{fin}}$). Progress rate is the reduction in the distance of the population center to the global minimum ($\boldsymbol{x}^*$) after one iteration: $\phi = \|\boldsymbol{x}_{\mathbf{C}}^{\text{ini}} - \boldsymbol{x}^*\| - \|\boldsymbol{x}_{\mathbf{C}}^{\text{fin}} - \boldsymbol{x}^*\|$.

Fig. 1 depicts the calculation of the progress rate schematically. Twelve solutions were generated by mutation of the current population center ($\boldsymbol{x}_{\mathbf{C}}^{\text{ini}}$). Three solutions with the best values were then selected as the parents and the new population center, $\boldsymbol{x}_{\mathbf{C}}^{\text{fin}}$, is calculated as the centroid of the selected parents. The progress rate is the difference between the lengths of the illustrated vectors.

IV. Methodology

Analytical methods can provide useful information on intervening factors. Some well-known heuristics were initially developed using theoretical analyses of simple problems [31]. Theoretical models, however, are limited to simple problems like the sphere function [32] or the OneMax problem [22], [37]. They may also use some simplifications, such as very large problem dimensionality or linearization of the landscape [32]. Another limitation of theoretical methods is that although the derived equations show the impact of each factor, it does not provide an upper-level insight into what happens and why it happens during the population evolution which results in this impact. Such insight can be very helpful to predict whether the findings from simple problems can be generalized to different and/or more complex problems.

This study employs highly controlled numerical experiments in which the rate of improvement, the improvement over one iteration, is measured and analyzed. Such a measure for improvement has two advantages over the measures that are defined over a longer course of evolution, e.g., the steady state behavior, which has been used in [2], [27], [20]:

- It isolates the effect of the employed mechanism for step size adjustment since an exhaustive search can be performed to find a near-optimal step size.
- the optimal value of κ is likely to change during the optimization process. For example, for a noisy problem with additive Gaussian noise, the difference between the true values of sampled solutions (signal) is great when the step size is large; therefore, resampling is not a rational choice since selection noise is already insignificant. In contrast, the step size is generally small when the population center approaches the (global) minimum. In this case, the signal is small and resampling becomes a reasonable choice since it can reduce the noise and thus improves the selection's reliability.

The rate of improvement has the limitation of measuring local performance; however, for spherical problems, which include all the problems considered in this study, local and global improvement are monotonously related.

For a given *FEpI*, we search for the optimal trade-off between implicit and explicit averaging. The trade-off parameter $0 \leq c_\kappa \leq 1$ is defined as follows:

$$c_\kappa = \frac{\log_2(\mathit{FEpI}/\lambda)}{\log_2(\mathit{FEpI}/\lambda_{\min})} = \frac{\log_2 \kappa}{\log_2 \kappa_{\max}}, \tag{7}$$

in which $\lambda_{\min} = 4$ is the smallest reasonable value for λ when the truncation ratio is 0.25 and $\kappa_{\max} = \mathit{FEpI}/\lambda_{\min}$. $c_\kappa = 0$ and $c_\kappa = 1$ means completely implicit and completely explicit averaging, respectively. Any other value indicates a trade-off between implicit and explicit averaging, where a greater c_κ is associated with a more explicit averaging and a greater sample size.

Static sampling is used in this study, according to which every solution is evaluated κ times. The optimal c_κ is the one that maximizes the progress rate ϕ for the given *FEpI*. It should be noted that a greater κ means a smaller λ since $\lambda\kappa = \mathit{FEpI}$. The true fitness is estimated as the mean of these κ independent evaluations. In particular, we are interested in investigating the optimality of $c_\kappa = 0$ when special features of the sphere problem are disturbed, which will be discussed in Section V.

V. Controlled Numerical Experiments

As discussed in Section I, theoretical findings strongly favor implicit averaging for the sphere problem with additive noise given that mutation strength is set to its optimal value. This can be explained as follows: For a greater λ, the optimal σ is greater [36], resulting in a greater variance of the sampled solutions. The true difference between the solutions, the signal, is thus greater while the noise strength is constant. Therefore, a higher λ is associated with a greater SNR. In contrast, explicit averaging aims at reducing noise to increase SNR. We speculate that the reported advantages of implicit averaging rely on exploiting specific features of the sphere or quadratic problem. More specifically:

- for the sphere problem, the objective function increases fast (quadratically) with respect to the distance from the global minimum. This results in a stronger signal if sample solutions are farther from the global minimum.
- the noise is independent of the solution
- spherical problems are symmetric around the global minimum.
- the (near-) optimal σ is provided for the population.

This section hypothesizes how each feature contributes to the optimality of implicit averaging according to theoretical findings. It designs controlled simulations to explore whether

$c_\kappa = 0$ remains as the best choice if these specific features, one at a time, are disturbed.

For a predefined *FEpI*, the progress rate is calculated for different values for $0 \leq c_\kappa \leq 1$ when $\boldsymbol{x}_\mathbf{C} = \mathbf{1}$. The tested values of c_κ correspond to $\kappa = 1, 2, 4, 8, \ldots, \kappa_{\max}$. For every value of c_κ, σ is set to its near-optimal value denoted by σ_{best} unless otherwise stated. This near-optimal value is calculated by performing an exhaustive search such that σ_{best} is estimated with 10% accuracy, assuming that ϕ is a unimodal function of σ. The lowest value for σ_{best} is 0.01. Each numerical simulation is repeated 5000 times and the results are averaged to calculate σ_{best} and ϕ. For all the simulations, $D = 10$ and *FEpI* $= 1024$ unless otherwise stated.

