Curse of Scale-Freeness: Intractability of Large-Scale Combinatorial Optimization with Multi-Start Methods

Hiroyuki Masuyama\textsuperscript{a}

Graduate School of Management, Tokyo Metropolitan University
Tokyo 192–0397, Japan

Hiroshige Dan\textsuperscript{b}

School of Creative Science and Engineering, Shinjuku-ku, Tokyo, 169–8555, Japan

Shunji Umetani\textsuperscript{c}

Department of Information and Physical Sciences, Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, 565–0871, Japan

Abstract

This paper investigates the intractability of large-scale optimization using multi-start methods. For the theoretical performance analysis, we focus on random multi-start (RMS), one of the representative multi-start methods, including RMS local search and greedy randomized adaptive search procedure (GRASP). Our primary theoretical contribution is to derive, using extreme value theory, power-law formulas for the two quantities: (i) the expected improvement rate of the best empirical objective value (EOV); (ii) the expected relative gap between the best EOV and the supremum of empirical objective values. Notably, the expected relative gap exhibits scale-freeness as a function of the number of iterations. Consequently, the half-life of the expected relative gap is ultimately proportional to the number of iterations completed by an RMS algorithm. This result can be interpreted as the curse of scale-freeness—a Zeno’s paradox-like phenomenon, encapsulated by the metaphor “reaching for the goal makes it slip away.” Through numerical experiments, we observe that applying several RMS algorithms to traveling salesman problems is subject to the curse of scale-freeness. Furthermore, we show that overcoming this curse requires developing a powerful LS algorithm equipped with a diversification mechanism that is exponentially more effective than RMS.

Keywords: multi-start method; extreme value theory (EVT); scale-free; power law; traveling salesman problem (TSP); Zeno’s paradox

Mathematics Subject Classification: 90C27; 60G70; 62G32

\textsuperscript{a}E-mail: masuyama@tmu.ac.jp
\textsuperscript{b}E-mail: dan@waseda.jp
\textsuperscript{c}E-mail: umetani@ist.osaka-u.ac.jp
1 Introduction

The purpose of this paper is to statistically reveal the intractability of large-scale optimization using the multi-start method (Martí et al., 2013, 2018). The multi-start method is a standard metaheuristic strategy that repeatedly restarts a local search (LS) algorithm with a diversification (or exploration) mechanism. In terms of diversification mechanisms, multi-start methods can be classified into two types: (i) random multi-start (RMS) methods (see Algorithm 1), such as random multi-start local search and greedy randomized adaptive search procedure (GRASP) (Resende and Rebeiro, 2016); and (ii) random perturbation methods (see Algorithm 2), such as iterated local search (ILS) (Lourenço et al., 2019), variable neighborhood search (Hansen and Mladenović, 2001), and memetic algorithms (Neri et al., 2012). An RMS algorithm generates an initial solution from scratch using a randomized construction algorithm, whereas random perturbation methods do so by perturbing an existing good (often locally optimal) solution. In general, the multi-start method achieves relatively good performance despite its simple implementation, making it a standard benchmark for designing heuristic algorithms in many real-world applications.

Algorithm 1: Random multi-start (RMS) for maximization problems

1: $t ← 1$ \hfill $\triangleright \ t$ is the number of trials.
2: $x^t ← -\infty$ \hfill $\triangleright \ x^t$ is the variable for the best empirical objective value (EOV).
3: while $t ≤ T$ do
4: \hspace{1em} $s^0 ← \text{Random}$ \hfill $\triangleright$ Randomly generate an initial solution $s^0$.
5: \hspace{1em} $s ← \text{LS}(s^0)$ \hfill $\triangleright$ Generate an empirical solution $s$ from $s^0$ by LS.
6: \hspace{1em} if $x(s) > x^t$ then \hfill $\triangleright \ x(s)$ is the objective value of the empirical solution $s$.
7: \hspace{1.5em} $s^t ← s$ \hfill $\triangleright$ Update the best empirical solution $s^t$ with $s$.
8: \hspace{1.5em} $x^t ← x(s^t)$ \hfill $\triangleright$ Update the best EOV $x^t$ with $x(s^t)$.
9: \hspace{1em} end if
10: \hspace{1em} $t ← t + 1$
11: \hspace{1em} end while
12: return $s^t$

This paper clarifies the intractability of large-scale optimization using the multi-start method in the following way. First, we apply Extreme Value Theory (EVT) to the process of solving a large-scale optimization problem using an RMS algorithm (Algorithm 1), which is mathematically easy to analyze. We then derive power-law formulas for the improvement of the empirical objective values (EOVs)—where “objective value” is an abbreviation for “objective function value”—generated by repeatedly running the RMS algorithm. These power-law formulas provide a theoretical basis for the intractability of large-scale optimization, which we figuratively express as the curse of scale-freeness. Furthermore, through numerical experiments—not theoretical analysis using EVT—we confirm that even the ILS algorithm (Algorithm 2), commonly known as a high-performance random perturbation method, cannot break the curse of scale-freeness.
Algorithm 2: Random perturbation for maximization problems

1: \( t \leftarrow 1 \)  
2: \( s^0 \leftarrow \text{Initialize} \)  
3: \( s^\dagger \leftarrow \text{LS}(s^0) \)  
4: \( x^\dagger \leftarrow x(s^\dagger) \)  
5: while \( t \leq T \) do  
6: \( s^0 \leftarrow \text{Kick}(s^\dagger) \)  
7: \( s \leftarrow \text{LS}(s^0) \)  
8: if \( x(s) > x^\dagger \) then  
9: \( s^\dagger \leftarrow s \)  
10: \( x^\dagger \leftarrow x(s^\dagger) \)  
11: end if  
12: \( t \leftarrow t + 1 \)  
13: end while  
14: return \( s^\dagger \)

As in [Giddings et al., 2014], this paper focuses on a large-scale mixed-integer programming problem with a feasible domain \( S \), a finite set of decision variables \( s \in S \), and a nonnegative objective function \( x(s) \) to be maximized:

\[
\begin{align*}
\text{maximize} & \quad x(s) \in \mathbb{R}_+ := [0, \infty) \\
\text{subject to} & \quad s \in S.
\end{align*}
\]

(1.1a) (1.1b)

The following assumptions are maintained throughout this paper.

