



Optimal Restart Strategies for Parameter-dependent Optimization Algorithms

Lisa Schönenberger

lisa.schoenenberger@uni-ulm.de

Ulm University, Institute of Theoretical Computer Science
Ulm, Germany
Vorarlberg University of Applied Sciences
Research Center Business Informatics
Dornbirn, Austria

Hans-Georg Beyer

hans-georg.beyer@fhv.at

Vorarlberg University of Applied Sciences
Research Center Business Informatics
Dornbirn, Austria

Abstract

This paper examines restart strategies for algorithms whose successful termination depends on a parameter λ . After each restart, λ is increased, until the algorithm terminates successfully. It is assumed that there is an unknown, optimal value for λ . For the algorithm to run successfully, this value must be surpassed. The key question is whether there exists an optimal strategy for selecting λ after each restart taking into account that the computational costs increase with λ . Potential restart strategies are classified into parameter-dependent strategy types. A loss function is introduced to quantify the wasted computational cost relative to the optimal strategy. A crucial requirement for any efficient restart strategy is that its loss, relative to the optimal λ , remains bounded. To this end, upper and lower bounds of the loss are derived. Using these bounds it will be shown that not all strategy types are bounded. However, for a particular strategy type, where λ is increased multiplicatively by a constant factor ρ , the relative loss function is bounded. Furthermore, it will be demonstrated that within this strategy type, there exists an optimal value $\rho = 2$ that minimizes the maximum relative loss. In the asymptotic limit, this optimal choice does not depend on the unknown optimal λ . While the multiplicative strategy with $\rho = 2$ was already used in implementations of evolutionary algorithms to control the population size showing acceptable performance in applications, a formal proof of its optimality is presented and the underlying conditions are discussed in this paper the first time.

CCS Concepts

- Theory of computation → Random search heuristics.

Keywords

Restart Strategy, Universal Optimal Strategy, Loss Function, Relative Loss, Evolution Strategy

ACM Reference Format:

Lisa Schönenberger and Hans-Georg Beyer. 2025. Optimal Restart Strategies for Parameter-dependent Optimization Algorithms. In *Foundations of Genetic Algorithms XVIII (FOGA '25)*, August 27–29, 2025, Leiden, Netherlands. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3729878.3746697>



This work is licensed under a Creative Commons Attribution 4.0 International License.
FOGA '25, August 27–29, 2025, Leiden, Netherlands

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1859-5/2025/08

<https://doi.org/10.1145/3729878.3746697>

1 Introduction

Optimization algorithms are often confronted with various challenges, such as becoming trapped in a local optimum or very long runtimes. Restart strategies have proven to be an effective way to overcome these obstacles. They can significantly enhance the performance and robustness of optimization algorithms. In the field of Evolution Strategies (ES), restart strategies are often used to reach the global optimum of objective functions with many local optima. For example, the IPOP-CMA-ES [1] performs restarts where the population size is doubled at each restart. A mixture of population size doubling and repeated restarts (with constant population size) is used in Bi-Population-ES [2, 3] to reach global optimal solutions.

Restart strategies regularly stop the current search process and restart the optimization algorithm. Often a different starting point is chosen, or an algorithmic parameter is changed. The specific implementation of restart strategies can vary considerably, ranging from simple random restarts to more sophisticated techniques that adapt to the particular algorithm and problem characteristics. In general, it is not clear what the optimal implementation of a restart strategy is. Obviously, this question cannot be answered in a general way for each type of optimization algorithm. Therefore, assumptions and constraints must be made.

The optimal implementation of a restart strategy has already been studied for Las Vegas algorithms, which are always successful, but whose running time is a random variable. In [4] an optimal restart strategy has been derived under the assumption that the distribution of the running time is known. Specifically, the optimal strategy prescribes terminating each trial after a fixed time t^* , where t^* is determined by the distribution of the running time. In scenarios where the distribution of the running time is not known, Luby et al. [4] proposes a universal restart strategy by the sequence $\mathbf{L} = (1, 1, 2, 1, 1, 2, 4, 1, 1, 2, 1, 1, 2, 4, 8, \dots)$, where the k th entry represents the running time of the k th restart. The performance of the Luby sequence \mathbf{L} is only marginally inferior to the optimal strategy (t^*, t^*, t^*, \dots) by a logarithmic factor. This topic was also explored in [5] within the context of a continuous setting. An optimal restart strategy was derived for a restricted class of continuous probability distributions. In [6] the optimal strategy from [4] was implemented in a (1+1)-EA (Evolutionary Algorithm) applied to a set of (theoretically treatable) test functions. The results demonstrated that this restart strategy can identify the optimum in polynomial time, whereas the conventional (1+1)-EA requires an exponential runtime. Furthermore, it was also demonstrated that restarting the (1+1)-EA outperforms the (μ +1)-EA.

In [7], the Luby sequence was integrated into a Bet-and-Run Strategy. Bet-and-Run Strategies [8] involve executing multiple independent runs, each with a predetermined time budget, after which only the best-performing run is continued until the total budget is exhausted. In [7] it was shown that a Luby sequence utilized in the initial phase of a Bet-and-Run Strategy can be outperformed by simpler strategies with uniform execution times across all runs in the initial phase. However, these simpler strategies are problem-specific, requiring parameter tuning, whereas the universal Luby sequence eliminates the need for such adjustments.

It is crucial to acknowledge that not all of the aforementioned considerations concerning Las Vegas algorithms are universally applicable, given that not every algorithm succeeds after a certain time. This is particularly evident in the case of deterministic algorithms, such as gradient-based methods, that are incapable of being improved by simple restarts. In such instances, restarts can be employed by varying other parameters, such as the initial starting point in gradient-based methods. Also in the context of non-deterministic algorithms, a longer runtime does not necessarily guarantee success. To illustrate this point, consider the case of Evolutionary Algorithms (EA) in multimodal landscapes where the population size exerts a more significant influence on the success of the algorithm than the runtime. When the population size is substantially smaller than the population size threshold, the algorithm invariably converges to a local minimum, regardless of runtime. In this context, restart strategies entail modifying the population size across restarts. However, enhanced success rates are only observed when the population size increases. Consequently, restarting the algorithm with the same or a smaller population size is ineffective in such scenarios, making the universal Luby sequence \mathbf{L} suboptimal for restarting EA.

All of the following theoretical considerations should be applicable to EA, or more generally, to algorithms whose success depends on a certain parameter λ . This parameter can be, for example, the population size in EA or particle swarm optimization, or a predefined runtime. An additional assumption is that the optimization algorithm is successful¹ if λ exceeds some unknown λ threshold and is unsuccessful otherwise. This means that there is an optimal choice $\hat{\lambda}$ for this parameter.

As for the scenario considered in this work, it is provided that if the λ parameter is equal to or greater than the (unknown) $\hat{\lambda}$ the EA terminates successfully, i.e., it approaches the global optimizer with a success probability that is arbitrarily close to one. However, if $\lambda < \hat{\lambda}$ the success probability drops quickly with decreasing λ . This scenario is observed in Evolution Strategies (ES) [9] optimizing certain highly multimodal objective functions such as Rastrigin, Bohachevsky, and Ackley to name a few. The λ parameter in ES refers to the offspring population size. On the basis of several multimodal test functions, it has already been shown [10] that the λ interval of global convergence uncertainty is rather small. Therefore, a tractable model for a theoretical analysis of these restart strategies assumes a fixed population size $\lambda = \hat{\lambda}$ that must be reached by the restart algorithm in an efficient manner.

¹By "successful" it is meant that the algorithm has reached the domain of the global optimizer in contrast to the convergence to a local optimum.

A restart strategy defines a start value λ_0 and determines how λ is changed after each restart. Thus, the restart strategy can be defined as $\mathcal{R} = (\lambda_0, \lambda_1, \lambda_2, \dots)$. Often used for such problems is $\mathcal{R}^* = (\lambda_0, \lambda_0\rho, \lambda_0\rho^2, \lambda_0\rho^3, \dots)$ ² where $\rho > 1$. In [1] this was applied to ES where the population size was increased after each restart with the above scaling rule using a restart parameter of $\rho = 2$. It was mentioned that experiments indicate that the optimal value for the restart parameter ρ is between 2 and 3. In [11] the above restart strategy was applied to the $(1+1)$ -EA where the maximum number of generations was increased after each restart. Also in this study, a restart parameter of 2 was chosen. Besides \mathcal{R}^* , an additive restart strategy was also considered, where a restart parameter v was added after each restart, i.e., $\mathcal{R}^+ = (\lambda_0, \lambda_0 + v, \lambda_0 + 2v, \lambda_0 + 3v, \dots)$. In contrast to [1], the EA in [11] was predicated on the assumption that the EA would always reach a successful outcome after a sufficiently protracted period, as it was the case for the Las Vegas Algorithm in [4]. In [12], a multiplicative restart strategy has been considered in the context of multi-objective optimization. There it was found that $\rho = 2$ yields optimal performance.

The goal of this investigation is to find the optimal choice of the restart parameter within the context of the parameter-dependent setting previously delineated, as well as across different types of restart strategies. This paper is organized as follows: The general definition of a restart strategy is given in Section 2 and the different types of restart strategies are introduced. Section 3 introduces the loss function, which indicates how much computational resources are wasted compared to the optimal strategy. Upper and lower bounds for the loss function will be derived for each strategy type. A relative loss function is introduced in Section 4. Being based on this relative loss function the optimal choice of the restart parameter is examined that minimizes the maximal relative loss. In Section 5 the multiplicative \mathcal{R}^* -RS is applied to Evolution Strategies. Experimental results are presented and compared with the theoretical predictions. Finally, in Section 6, a summary of the results and an outlook on future research are given.

2 Restart Strategies

The restart strategy (RS) under consideration is applicable to algorithms whose success depends on an algorithmic parameter $\lambda \in \mathbb{N}$. The algorithm \mathbf{A} is only successful if this algorithmic parameter exceeds a certain bound. Formally, this can be expressed as

$$\begin{aligned} \mathbf{A}(\lambda) \text{ is successful if } \lambda &\geq \hat{\lambda} \\ \mathbf{A}(\lambda) \text{ is unsuccessful if } \lambda &< \hat{\lambda}. \end{aligned} \quad (1)$$

It is customary to measure the computational effort of black-box optimization algorithms, such as Evolution Strategies, by the number $F_E(\lambda)$ of objective function evaluations that $\mathbf{A}(\lambda)$ uses until termination. As it is the case for the majority of algorithms, it can be assumed that $F_E(\lambda)$ increases with λ . Consequently, the optimal choice is to execute the algorithm with $\hat{\lambda}$. In this context, $\hat{\lambda}$ is also referred to as the *optimal* λ .