For the simulations involving additive noise only, we evaluated each solution one time but set the noise parameter to $\epsilon_\text{A}/\sqrt{\kappa}$. This is equivalent to the mean of κ independent evaluations of a solution when the noise strength is ϵ_A, but the former is computationally κ times cheaper.

A. Rate of Increase in Signal for Farther Solutions

In the sphere problem, the true values of the solutions far from the global optimum increase quadratically as we move away from the global minimum. This means that the signal, and thus SNR, increases rapidly while the noise strength does not change if it is additive noise. To study the impact of this factor, the following test problem is considered:

$$f(\boldsymbol{x}; p) = \|\boldsymbol{x}\|^p. \tag{8}$$

Parameter p controls how fast the objective value increases for farther solutions. $p = 2$ results in the well-known sphere function. The additive noise model is considered for this problem.

Fig. 2 illustrates ϕ as a function of c_κ and ϵ_A for different values of p. The values of ϵ_A were selected to have diverse and reasonable values for ϕ when $c_\kappa = 0$ (completely implicit averaging). It can be observed that:

- for $p = 2$, a greater λ (a smaller c_κ) is always a better choice. The slope of ϕ is also negative at $c_\kappa = 0$, which indicates the high contribution of a smaller c_κ. This observation completely agrees with the theoretical findings.
- for $p = 1$ and $p = 4$, a smaller c_κ is still a better choice for all the tested values of ϵ_A; however, when compared with $p = 2$, there is a noticeable difference in the slope of the graphs: For a greater p, the progress rate declines much faster when c_κ is increased. This trend is more noticeable for a greater ϵ_A. In contrast, when $p = 1$ and $\epsilon_\text{A} \geq 1.5$, the gain from reduction of c_κ is minor, especially when $c_\kappa \leq 0.25$.
- for $p = 0.5$, $p = 0.25$, and $p = 0.125$, there is a detectable optimal value for c_κ, which is off the extreme values, unless ϵ_A is very small. For a sufficiently small value of ϵ_A, the effect of the noise is not strong enough to (significantly) mislead the ranking of the solutions, and thus, reevaluation does not provide any remarkable contribution. As ϵ_A increases, the optimal c_κ increases as well.
- As expected, σ_{best} is greater for a smaller c_κ; however, it is also observed for small values of c_κ, σ_{best} reduces when p decreases. This can be explained as follows: sampling solutions farther from the global minimum improves the signal less for a smaller p, and thus, the improvement in SNR is less. Therefore, the advantage of a greater σ diminishes when p is reduced.
- When $c_\kappa = 0$ (implicit averaging) and $p = 4$, there is a noticeable direct relationship between σ_{best} and ϵ_A. Although not shown here, based on similar simulations, we calculated $\sigma_{\text{best}} = 3.68$ for the noiseless problem ($\epsilon_\text{A} = 0$). This indicates that for a greater ϵ_A, the benefits of an increased signal pay off for the drawbacks of the deviation from the optimal step size of the noiseless problem. This trend diminishes or disappears for smaller values of $p = 2$

The obtained results from this experiment confirm that the optimality of $c_\kappa = 0$ relies on a fast increase in the objective function when moving away from the global minimum. When this increase is not that fast (e.g., the objective function increases sub-linearly with the distance to the global minimum), the optimal c_κ increases with ϵ_A.