Assumption 1.1 An RMS algorithm generates a single empirical solution \( s_n \in S \) of Problem (1.1) for each trial \( n \in \mathbb{N} := \{1, 2, \ldots\} \), forming the set of good (typically locally optimal) solutions \( S_G = \{s_n; n = 1, 2, \ldots\} \) (see Figure 1). We assume that \( S \) and its subset \( S_G \) are infinite or sufficiently large to be considered infinite. Thus, the EOVs \( X_n := x(s_n), n \in \mathbb{N}, \text{are independent and identically distributed (i.i.d.), implying that an RMS algorithm is a random sampler from } S_G \) (Golden 1977, 1978; Golden and Alt 1979). Let \( X \) denote a generic random variable from the i.i.d. sequence \( \{X_n\} \) with distribution function \( F \):

\[
F(x) = P(X \leq x) = P(X_n \leq x), \quad x \in \mathbb{R}_+.
\]

Let \( x^* \) denote the right endpoint of \( F \), i.e.,

\[
x^* = \sup\{x \in \mathbb{R}_+ : F(x) < 1\},
\]

which is the supremum of EOVs and is assumed to be positive.

Remark 1.2 If the feasible domain \( S \) is finite and the used RMS algorithm is powerful enough to reach the (global) optimal solution given a suitable initial solution, the supremum \( x^* \) of EOVs coincides with the optimal value. Otherwise, \( x^* \) may be infinite in some cases, or finite but not equal to the optimal value.
Remark 1.3 Each time a locally optimal solution is reached, an RMS algorithm restarts the local search with an initial solution generated in the same stochastic manner, independent of the computational history. Consequently, it is reasonable to assume that the EOVs generated by the RMS algorithm satisfy the i.i.d. assumption.

A related topic to our research is the estimation of the optimal value, which may or may not be identical to the supremum $x^*$ of the EOVs (see Remark 1.2). Estimation methods for optimal values can be broadly classified into two types. The first is the relaxation method (see, e.g., Held and Karp 1970, Fisher et al. 1975), which establishes bounds on the optimal value by solving relaxation problems. The second is the statistical method, which provides point and interval estimates of the optimal value using statistical tools such as EVT. Pioneering work on optimal value estimation using EVT was conducted by Golden (1977) (see also Golden 1978, Golden and Alt 1979). Giddings et al. (2014) provides a comprehensive survey not only on optimal value estimation using EVT but also on other statistical methods. Following the publication of Giddings et al. (2014), Carling and Meng (2015, 2016) evaluated the quality of heuristic solutions by estimating the optimal value using the Weibull distribution. Several other papers discuss the estimation of optimal values using extreme value theory for specific optimization problems; see Section 1 of Velasco et al. (2022). However, Velasco et al. (2022) raises concerns about the reliability of optimal value estimation using EVT.

Note that even an accurate estimate of the gap between the best EOV and the optimal value does not necessarily lead to knowing how much effort is required to close that gap. In many cases, the gap narrows rapidly in the early stages of computation, but relatively quickly reaches a “steady state” where further reduction rarely occurs. Figure 2a illustrates a typical example of this behavior.

Figure 2b (the log-log plot of Figure 2a) shows that the evolution of the relative gap of the best EOV is approximately linear. Thus, the power-law behavior and scale-freeness appear to be evident in the evolution of the relative gap of the best EOV. In other words, the number of additional trials needed to halve the relative gap increases asymptotically in proportion to the number of trials completed. This Zeno’s paradox-like phenomenon can be figuratively described as the curse of scale-freeness: “reaching for the goal makes it slip away.” As discussed later, the curse of scale-freeness essentially represents the intractability of large-scale optimization (at least when using the RMS method).

The above considerations are supported by the power-law formulas associated with the evolution of the relative gap of the best EOV, which are summarized below.
Figure 2: Evolution of the relative gap (error) between the best EOV and the optimal value for an RMS algorithm applied to 100 random TSP (Traveling Salesman Problem) instances. Figure 2b is identical to Figure 3d, while Figure 2a shows the same data on a normal scale.

Summary of the power-law formulas: Let $\forall_{i=k}^{\ell}X_i = \max(X_k, X_{k+1}, \ldots, X_{\ell})$ for $k, \ell \in \mathbb{N}$ such that $k < \ell$. Since $\{X_n\}$ is i.i.d., the random variable $\forall_{i=N+1}^{N+n}X_i$ is independent of $Z_N := \forall_{i=1}^{N}X_i$ and has the same distribution as $Z_n := \forall_{i=1}^{n}X_i$. Let

$$R_n(x) = \frac{(Z_n - x)_+}{x}, \quad 0 < x < x^*, \; n \in \mathbb{N},$$

(1.3)

where $(x)_+ = \max(x, 0)$ for $x \in \mathbb{R} := (-\infty, \infty)$. Let

$$\Delta_n(x) = \frac{x^* - \max(Z_n, x)}{x^*} = \frac{x^* - x}{x^*} - \frac{(Z_n - x)_+}{x^*}, \quad 0 < x < x^*, \; n \in \mathbb{N},$$

(1.4)

where $x^* < \infty$ is always assumed when considering $\Delta_n(x)$. It follows from (1.3) and (1.4) that $R_n(x)$ and $\Delta_n(x)$ are related by the affine equation:

$$\Delta_n(x) = \frac{x^* - x}{x^*} - \frac{x}{x^*}R_n(x).$$

(1.5)

Note that, given that the best EOV is $x$, (i) $\mathbb{E}[R_n(x)]$ represents the (conditional) expected improvement rate of the best EOV after $n$ additional trials of an RMS algorithm, and (ii) $\mathbb{E}[\Delta_n(x)]$ represents the (conditional) expected relative gap of the best EOV after $n$ additional trials, measured from the supremum $x^* < \infty$ of the EOVs. The two expectations, $\mathbb{E}[R_n(x)]$ and $\mathbb{E}[\Delta_n(x)]$, follow the power-law formulas given below:

$$\mathbb{E}[R_n(x)] = \frac{x^* - x}{x} - \frac{x^*}{x}\mathbb{E}[\Delta_n(x)], \quad 0 < x < x^*,$$

(1.6a)

$$\mathbb{E}[\Delta_n(x)] = L(n)n^\xi,$n^\xi}, \quad 0 < x < x^*,$$

(1.6b)

where $\xi < 0$ is a constant and $L$ is a slowly varying function (see Definition 2.4). The above results are presented more rigorously and in detail in Section 3, along with other related results.