Evolution Strategies in multimodal landscapes meet this requirement to a satisfactory extent. This has been demonstrated in [10] for several multimodal test functions. In this case, the parameter λ represents the population size. There is an interval for λ where

²If $\lambda \in \mathbb{N}$ is a prerequisite, the values must be rounded accordingly.

it is possible to achieve a positive success probability of less than 1. This interval, however, is relatively small in comparison to the population size required to achieve a positive success probability. If λ exceeds this interval, the success rate will remain constant arbitrarily close to one.

To approach the optimal choice of the algorithmic parameter $\hat{\lambda}$, which is generally unknown, a restart strategy can be used. Restart strategies are defined by an unbounded sequence

$$\mathcal{R} = (\lambda_0, \lambda_1, \lambda_2, \dots), \quad \lambda_i \in \mathbb{N}, \quad (2)$$

where λ_k represents the algorithmic parameter of the k th run. The k th run of the RS is stopped when a local stopping criterion is fulfilled. Then, an independent algorithm with parameter λ_{k+1} is executed. This process is repeated until the algorithm is successful. This raises the question of how to choose λ_k . Because of condition (1), it is clear that λ should be increased after each restart, but the optimal amount of increase is unknown. A common choice for λ_k is $\lambda_k = \lambda_0 2^k$ (see for example [1] or [11]), however, no criterion of optimality exists up until now.

In principle there are infinitely many restart strategies. Therefore, possible restart strategies are classified into parameter-dependent groups, called *strategy types*. The corresponding parameter is called *restart parameter*. Different strategy types will be considered. The first type increases the algorithmic parameter by a constant amount, i.e.,

$$\begin{aligned} \mathcal{R}^+ &= (\lambda_0, \lambda_1, \lambda_2, \dots) \\ \lambda_k &= \lambda_{k-1} + v = \lambda_0 + kv, \quad k \geq 1, \end{aligned} \quad (3)$$

where $v \in \mathbb{N} \setminus \{0\}$ is the restart parameter. Another way to increase λ is to use a multiplicative change after each restart, i.e.,

$$\begin{aligned} \mathcal{R}^* &= (\lambda_0, \lambda_1, \lambda_2, \dots) \\ \lambda_k &= \lceil \lambda_{k-1} \rho \rceil = \lceil \lceil \lceil \lambda_0 \rho \rceil \rho \rceil \dots \rho \rceil, \quad k \geq 1. \end{aligned} \quad (4)$$

Because of the assumption that λ increases after each restart, it can be assumed that $\rho > 1$. For \mathcal{R}^* -RS, λ_k is determined based on the previous rounded-up values. Alternatively, one can consider

$$\begin{aligned} \mathcal{R}^\times &= (\lambda_0, \lambda_1, \lambda_2, \dots) \\ \lambda_k &= \lceil \lambda_0 \rho^k \rceil. \end{aligned} \quad (5)$$

The \mathcal{R}^* -RS and \mathcal{R}^\times -RS are also called *multiplicative strategy types*. A third type of restart strategies obeys a power law with constant α defined by

$$\begin{aligned} \mathcal{R}^\# &= (\lambda_0, \lambda_1, \lambda_2, \dots) \\ \lambda_k &= \lceil \lambda_0 (k+1)^\alpha \rceil. \end{aligned} \quad (6)$$

It is assumed $\lambda_k > \lambda_{k-1}$, which implies that $\alpha \geq 1$.

There is no clear indication that one strategy type is more effective than the other. Furthermore, it is also not clear how to choose the restart parameters v , ρ , or α for any given strategy type. The following sections investigates how the number of function evaluations is affected by the restart parameters. The objective of these investigations is to identify the optimal restart parameter for each strategy type. Furthermore, the criteria for a suitable strategy type will be defined.

3 The Loss Function

When selecting the restart parameters v , ρ , or α , it is important to avoid choosing values that are too small as this will result in many restarts being necessary. Conversely, if the restart parameter is set to a very large value, λ will also become very large after just a few restarts. This can lead to λ being much larger than necessary, requiring more function evaluations than necessary. Let $\hat{\lambda}$ be the minimal algorithmic parameter needed such that algorithm **A** is successful. The loss ΔF_E of an \mathcal{R} -RS is defined by³

$$\Delta F_E(\hat{\lambda}, \rho) = \sum_{k=0}^{\hat{k}(\hat{\lambda})} F_E(\lambda_k) - F_E(\hat{\lambda}). \quad (7)$$

$F_E(\lambda_k)$ denotes the number of function evaluations that algorithm **A** (λ_k) uses until termination. It holds $F_E(\lambda_k) = \lambda_k g_k$, where g_k is the number of generations used in the k th run. The \hat{k} denotes the minimum number of restarts required to attain a λ larger than or equal to $\hat{\lambda}$, i.e.,

$$\hat{k}(\hat{\lambda}) = \arg \min \{k \mid \lambda_k \geq \hat{\lambda}\}. \quad (8)$$

Assuming that each run requires the same number of generations g until termination, (7) becomes

$$\Delta F_E(\hat{\lambda}, \rho) = \left(\sum_{k=0}^{\hat{k}(\hat{\lambda})} \lambda_k - \hat{\lambda} \right) g. \quad (9)$$

The validity of this simplification does hold approximately for Evolution Strategies on certain highly multimodal objective functions as has been shown in [13]⁴. It is used here as a model assumption. As a result, g can be dropped in the following considerations, thus, a reduced loss function

$$\mathcal{L}(\hat{\lambda}, \rho) = \sum_{k=0}^{\hat{k}(\hat{\lambda})} \lambda_k - \hat{\lambda} \quad (10)$$

will be used. (10) can be calculated numerically using Alg. 1, where the update of λ is denoted by r and depends on the specific strategy type. It holds that

$$r(\lambda) = \lambda + v \quad \text{for } \mathcal{R}^+ \quad (11)$$

$$r(\lambda) = \lceil \lambda \rho \rceil \quad \text{for } \mathcal{R}^* \quad (12)$$

$$r(\lambda) = \lceil \lambda_0 \rho^k \rceil \quad \text{for } \mathcal{R}^\times \quad (13)$$

$$r(\lambda) = \lceil \lambda_0 (k+1)^\alpha \rceil \quad \text{for } \mathcal{R}^\#, \quad (14)$$

where k is the number of restarts. Alg. 1 has been used for the generation of numerical data in all figures below.

If $\hat{\lambda} = \lambda_k$ exactly k restarts are necessary, i.e., $\hat{k}(\lambda_k) = k$. If $\hat{\lambda} = \lambda_k + 1$ an additional restart is necessary and it holds that $\hat{k}(\lambda_k + 1) = k + 1$. Therefore the loss function (10) jumps between λ_k and λ_{k+1} . If the number of restarts \hat{k} is the same, the first expression in (10) does

³The ρ is used as a substitute to indicate the dependency of ΔF_E on the respective restart parameter.

⁴In general, an ES is terminated when there is an almost complete absence of progress in the search space. For instance, in the event that the mutation strength falls below a certain threshold. For a more thorough examination of termination criteria, consult the IPOP-CMA-ES [1].

Algorithm 1 Numerical Calculation of the Loss Function (10)

```

1: Initialize ( $\lambda = \lambda_0, F_E = \lambda_0, k = 0$ )
2: while  $\lambda < \hat{\lambda}$  do
3:    $k = k + 1$ 
4:    $\lambda = r(\lambda)$             $\triangleright$  update  $\lambda$ , depending on strategy type
5:    $F_E = F_E + \lambda$ 
6: end while
7:  $\mathcal{L} = F_E - \hat{\lambda}$ 

```

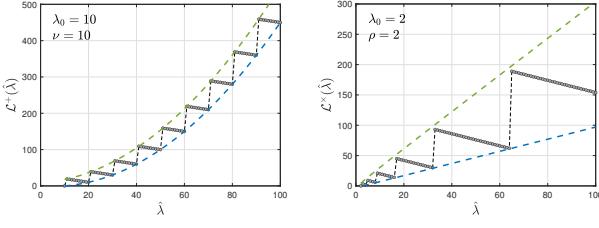


Figure 1: Loss function (10) for \mathcal{R}^+ (left) and the \mathcal{R}^\times (right) depending on $\hat{\lambda}$. Gray markers represent the numerical values. The green dashed lines show the upper bounds (20) (left) and (21) (right). The blue dashed lines show the lower bounds (16) (left) and (26) (right).

not depend on $\hat{\lambda}$. Hence, \mathcal{L} decreases linearly for $\lambda_{k-1} + 1 \leq \hat{\lambda} \leq \lambda_k$, i.e., it holds for all $k \geq 1$ that

$$\mathcal{L}(\hat{\lambda}) > \mathcal{L}(\lambda_k) \quad \text{for } \lambda_{k-1} + 1 \leq \hat{\lambda} < \lambda_k. \quad (15)$$

These observations, which are independent of the specific strategy type, lead to the typical saw tooth shape of the loss function. This is visualized for the different strategy types in Figs. 1, 2, and 4 where the loss function is represented by the gray markers.

For further investigation, this saw tooth function is difficult to handle. The following subsections derive upper and lower bounds of the loss function for each strategy type. The objective is to identify sharp bounds that can be represented explicitly as a function of $\hat{\lambda}$ and the restart parameter.

3.1 The Loss Function for the \mathcal{R}^+ -RS

The loss function of the \mathcal{R}^+ -RS is denoted as $\mathcal{L}^+(\hat{\lambda}, \nu)$ and represented by the gray markers in the left plot of Fig. 1.

THEOREM 3.1. Let $\mathcal{L}^+(\hat{\lambda}, \nu)$ be the loss function (10) of the \mathcal{R}^+ -RS with restart parameter $\nu \in \mathbb{N} \setminus \{0\}$ and $\lambda_k = \lambda_0 + kv$. Let

$$\mathcal{L}_{\text{low}}^+(\hat{\lambda}, \nu) := \frac{1}{2} (\hat{\lambda} - \lambda_0) \left(\frac{\hat{\lambda} + \lambda_0}{\nu} - 1 \right). \quad (16)$$

Then $\mathcal{L}_{\text{low}}^+(\hat{\lambda}, \nu)$ is a lower bound of $\mathcal{L}^+(\hat{\lambda}, \nu)$.

