A Review on Population Sizing Models for Evolution Strategies in Multimodal Optimization

Zhao ZHAO
Southern University of Science and Technology
Shenzhen, China
12541018@mail.sustech.edu.cn

***Abstract*—Optimizing multimodal landscapes with numerous local optima remains a challenge in computational intelligence. Evolution Strategies (ES) excel in such scenarios, but their success critically depends on appropriate population sizing. This paper introduces a "Frozen Noise Model" to analyze ES convergence on multimodal functions, decomposing complex landscapes into smooth global structures and local fluctuating noise. The framework derives population scaling laws for three benchmark functions: Rastrigin, Bohachevsky, and Ackley. Our analysis reveals that ES performs a "circling search" around global attractor regions rather than gradient descent, with population requirements scaling sublinearly with problem dimension. The results provide theoretical guidance for parameter configuration in multimodal optimization.**

***Index Terms*—Evolution strategies (ES), global convergence, global optimization, multimodal objective function, population sizing**

I. Introduction

Optimizing highly multimodal fitness landscapes with numerous local optima poses a significant challenge in computational intelligence. Traditional gradient-based optimization methods often fail to locate global optima, while restart strategies may become impractical due to exponentially increasing function evaluations. Evolution Strategies (ES), owing to their inherent stochastic search characteristics, have demonstrated robust performance on multimodal optimization problems. However, the success rate of ES critically depends on the selection of population size: populations that are too small struggle to escape local optima, whereas excessively large populations incur prohibitive computational costs. This paper introduces a "Frozen Noise Model" for analyzing the convergence behavior of standard ES in multimodal landscapes and derives the scaling relationship between the required population size and the search space dimension for three canonical multimodal functions: Rastrigin, Bohachevsky, and Ackley. The proposed framework provides a theoretical foundation and practical guidance for population sizing in ES.

II. Background

This paper focuses on multi-recombinant Evolution Strategies $(\mu/\mu_I, \lambda)-ES$, including variants using σ self-adaptation (σSA) and cumulative step-size adaptation (CSA). The investigation centers on three classic multimodal benchmark functions:

1) Rastrigin Function: Composed of a global quadratic component and a periodic oscillatory component, exhibiting an exponential number of local optima.
2) Bohachevsky Function: Structurally similar to Rastrigin but asymmetric across dimensions, presenting a hyper-ellipsoidal global structure.
3) Ackley Function: Characterized by a funnel-shaped surface with infinitely many local optima and a nonlinear decaying global component.

Empirical studies have shown that the convergence success rate of ES on these functions strongly depends on population size. While preliminary modeling has been conducted for the Rastrigin function, this paper extends the analysis to Bohachevsky and Ackley functions, establishing a unified framework for population size prediction.

III. Proposed Algorithm: Frozen Noise Model and Population Sizing Derivation

A. Core Algorithm Framework

The ES algorithms studied in this paper are outlined in the pseudocode Algorithm 1 and Algorithm 2:

Algorithm 1 $(\mu/\mu_I, \lambda)$-σSA-ES

Require: $\mu, \lambda, \tau, \sigma^{(0)}, \mathbf{y}^{(0)}$
Ensure: Optimal solution $\mathbf{y}$
$g \leftarrow 0$
repeat
  for $k = 1$ to λ **do**
    $\sigma_k \leftarrow \sigma^{(g)} \exp(\tau \cdot N(0,1))$ ▷ Step-size adaptation
    $\mathbf{z}_k \leftarrow N(0, \mathbf{I})$ ▷ Isotropic Gaussian mutation
    $\mathbf{y}_k \leftarrow \mathbf{y}^{(g)} + \sigma_k \cdot \mathbf{z}_k$
    $F_k \leftarrow F(\mathbf{y}_k)$ ▷ Evaluate fitness
  end for
  Select μ best individual indices I_{best}
  $\mathbf{y}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{k \in I_{\text{best}}} \mathbf{y}_k$ ▷ Intermediate recombination
  $\sigma^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{k \in I_{\text{best}}} \sigma_k$
  $g \leftarrow g + 1$
until termination condition satisfied

B. Frozen Noise Model Construction

To analyze the behavior of Evolution Strategies (ES) on multimodal functions, this paper proposes the "Frozen Noise

Algorithm 2 $(\mu/\mu_I, \lambda)$-CSA-ES

Require: $\mu, \lambda, c, \sigma^{(0)}, \mathbf{y}^{(0)}$
Ensure: Optimal solution $\mathbf{y}$