The power-law formulas in (1.6) provide the theoretical foundation for the curse of scale-freeness. Equation (1.6) shows that additional trials increase $\mathbb{E}[R_n(x)]$ toward its supremum
\( (x^* - x)/x \) at the same rate as the decrease in \( (x^*/x)E[\Delta_n(x)] \), which decays toward zero at a power-law rate of order \( n^\xi \). Notably, (1.6b) shows that \( E[\Delta_n(x)] \) is asymptotically independent of \( x \) and has a scale-free property: For \( x \in (0, x^*) \),

\[
\lim_{n \to \infty} \frac{E[\Delta_{cn}(x)]}{E[\Delta_n(x)]} = c^\xi, \quad 0 < x < x^*, \ c \in \mathbb{N},
\]

which yields

\[
\lim_{n \to \infty} \frac{E[\Delta_{[2^{-1/\xi}]n}(x)]}{E[\Delta_n(x)]} = \left[2^{-1/\xi}\right]^\xi \leq \frac{1}{2} \leq \left[2^{-1/\xi}\right]^\xi = \lim_{n \to \infty} \frac{E[\Delta_{[2^{-1/\xi}]n}(x)]}{E[\Delta_n(x)]}, \tag{1.7}
\]

where \( \lfloor \cdot \rfloor \) and \( \lceil \cdot \rceil \) denote the floor and ceiling functions, respectively. Although suggested by Figure 2b, Equation (1.7) implies that the half-life of the expected relative gap is asymptotically proportional to the number of trials completed. This would deserve to be described as “reaching for the goal makes it slip away.”

Our power-law formulas (1.6) imply that overcoming the curse of scale-freeness requires developing a powerful LS algorithm equipped with a diversification mechanism that is exponentially more effective than RMS (Section 4.2). To validate this intractability, we apply three RMS and three ILS algorithms to the five instances (Table 1) of the Traveling Salesman Problem (TSP) taken from TSPLIB (see Appendix B for details) and summarize the numerical results in Figures 5–7. From these figures, we observe that the relative gap between the best EOV and the optimal value decays at most at a power-law rate. Thus, the ILS algorithms cannot overcome the curse of scale-freeness. It is well known that ILS is one of the most powerful metaheuristics among multi-start methods. Therefore, other multi-start algorithms are also likely to be trapped by the curse of scale-freeness.

The rest of this paper is organized as follows. Section 2 provides the preliminaries for our EVT analysis of empirical solutions generated by an RMS algorithm. Section 3 presents the main theoretical results of this paper, including the power-law formulas mentioned above. Based on these theoretical results, Section 4 discusses the intractability of large-scale optimization with multi-start methods. Finally, Section 5 offers concluding remarks.

## 2 Preliminaries

This section consists of two subsections. Section 2.1 introduces the fundamental assumption of EVT, which forms the basis for the theoretical arguments presented later. Section 2.2 presents preliminary results on regularly varying functions, which are necessary to derive the power-law formulas for the expected improvement rate and expected relative gap discussed in the next section.

For later reference, we introduce some symbols for functions and sets of numbers. For any eventually positive function \( f \) on \( \mathbb{R} \), the notation \( g(x) = o(f(x)) \) represents \( \lim_{x \to \infty} |g(x)|/f(x) = 0 \). In addition, for any non-decreasing function \( f \), let \( f^- \) denote the left-continuous inverse of \( f \), i.e.,

\[
f^-(x) = \inf\{y \in \mathbb{R} : f(y) \geq x\}.
\]
Finally, let $1(\cdot)$ denote the indicator function so that it equals one if the statement in the parentheses is true; otherwise, it equals zero.

### 2.1 The fundamental assumption of EVT

This section introduces the fundamental assumption of EVT (de Haan and Ferreira, 2006, Section 1.1) in addition to Assumption 1.1. The fundamental assumption is reformulated as two equivalent conditions. These form the basis for deriving the theoretical results presented in the following sections.

The fundamental assumption of EVT is as follows:

**Assumption 2.1** There exist two sequences of constants $\{a_n > 0; n \in \mathbb{N}\}$ and $\{b_n \in \mathbb{R}; n \in \mathbb{N}\}$ and a non-degenerate distribution $G$ such that

$$
\lim_{n \to \infty} P \left( \frac{Z_n - b_n}{a_n} \leq z \right) = \lim_{n \to \infty} [F(a_n z + b_n)]^n = G(z),
$$

for every continuity point $z$ of $G$. In this case, $F$ is said to be in the maximum domain of attraction (MDA) of $G$, denoted by $F \in \text{MDA}(G)$.

For later discussion, we present two conditions equivalent to Assumption 2.1. To describe them, we introduce a specific function $U$ and the standard GEV distribution as follows. Let $U$ denote a function such that

$$
U(t) = \left( \frac{1}{1 - F} \right)^\left( t \right) = \inf \left\{ x \in \mathbb{R} : \frac{1}{1 - F(x)} \geq t \right\}, \quad t \geq 1.
$$

Equations (2.1) and (1.2) result in

$$
\lim_{t \to \infty} U(t) = x^*.
$$

Furthermore, let $G_\xi$ denote the *standard generalized extreme value (GEV) distribution (function)*, defined as follows:

$$
G_\xi(z) = \exp \left\{ -(1 + \xi z)^{-1/\xi} \right\}, \quad z \in \mathbb{R},
$$

where $(1 + \xi z)^{-1/\xi}$ is interpreted as $e^{-z}$ if $\xi = 0$ (this convention comes from de Haan and Ferreira, 2006, Theorem 1.1.3, and it may be used in the following) and thus

$$
G_0(z) = \exp\{ -e^{-z} \}, \quad z \in \mathbb{R}.
$$

Note that the mean $\bar{m}_\xi := \int_{-\infty}^{\infty} zG_\xi(z)dz$ is given by

$$
\bar{m}_\xi = \begin{cases} 
- \Gamma(-\xi) - \frac{1}{\xi}, & \xi \neq 0, \xi < 1, \\
\gamma, & \xi = 0,
\end{cases}
$$

where $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ (i.e., the gamma function) and $\gamma := -\Gamma'(1)$ is Euler’s constant.