PROOF. The number of required restarts at $\hat{\lambda} = \lambda_k$ is $\hat{k}(\lambda_k) = k$. Because

$$\lambda_k = \lambda_0 + kv \Leftrightarrow k = \frac{\lambda_k - \lambda_0}{\nu}, \quad (17)$$

it follows for all $k \geq 0$

$$\begin{aligned} \mathcal{L}^+(\lambda_k, \nu) &= \sum_{j=0}^k \lambda_j - \lambda_k = \sum_{j=0}^{k-1} \lambda_j = \sum_{j=0}^{k-1} (\lambda_0 + j\nu) = k\lambda_0 + \nu \sum_{j=0}^{k-1} j \\ &= k \left(\lambda_0 + \frac{\nu}{2} (k-1) \right) = \frac{\lambda_k - \lambda_0}{\nu} \left(\lambda_0 + \frac{\nu}{2} \left(\frac{\lambda_k - \lambda_0}{\nu} - 1 \right) \right) \\ &= \frac{\lambda_k - \lambda_0}{\nu} \left(\frac{\lambda_k + \lambda_0}{2} - \frac{\nu}{2} \right) = \mathcal{L}_{\text{low}}^+(\lambda_k, \nu). \end{aligned} \quad (18)$$

In (15) it was shown that $\mathcal{L}^+(\hat{\lambda}, \nu)$ decreases between $\lambda_{k-1} + 1$ and λ_k . Therefore, it follows for $\lambda_{k-1} + 1 \leq \hat{\lambda} < \lambda_k$ and $k \geq 1$ by using (18) that

$$\begin{aligned} \mathcal{L}^+(\hat{\lambda}, \nu) &> \mathcal{L}^+(\lambda_k, \nu) = \mathcal{L}_{\text{low}}^+(\lambda_k, \nu) = \frac{1}{2} (\lambda_k - \lambda_0) \left(\frac{\lambda_k + \lambda_0}{\nu} - 1 \right) \\ &> \frac{1}{2} (\hat{\lambda} - \lambda_0) \left(\frac{\hat{\lambda} + \lambda_0}{\nu} - 1 \right) = \mathcal{L}_{\text{low}}^+(\hat{\lambda}, \nu). \end{aligned} \quad (19)$$

□

The lower bound (16) is represented by the blue dashed line in the left plot of Fig. 1.

THEOREM 3.2. Let $\mathcal{L}^+(\hat{\lambda}, \nu)$ be the loss function (10) of the \mathcal{R}^+ -RS with restart parameter $\nu \in \mathbb{N} \setminus \{0\}$ and $\lambda_k = \lambda_0 + kv$. Let

$$\mathcal{L}_{\text{up}}^+(\hat{\lambda}, \nu) = \frac{1}{2} (\hat{\lambda} - \lambda_0 - 1) \left(\frac{\hat{\lambda} + \lambda_0 - 1}{\nu} + 1 \right) + \lambda_0 + \nu - 1. \quad (20)$$

Then $\mathcal{L}_{\text{up}}^+(\hat{\lambda}, \nu)$ is an upper bound of $\mathcal{L}^+(\hat{\lambda}, \nu)$.

PROOF. The proof is similar to that of Theorem 3.1. It is given in the supplementary material. □

The upper bound (20) is represented by the green dashed line in the left plot of Fig. 1.

3.2 The Loss Function for the \mathcal{R}^\times -RS

The loss function of the \mathcal{R}^\times -RS is denoted as $\mathcal{L}^\times(\hat{\lambda}, \rho)$ and represented by the gray markers in the right plot of Fig. 1.

THEOREM 3.3. Let $\mathcal{L}^\times(\hat{\lambda}, \rho)$ be the loss function (10) of the \mathcal{R}^\times -RS with restart parameter $\rho > 1$ and $\lambda_k = \lceil \lambda_0 \rho^k \rceil$. Let

$$\mathcal{L}_{\text{up}}^\times(\hat{\lambda}, \rho) := \lambda_0 \rho + (\hat{\lambda} - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right) + \frac{\ln(\hat{\lambda}/\lambda_0)}{\ln(\rho)}. \quad (21)$$

Then $\mathcal{L}_{\text{up}}^\times(\hat{\lambda}, \rho)$ is an upper bound of $\mathcal{L}^\times(\hat{\lambda}, \rho)$.

PROOF. The number of required restarts at $\hat{\lambda} = \lambda_k + 1$ is given by $\hat{k}(\lambda_k + 1) = k + 1$. It holds that

$$\lambda_0 \rho^k \leq \lceil \lambda_0 \rho^k \rceil = \lambda_k < \lambda_0 \rho^k + 1. \quad (22)$$

Therefore,

$$\begin{aligned} \mathcal{L}^{\times}(\lambda_k + 1, \rho) &= \sum_{j=0}^{k+1} \lambda_j - \lambda_k - 1 = \sum_{j=0}^{k-1} \lambda_j + \lambda_{k+1} - 1 \\ &< \sum_{j=0}^{k-1} (\lambda_0 \rho^j + 1) + \lambda_0 \rho^{k+1} + 1 - 1 \\ &= \lambda_0 \sum_{j=0}^{k-1} \rho^j + k + \lambda_0 \rho^{k+1} = \lambda_0 \frac{\rho^k - 1}{\rho - 1} + k + \lambda_0 \rho^{k+1}. \end{aligned} \quad (23)$$

It follows from (22) that

$$\lambda_0 \rho^k < \lambda_k + 1 \Leftrightarrow \rho^k < \frac{\lambda_k + 1}{\lambda_0} \Leftrightarrow k < \frac{\ln((\lambda_k + 1)/\lambda_0)}{\ln(\rho)}. \quad (24)$$

Inserting this into (23), then it follows for all $k \geq 0$ that

$$\begin{aligned} \mathcal{L}^{\times}(\lambda_k + 1, \rho) &< \lambda_0 \frac{\frac{\lambda_{k+1}}{\lambda_0} - 1}{\rho - 1} + \frac{\ln((\lambda_k + 1)/\lambda_0)}{\ln(\rho)} + \lambda_0 \frac{\lambda_k + 1}{\lambda_0} \rho \\ &= \frac{\lambda_k + 1}{\rho - 1} - \frac{\lambda_0}{\rho - 1} + (\lambda_k + 1) \rho + \frac{\ln((\lambda_k + 1)/\lambda_0)}{\ln(\rho)} \\ &= \mathcal{L}_{\text{up}}^{\times}(\lambda_k + 1, \rho). \end{aligned} \quad (25)$$

For $\hat{\lambda} = \lambda_0$ it holds that $\mathcal{L}^{\times}(\lambda_0, \rho) = 0 < \lambda_0 \rho = \mathcal{L}_{\text{up}}^{\times}(\lambda_0, \rho)$. Because $(\rho + 1/(\rho - 1)) > 0$ for $\rho > 1$ it holds that $\mathcal{L}_{\text{up}}^{\times}$ increases with $\hat{\lambda}$. Therefore, (15) implies $\mathcal{L}^{\times}(\hat{\lambda}, \rho) < \mathcal{L}_{\text{up}}^{\times}(\hat{\lambda}, \rho)$ for all $\hat{\lambda} \geq \lambda_0$. \square

The upper bound (21) is represented by the green dashed line in the right plot of Fig. 1.

THEOREM 3.4. *Let $\mathcal{L}^{\times}(\hat{\lambda}, \rho)$ be the loss function (10) of the \mathcal{R}^{\times} -RS with restart parameter $\rho > 1$ and $\lambda_k = \lceil \lambda_0 \rho^k \rceil$. Let*

$$\mathcal{L}_{\text{low}}^{\times}(\hat{\lambda}, \rho) := \frac{\hat{\lambda} - 1 - \lambda_0}{\rho - 1}. \quad (26)$$

Then $\mathcal{L}_{\text{low}}^{\times}(\hat{\lambda}, \rho)$ is a lower bound of $\mathcal{L}^{\times}(\hat{\lambda}, \rho)$.

PROOF. The proof is similar to that of Theorem 3.3. It is given in the supplementary material. \square

The lower bound (26) is represented by the blue dashed line in the right plot of Fig. 1.

3.3 The Loss Function for the \mathcal{R}^* -RS

The loss function of the \mathcal{R}^* -RS is denoted as $\mathcal{L}^*(\hat{\lambda}, \rho)$ and represented by the gray markers in Fig. 2.

LEMMA 3.5. *For $\rho > 1$, $k \in \mathbb{N}$ and $\lambda_{k+1} = \lceil \lambda_k \rho \rceil$ let*

$$\mathcal{F}_{\text{up}}(k) := \lambda_0 \rho + k + (\lambda_k - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right), \quad (27)$$

$$\mathcal{F}(k) := \sum_{j=0}^{k+1} \lambda_j - \lambda_k - 1. \quad (28)$$

Then it holds that $\mathcal{F}_{\text{up}}(k) > \mathcal{F}(k)$.

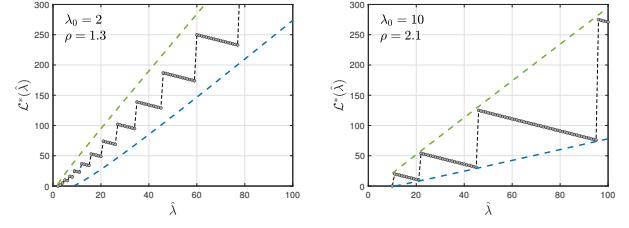


Figure 2: Loss function (10) for \mathcal{R}^* depending on $\hat{\lambda}$. Gray markers represent the numerical values. The green dashed line shows the upper bound (32) and the blue dashed line shows the lower bound (39).

PROOF. It holds for all $k \geq 0$ that

$$\lambda_k \rho + 1 > \lambda_{k+1} = \lceil \lambda_k \rho \rceil \geq \lambda_k \rho > \lceil \lambda_k \rho \rceil - 1. \quad (29)$$

It follows for $k = 0$ that

$$\mathcal{F}_{\text{up}}(0) = \lambda_0 \rho > \lceil \lambda_0 \rho \rceil - 1 = \lambda_1 - 1 = \sum_{j=0}^1 \lambda_j - \lambda_0 - 1 = \mathcal{F}(0). \quad (30)$$

Assume that the condition $\mathcal{F}_{\text{up}}(k) > \mathcal{F}(k)$ holds for k , then it follows by induction⁵

$$\begin{aligned} \mathcal{F}_{\text{up}}(k+1) &= \lambda_0 \rho + k + 1 + (\lambda_{k+1} - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right) \\ &= \mathcal{F}_{\text{up}}(k) + (\lambda_{k+1} - \lambda_k) \left(\rho + \frac{1}{\rho - 1} \right) + 1 \\ &> \mathcal{F}(k) + (\lambda_{k+1} - \lambda_k) \left(\rho + \frac{1}{\rho - 1} \right) + 1 \\ &= \sum_{j=0}^{k+1} \lambda_j - \lambda_k - 1 + (\lambda_{k+1} - \lambda_k) \left(\rho + \frac{1}{\rho - 1} \right) + 1 \\ &= \mathcal{F}(k+1) - \lambda_{k+2} + \lambda_{k+1} - \lambda_k + \lambda_{k+1} \rho - \lambda_k \rho + \frac{\lambda_{k+1} - \lambda_k}{\rho - 1} + 1 \\ &> \mathcal{F}(k+1) + \lambda_k \rho - \lambda_k - \lambda_k \rho + \frac{\lambda_k \rho - \lambda_k}{\rho - 1} = \mathcal{F}(k+1), \end{aligned} \quad (31)$$

where (29) was used for the last inequality. \square

Using this result it is possible to derive an upper bound for the loss function of the \mathcal{R}^* -RS.