B. Noise Type Effect

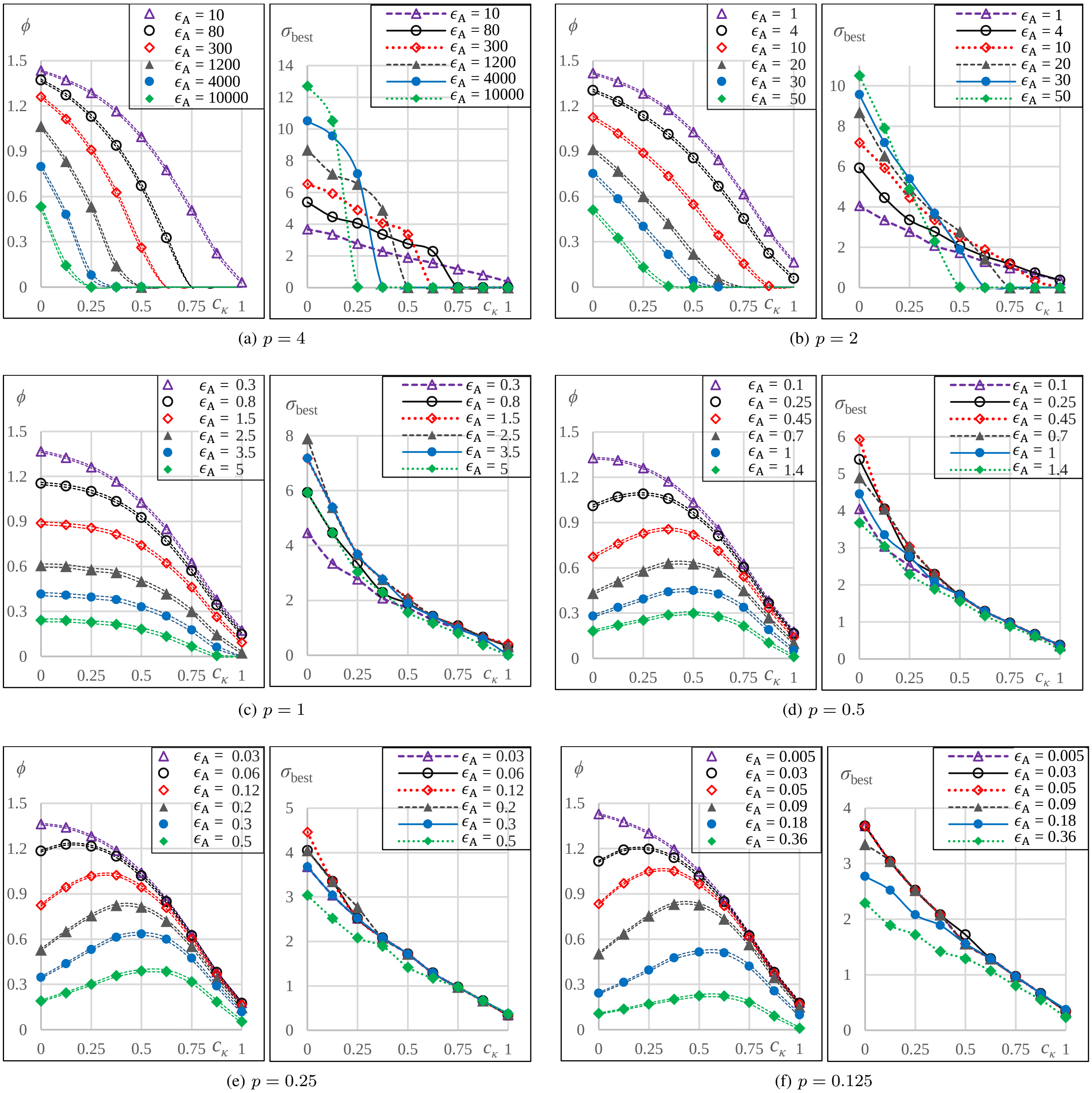
The previous simulation revealed that the success of $c_\kappa = 0$ could be explained by the rapid increase in the signal when the noise strength remains constant. This condition may vanish if the noise strength grows with the signal. In such situations, a greater σ is associated with not only a greater signal but also a greater noise; therefore, it does not improve SNR. To check the effect of this factor, we employ the sphere function (test problem in (8) with $p = 2$) with fitness proportionate noise (see III-B). Fig. 3 illustrates ϕ (mean and 95% confidence interval) and σ_{best} as a function of ϵ_P and c_κ. It can be observed that:

- unless ϵ_P is very small, the optimal value of c_κ is greater than zero. More importantly, this value is greater for a greater ϵ_P. For example, for $\epsilon_\text{P} = 0.3$ and $\epsilon_\text{P} = 4$, the optimal c_κ is about 0.125 and 0.75, respectively.
- for $c_\kappa = 0$, σ_{best} rapidly reduces when ϵ_P increases such that for $\epsilon_\text{P} \geq 2$, $\sigma^{\text{best}} \approxeq 0$, indicating that no positive progress can be made. This indicates that a greater σ does not help the implicit averaging anymore. For a greater value of c_κ, σ_{best} is less affected by ϵ_P.

In summary, when the noise also grows with the signal, $c_\kappa = 0$ is generally not the best choice. A greater ϵ_P increases the optimal c_κ. Although the case of fitness-proportionate noise was tested, a qualitatively similar trend is predictable for other noise types in which the noise strictly increases with fitness. It is also predicted that a noise model in which the noise strength grows faster with fitness favors a greater c_κ.

C. Symmetry in the Landscape

One remarkable feature of the spherical problems is the symmetry of their landscapes around the global minimum. A selection strategy that averages a large number of parents can exploit this specific feature.



(a) $p = 4$

(b) $p = 2$

(c) $p = 1$

(d) $p = 0.5$

(e) $p = 0.25$

(f) $p = 0.125$

Fig. 2. Progress rate ϕ (mean and 95% confidence interval) and near-optimal mutation strength (σ_{best}) as a function of c_κ and noise strength (ϵ_{A}) for different values of p (see (8)). For ϕ, the dashed lines delineate the 95% confidence interval.

Let us consider a simple problem to analyze this effect. For this example, we generalize the test problem defined in (8) so that it can simulate a skewed landscape:

$$f(\boldsymbol{x}; p, k_{\text{skew}}) = \|\boldsymbol{y}\|^p, \quad y_i = \begin{cases} (1+k_{\text{skew}})x_i & \text{if} \quad x_i < 0 \\ x_i & \text{if} \quad x_i \geq 0 \end{cases}, \tag{9}$$

in which k_{skew} controls the skewness of the fitness landscape around the global minimum. This problem reduces to (8) for $k_{\text{skew}} = 0$.