Proposition 2.2 presents the two conditions that are equivalent to Assumption 2.1.
Proposition 2.2 [de Haan and Ferreira 2006, Theorems 1.1.3 and 1.1.6] Assumption 2.1 is equivalent to each of the conditions (a) and (b):

(a) There exist some constant \( \xi \in \mathbb{R} \) and some positive function \( a(\cdot) \) such that

\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\xi - 1}{\xi}, \quad x > 0, \tag{2.5}
\]

where \( (x^\xi - 1)/\xi \) is interpreted as \( \log x \) if \( \xi = 0 \) (according to the convention introduced after (2.3)).

(b) For some \( \xi \in \mathbb{R} \), \( F \in \text{MDA}(G_\xi) \), that is, there exist two sequences of constants \( \{a_n > 0; n \in \mathbb{N}\} \) and \( \{b_n \in \mathbb{R}; n \in \mathbb{N}\} \) such that

\[
\lim_{n \to \infty} P \left( \frac{Z_n - b_n}{a_n} \leq z \right) = \exp \left\{ -(1 + \xi z)^{-1/\xi} \right\} = G_\xi(z), \quad z \in \mathbb{R}. \tag{2.6}
\]

Note that the constant \( \xi \) is the same in (2.5) and (2.6). In addition, (2.6) holds with \( a_n = a(n) \) and \( b_n = U(n) \) for \( n \in \mathbb{N} \).

Remark 2.3 The standard GEV distribution \( G_\xi \) is classified into three cases based on the value of \( \xi \): (i) \( \xi > 0 \), (ii) \( \xi = 0 \), and (iii) \( \xi < 0 \). The MDAs of these three cases cover almost all continuous distributions commonly used in statistics (see, e.g., Embrechts et al. 1997, Sections 3.3 and 3.4). Therefore, Assumption 2.1, which is equivalent to \( F \in \text{MDA}(G_\xi) \), is not overly restrictive.

2.2 Regularly varying functions and their scale-freeness in EVT

This subsection is dedicated to preliminary content related to regularly varying functions and their scale-freeness in EVT. First, we present the definition of regularly varying functions by (asymptotic) scale-freeness (which is a standard way). We then summarize the regularly varying (i.e., scale-free) properties of the functions \( a \) and \( U \), which are essential for deriving our power-law formulas for the expected improvement rate and expected relative gap.

The definition of regularly varying functions is as follows:

Definition 2.4 An eventually positive and measurable function \( f \) on \( \mathbb{R}_+ \) is said to be regularly varying (at infinity) with index \( \alpha \in (-\infty, \infty) \) if \( f \) is (asymptotically) scale-free, i.e.,

\[
\lim_{t \to \infty} \frac{f(ct)}{f(t)} = c^\alpha, \quad c > 0. \tag{2.7}
\]

The symbol \( \mathcal{R}_\alpha \) denotes the set of regularly varying functions with index \( \alpha \). In particular, if \( f \in \mathcal{R}_0 \), then \( f \) is said to be slowly varying.
Remark 2.5 Equation \((2.7)\), the definition of regularly varying functions, shows that the form of the function \(f(t)\) is asymptotically similar to the form of \(\tilde{f}(t) := f(ct)\). Here, \(\tilde{f}(t) = f(ct)\) is a function where the measure of the quantity represented by the independent variable \(t\) is scaled by a factor of \(c\) compared to the original function \(f(t)\). Nevertheless, \((2.7)\) states that \(\tilde{f}(t) = f(ct)\) asymptotically retains the same form as the original function \(f(t)\). In this sense, a regularly varying function may have to be called “asymptotically scale-free,” but for simplicity, we simply call it “scale-free.” Note that a typical example of an “exactly” scale-free function is the power function \(t^\alpha, (\alpha \in \mathbb{R})\).

Proposition 2.6 below presents the fundamental properties of regularly varying functions.

Proposition 2.6 The following statements hold:

(i) \(f \in R_\alpha\) if and only if there exists some \(L \in R_0\) such that \(f(t) = L(t)t^\alpha\) for \(t \in R_+\) (see, e.g., Bingham et al. 1989, Theorem 1.4.1).

(ii) If an eventually positive and measurable function \(g\) on \(R_+\) satisfies \(\lim_{t \to \infty} g(t)/f(t) = c\) for some \(c > 0\) and \(f \in R_\alpha\), then \(g \in R_\alpha\) (this statement follows directly from Definition 2.4).

Remark 2.7 A positive and measurable function \(f\) on \(R_+\) is said to be regularly varying at the origin with index \(\alpha \in (-\infty, \infty)\) if

\[
\lim_{t \downarrow 0} \frac{f(ct)}{f(t)} = c^\alpha, \quad c > 0.
\]

By definition, \(f(t)\) is regularly varying at the origin with index \(\alpha\) if and only if \(f(1/t) \in R_{-\alpha}\), i.e., \(f(1/t)\) is regularly varying at infinity with index \(-\alpha\). Therefore, by Proposition 2.6 if the former holds, then \(f(t) = L(1/t)t^{-\alpha}\) for some \(L \in R_0\) and \(f\) exhibits power-law behavior near the origin.

The asymptotic properties of \(a(t)\) and \(U(t)\) as \(t \to \infty\) play an important role in the analysis of EVT. These properties are summarized in the following proposition (see de Haan and Ferreira 2006, Lemma 1.2.9 and Theorem B.2.1).