THEOREM 3.6. *Let $\mathcal{L}^*(\hat{\lambda}, \rho)$ be the loss function (10) of the \mathcal{R}^* -RS with restart parameter $\rho > 1$ and $\lambda_k = \lceil \lambda_{k-1} \rho \rceil$ for $k \geq 1$. Let*

$$\mathcal{L}_{\text{up}}^*(\hat{\lambda}, \rho) := \lambda_0 \rho + (\hat{\lambda} - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right) + \frac{\ln(\hat{\lambda}/\lambda_0)}{\ln(\rho)}. \quad (32)$$

Then $\mathcal{L}_{\text{up}}^(\hat{\lambda}, \rho)$ is an upper bound of $\mathcal{L}^*(\hat{\lambda}, \rho)$.*

PROOF. Because $\lambda_k = \lceil \lambda_{k-1} \rho \rceil = \lceil \lceil \lambda_0 \rho \rceil \dots \rho \rceil \geq \lambda_0 \rho^k$ it follows that

$$\frac{\lambda_k}{\lambda_0} \geq \rho^k \Leftrightarrow \frac{\ln(\lambda_k/\lambda_0)}{\ln(\rho)} \geq k. \quad (33)$$

⁵A more comprehensive description of this induction step is provided in the supplementary material.

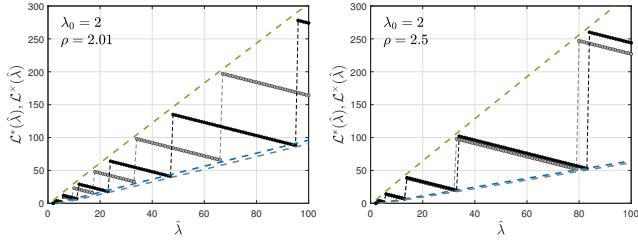


Figure 3: Loss function (10) depending on $\hat{\lambda}$. Markers represent the numerical values, black for the \mathcal{R}^* -RS and gray for the \mathcal{R}^X -RS. The green line shows the upper bound (32). The gray dashed line shows the lower bound (39) of the \mathcal{R}^* -RS and the blue dashed line shows the lower bound (26) of the \mathcal{R}^X -RS.

By using (33) and Lemma 3.5 it follows for $\lambda_k + 1 \leq \hat{\lambda} \leq \lambda_{k+1}$ and for all $k \geq 0$ that

$$\begin{aligned} \mathcal{L}_{\text{up}}^*(\hat{\lambda}, \rho) &> \lambda_0 \rho + \frac{\ln(\lambda_k/\lambda_0)}{\ln(\rho)} + (\lambda_k - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right) \\ &\geq \lambda_0 \rho + k + (\lambda_k - \lambda_0) \left(\rho + \frac{1}{\rho - 1} \right) \\ &= \mathcal{F}_{\text{up}}(k) > \mathcal{F}(k) = \mathcal{L}^*(\lambda_k + 1, \rho) \geq \mathcal{L}^*(\hat{\lambda}, \rho). \end{aligned} \quad (34)$$

The last inequality follows from (15). Because $\mathcal{L}_{\text{up}}^*(\lambda_0, \rho) = \lambda_0 \rho > 0 = \mathcal{L}^*(\lambda_0, \rho)$ it follows for all $\hat{\lambda} \geq \lambda_0$ that $\mathcal{L}_{\text{up}}^*(\hat{\lambda}, \rho) > \mathcal{L}^*(\hat{\lambda}, \rho)$. \square

The upper bound (32) is represented in Fig. 2 by the green dashed line. In the left figure, where λ_0 and ρ are small, the differences between (32) and the upper corners of the loss function are clearly visible. In the right figure for larger values of λ_0 and ρ the upper bound (32) is only slightly larger than the upper corners of the loss function.

A comparison of (32) with (21) reveals that the upper bounds of the \mathcal{R}^X -RS and the \mathcal{R}^* -RS are identical. In Fig. 3, the \mathcal{R}^* -RS is compared with the \mathcal{R}^X -RS. The figure illustrates that the loss functions of the two strategy types are nested within one another.

LEMMA 3.7. For $\rho > 1$, $k \in \mathbb{N} \setminus \{0\}$ and $\lambda_{k+1} = \lceil \lambda_k \rho \rceil$ let

$$\mathcal{F}_{\text{low}}(k) := \frac{1}{\rho - 1} (\lambda_k - \lambda_0 - k), \text{ and } \mathcal{F}(k) := \sum_{j=0}^k \lambda_j - \lambda_k. \quad (35)$$

Then it holds that $\mathcal{F}_{\text{low}}(k) < \mathcal{F}(k)$.

PROOF. For $k = 1$ it holds that

$$\begin{aligned} \mathcal{F}_{\text{low}}(1) &= \frac{1}{\rho - 1} (\lambda_1 - \lambda_0 - 1) < \frac{1}{\rho - 1} (\lambda_0 \rho - \lambda_0) \\ &= \lambda_0 = \sum_{j=0}^1 \lambda_j - \lambda_1 = \mathcal{F}(1). \end{aligned} \quad (36)$$

Assume that the condition $\mathcal{F}_{\text{low}}(k) < \mathcal{F}(k)$ holds for k , then it follows by induction that⁶

$$\mathcal{F}_{\text{low}}(k+1) = \frac{1}{\rho - 1} (\lambda_{k+1} - \lambda_0 - k - 1) \quad (37)$$

$$\begin{aligned} &= \mathcal{F}_{\text{low}}(k) + \frac{1}{\rho - 1} (\lambda_{k+1} - \lambda_k - 1) \\ &< \mathcal{F}(k) + \frac{1}{\rho - 1} (\lambda_{k+1} - \lambda_k - 1) \end{aligned}$$

$$\begin{aligned} &= \mathcal{F}(k+1) - \lambda_k + \frac{1}{\rho - 1} (\lambda_{k+1} - \lambda_k - 1) \\ &< \mathcal{F}(k+1) - \lambda_k + \frac{1}{\rho - 1} (\lambda_k \rho + 1 - \lambda_k - 1) = \mathcal{F}(k+1), \end{aligned} \quad (38)$$

where (29) was used for the last inequality. \square

Using Lemma 3.7, the lower bound of \mathcal{R}^* is given below.

THEOREM 3.8. Let $\mathcal{L}^*(\hat{\lambda}, \rho)$ be the loss function (10) of the \mathcal{R}^* -RS with restart parameter $\rho > 1$ and $\lambda_k = \lceil \lambda_{k-1} \rho \rceil$ for $k \geq 1$. Let

$$\mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho) := \frac{1}{\rho - 1} \left(\hat{\lambda} - \lambda_0 - \frac{\ln(\hat{\lambda}/\lambda_0)}{\ln(\rho)} - 1 \right). \quad (39)$$

Then $\mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho)$ is a lower bound of $\mathcal{L}^*(\hat{\lambda}, \rho)$.

PROOF. It holds for $\lambda_{k-1} + 1 \leq \hat{\lambda} \leq \lambda_k$ and all $k \geq 1$ that

$$\begin{aligned} \mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho) &= \frac{1}{\rho - 1} \left(\hat{\lambda} - \lambda_0 - \frac{\ln(\hat{\lambda}/\lambda_0)}{\ln(\rho)} - 1 \right) \\ &< \frac{1}{\rho - 1} \left(\lambda_k - \lambda_0 - \frac{\ln(\lambda_{k-1}/\lambda_0)}{\ln(\rho)} - 1 \right). \end{aligned} \quad (40)$$

From (33) it follows that $k - 1 \leq \ln(\lambda_{k-1}/\lambda_0)/\ln(\rho)$ for all $k \geq 1$. Using this, Lemma 3.7 and (15), one gets for $\lambda_{k-1} + 1 \leq \hat{\lambda} \leq \lambda_k$ and $k \geq 1$ that

$$\begin{aligned} \mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho) &< \frac{1}{\rho - 1} (\lambda_k - \lambda_0 - (k - 1) - 1) = \mathcal{F}_{\text{low}}(k) \\ &< \mathcal{F}(k) = \mathcal{L}^*(\lambda_k, \rho) \leq \mathcal{L}^*(\hat{\lambda}, \rho). \end{aligned} \quad (41)$$

For $\hat{\lambda} = \lambda_0$ it holds that $\mathcal{L}_{\text{low}}^*(\lambda_0, \rho) < 0 = \mathcal{L}^*(\lambda_0, \rho)$ and therefore, $\mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho) < \mathcal{L}^*(\hat{\lambda}, \rho)$ for all $\hat{\lambda} \geq \lambda_0$. \square

The lower bound (39) is represented in Fig. 2 by the blue dashed line. In the left figure, where λ_0 and ρ are small, there are clear discrepancies between (39) and the lower corners of the loss function. In the right figure for larger values of λ_0 and ρ , the lower bound (39) is only slightly smaller than the lower corners of the loss function.

In Fig. 3 the lower bound (39) of the \mathcal{R}^* -RS is compared with the lower bound (26) of the \mathcal{R}^X -RS. The lower bound (39) of \mathcal{L}^* is slightly smaller than the lower bound (26) of \mathcal{L}^X and is a lower bound for both loss functions. The lower bound (26) of \mathcal{L}^X intersects \mathcal{L}^* .

⁶A more comprehensive description of this induction step is provided in the supplementary material.

3.4 The Loss Function for the $\mathcal{R}^\#$ -RS

The loss function of the $\mathcal{R}^\#$ -RS is denoted as $\mathcal{L}^\#(\hat{\lambda}, \alpha)$ and represented by the gray markers in Fig. 4.