It is already known that the skewness of the fitness landscape around the global minimum challenges population-based methods, even for noiseless problems [34], [38]. The skewness pushes the population center away from the global minimum since sampled solutions on one side of the global minimum are more likely to be selected for recombination. A simple 1-D problem is illustrated in Fig. 4 for a better visualization of this effect, assuming that the population center lies on the global minimum. Fig. 4a illustrates the plot of this function when $p = 1$ and $k_{\text{skew}} = 0$. Forty solutions (plus and cross marks)

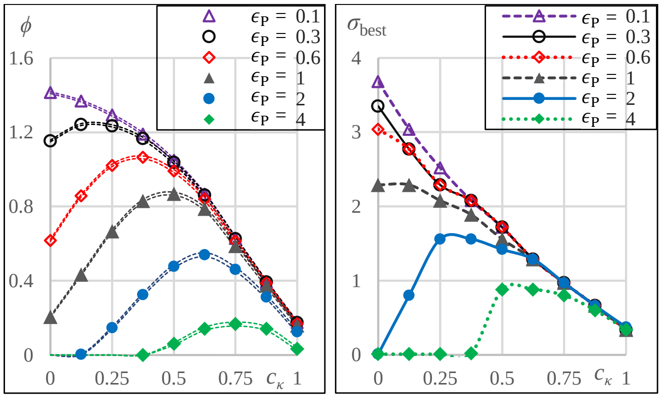
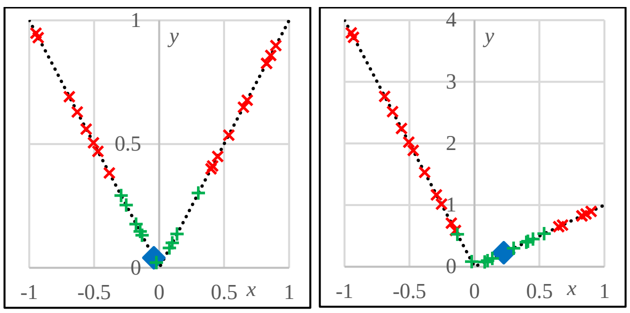


Fig. 3. ϕ and σ_{best} for different values of c_κ and ϵ_{P} for the standard sphere problem. The dashed lines delineate the 95% confidence interval for ϕ.



(a) (b)

Fig. 4. Effect of the landscape skewness on the new $\boldsymbol{x}_{\mathbf{C}}$ (solid square) when 40 solutions (plus and cross marks) have been randomly generated from $\mathcal{N}(0, 1)$. The new $\boldsymbol{x}_{\mathbf{C}}$ is the centroid of the selected solutions (plus marks).

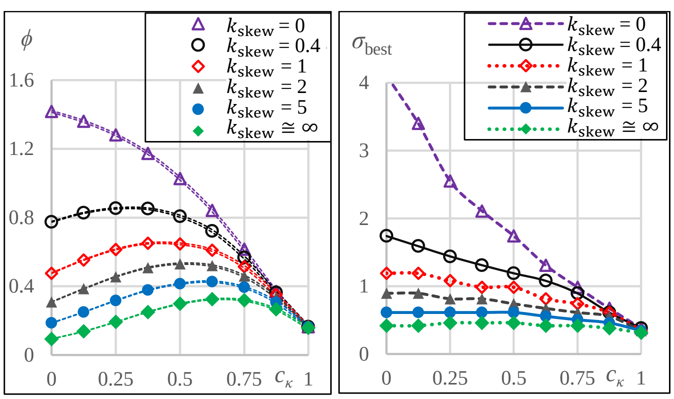


Fig. 5. ϕ and σ_{best} for different values of κ and k_{skew} for the test problem formulated in (9) with additive noise ($\epsilon_{\text{A}} = 1$). For ϕ, the dashed lines delineate the 95% confidence interval.

were generated when $\sigma = 1$. The 10-best solutions were selected as the parents (plus marks) to calculate the new center (solid square). As can be observed, the new center remains close to the global minimum. The slight negative progress is caused by the randomness in the sampling process. Fig. 4b illustrates the outcome of the same process for the same 40 solutions when $k_{\text{skew}} = 3$. The new center now falls on the right side of the global minimum. This negative progress will intensify if k_{skew} is increased because more solutions on the right side of the global minimum will be selected as parents.

This illustration shows that for the skewed landscape, the population center is pushed away from the global minimum if σ is large. This observation raises a question regarding the benefits of implicit averaging ($c_\kappa = 0$) for non-symmetric noisy problems, as it uses a larger σ in symmetric spherical landscapes (see Fig. 2), to arguably increase the signal.

To investigate the dependency of the optimal c_κ on skewness, we consider the problem in (9) with additive noise, 10 variables and $p = 2$. Since the population center $\boldsymbol{x}_{\mathbf{C}} = \mathbf{1}$ is in the first hyperoctant of the search space, and the objective increases faster on the negative sides, we can safely assume that the population center makes progress towards the global minimum from the right, and thus, it remains in the first hyperoctant. Therefore, although the test problem is not spherical, the progress rate still provides a reliable indicator of improvement since the fitness landscape in the first hyperoctant is still a spherical function.