$g \leftarrow 0$, $\mathbf{p}_\sigma \leftarrow \mathbf{0}$
repeat
  for $k = 1$ to λ **do**
    $\mathbf{z}_k \leftarrow N(0, \mathbf{I})$ ▷ Isotropic Gaussian mutation
    $\mathbf{y}_k \leftarrow \mathbf{y}^{(g)} + \sigma^{(g)} \cdot \mathbf{z}_k$
    $F_k \leftarrow F(\mathbf{y}_k)$ ▷ Evaluate fitness
  end for
  Select μ best individual indices I_{best}
  $\mathbf{z}_{\text{avg}} \leftarrow \frac{1}{\mu} \sum_{k \in I_{\text{best}}} \mathbf{z}_k$
  $\mathbf{p}_\sigma \leftarrow (1-c)\mathbf{p}_\sigma + \sqrt{c(2-c)\mu} \cdot \mathbf{z}_{\text{avg}}$
  $\mathbf{y}^{(g+1)} \leftarrow \mathbf{y}^{(g)} + \sigma^{(g)} \cdot \mathbf{z}_{\text{avg}}$
  $\sigma^{(g+1)} \leftarrow \sigma^{(g)} \exp\left(\frac{\|\mathbf{p}_\sigma\|^2 - N}{2N\sqrt{N}}\right)$
  $g \leftarrow g + 1$
until termination condition satisfied

Model." The basic idea of this model is to decompose a complex multimodal function into two parts: a globally smooth "skeleton" component and a series of local fluctuating "noise" components. This is mathematically expressed as:

$$F(\mathbf{y}) = G(\mathbf{y}) + C(\mathbf{y})$$

Here, $G(\mathbf{y})$ represents the global part, which can typically be approximated as a sphere model $G(R) = G(\|\mathbf{y}\|)$; $C(\mathbf{y})$ represents the residual part, which includes all local oscillations and fluctuations.

Taking the Rastrigin function as an example, its global part is a simple quadratic function $G_{\mathcal{R}}(\mathbf{y}) = R^2$, with the minimum at the origin. The local noise part contains all oscillations: $C_{\mathcal{R}}(\mathbf{y}) = NA - A\sum_{i=1}^{N}\cos(\alpha y_i)$.

Through this decomposition, the originally complex problem is transformed into a relatively simple sphere model optimization problem, with the sphere model superimposed by distance-dependent noise:

$$\tilde{F}(\mathbf{y}) = G(R) + \sigma_{ES}(R)\mathcal{N}(0,1)$$

This allows us to utilize existing noisy sphere model theory to analyze the behavior of ES.

C. Steady-State Behavior Analysis

Research shows that when ES operates on multimodal functions, it does not continuously converge to the optimal solution but reaches a stable state at a certain distance from the optimum. This stable distance depends on multiple factors: noise intensity, population size, and problem dimensionality.

The noise intensity $\sigma_{ES}(R)$ is a function of the distance R, representing the "noise" strength caused by local fluctuations at a distance R from the global optimum.

The steady-state distance R_{st} can be approximated by the following formula:

$$R_{st} \simeq \sqrt{\frac{\sigma_{ES} N}{4a\mu c_{\mu/\mu,\sigma}}}$$

Here, a is the sphere model coefficient ($a = 1$ for the Rastrigin function, $a = 3$ for the Bohachevsky function), and $c_{\mu/\mu,\sigma}$ is a constant related to the recombination strategy.

In simple terms, stronger noise makes it harder for ES to approach the optimum; larger population sizes allow ES to better resist noise effects and thus get closer to the optimum. Mathematical analysis indicates that ES behavior at steady state resembles searching on a noisy spherical surface.

D. Derivation of Success Probability

The key to ES finding the global optimum lies in its ability to enter the "global attractor region." The attractor region is a special area surrounding the global optimum where all gradient flows point toward the global optimum, with no interference from local optima.

For the Rastrigin function, this attractor region is a hypercube, whose boundary distance $\Delta_{\mathcal{R}}$ can be calculated as:

$$\Delta_{\mathcal{R}} \simeq \frac{A\alpha\pi}{A\alpha^2 - 2}$$

If the parent individuals of ES can enter this attractor region, they will quickly converge to the global optimum.

The success probability is the probability that ES enters the attractor region. Since ES components at steady state approximately follow a normal distribution in each dimension, we can calculate this probability based on statistical principles.

Analysis shows that the success probability mainly depends on two factors: the size of the attractor region and the search range of ES at steady state.

E. Population Size Prediction Formulas

Through mathematical derivation, we have reached an important conclusion: to achieve a specified success probability, there exists a specific relationship between the required ES population size and problem dimensionality.

For the Rastrigin and Bohachevsky functions, this relationship is sublinear. Specifically, the growth rate of population size is approximately the square root of the problem dimension multiplied by a logarithmic term. This means that even if the problem dimension increases significantly, the required population size will not grow exponentially, which is of great practical importance.

For the Ackley function, the situation is different. Due to the special structure of this function, if the initial point is appropriately chosen, the required population size depends only weakly on dimensionality. However, if the initial point is too far from the global optimum, the situation becomes more complex and may even exhibit exponential difficulty growth.

IV. Experiment

A. Noise distribution validation

Figure 1 is the histogram of residual parts for 300 ($50/50_I$, 100)-CSA-ES runs with $c = 1/\sqrt{N}$ and $N = 100$ in the Bohachevsky landscape at $R_{st}^{\mathcal{B}} = 1.02$, (left plot) and in the Ackley landscape at R = 0.22, (middle plot) and R = 0.27 (right plot). Bold solid lines show the pdf of normally distributed variates. Residual distributions approximate Gaussian, supporting model assumptions.