THEOREM 3.9. Let $\mathcal{L}^\#(\hat{\lambda}, \alpha)$ be the loss function (10) of the $\mathcal{R}^\#$ -RS with restart parameter $\alpha \geq 1$ and $\lambda_k = \lceil \lambda_0(k+1)^\alpha \rceil$. Let

$$\mathcal{L}_{\text{low}}^\#(\hat{\lambda}, \alpha) := \begin{cases} \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\hat{\lambda}-1}{\lambda_0}} - 1 \right)^{\alpha+1} & \text{for } \hat{\lambda} > \lambda_0 \\ 0 & \text{for } \hat{\lambda} = \lambda_0 \end{cases}. \quad (42)$$

Then $\mathcal{L}_{\text{low}}^\#(\hat{\lambda}, \alpha)$ is a lower bound of $\mathcal{L}^\#(\hat{\lambda}, \alpha)$.

PROOF. The loss function jumps at $\lambda_k + 1$ and the number of restarts is $\hat{k}(\lambda_k + 1) = k + 1$. It holds that

$$\lambda_0(k+1)^\alpha \leq \lceil \lambda_0(k+1)^\alpha \rceil = \lambda_k, \quad (43)$$

which implies that for $\lambda_k > \lambda_0$

$$(k+1)^\alpha > \frac{\lambda_k - 1}{\lambda_0} \Leftrightarrow k > \sqrt[\alpha]{\frac{\lambda_k - 1}{\lambda_0}} - 1. \quad (44)$$

Using

$$\sum_{j=1}^k j^\alpha > \int_0^k x^\alpha dx = \frac{k^{\alpha+1}}{\alpha+1}, \quad (45)$$

as visualized in the right plot of Fig. 4, it follows with (43) and (44)

$$\begin{aligned} \mathcal{L}^\#(\lambda_k, \alpha) &= \sum_{j=0}^k \lambda_j - \lambda_k = \sum_{j=0}^{k-1} \lambda_j = \sum_{j=0}^{k-1} \lceil \lambda_0(j+1)^\alpha \rceil \\ &\geq \lambda_0 \sum_{j=0}^{k-1} (j+1)^\alpha = \lambda_0 \sum_{i=1}^k i^\alpha > \lambda_0 \frac{k^{\alpha+1}}{\alpha+1} \\ &> \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\lambda_k - 1}{\lambda_0}} - 1 \right)^{\alpha+1} = \mathcal{L}_{\text{low}}^\#(\lambda_k, \alpha) \end{aligned} \quad (46)$$

for all $k \geq 1$ ⁷. It follows for $\lambda_{k-1} + 1 \leq \hat{\lambda} < \lambda_k$ by using (15) and (46) that

$$\begin{aligned} \mathcal{L}^\#(\hat{\lambda}, \alpha) &> \mathcal{L}^\#(\lambda_k, \alpha) > \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\lambda_k - 1}{\lambda_0}} - 1 \right)^{\alpha+1} \\ &> \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\hat{\lambda} - 1}{\lambda_0}} - 1 \right)^{\alpha+1} = \mathcal{L}_{\text{low}}^\#(\hat{\lambda}, \alpha) \end{aligned} \quad (47)$$

which holds for all $k \geq 1$. Because $\mathcal{L}^\#(\lambda_0, \alpha) = 0 = \mathcal{L}_{\text{low}}^\#(\lambda_0, \alpha)$ it holds for all $\hat{\lambda} \geq \lambda_0$ that $\mathcal{L}^\#(\hat{\lambda}, \alpha) \geq \mathcal{L}_{\text{low}}^\#(\hat{\lambda}, \alpha)$. \square

The lower bound (42) is represented by the blue dashed line in the left plot of Fig. 4.

THEOREM 3.10. Let $\mathcal{L}^\#(\hat{\lambda}, \alpha)$ be the loss function (10) of the $\mathcal{R}^\#$ -RS with restart parameter $\alpha \geq 1$ and $\lambda_k = \lceil \lambda_0(k+1)^\alpha \rceil$. Let

$$\mathcal{L}_{\text{up}}^\#(\hat{\lambda}, \alpha) := \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\hat{\lambda}-1}{\lambda_0} + 2} \right)^{\alpha+1} + \sqrt[\alpha]{\frac{\hat{\lambda}-1}{\lambda_0}} - \frac{\lambda_0}{\alpha+1}. \quad (48)$$

⁷In the case of $\hat{\lambda} = \lambda_0$, the power term in (46) may be imaginary.

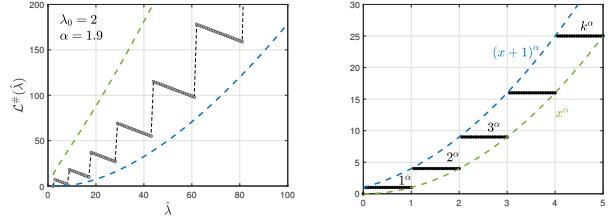


Figure 4: Left plot: Loss function (10) for $\mathcal{R}^\#$ depending on $\hat{\lambda}$. Gray markers represent the numerical values. The green dashed line shows the upper bound (48) and the blue dashed line shows the lower bound (42). Right plot: Visualization for (45) and (49).

Then $\mathcal{L}_{\text{up}}^\#(\hat{\lambda}, \alpha)$ is an upper bound of $\mathcal{L}^\#(\hat{\lambda}, \alpha)$.

PROOF. Using the inequality

$$\sum_{j=1}^k j^\alpha < \int_0^k (x+1)^\alpha dx = \frac{(k+1)^{\alpha+1}}{\alpha+1} - \frac{1}{\alpha+1}, \quad (49)$$

the proof is similar to that of Theorem 3.9. It is given in the supplementary material. \square

The upper bound (48) is represented by the green dashed line in the left plot of Fig. 4.

Figure 4 shows that (42) and (48) provide only very rough bounds. The reason for this is that the estimate (49) and (45) give much larger or smaller values. Nevertheless, the subsequent section will demonstrate that the $\mathcal{R}^\#$ -RS fails to satisfy a fundamental criterion of optimality. Consequently, there is no necessity for a more precise estimation of the power sums in (45) and (49).

4 The Relative Loss Function

For a fixed value of the restart parameter, the loss is unbounded for $\hat{\lambda}$ and tends to infinity. This holds for all strategy types, as evidenced by Figs. 1, 2, and 4, presented in the previous section. To further characterize the restart effort, it is useful to introduce the relative loss. It measures the loss w.r.t. the minimal $\lambda = \hat{\lambda}$ needed to complete the algorithm successfully. The relative loss is defined by⁸

$$\ell(\hat{\lambda}, \rho) := \frac{\mathcal{L}(\hat{\lambda}, \rho)}{\hat{\lambda}}. \quad (50)$$

If $\ell_{\text{up}}(\hat{\lambda}, \rho)$ is an upper bound of the relative loss function, the value

$$\bar{\ell}_{\text{up}}(\rho) := \lim_{\hat{\lambda} \rightarrow \infty} \ell_{\text{up}}(\hat{\lambda}, \rho) \quad (51)$$

is called the *asymptotic upper bound* of the relative loss function. The question to be addressed here is whether a finite asymptotic upper bound exists for a given strategy type. If a finite asymptotic upper bound $\bar{\ell}_{\text{up}}(\rho)$ exists for a given restart parameter then the RS is termed to be *bounded*. Conversely, if no finite upper bound exists, the strategy is called *unbounded*.

⁸ ρ is used as a substitute to show the dependence of the restart parameter. It can be replaced by v and α , respectively.

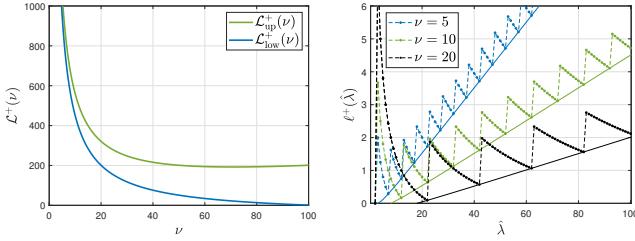


Figure 5: Left plot: lower bound (16) and upper bound (20) of the \mathcal{R}^+ loss function depending on ν for $\lambda_0 = 2$ and $\hat{\lambda} = 100$. Right plot: markers show the relative loss function for $\lambda_0 = 2$ and solid lines represent the corresponding lower bound (55).

The magnitude of the loss depends strongly on the restart parameters. It can therefore be assumed that the asymptotic upper bound also depends on the choice of the restart parameter. In the case of a bounded RS, it is possible to search for the optimal restart parameter, that is, the one which yields the minimal asymptotic upper bound. A bounded RS is referred to as an *asymptotically optimal RS* if there exists a restart parameter for which the asymptotic upper bound is minimal. This value is called the *optimal choice* for the restart parameter, which is independent of $\hat{\lambda}$. The optimal choice of the restart parameter is also denoted with $\hat{\rho}$.

Similarly, one can define an *asymptotic lower bound* and an optimal choice of the restart parameter with respect to the lower bound. If there exist an infinite asymptotic lower bound, the RS is called *strictly unbounded*. It is evident that strict unboundedness sufficiently implies unboundedness. Nevertheless, the converse is not necessarily true.

4.1 Relative Loss and Optimal ν of the \mathcal{R}^+ -RS

Assuming that $\hat{\lambda}$ is known, then the loss functions are functions that depend on the restart parameter. The left plot of Fig. 5 represents this for the \mathcal{R}^+ -RS. The blue line represents the lower bound (16) and the green line represents the upper bound (20). The question is whether there is an optimal choice for ν that minimizes the loss. The derivative of the lower bound (16) is

$$\frac{d}{d\nu} \mathcal{L}_{\text{low}}^+(\nu) = -\frac{(\hat{\lambda} + \lambda_0)(\hat{\lambda} - \lambda_0)}{2\nu^2}, \quad (52)$$

which is non-zero for all ν and for $\hat{\lambda} > \lambda_0$. There is no optimal value of ν that minimizes the lower bound of the \mathcal{R}^+ loss function. This is also visible in the left plot of Fig. 5. The derivative of the upper bound (20) is

$$\frac{d}{d\nu} \mathcal{L}_{\text{up}}^+(\nu) = -\frac{(\hat{\lambda} - \lambda_0 - 1)(\hat{\lambda} + \lambda_0 - 1)}{2\nu^2} + 1 = -\frac{(\hat{\lambda} - 1)^2 - \lambda_0^2}{2\nu^2} + 1. \quad (53)$$

Setting the derivative to zero yields

$$\hat{\nu} = \sqrt{\frac{1}{2}((\hat{\lambda} - 1)^2 - \lambda_0^2)}, \quad (54)$$

which minimizes the upper bound of the loss function. This is evident in the left plot of Fig. 5, where the minimum of the upper bound occurs at $\nu \approx 70$. It is important to note that the value of $\hat{\nu}$

(54) strongly depends on $\hat{\lambda}$, which is generally unknown. As a result, there does not exist an optimal $\hat{\nu}$ independent of the unknown $\hat{\lambda}$. Therefore, \mathcal{R}^+ cannot be an asymptotically optimal restart strategy. Moreover, \mathcal{R}^+ is strictly unbounded, which is shown in the following theorem.