Fig. 5 plots the calculated values of ϕ (mean and 95% confidence interval) and σ_{best} for different values of c_κ and k_{skew}. It can be observed that:

- For a symmetric landscape ($k_{\text{skew}} = 0$), the optimal c_κ is zero. However, when k_{skew} increases, the optimal c_κ also increases.
- A greater skewness diminishes the progress rate regardless of the value of c_κ, but this decline in the progress rate is more drastic for a smaller c_κ.
- For a small k_{skew}, σ_{best} drastically reduces when k_{skew} increases. It should be highlighted that a greater k_{skew} is associated with a greater or equal signal and the noise is additive, so SNR still substantially improves if a greater σ is used. This indicates that the detrimental effect of skewness is stronger than the positive effect of an increased SNR.

This simulation indicates that the success of $c_\kappa = 0$ strongly relies on the presence of symmetry around the global minimum, a special condition that may not always exist in the problem.

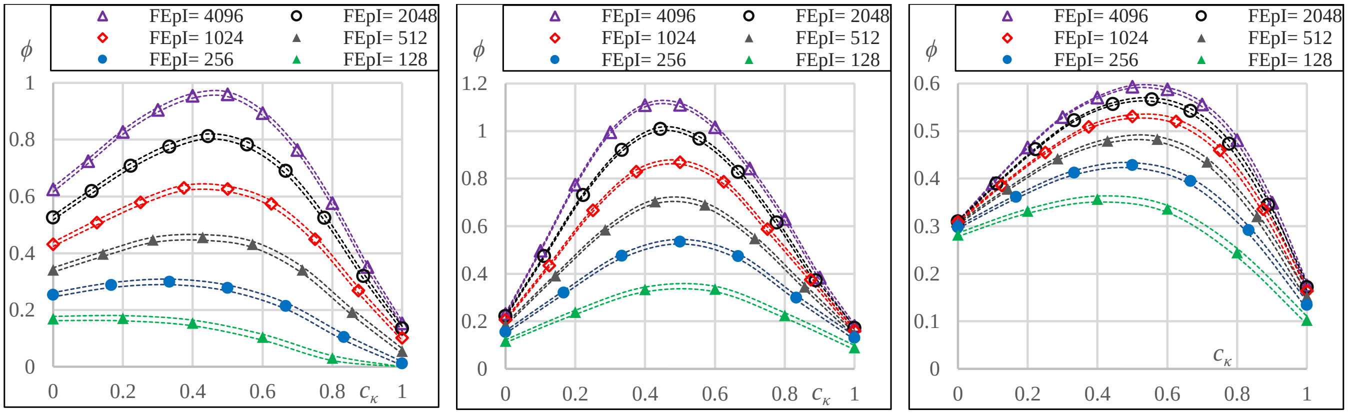
D. Effect of *FEpI*

To check the effect of *FEpI* on the optimal value of c_κ, we select three different problem settings, one from each previous simulation, such that for $FEpI = 1024$, the optimal c_κ lies off the extremes:

- Problem P1: Spherical problem defined in (8) with $p = 0.5$ when $\epsilon_{\text{A}} = 0.7$ (additive noise)
- Problem P2: Spherical problem in (8) with $p = 2$ and $\epsilon_{\text{P}} = 0.7$ (fitness proportionate noise)
- Problem P3: Skewed problem in (9) with $p = k_{\text{skew}} = 2$ and $\epsilon_{\text{A}} = 1$ (additive noise)

Fig. 6 illustrates the effect of *FEpI* and c_κ on ϕ for each problem. As observed,

- for a large range of the tried values for *FEpI*, the optimal c_κ remains off the extremes.
- the optimal c_κ is generally smaller for a smaller *FEpI*. This was predictable since evolutionary algorithms need diversity in their populations; otherwise their exploration



(a) Problem P1 (b) Problem P2 (c) Problem P3

Fig. 6. Effect of *FEpI* and c_κ on ϕ for three problems: P1, P2, and P3. The dashed lines delineate the 95% confidence interval for ϕ.

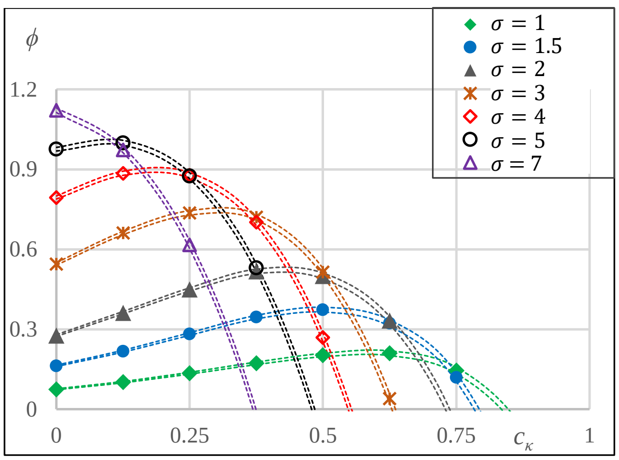


Fig. 7. ϕ for different values of σ and c_κ. The test problem is the one defined in (8) when $\epsilon_A = 10$ and $p = 2$. The dashed lines delineate the 95% confidence interval for ϕ.

will be compromised. Surprisingly, this is not the case for Problem P2 (Fig. 6b). For this problem, the optimal c_κ is almost independent of *FEpI*.