Proposition 2.8 Suppose that Assumption 2.1 is satisfied, or equivalently, \((2.5)\) holds for some constant \(\xi \in \mathbb{R}\) and some positive function \(a(\cdot)\). Under this condition, the following statements are true:

(i) The positive function \(a(\cdot)\) in \((2.5)\) belongs to the class \(R_\xi\).

(ii) If \(\xi > 0\), then

\[
x^* = \lim_{t \to \infty} U(t) = \infty, \quad \lim_{t \to \infty} \frac{U(t)}{a(t)} = \frac{1}{\xi}.
\]
(iii) If \( \xi < 0 \), then
\[
x^* = \lim_{t \to \infty} U(t) < \infty, \quad \lim_{t \to \infty} \frac{x^* - U(t)}{a(t)} = -\frac{1}{\xi}.
\]

(iv) If \( \xi = 0 \), then \( U \in \mathcal{R}_0 \) and
\[
\lim_{t \to \infty} \frac{a(t)}{U(t)} = 0.
\]

Furthermore, if \( x^* = \lim_{t \to \infty} U(t) < \infty \), then \( U(t) = x^* - L(t) \) for some \( L \in \mathcal{R}_0 \) such that \( \lim_{t \to \infty} L(t) = 0 \), and
\[
\lim_{t \to \infty} \frac{a(t)}{x^* - U(t)} = 0.
\]

3 Power-law Formulas behind the Curse of Scale-Freeness

This section discusses the power-law behavior of empirical solutions generated by an RMS algorithm. First, we present the main theorem of this paper, Theorem 3.1, which establishes the fundamental formulas underlying the power-law properties of the expected improvement rate, \( \mathbb{E}[R_n(x)] \), and the expected relative gap, \( \mathbb{E}[\Delta_n(x)] \). We then derive simplified versions of these formulas to highlight the power-law characteristics in \( \mathbb{E}[R_n(x)] \) and \( \mathbb{E}[\Delta_n(x)] \). Finally, we validate these theoretical results through numerical experiments (see Section B for the setup of the numerical experiments).

The following is the main theorem of this paper, presenting the power-law formulas for \( \mathbb{E}[R_n(x)] \) and \( \mathbb{E}[\Delta_n(x)] \).

**Theorem 3.1** Suppose that \( \mathbb{E}[X] = \int_0^\infty x dF(x) < \infty \) and Assumption 2.1 holds for \( G = G_\xi \) with \( \xi < 1 \). Then, the following statements are then true:

(i) If \( 0 < \xi < 1 \),
\[
\lim_{n \to \infty} \frac{1}{a(n)} \mathbb{E}[R_n(x)] = -\frac{\Gamma(-\xi)}{x} > 0, \quad x > 0,
\]
where \( a \in \mathcal{R}_\xi \) (due to Proposition 2.8 (i)).

(ii) If \( \xi = 0 \) and \( x^* = \infty \), then
\[
\lim_{n \to \infty} \frac{1}{U(n)} \mathbb{E}[R_n(x)] = \frac{1}{x}, \quad x > 0,
\]
where \( U \in \mathcal{R}_0 \) (due to Proposition 2.8 (iv)).
(iii) If $\xi = 0$ and $x^* < \infty$, then

$$
\lim_{n \to \infty} \frac{1}{x^* - U(n)} \left( \frac{x^* - x}{x} - E[R_n(x)] \right) = \frac{1}{x}, \quad 0 < x < x^*,
$$

$$
\lim_{n \to \infty} \frac{1}{x^* - U(n)} E[\Delta_n(x)] = \frac{1}{x^*}, \quad 0 < x < x^*,
$$

where $U(t) = x^* - L(t)$ for some $L \in \mathcal{R}_0$ such that $\lim_{t \to \infty} L(t) = 0$ (due to Proposition 2.8 (iv)).

(iv) If $\xi < 0$, then

$$
\lim_{n \to \infty} \frac{1}{a(n)} \left( \frac{x^* - x}{x} - E[R_n(x)] \right) = \Gamma(-\xi) \frac{x}{x^*}, \quad 0 < x < x^*,
$$

$$
\lim_{n \to \infty} \frac{1}{a(n)} E[\Delta_n(x)] = \Gamma(-\xi) \frac{x^*}{x^*}, \quad 0 < x < x^*,
$$

where $a \in \mathcal{R}_\xi$ (due to Proposition 2.8 (i)).

See Appendix A.

Remark 3.2 Equations (3.3) and (3.4) reflect the affine equation (1.5).

Remark 3.3 We confirm that the right-hand sides of (3.3), (3.4a), and (3.4b) are positive because $\Gamma(-\xi) > 0$ for $\xi < 0$ and $-\Gamma(-\xi) > 0$ for $0 < \xi < 1$. The former follows from the definition of the gamma function $\Gamma$, and the latter follows from Euler's reflection formula, which states that $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ for $z \notin \{0, \pm 1, \pm 2, \ldots\}$.

Theorem 3.1 in conjunction with Proposition 2.6 (i), yields a simplified version of the power-law formulas for $E[R_n(x)]$ and $E[\Delta_n(x)]$. This simplification clarifies the power-law characteristics of the two expectations as functions of the number of trials.

Corollary 3.4 Under the conditions of Theorem 3.1, the following statements are true:

(i) If $0 \leq \xi < 1$ and $x^* = \infty$, then

$$
E[R_n(x)] = L(n)n^\xi, \quad x > 0,
$$

for some $L \in \mathcal{R}_0$.

(ii) If $\xi \leq 0$ and $x^* < \infty$, then

$$
E[R_n(x)] = \frac{x^* - x}{x} - \frac{x^*}{x} L(n)n^\xi, \quad 0 < x < x^*,
$$

$$
E[\Delta_n(x)] = L(n)n^\xi, \quad 0 < x < x^*,
$$

for some $L \in \mathcal{R}_0$. 