THEOREM 4.1. *Let $\ell^+(\hat{\lambda}, \nu)$ be the relative loss function (50) for the \mathcal{R}^+ -RS. Then $\ell^+(\hat{\lambda}, \nu)$ is strictly unbounded for all $\nu \in \mathbb{N} \setminus \{0\}$.*

PROOF. The lower bound (16) of the loss function implies that

$$\ell_{\text{low}}^+(\hat{\lambda}, \nu) := \frac{\mathcal{L}_{\text{low}}^+(\hat{\lambda}, \nu)}{\hat{\lambda}} = \frac{1}{2} \left(1 - \frac{\lambda_0}{\hat{\lambda}}\right) \left(\frac{\hat{\lambda} + \lambda_0}{\nu} - 1\right) \quad (55)$$

is a lower bound of ℓ^+ , which tends to infinity for $\hat{\lambda} \rightarrow \infty$. Therefore, an infinite asymptotic lower bound exists. \square

The relative loss function ℓ^+ is illustrated in the right plot of Fig. 5 for varying values of ν . It can be seen that for a fixed ν the relative loss exhibits a linear trend with $\hat{\lambda}$ and approaches infinity. Furthermore, for large $\hat{\lambda}$ values the relative loss increases for smaller ν values. The solid lines in the right plot of Fig. 5 show the lower bound (55) of the relative loss function.

4.2 Relative Loss and Optimal ρ of the Multiplicative Strategy Types

The relative loss (50) of the \mathcal{R}^* -RS is given by $\ell^*(\hat{\lambda}, \rho) = \mathcal{L}^*(\hat{\lambda}, \rho)/\hat{\lambda}$ and is represented in Fig. 6 as a function of $\hat{\lambda}$. In contrast to the \mathcal{R}^+ -RS (see the right plot of Fig. 5) the relative loss exhibits an upper bound. In order to derive the upper bound for the relative loss, start with the upper bound of the loss function (32) and divide it by $\hat{\lambda}$, i.e.,

$$\ell_{\text{up}}^*(\hat{\lambda}, \rho) := \frac{\mathcal{L}_{\text{up}}^*(\hat{\lambda}, \rho)}{\hat{\lambda}} = \rho + \frac{1}{\rho - 1} - \frac{\lambda_0}{\hat{\lambda}(\rho - 1)} + \frac{\ln(\hat{\lambda}/\lambda_0)}{\hat{\lambda} \ln(\rho)}. \quad (56)$$

The corresponding asymptotic upper bound is

$$\bar{\ell}_{\text{up}}^*(\rho) = \lim_{\hat{\lambda} \rightarrow \infty} \ell_{\text{up}}^*(\hat{\lambda}, \rho) = \rho + \frac{1}{\rho - 1}. \quad (57)$$

It can be seen that this expression is finite for all $\rho > 1$. It can thus be concluded that the \mathcal{R}^* -RS is bounded. The upper bound (56) is illustrated in Fig. 6 by the solid lines. It can be observed that for sufficient large $\hat{\lambda}$, the upper bound provides a satisfactory approximation of the upper corners of the relative loss function.

The left plot of Fig. 6 shows that the maximum relative loss increases with smaller ρ , while the middle plot shows that the maximum relative loss increases with larger ρ . Therefore, it can be assumed that there is a value of ρ where the maximal relative loss is minimal. In both figures, the smallest relative loss occurs at about $\rho = 2$. For sufficiently large $\hat{\lambda}$, these observations are independent of λ_0 . This can be seen in the right plot of Fig. 6, which shows the loss functions for $\lambda_0 = 10$. The maximum values of the asymptotic behavior are identical to those for $\lambda_0 = 2$. These observations are confirmed by the following theorem:

THEOREM 4.2. *The \mathcal{R}^* -RS is an asymptotic optimal RS with $\hat{\rho} = 2$. Furthermore, it holds for the asymptotic upper bound (57) $\bar{\ell}_{\text{up}}^*(\hat{\rho}) = 3$.*

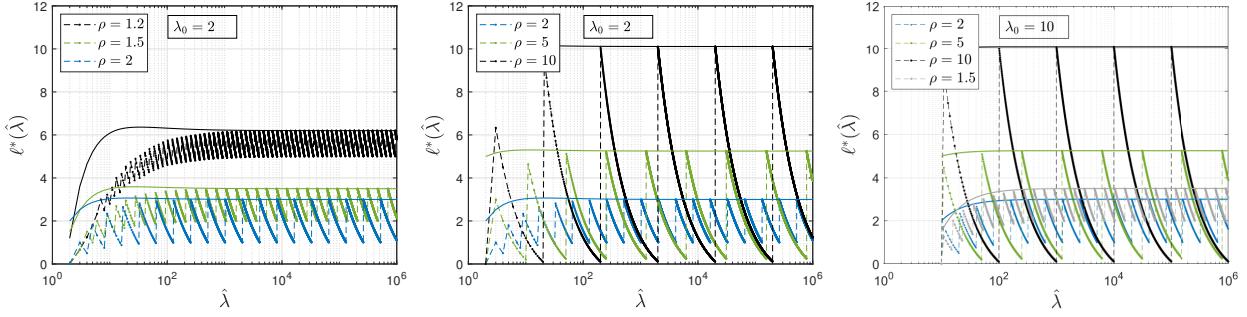


Figure 6: Markers with dashed lines: relative loss function of the \mathcal{R}^* -RS depending on $\hat{\lambda}$. Solid lines: upper bound of the relative loss function.

PROOF. Setting the derivative of the asymptotic upper bound (57) to zero yields

$$\frac{d}{d\rho} \bar{\ell}_{\text{up}}^*(\rho) = 1 - \frac{1}{(\rho-1)^2} = \frac{\rho^2 - 2\rho}{(\rho-1)^2} = 0 \Leftrightarrow \rho = 2, \quad (58)$$

thus indicating that the optimal restart parameter is $\hat{\rho} = 2$. Inserting this into the asymptotic upper bound (57), it follows that $\bar{\ell}_{\text{up}}^*(\hat{\rho}) = 3$. \square

Figure 6 illustrates that the asymptotic lower bound of the relative loss function is not minimal when $\rho = 2$. Instead, it is a monotonously decreasing function. The lower bound (39) implies that

$$\ell_{\text{low}}^*(\hat{\lambda}, \rho) = \frac{\mathcal{L}_{\text{low}}^*(\hat{\lambda}, \rho)}{\hat{\lambda}} = \frac{\hat{\lambda} - \lambda_0 - 1}{\hat{\lambda}(\rho-1)} - \frac{\ln(\hat{\lambda}/\lambda_0)}{\hat{\lambda}(\rho-1)\ln(\rho)} \quad (59)$$

is a lower bound of the relative loss function of the \mathcal{R}^* -RS. Consequently

$$\bar{\ell}_{\text{low}}^*(\rho) = \lim_{\hat{\lambda} \rightarrow \infty} \ell_{\text{low}}^*(\hat{\lambda}, \rho) = \frac{1}{\rho-1} \quad (60)$$

is an asymptotic lower bound of the relative loss function of the \mathcal{R}^* -RS. This expression is minimized when $\rho \rightarrow \infty$. Therefore, an asymptotic lower bound exist for all ρ , but there is not an optimal choice of ρ w.r.t. the lower bound.

The relative loss (50) of the \mathcal{R}^\times -RS is represented by the markers with dashed lines in the left plot of Fig. 7. It is evident that for integer values of ρ , the relative loss curves are identical to those of Fig. 6. However, even for non-integer numbers, the differences between the relative loss of the \mathcal{R}^\times -RS and the \mathcal{R}^* -RS are only barely visible. This behavior is expected, as demonstrated in Section 3, where it was shown that the upper bound (32) of the \mathcal{R}^* -RS is identical to that of the \mathcal{R}^\times -RS. It can thus be concluded that the upper bound of the relative loss function and the asymptotic upper bounds are identical, i.e.,

$$\bar{\ell}_{\text{up}}^\times(\rho) = \rho + \frac{1}{\rho-1}. \quad (61)$$

In consequence, Theorem 4.2 holds also for the \mathcal{R}^\times -RS.

THEOREM 4.3. *The \mathcal{R}^\times -RS is an asymptotic optimal RS with $\hat{\rho} = 2$. Furthermore, it holds for the asymptotic upper bound (61) $\bar{\ell}_{\text{up}}^\times(\hat{\rho}) = 3$.*

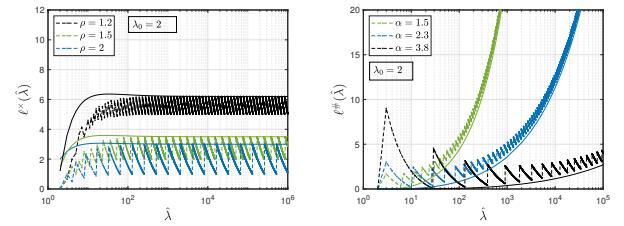


Figure 7: Markers with dashed lines: relative loss function of the \mathcal{R}^\times -RS (left) and the $\mathcal{R}^\#$ -RS (right) depending on $\hat{\lambda}$. Solid lines: upper bound (56) (left) and lower bound (62) (right) of the relative loss function.

PROOF. The proof can be done similar to those of Theorem 4.2. \square

4.3 Relative Loss of the $\mathcal{R}^\#$ -RS

The relative loss of the $\mathcal{R}^\#$ -RS is defined by $\ell^\#(\hat{\lambda}, \alpha) := \mathcal{L}^\#(\hat{\lambda}, \alpha)/\hat{\lambda}$ and is represented in the right plot of Fig. 7. It can be observed that for a fixed α , the relative loss tends to infinity with $\hat{\lambda}$. This leads to the hypothesis that $\mathcal{R}^\#$ is a strictly unbounded RS.