E. Optimality of the Mutation Strength

As discussed in Subsection III-D, evolution strategies employ a (self-) adaptation mechanism to dynamically adjust σ during the optimization process. These mechanisms, however, might not necessarily set σ to its optimal value. For example, self-adaptation suffers from *opportunism* [39], a tendency to set σ to a value smaller than the optimal one, especially for badly-scaled problems. Moreover, the update of σ during the optimization process is gradual and the optimal σ changes dynamically during the optimization process; therefore, even if the (self-) adaptation mechanism can learn the optimal σ, the adjusted σ might always differ from it.

It is already known that for quadratic problems with additive noise, implicit averaging is advantageous if σ is set to the optimal value [10], [1]. This raises a question regarding the dependency of the optimal c_κ on a given σ, which is not a near-optimal value. Fig. 7 illustrates ϕ for different values of c_κ and σ for the problem formulated in (8) when $\epsilon_A = 10$ and $p = 2$. From previous simulations, we already know that when $p = 2$, $c_\kappa = 0$ is the best choice provided that σ is set to its near-optimal value. However, when σ is smaller than σ_{best}, the optimal c_κ will be much greater.

It is noteworthy that a fixed σ means an identical distribution of sampled solutions for all values of c_κ. Therefore, this simulation indicates that more data with less accuracy is not necessarily a better option than less data with high accuracy since the spread of these data also plays a decisive role.

VI. Discussion and Conclusion

For the case of the sphere problem with additive noise and given optimal mutation strength, our simulations completely agree with existing theoretical findings: Implicit averaging is the best choice. However, this study demonstrated that the optimality of this choice relies on the presence of certain and specific features in the sphere problem: i) a rapid increase in the signal (difference between true values of solutions) for regions farther from the global minimum, ii) independence of the noise strength from the objective function (i.e., additive noise), and iii) symmetry of the objective function around the global minimum.

The controlled simulations in this study revealed that if *any* of these features is suppressed, implicit averaging might not remain as the best choice. Furthermore, for a wide range of noise strength and evaluation budget per iteration, the more a problem deviates from these properties, the greater the optimal c_κ will be, which means the optimal averaging strategy becomes more explicit. These trends are qualitatively the same for a large range of *FEpI*, which may be static, dynamic, or adapted during optimization. Besides, as demonstrated in this study, if the available mutation strength is smaller than the optimal one, the optimal c_κ becomes greater. This is particularly important because existing (self-) adaptation strategies have some delay in learning the optimal mutation strength and cannot learn its precise value in certain landscapes. We speculate that suppressing two or more of these factors simultaneously will have similar or even stronger impacts on the optimal c_κ although this study did not consider numerical simulations for this purpose. Nevertheless, if there is no knowledge on the best choice for c_κ before or during the optimization process and a fixed sample size should be used for every solution, our results suggest that $c_\kappa = 0$ is the safest choice.

There are two other factors that favor explicit averaging and are worth further investigation. First, using the median instead of the mean of independent evaluations can enable explicit averaging to deal with noise models with heavy tails [12], [11], a noise model in which explicit averaging fails if the conventional mean indicator is used [14]. Second, it is not necessary to resample all solutions κ times. More efficient strategies that sequentially resample some solutions selectively can save a lot of unnecessary reevaluations, e.g., adaptive sampling techniques [16], [37]. This is particularly useful for global intermediate recombination in which parents participate with equal weights in the recombination process. The selection operator should only determine whether or not a solution is among the μ-best ones with sufficient reliability.

The findings from these simulations reveal a need for efficient and reliable heuristics that can learn the optimal c_κ during the optimization process since, in general, a problem may lack the specific features of the sphere problems, or the noise strength may depend on the solution value. The simulations performed in this study can serve as a checkpoint to confirm the capability of a candidate heuristic in learning the optimal trade-off between implicit and explicit averaging.

The findings in this study also cast doubt on the merits of another theoretically supported strategy for noisy optimization: *re-scaled mutation* [40]. This strategy uses an increased mutation strength for sampling solutions but proportionally reduces the change in the population center. For the sphere problem with additive noise, theoretical findings indicate the positive and drastic impact of this strategy when the noise is sufficiently strong [41]; however, as demonstrated in this study, benefits of a greater diversity in the sampled solutions is contingent upon the presence of specific features in the problem as enumerated and explained in this study. Consequently, this study motivates reevaluation of the advantages of re-scaled mutation when these specific features are suppressed to clarify how this strategy performs in more general situations.

Acknowledgments

This research has been funded by the Australian Research Council Discovery Project DP190102637. Computational resources for this study were provided by the National Computational Infrastructure (NCI), which is supported by the Australian Government. The last author gratefully acknowledges support from CONACyT grant no. 2016-01-1920 (*Investigación en Fronteras de la Ciencia 2016*) and from a SEP-Cinvestav grant (application no. 4). He was also partially supported by the Basque Government through the BERC 2022-2025 program and by the Spanish Ministry of Science, Innovation and Universities: BCAM Severo Ochoa accreditation SEV-2017-0718.

References

[1] H.-G. Beyer and B. Sendhoff, "Robust optimization–a comprehensive survey," *Computer methods in applied mechanics and engineering*, vol. 196, no. 33-34, pp. 3190–3218, 2007.

[2] ——, "Evolutionary algorithms in the presence of noise: To sample or not to sample," in *2007 IEEE Symposium on Foundations of Computational Intelligence*. IEEE, 2007, pp. 17–24.