The simplified formulas in Corollary 3.4 lead to Proposition 3.5 below, which supports the perspective introduced in Section 1: the evolution of the relative gap of the best EOV is subject to the curse of scale-freeness (see (1.7)) and is approximately linear on a log-log scale as a function of the number of trials $n$ (see Figure 2b).

**Proposition 3.5** Suppose that the conditions of Theorem 3.1 hold. Furthermore, suppose that $\xi \leq 0$ and $x^* < \infty$. We then have

$$
\lim_{n \to \infty} \frac{E[\Delta cn(x)]}{E[\Delta n(x)]} = c^\xi, \quad 0 < x < x^*, \ c \in \mathbb{N}, \quad (3.6)
$$

$$
\lim_{n \to \infty} \frac{\log E[\Delta n(x)]}{\log n} = \xi, \quad 0 < x < x^*. \quad (3.7)
$$

First, we prove (3.6). It follows from (3.5), Definition 2.4, and Proposition 2.6 that, for all $c \in \mathbb{N}$,

$$
\lim_{n \to \infty} \frac{E[\Delta cn(x)]}{E[\Delta n(x)]} = \lim_{n \to \infty} \frac{L(cn)(cn)^\xi}{L(n)n^\xi} = c^\xi \lim_{n \to \infty} \frac{L(cn)}{L(n)} = c^\xi, \quad 0 < x < x^*,
$$

which shows that (3.6) holds.

Next, we prove (3.7). It follows from (3.5) that

$$
\log E[\Delta n(x)] = \log L(n) + \xi \log n. \quad (3.8)
$$

Since $L \in \mathcal{R}_0$, $\log L(t)$ is expressed in the following form, known as the Karamata representation (see de Haan and Ferreira 2006, Theorem B.1.6):

$$
\log L(t) = \log \eta(t) + \int_{t_0}^t \frac{\theta(s)}{s} ds, \quad t > t_0, \quad (3.9)
$$

for some $t_0 > 0$, where $\eta : \mathbb{R}_+ \to \mathbb{R}$ and $\theta : \mathbb{R}_+ \to \mathbb{R}$ are measurable functions such that $\lim_{t \to \infty} \eta(t) =: \eta(\infty) \in (0, \infty)$ and $\lim_{t \to \infty} \theta(t) = 0$. Applying these limits to (3.9) yields

$$
\lim_{t \to \infty} \frac{\log L(t)}{\log t} = \lim_{t \to \infty} \frac{\eta(t)}{\log t} + \lim_{t \to \infty} \frac{1}{\log t} \int_{t_0}^t \frac{\theta(s)}{s} ds
$$

$$
= \lim_{t \to \infty} \frac{1}{t} \frac{\theta(t)}{t} = \lim_{t \to \infty} \theta(t) = 0,
$$

where the second equality holds by L’Hôpital’s rule. Combining the above equation and (3.8) leads to (3.7).

Proposition 3.5 implies the curse of scale-freeness, which characterizes the intractability of large-scale optimization with multi-start methods. As mentioned in Section 1, Equation (3.6) leads to (1.7), which indicates that the number of additional trials required to halve the expected relative gap is approximately $(2^{-1/\xi} - 1)$ times the number of completed trials. This approximation is likely to become more accurate as the number of completed trials increases. Such a situation deserves to be figuratively described as "Reaching for the goal makes it slip away."
Finally, we perform numerical experiments to confirm the occurrence of the curse of scale-freeness in large-scale optimization with multi-start methods. To achieve this, we solve 100 random instances of TSP with 1000 cities using six different RMS algorithms (see Table 2). For each instance, each algorithm generates $10^6$ locally optimal solutions (some of which are duplicates). Figure 3 shows the log-log plots of the evolution of the relative gap of the best EOV as the six RMS algorithms are applied to 100 random TSP instances (see Appendix B.3 for the details of running the algorithm to generate this figure). The results show that the relative gap of the best EOV decreases with the number of iterations at a power-law rate, supporting the presence of the curse of scale-freeness.

Figure 3: Evolution of the relative gap of the best EOV over $10^6$ iterations for each of the six RMS algorithms listed in Table 2, applied to 100 random TSP instances. The blue, red, and green lines represent the 90th percentile, mean, and 10th percentile behaviors, respectively, based on the results from the 100 instances.
4 Understanding The Curse of Scale-Freeness

This section is divided into two subsections and delves into the intractability of large-scale optimization through the lens of the curse of scale-freeness. Section 4.1 examines how the curse of scale-freeness is linked to the ratio of good solutions. Section 4.2 addresses the intractability of overcoming this curse.

4.1 The curse of scale-freeness in connection to the ratio of good solutions

This subsection discusses the connection between the curse of scale-freeness and the ratio of good (empirical) solutions. First, we introduce the good-solution-ratio function $r(\varepsilon)$ for $\varepsilon \in (0, 1)$. We then prove that the curse of scale-freeness is caused by the scale-freeness (i.e., power-law-type decay) of the function $r(\varepsilon)$ as $\varepsilon$ approaches zero. This theoretical result is supported by numerical experiments. Furthermore, we show that the situation worsens if the left tail of $r(\varepsilon)$ decays faster than the power-law rate.

We introduce a function related to the ratio of good (empirical) solutions generated by an RMS algorithm. Assuming that the supremum $x^*$ of EOVs is finite, we define $r(\varepsilon)$ for $0 < \varepsilon < 1$ as follows:

$$r(\varepsilon) = P \left( \frac{x^* - X}{x^*} < \varepsilon \right) = P(X > x^* - \varepsilon x^*) = 1 - F(x^* - \varepsilon x^*).$$

(4.1)

In general, $\varepsilon \in (0, 1)$ is assumed to be small. Since the EOVs $X_i$ are i.i.d., $r(\varepsilon)$ can be interpreted as the ratio of good solutions whose objective values are within a relative gap (error) $\varepsilon$ of the supremum $x^*$ of the EOVs. Thus, $r(\varepsilon)$ is referred to as the good-solution-ratio function.