THEOREM 4.4. *The $\mathcal{R}^\#$ is a strictly unbounded strategy type.*

PROOF. By using the lower bound (42) of the loss function, it follows that

$$\begin{aligned} \ell_{\text{low}}^\#(\hat{\lambda}, \alpha) &= \frac{\mathcal{L}_{\text{low}}^\#(\hat{\lambda}, \alpha)}{\hat{\lambda}} = \frac{1}{\hat{\lambda}} \frac{\lambda_0}{\alpha+1} \left(\sqrt[\alpha]{\frac{\hat{\lambda}-1}{\lambda_0}} - 1 \right)^{\alpha+1} \\ &= \frac{\lambda_0}{\alpha+1} \frac{1}{\hat{\lambda}} \sqrt[\alpha]{\frac{\hat{\lambda}-1}{\lambda_0}}^{\alpha+1} \left(1 - \sqrt[\alpha]{\frac{\lambda_0}{\hat{\lambda}-1}} \right)^{\alpha+1} \end{aligned} \quad (62)$$

is a lower bound of the relative loss function for $\hat{\lambda} > \lambda_0$. For $\hat{\lambda} \rightarrow \infty$ the expression within the parentheses approaches 1, while the first terms diverges. Consequently, the asymptotic lower bound is infinite indicating that $\mathcal{R}^\#$ is a strictly unbounded RS. \square

The solid lines in the right plot of Fig. 7 illustrate the lower bound (62) of the relative loss function.

5 Comparison with Experiments

This section presents comparative experimental results for restart optimization algorithms to which the aforementioned theoretical results can be applied. Based on the above considerations, a multiplicative restart is preferable to the other strategy types and the two multiplicative types can be considered as interchangeable. Therefore, only the \mathcal{R}^* -RS will be considered in the following experiments.

As previously mentioned, ES in multimodal landscapes fulfill the above assumptions with regard to the underlying optimization algorithm A. Therefore, in the following experiments a σ SA-ES (see Alg. 2) will be restarted according to the \mathcal{R}^* -RS. This will henceforth be abbreviated as \mathcal{R}^* - σ SA-ES.

5.1 Experimental Setup

The data presented in Fig. 8 has been calculated using the $(\mu/\mu_I, \lambda)$ - σ SA-ES, which is outlined in detail in Alg. 2. Within each generation, the population comprises μ parents and λ offspring individuals with a truncation ratio $\vartheta := \mu/\lambda$. The offspring individuals are generated by isotropic Gaussian mutations with mutation strength σ (Line 5). The mutation strength of each offspring was sampled from a log-normal distribution with *learning parameter* τ . In the subsequent step, the fitness value of each offspring individual is evaluated (Line 6) and the individuals are sorted in accordance with their fitness values (Line 8). The individuals with the best fitness and their corresponding mutation strengths are then recombined to obtain a new parent centroid $\mathbf{y}^{(g+1)}$ and a new parental mutation strength $\sigma^{(g+1)}$ (Lines 10 and 11). The subscript $m; \lambda$ denotes the selection of the $m = 1, \dots, \mu$ best individuals.

Algorithm 2 The $(\mu/\mu_I, \lambda)$ - σ SA Evolution Strategy

```

1: Initialize  $(\mathbf{y}^{(0)}, \sigma^{(0)}, \mathbf{s} = \mathbf{1}, g_{\max}, g = 0)$ 
2: repeat
3:   for  $l = 1$  to  $\lambda$  do
4:      $\tilde{\sigma}_l = \sigma^{(g)} e^{\tau N(0,1)}$                                 ▷ mutate parental  $\sigma$ 
5:      $\tilde{\mathbf{y}}_l = \mathbf{y}^{(g)} + \tilde{\sigma}_l (N(0, 1), \dots, N(0, 1))$     ▷ mutate  $\mathbf{y}$ 
6:      $\tilde{F}_l = F(\tilde{\mathbf{y}}_l)$                                          ▷ evaluate offspring
7:   end for
8:   Sort Individuals  $\tilde{\mathbf{y}}$  Ascendingly w.r.t. Fitness  $\tilde{F}$ 
9:    $g = g + 1$ 
10:   $\mathbf{y}^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m; \lambda}$                       ▷ recombine the  $\mu$  best  $\tilde{\mathbf{y}}$ 
11:   $\sigma^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m; \lambda}$                       ▷ recombine the  $\mu$  best  $\tilde{\sigma}$ 
12: until  $g \geq g_{\max}$ 

```

The ES, Alg. 2, stops upon attaining a maximum number of generations g_{\max} . Consequently, all runs utilize precisely the same number of generations and the number of function evaluations is linear in λ , i.e., $F_E(\lambda) = g\lambda$. The ES-run is deemed successful if $F(\mathbf{y}^{(g_{\max})}) - F(\hat{\mathbf{y}}) < 10^{-3}$, where $\hat{\mathbf{y}}$ is the global optimizer.

All subsequent experiments were conducted with $\vartheta = 0.5$ and $\tau = 1/\sqrt{2N}$, a value of τ that is widely adopted as the standard for the learning parameter. It ensures optimal performance on the sphere model [14]. An elevated value of τ results in a faster adaptation. This usually results in a lower success rate in multimodal landscapes [10]. As for global optimization in multimodal landscapes

the optimal sphere τ is rather large. W.r.t. the restart strategy this means that a rather large $\hat{\lambda}$ is needed to get to the global optimizer. However, this high effort is partially compensated by a smaller g_{\max} due to the higher convergence rate. That is, success can be achieved for relatively small values of g_{\max} . This scenario is intended in the experiments since the theoretical results are valid for large $\hat{\lambda}$.

For the objective function to be optimized by the σ SA-ES, highly multimodal landscapes were selected, i.e., the well-known Rastrigin function and the Ackley function. The Rastrigin function for $\mathbf{y} \in \mathbb{R}^N$ is given by

$$F_R(\mathbf{y}) = \sum_{i=1}^N [y_i^2 + A(1 - \cos(\alpha y_i))]. \quad (63)$$

The parameter A is the oscillation amplitude and α is the frequency with standard values of $A = 10$ and $\alpha = 2\pi$. The global optimizer of the Rastrigin function is located at $\hat{\mathbf{y}} = \mathbf{0}$. The number of local minima increases exponentially with the search space dimensionality N .

The Ackley function for $\mathbf{y} \in \mathbb{R}^N$ is given by

$$F_A(\mathbf{y}) = C_1 - C_1 e^{-C_2 \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}} + e^{-e^{\frac{1}{N} \sum_{i=1}^N \cos(\gamma y_i)}}. \quad (64)$$

In the $N = 2$ case it represents a funnel-shaped surface with an infinite number of local minima and a global optimizer at $\hat{\mathbf{y}} = \mathbf{0}$. In all experiments, the standard Ackley parameters $C_1 = 20$, $C_2 = 0.2$ and $\gamma = 2\pi$ are used.

5.2 Experimental Results

The impact of the restart parameter ρ on the relative loss of an \mathcal{R}^* - σ SA-ES is illustrated in the left and middle plot of Fig. 8. For each value of ρ , 51 independent \mathcal{R}^* - σ SA-ES runs were performed. Subsequently, for each run the total number of function evaluations F_E and the value of λ at the last restart were recorded. The markers show the median value over all 51 runs. The green dots show the value of λ at the last restart. They are represented by the right y -axis. The black stars show the experimental relative loss, which is represented by the left y -axis. The sum in the loss function (10) can be determined experimentally by F_E/g_{\max} . Therefore the relative loss can be modeled by

$$\ell^*(\rho) = \frac{F_E/g_{\max} - \hat{\lambda}}{\hat{\lambda}}, \quad (65)$$

The values of $\hat{\lambda}$ were determined experimentally in the corresponding landscape⁹.

The gray dashed-dotted lines in Fig. 8 shows the theoretical upper and lower bound of the relative loss function. In the majority of cases, the experimental values of the relative loss fall between these bounds. In certain instances, experimental values are slightly larger or smaller than the predicted bounds of the relative loss. However, these deviations are not unexpected given the relatively small sample size of 51 runs.

As demonstrated in Fig. 8, the minimum of ℓ^* is not attained at $\rho = 2$ or in its vicinity, as would be expected, since the global minimum is reached at a value of $\rho = \hat{\lambda}/\lambda_0$, for which only a single

⁹The experimental value of λ is obtained where the success rate exceeds 50%. This is the value where the median run is assumed to be successful. The individual success rates were determined with a minimum of 1 000 runs.

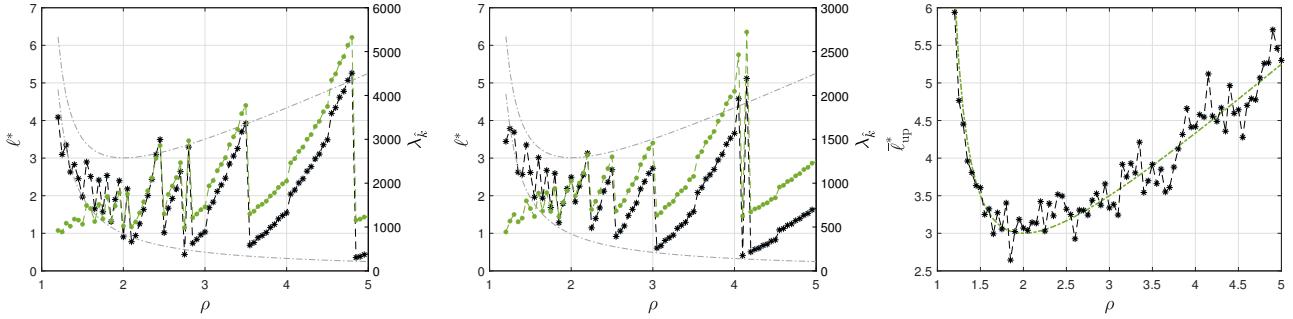


Figure 8: Population size of the last restart $\lambda_{\hat{k}}$ (green dots) and relative loss ℓ^* (65) (black stars) for the $\mathcal{R}^*\text{-}\sigma\text{SA-ES}$. Gray dashed-dotted lines show ℓ^*_{up} (56) and ℓ^*_{low} (59). Markers show the median value of 51 runs with $\lambda_0 = 2$. Left figure: Rastrigin landscape (63) with $N = 50$, $A = 1$, $\alpha = 6\pi$, $g_{\max} = 500$ and $\hat{\lambda} = 1074$. Middle figure: Ackley landscape (64) with $N = 100$, $g_{\max} = 2000$, initial start distance $R_{\text{init}} = 40\sqrt{N}$ and $\hat{\lambda} = 586$. Right figure: Black stars show the maximum relative loss (65) over 40 different Ackley and Rastrigin landscapes. Green dashed dotted line shows the asymptotic upper bound $\bar{\ell}_{\text{up}}^*(\rho)$ (57).

restart is necessary. The minimal value in the left plot occurs at $\rho = 4.85$, with the corresponding $\lambda_{\hat{k}} = 1155$. In the middle plot the minimal value of ℓ^* is attained at a value of $\rho = 4.1$, with $\lambda_{\hat{k}} = 624$. In both scenarios, only 4 restarts are required to exceed $\hat{\lambda}$. The values of ℓ^* and $\lambda_{\hat{k}}$ for adjacent values of ρ can vary significantly. However, it is observed that in the vicinity of $\rho = 2$ the discontinuities are relatively small.