[3] V. Gabrel, C. Murat, and A. Thiele, "Recent advances in robust optimization: An overview," *European journal of operational research*, vol. 235, no. 3, pp. 471–483, 2014.

[4] M. A. Valdebenito and G. I. Schuëller, "A survey on approaches for reliability-based optimization," *Structural and Multidisciplinary Optimization*, vol. 42, no. 5, pp. 645–663, 2010.

[5] T. Orosz, A. Rassõlkin, A. Kallaste, P. Arsénio, D. Pánek, J. Kaska, and P. Karban, "Robust design optimization and emerging technologies for electrical machines: Challenges and open problems," *Applied Sciences*, vol. 10, no. 19, p. 6653, 2020.

[6] P. Rakshit, A. Konar, and S. Das, "Noisy evolutionary optimization algorithms–a comprehensive survey," *Swarm and Evolutionary Computation*, vol. 33, pp. 18–45, 2017.

[7] C. Qian, Y. Yu, and Z.-H. Zhou, "Analyzing evolutionary optimization in noisy environments," *Evolutionary computation*, vol. 26, no. 1, pp. 1–41, 2018.

[8] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments-a survey," *IEEE Transactions on evolutionary computation*, vol. 9, no. 3, pp. 303–317, 2005.

[9] F. Hase, M. Aldeghi, R. Hickman, L. Roch, E. Liles, M. Christensen, J. Hein, and A. Aspuru-Guzik, "Olympus: a benchmarking framework for noisy optimization and experiment planning," *Machine Learning: Science and Technology*, 2021.

[10] N. Hansen, A. S. Niederberger, L. Guzzella, and P. Koumoutsakos, "A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 1, pp. 180–197, 2008.

[11] C. Bian, C. Qian, Y. Yu, and K. Tang, "On the robustness of median sampling in noisy evolutionary optimization," *Science China Information Sciences*, vol. 64, no. 5, pp. 1–13, 2021.

[12] B. Doerr and A. M. Sutton, "When resampling to cope with noise, use median, not mean," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019, pp. 242–248.

[13] J. M. Fitzpatrick and J. J. Grefenstette, "Genetic algorithms in noisy environments," *Machine learning*, vol. 3, no. 2, pp. 101–120, 1988.

[14] P. A. Bosman, J. Grahl, and D. Thierens, "Benchmarking parameter-free amalgam on functions with and without noise," *Evolutionary computation*, vol. 21, no. 3, pp. 445–469, 2013.

[15] M. Hellwig and H.-G. Beyer, "Evolution under strong noise: A self-adaptive evolution strategy can reach the lower performance bound-the pcCMSA-ES," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2016, pp. 26–36.

[16] E. Cantú-Paz, "Adaptive sampling for noisy problems," in *Genetic and Evolutionary Computation Conference*. Springer, 2004, pp. 947–958.

[17] J. Branke and C. Schmidt, "Selection in the presence of noise," in *Genetic and Evolutionary Computation Conference*. Springer, 2003, pp. 766–777.

[18] D. V. Arnold and H.-G. Beyer, "Local performance of the (μ/μi, λ)-es in a noisy environment," in *Foundations of Genetic Algorithms 6*. Elsevier, 2001, pp. 127–141.

[19] ——, "A general noise model and its effects on evolution strategy performance," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 380–391, 2006.

[20] H.-G. Beyer and B. Sendhoff, "Functions with noise-induced multimodality: a test for evolutionary robust optimization-properties and performance analysis," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 507–526, 2006.

[21] N. Hansen, "Benchmarking a BI-population CMA-ES on the BBOB-2009 noisy testbed," in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, 2009, pp. 2397–2402.

[22] C. Qian, Y. Yu, K. Tang, Y. Jin, X. Yao, and Z.-H. Zhou, "On the effectiveness of sampling for evolutionary optimization in noisy environments," *Evolutionary computation*, vol. 26, no. 2, pp. 237–267, 2018.

[23] J. Branke, "Creating robust solutions by means of evolutionary algorithms," in *International Conference on Parallel Problem Solving from Nature*. Springer, 1998, pp. 119–128.

[24] P. Stagge, "Averaging efficiently in the presence of noise," in *International Conference on Parallel Problem Solving from Nature*. Springer, 1998, pp. 188–197.

[25] P. Rakshit and A. Konar, *Principles in Noisy Optimization*. Springer, 2018.

[26] H.-G. Beyer and M. Hellwig, "Analysis of the pccmsa-es on the noisy ellipsoid model," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2017, pp. 689–696.

[27] H.-G. Beyer and B. Sendhoff, "Toward a steady-state analysis of an evolution strategy on a robust optimization problem with noise-induced multimodality," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 629–643, 2017.

[28] N. Hansen, S. Finck, R. Ros, and A. Auger, "Real-Parameter Black-Box Optimization Benchmarking 2009: Noisy Functions Definitions," INRIA, Research Report RR-6869, 2009. [Online]. Available: https://hal.inria.fr/inria-00369466

[29] H.-G. Beyer and B. Sendhoff, "Evolution strategies for robust optimization," in *2006 IEEE International Conference on Evolutionary Computation*. IEEE, 2006, pp. 1346–1353.