In what follows, we consider the relationship between the decay speed of the expected relative gap $E[\Delta_n(x)]$ and that of $r(\varepsilon)$. Note that the decay of $r(\varepsilon)$ refers to its behavior as $\varepsilon$ decreases toward zero. Therefore, the part of $r(\varepsilon)$ with $\varepsilon$ close to 0 is referred to as the left tail, or simply the tail, of $r(\varepsilon)$.

We analyze the relationship between the curse of scale-freeness and the (left) tail decay of the good-solution-ratio function $r(\varepsilon)$, under the assumption that $r(\varepsilon)$ exhibits regular variation at the origin $\varepsilon = 0$ and hence power-law behavior near the origin (see Remark 2.7). The results of this analysis are presented in the following proposition.

**Proposition 4.1** Suppose that $x^* < \infty$ and $E[X] = \int_0^\infty xdF(x) < \infty$. Furthermore, if there exist some $\psi > 0$ and $L \in \mathcal{R}_0$ such that $r(\varepsilon) = L(1/\varepsilon)\varepsilon^\psi$, or equivalently,

$$\lim_{\varepsilon \downarrow 0} \frac{r(c\varepsilon)}{r(\varepsilon)} = c^\psi, \quad c > 0,$$

(4.2)

then $F \in \text{MDA}(G_\xi)$ with $\xi = -1/\psi < 0$, and thus both (3.6) and (3.7) hold.
Using (4.1), we rewrite (4.2) as follows:

\[ c \psi = \lim_{\epsilon \to 0} \frac{P(X > x^* - c\epsilon x^*)}{P(X > x^* - \epsilon x^*)} = \lim_{t \to 0} \frac{P(X > x^* - ct)}{P(X > x^* - t)}, \quad c > 0, \]

which implies that \( F \in \text{MDA}(G_\xi) \) with \( \xi = -1/\psi < 0 \) (de Haan and Ferreira 2006, Theorem 1.2.1). Thus, Assumption 2.1 holds for \( G = G_\xi \) with \( \xi = -1/\psi < 0 \) (see Proposition 2.2), and \( x^* < \infty \) by Proposition 2.8(iii). Therefore, the conditions of Proposition 4.1 imply those of Proposition 3.5, and hence, (3.6) and (3.7) hold with \( \xi = -1/\psi < 0 \).

Proposition 4.1 provides a valuable insight into the curse of scale-freeness, expressed mathematically as the scale-freeness of \( \mathbb{E}[\Delta_n(x)] \) in (3.7). This intractable phenomenon occurs when the good-solution-ratio function \( r(\epsilon) \) exhibits scale-freeness, or equivalently, when the distribution \( F \) of EOVs has a power-law tail. The power-law tail of \( F \) would generally be interpreted as indicating that its tail decays slowly. Thus, more intuitively, we can say that the curse of scale-freeness manifests itself in situations where “even if you find a good solution, there may still be many better ones”.

To support the insight derived from Proposition 4.1, we present Figure 4, which is based on the same data used to generate Figure 3. Figure 4 plots the good-solution-ratio function \( r(\epsilon) \) on a log-log scale. The figure shows that \( r(\epsilon) \) eventually becomes linear, suggesting that it follows a power law; in other words, \( r(\epsilon) \) exhibits regular variation. These results from our numerical experiments confirm that the insight derived from Proposition 4.1 is well-founded.

Let us consider the situation where the tail decay of the good-solution-ratio function is faster, specifically of the exponential type, such as \( \exp\{-\epsilon^{-\alpha}\} \) \( (\alpha > 0) \). For this case, the results of the analysis are presented in the following proposition, which complements Proposition 4.1.

**Proposition 4.2** Suppose that \( x^* < \infty \) and the good-solution-ratio function \( r(\epsilon) \) has the following representation:

\[ r(\epsilon) = c_0 \exp\{-g(1/\epsilon)\}, \quad 0 < \epsilon < 1, \quad \tag{4.3} \]

where \( c_0 > 0 \) is a constant and \( g: \mathbb{R}_+ \to (0, \infty) \) is a twice differentiable function such that \( g \in \mathcal{R}_\alpha \) for some \( \alpha > 0 \) and

\[ \lim_{t \to \infty} tg'(t) = \infty, \quad \tag{4.4} \]

\[ \lim_{t \to \infty} \frac{g''(t)}{|g'(t)|^2} = 0. \quad \tag{4.5} \]

We then have \( F \in \text{MDA}(G_0) \) and

\[ \mathbb{E}[\Delta_n(x)] = \frac{x^* - U(n)}{x^*} + o(x^* - U(n)), \quad \text{as } n \to \infty, \quad \tag{4.6} \]

for \( x \in (0, x^*) \), where \( U(t) = x^* - L(t) \) for some \( L \in \mathcal{R}_0 \) such that \( \lim_{t \to \infty} L(t) = 0 \).
Substituting \((4.3)\) into \((4.1)\), and rearranging the terms, we obtain
\[
1 - F(x) = c_0 \exp \left\{ -g \left( \frac{x^*}{x^* - x} \right) \right\}, \quad 0 \leq x < x^*. \tag{4.7}
\]

We assume the existence of a function \(f : \mathbb{R}_+ \to (0, \infty)\) such that
\[
1 - F(x) = c_0 \exp \left\{ - \int_0^x \frac{ds}{f(s)} \right\}, \quad 0 \leq x < x^*. \tag{4.8}
\]

It follows from \((4.7)\) and \((4.8)\) that
\[
\int_0^x \frac{ds}{f(s)} = g \left( \frac{x^*}{x^* - x} \right), \quad 0 \leq x < x^*.
\]

Differentiating both sides of the above equation yields
\[
\frac{1}{f(x)} = \frac{x^*}{(x^* - x)^2} g' \left( \frac{x^*}{x^* - x} \right), \quad 0 < x < x^*.
\]

Thus, we have
\[
f(x) = \frac{(x^* - x)^2}{x^*} \frac{1}{g' \left( \frac{x^*}{x^* - x} \right)}, \quad 0 < x < x^*, \tag{4.9}
\]
\[
f'(x) = - \frac{2(x^* - x)}{x^*} \frac{1}{g' \left( \frac{x^*}{x^* - x} \right)} - \frac{g'' \left( \frac{x^*}{x^* - x} \right)}{\left[ g' \left( \frac{x^*}{x^* - x} \right) \right]^2}, \quad 0 < x < x^*. \tag{4.10}
\]