In consideration of the aforementioned reasoning, it is not possible to experimentally validate a general optimum for an RS when considering a single problem. Therefore, a multitude of landscapes will be considered in order to cover a wide range of possible $\hat{\lambda}$. For ES in the Rastrigin landscape, the optimal population size depends strongly on both the dimension and the Rastrigin parameters A and α . As demonstrated in [15], the scaling behavior of $\hat{\lambda}$ with respect to A is linear, while it is quadratic with respect to α . For the dimension N , the scaling behavior is given by $\sqrt{N} \ln(N)$. For the Ackley function, the dependence on N is approximately $\ln(N)$ [10]. It is much weaker compared to the case of the Rastrigin function. In contrast to the Rastrigin function, the distance between the starting point and the optimizer is decisive for the success in the Ackley landscape. It has been demonstrated that, from a certain distance to the optimizer, the population size required for successful execution increases exponentially [10]. For $N = 100$ the critical distance is approximately at 300. Variation of these parameters will therefore ensure that a large range of different values for $\hat{\lambda}$ is covered. The largest relative loss over all landscapes for any given ρ can then serve as a measure for the maximal relative loss.

It is important to acknowledge the limitation of theoretical results concerning the relative loss, which are only asymptotically exact for $\hat{\lambda} \rightarrow \infty$. This should be kept in mind when trying to empirically check the predictions. Therefore, experimental setups are required to provide sufficiently large values of $\hat{\lambda}$. However, as illustrated in Fig. 6 the asymptotic behavior is nearly reached for $\hat{\lambda} = 100$.

The experimental results in the right plot of Fig. 8 are the results from 40 different Rastrigin and Ackley landscapes. In the Rastrigin

landscape the dimension and Rastrigin parameters A and α were varied. This resulted in $\hat{\lambda}$ -values ranging from 74 to 1084. In the Ackley landscape, the dimension and the initial start distance were varied, resulting in λ values between 28 and 2800¹⁰. For each value of ρ the worst value of ℓ^* was taken across all landscapes. The results are shown in the right plot of Fig. 8. Due to the small sample size, there are some deviations between the theoretical and experimental values. However, the general trend is clearly visible. For $\rho = 2$ the maximum relative loss is approximately 3. Away from this value, there is a tendency for the loss to increase sharply.

6 Conclusion and Outlook

This work examined restart strategies (RS) for algorithms that rely on an algorithmic parameter, denoted by λ , to achieve success. The optimal choice of this algorithmic parameter for each restart is a problem that has not been investigated so far. To estimate and compare different restart strategies, the set of restart strategies was divided into parameter-dependent subsets, which were classified as strategy types. The objective was to evaluate the impact of varying restart strategies and to estimate the influence of the restart parameter. For this purpose, a loss function was introduced. It measures the number of function evaluations of an RS compared to the number of function evaluations of the optimal strategy. Due to the complexity of the loss function, upper and lower bounds of the loss function were derived for each strategy type under consideration. These bounds have been expressed as a function of the optimal λ and the restart parameter.

To further examine the restart effort, the relative loss has been introduced as a measure of the loss relative to the optimal λ . A prerequisite for an appropriate RS is that there exists a finite upper bound for the relative loss function. Strategy types that satisfy this criterion are called bounded. For strategy types whose relative loss functions are upper bounded it is possible to minimize the upper bound according to the parameter of the strategy type. The analyses in this paper have shown that the strategy types \mathcal{R}^+ and $\mathcal{R}^\#$ are not

¹⁰Further information regarding the utilized landscapes can be found in the supplementary material.

well-suited as restart strategies. In the case of the strategy type \mathcal{R}^+ , where λ is increased additively by the same amount v after each restart, it has been demonstrated that the relative loss function is unbounded. This is also the case for $\mathcal{R}^\#$. For this strategy type, λ is determined according to a power law.

In the case of the multiplicative strategy type \mathcal{R}^* -RS, where the algorithmic parameter λ is multiplied by a restart parameter ρ for each restart, it was demonstrated that the relative loss function is bounded. Examining this strategy type, it was demonstrated that there exists an optimal choice of ρ which minimizes the asymptotic upper bound of the relative loss function. This value was found to be $\hat{\rho} = 2$. It is independent of the start value λ_0 . Furthermore, it was shown that there is no value of ρ that minimizes the asymptotic lower bound of the relative loss function. The same results have been derived for the \mathcal{R}^\times -RS which is also a multiplicative strategy type.

Using $\rho = \hat{\rho}$, the maximum relative loss w.r.t. $\hat{\lambda}$ is 3. This result is remarkable indicating that even in the worst case the performance of the strategies degrades by only a (constant) factor of three. This is in contrast to Luby's universal restart strategy. As has been shown in [4], the effort of the Luby sequence \mathbf{L} increases according to $O(\lambda \ln(\lambda))$. Thus, the relative loss is in $O(\ln(\lambda))$, i.e., it increases logarithmically while the \mathcal{R}^* - and \mathcal{R}^\times -RS guarantee (theoretically) a constant factor of at most 3 independent of λ . Of course this advantage is bought at the price of a priori information that there exists an unknown $\hat{\lambda}$ above which the algorithm successfully reaches the global attractor. Values $\lambda < \hat{\lambda}$ are assumed to be never successful. That is why, the repetitions in the Luby sequence \mathbf{L} are wasted in such a scenario. The investigation in this paper is limited to a specific scenario that may not apply to all optimization algorithms and objective functions. These conditions are applicable to ES, albeit in a limited capacity. Condition (1), for instance, is not met by all objective functions, however, it is applicable to a class of highly multimodal test functions. Additionally, the representability of the loss function as in (10) depends on the termination criteria. While these conditions indicate a rather narrow scope of the multiplicative restart strategy, the analysis presented in this paper provides at least a first theoretical justification of the common use of such strategies in EA implementations such as the IPOP-CMA-ES.

In this work, the worst case scenario has been investigated. Alternatively, some kind of amortized analysis seems to be possible, i.e., the average relative loss can also be considered. The restart parameter that minimizes the average relative loss must be necessarily larger than $\hat{\rho}$. This is a topic that is currently under investigation.

While it has been demonstrated in this paper that the multiplicative strategy types are preferable to the other strategy types under consideration, it is currently unclear whether there exist other strategy types whose relative loss is bounded as it is the case for the multiplicative strategy types. This is a topic for future research.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF) under grant P33702-N.

References

- [1] A. Auger and N. Hansen. 2005. A Restart CMA Evolution Strategy with Increasing Population Size. In *IEEE Congress on Evolutionary Computation*. Vol. 2, pp. 1769–1776. DOI: 10.1109/CEC.2005.1554902
- [2] N. Hansen. 2009. Benchmarking a BI-Population CMA-ES on the BBOB-2009 Function Testbed. In *Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference*. Association for Computing Machinery, pp. 2389–2395.
- [3] H.-G. Beyer and S. Finck. 2012. On the Design of Constraint Covariance Matrix Self-Adaptation Evolution Strategies Including a Cardinality Constraint. *IEEE Transactions on Evolutionary Computation*. Vol. 6, no. 4, pp. 578–596. DOI: 10.1109/TEVC.2011.2169967
- [4] M. Luby, A. Sinclair, and D. Zuckerman. 1993. Optimal speedup of Las Vegas algorithms. In *[1993] The 2nd Israel Symposium on Theory and Computing Systems*. pp. 128–133. DOI: 10.1109/ISTCS.1993.253477
- [5] J.-H. Lorenz. 2021. Restart Strategies in a Continuous Setting. *Theory of Computing Systems*. Vol. 65, pp. 1143–1164. DOI: 10.1007/s00224-021-10041-0
- [6] T. Friedrich, T. Kötzing, F. Quinzan, and A. M. Sutton. 2018. Improving the Run Time of the (1 + 1) Evolutionary Algorithm with Luby Sequences. In *GECCO '18, Kyoto, Japan: Proceedings of the Genetic and Evolutionary Computation Conference*. pp. 301–308. DOI: 10.1145/3205455.3205525
- [7] T. Friedrich, T. Kötzing, and M. Wagner. 2017. A generic bet-and-run strategy for speeding up stochastic local search. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31.
- [8] M. Fischetti and M. Monaci. 2014. Exploiting erraticism in search. *Operations Research*. Vol. 62, pp. 114–122.
- [9] H.-G. Beyer and H.-P. Schwefel. 2002. Evolution Strategies: A Comprehensive Introduction. *Natural Computing*. Vol. 1, pp. 3–52.
- [10] L. Schönenberger and H.-G. Beyer. 2024. On a Population Sizing Model for Evolution Strategies in Multimodal Landscapes. *IEEE Transactions on Evolutionary Computation*. DOI: 10.1109/TEVC.2024.3419931
- [11] T. Jansen. 2002. On the Analysis of Dynamic Restart Strategies for Evolutionary Algorithms. In *Parallel Problem Solving from Nature – PPSN VII*. Springer Berlin Heidelberg, pp. 33–43.
- [12] M. Buzdalov. 2014. A Switch-and-Restart Algorithm with Exponential Restart Strategy for Objective Selection and its Runtime Analysis. In *2014 13th International Conference on Machine Learning and Applications*. pp. 141–146.
- [13] A. Omeradzic and H.-G. Beyer. 2024. Self-Adaptation of Multi-Recombinant Evolution Strategies on the Highly Multimodal Rastrigin Function. *IEEE Transactions on Evolutionary Computation*. DOI: 10.1109/TEVC.2024.3400857
- [14] S. Meyer-Nieberg. 2007. *Self-Adaptation in Evolution Strategies*. Ph.D. Dissertation. University of Dortmund, CS Department.
- [15] L. Schönenberger and H.-G. Beyer. 2023. On a Population Sizing Model for Evolution Strategies Optimizing the Highly Multimodal Rastrigin Function. In *GECCO '23*. Association for Computing Machinery, pp. 848–855.