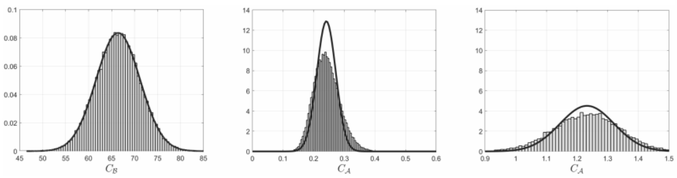

Fig. 1. Residual Distributions

B. Steady-state behavior verification

Figure 2 Left plot: R-dynamics of the noisy hyperellipsoid model (dark lines) and noisy sphere model (light lines). Black dashed line shows $R_{st}^{\mathcal{B}}$. Right plot: Histogram of all individual components y_i at distance $R_{st}^{\mathcal{B}}$. Bold solid line shows the pdf of the $\mathcal{N}(0, R_{st}^2/N)$ variate. Experiments executed for the (100/100$_I$, 200)-CSA-ES with $c = 1/\sqrt{N}$, N = 100, and $R_{st}^{\mathcal{B}} = 0.72$. The experiment shows that parent component distributions match $\mathcal{N}(0, R_{st}^2/N)$ predictions.

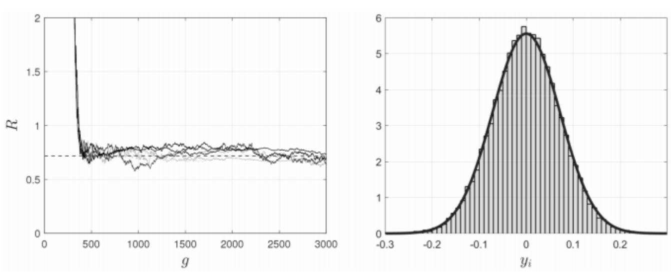

Fig. 2. Steady-state Behavior Verification

C. Initial condition sensitivity

Figure 3 Solid lines show μ^A with $\sigma^* = 10$ (left plot) and $\sigma^* = 20$ (middle and right plot). Vertical dashed-dotted lines in left and middle plot show R_{crit}. Ackley function has critical initial distance R_{crit} beyond which convergence probability drops sharply.

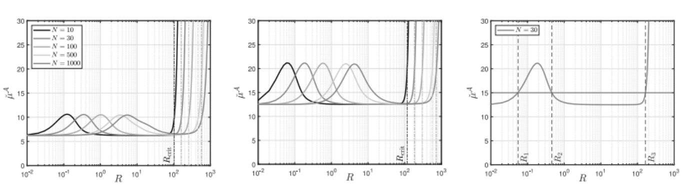

Fig. 3. Initial condition sensitivity

V. Discussions

A. Scaling Behavior Analysis

The required population size for the Rastrigin and Bohachevsky functions grows sublinearly as $O(N \times \ln N)$, which is significantly more efficient than exponential restart strategies. For the Ackley function, the population size depends only weakly on the dimension N; however, if the initial distance exceeds the critical value $R_c rit$, the required population size grows exponentially, highlighting the high sensitivity to initialization conditions.

B. Implications for Benchmark Design

Given the high similarity in population size behavior between Rastrigin and Bohachevsky, only one of them may be retained in benchmark suites to streamline testing. The convergence of the Ackley function heavily depends on the choice of initial points. Standard initialization intervals may mask algorithmic limitations; it is therefore recommended to adjust parameters (e.g., reduce C_2) or expand the initial search range to increase test difficulty and improve the discriminative power of algorithm evaluation.

C. Comparison with the Griewank Function

Although the Griewank function can also be decomposed into global and residual components, its noise variance decays rapidly with increasing N, making the problem easier as dimensionality grows. Furthermore, its noise does not follow a Gaussian distribution. Therefore, the proposed model is not applicable to such special functions, revealing the scope and boundaries of the theoretical model.

VI. Conclusion

This paper establishes, for the first time, a systematic theoretical framework for population sizing in Evolution Strategies (ES) applied to multimodal functions such as Rastrigin, Bohachevsky, and Ackley through the introduction of the frozen noise model. The findings indicate that ES does not follow gradient descent in multimodal optimization but rather performs a "circling search" outside the global attractor region, converging rapidly only after entering this region. The derived population sizing formulas exhibit sublinear scaling properties, providing theoretical guidance for ES parameter configuration and offering valuable insights for benchmark design, algorithm evaluation, and practical applications. Future research may extend to non-spherical global structures, dynamic optimization environments, and more complex real-world multimodal problems.

References

[1] L. Schönenberger and H.-G. Beyer, "On a population sizing model for evolution strategies in multimodal landscapes," *IEEE Transactions on Evolutionary Computation*, vol. 29, no. 5, pp. 1807–1819, Oct. 2025.