[30] Z. Li, S. Zhang, X. Cai, Q. Zhang, X. Zhu, Z. Fan, and X. Jia, "Noisy optimization by evolution strategies with online population size learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[31] D. V. Arnold and H.-G. Beyer, "Investigation of the (μ, λ)-ES in the presence of noise," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*, vol. 1. IEEE, 2001, pp. 332–339.

[32] ——, "On the benefits of populations for noisy optimization," *Evolutionary Computation*, vol. 11, no. 2, pp. 111–127, 2003.

[33] ——, "A comparison of evolution strategies with other direct search methods in the presence of noise," *Computational Optimization and Applications*, vol. 24, no. 1, pp. 135–159, 2003.

[34] N. Hansen and S. Kern, "Evaluating the CMA evolution strategy on multimodal test functions," in *International conference on parallel problem solving from nature*. Springer, 2004, pp. 282–291.

[35] H.-G. Beyer and B. Sendhoff, "Covariance matrix adaptation revisited–the CMSA evolution strategy–," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2008, pp. 123–132.

[36] Y. Akimoto, A. Auger, and N. Hansen, "Quality gain analysis of the weighted recombination evolution strategy on general convex quadratic functions," *Theoretical Computer Science*, vol. 832, pp. 42–67, 2020.

[37] C. Qian, C. Bian, Y. Yu, K. Tang, and X. Yao, "Analysis of noisy evolutionary optimization when sampling fails," *Algorithmica*, pp. 1–36, 2020.

[38] A. Ahrari and O. Kramer, "Finite life span for improving the selection scheme in evolution strategies," *Soft Computing*, vol. 21, no. 2, pp. 501–513, 2017.

[39] H.-G. Beyer and H.-P. Schwefel, "Evolution strategies–a comprehensive introduction," *Natural computing*, vol. 1, no. 1, pp. 3–52, 2002.

[40] H.-G. Beyer, "Mutate large, but inherit small! on the analysis of rescaled mutations in ($\tilde{1}$, $\tilde{\lambda}$)-es with noisy fitness data," in *International Conference on Parallel Problem Solving from Nature*. Springer, 1998, pp. 109–118.

[41] ——, "Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice," *Computer methods in applied mechanics and engineering*, vol. 186, no. 2-4, pp. 239–267, 2000.

Ali Ahrari received his Ph.D. degree in mechanical engineering from Michigan State University, Michigan, USA. He is now a postdoc research fellow at the University of Sydney, Sydney, Australia. His research concentrates on evolutionary algorithms and engineering optimization. He has won GECCO and CEC competitions on multimodal optimization in 2016 and 2020, and ISCSO student competition in structural optimization in 2017 and 2018. He is now a member of the IEEE CIS Task Force on Multimodal Optimization.

Saber Elsayed received the Ph.D.degree in Computer Science from the University of New South Wales Canberra, Australia, in 2012. Currently, Saber is a Senior Lecturer with the School of Engineering and Information Technology, University of New South Wales Canberra. His research interests include the areas of evolutionary algorithms, constraint-handling techniques for evolutionary algorithms, scheduling, big data and cybersecurity using computational intelligence. Saber won several IEEE-CEC competitions. Dr. Elsayed was serving as the chair of the IEEE Computational Intelligence Society (ACT Chapter) (2019-2020), and has held organizational roles in several conferences.

Ruhul Sarker received his Ph.D. in 1992 from Dalhousie University, Halifax, Canada. He is currently a Professor in the School of Engineering and IT, Canberra, Australia. His main research interests are Evolutionary Optimization and Applied Operations Research. He is the lead author of the book *Optimization Modelling: A Practical Approach*. Prof. Sarker is a member of IEEE and INFORMS.

Daryl Essam received his B.Sc. degree in computer science from University of New England, Australia in 1990 and his Ph.D. degree from University of New South Wales, Australia, in 2000. Since 1994, he has been with the Canberra campus, UNSW, where he is currently a Senior Lecturer. His research interests include genetic algorithms, with a focus on both evolutionary optimization and large scale problems.

Carlos A. Coello Coello (M'98-SM'04-F'11) received a PhD in computer science from Tulane University, USA, in 1996. He is currently Professor with Distinction (CINVESTAV-3F Researcher) at the Computer Science Department of CINVESTAV-IPN, in Mexico City, Mexico. Dr. Coello has authored and co-authored over 500 technical papers and book chapters. He has also co-authored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Second Edition, Springer, 2007). His publications currently report over 58,300 citations in Google Scholar (his h-index is 96). Currently, he is the Editor-in-Chief of the IEEE Transactions on Evolutionary Computation and serves in the editorial board of several other international journals. He received the 2007 National Research Award from the Mexican Academy of Sciences in the area of Exact Sciences, the Medal to the Scientific Merit 2009, granted by Mexico City's congress and the National Medal of Science and Arts in the area of Physical, Mathematical and Natural Sciences. He is also a recipient of the 2013 IEEE Kiyo Tomiyasu Award, of the 2016 The World Academy of Sciences (TWAS) Award in "Engineering Sciences," and of the 2021 IEEE Computational Intelligence Society Evolutionary Computation Pioneer Award. He is a Fellow of the IEEE, and a member of the ACM and the Mexican Academy of Science. His major research interests are: evolutionary multi-objective optimization and constraint-handling techniques for evolutionary algorithms.