Letting \(t = x^*/(x^* - x)\), it follows from \((4.4)\), \((4.5)\), \((4.9)\), and \((4.10)\) that
\[
\lim_{x \uparrow x^*} f(x) = \lim_{t \to \infty} t^2 g'(t) = 0,
\]
\[
\lim_{x \uparrow x^*} f'(x) = - \lim_{t \to \infty} \frac{2}{t g'(t)} - \lim_{t \to \infty} \frac{g''(t)}{[g'(t)]^2} = 0.
\]

The combination of these equations and \((4.8)\) implies that the distribution function \(F\) is a von Mises function in MDA(\(G_0\)) (see de Haan and Ferreira 2006, Theorem 1.2.6). Therefore, \((2.6)\) holds for \(\xi = 0\) due to Proposition 2.2. It thus follows from Theorem 3.1 (iii) that
\[
\lim_{n \to \infty} \frac{1}{x^* - U(n)} E[\Delta_n(x)] = \frac{1}{x^*}, \quad 0 < x < x^*,
\]

where \(U(t) = x^* - L(t)\) for some \(L \in R_0\) such that \(\lim_{t \to \infty} L(t) = 0\). As a result the statement of this proposition hold.

Proposition 4.2, together with Proposition 4.1, suggests that the faster the decay of the good-solution-ratio function, the more difficult it becomes to close the relative gap of the best
EOV. Specifically, (4.6) in Proposition 4.2 indicates that $E[\Delta_n(x)]$, when viewed as a function of $n$, is slowly varying, making its decay slower than any power function. A typical example of this case is as follows:

$$\lim_{n \to \infty} \frac{E[\Delta_n(x)]}{(\log n)^{-\beta}} = c_0 \quad \text{for some } \beta \leq 0 \text{ and } c_0 > 0.$$ 

Such behavior of $E[\Delta_n(x)]$ is even more problematic than the curse of scale-freeness. This problem likely arises from the challenge of finding better solutions among a sparse set of good solutions scattered across a large feasible domain.

To further illustrate the above discussion, we consider a specific example from the cases covered by Proposition 4.2 and derive a explicit limit formula for the expected relative gap. Assume that

$$r(\varepsilon) = c_0 \exp\{-\phi \varepsilon^{-\alpha}\}, \quad 0 < \varepsilon < 1,$$

where $c_0 > 0$, $\phi > 0$, and $\alpha > 0$. From (4.11) and (4.1), we have

$$1 - F(x) = c_0 \exp\left\{-\phi \left(\frac{x^*}{x^* - x}\right)^\alpha\right\}, \quad 0 \leq x < x^*.$$ 

From this equation and (2.1), we obtain

$$U(t) = x^* - x^*[\phi^{-1} \log(c_0 t)]^{-1/\alpha}.$$ 

(4.12)

Applying (4.12) to (4.6) yields

$$E[\Delta_n(x)] = [\phi^{-1} \log(c_0 n)]^{-1/\alpha} + o((\log n)^{-1/\alpha}), \quad \text{as } n \to \infty,$$

(4.13)

for $x \in (0, x^*)$. Equation (4.13) shows that the decay of the expected relative gap $E[\Delta_n(x)]$ is dominated by $(\log n)^{-1/\alpha}$, resulting in an extremely slow rate. As mentioned above, in this scenario, the good-solution-ratio function $r(\varepsilon)$ decays even slower than any power-law function.

Building on the above analysis, we conclude that the curse of scale-freeness—encapsulated by the metaphor “Reaching for the goal makes it slip away”—inevitably arises in large-scale optimization when using RMS methods, regardless of the decay rate of the good-solution-ratio function $r(\varepsilon)$. Notably, if $r(\varepsilon)$ decays faster than any power-law rate, the resulting intractability is even greater than that posed by the curse of scale-freeness.

4.2 Intractability of overcoming the curse of scale-freeness

This subsection explores the curse of scale-freeness, the intractability of large-scale optimization with multi-start methods. We first show that overcoming this curse requires developing a powerful LS algorithm equipped with a diversification mechanism that is exponentially more effective than RMS. We then present numerical experiments, demonstrating that even ILS [Lourenço et al., 2019], a widely recognized and powerful metaheuristic, is unable to overcome
the curse of scale-freeness. Finally, we summarize our theoretical and numerical results, highlighting what on earth the curse of scale-freeness is, which arises in large-scale optimization with multi-start methods.

The power-law formula (3.5) implies the difficulty of developing a metaheuristic algorithm to overcome the curse of scale-freeness in large-scale optimization. To illustrate this, we consider two types of accelerations of RMS: one is polynomial acceleration, and the other is exponential acceleration. The expected relative gaps of polynomially accelerated and exponentially accelerated RMSs are given by replacing $n$ in (3.5) with $n^K$ and $K^n$ ($K = 2, 3, \ldots$), respectively:

(Polynomial acceleration) \[ E[\Delta_{n^K}(x)] = L(n^K)n^{K\xi}, \quad \xi < 0, \quad (4.14a) \]

(Exponential acceleration) \[ E[\Delta_{K^n}(x)] = L(K^n)K^{n\xi}, \quad \xi < 0, \quad (4.14b) \]

where $L$ is a slowly varying function that may be different from the $L$ in (3.5). Equation (4.14) shows that scale-freeness persists in the polynomial acceleration case but not in the exponential acceleration case. In other words, overcoming the curse of scale-freeness is impossible without developing a powerful LS algorithm with a diversification mechanism that is exponentially more effective than RMS. However, achieving such exponential acceleration while maintaining a manageable cost per trial is challenging without a sufficiently tractable problem structure.

We perform numerical experiments with ILS (Lourenço et al., 2019), a widely recognized and powerful metaheuristic, to understand the intractability of overcoming the curse of scale-freeness. Figures 5–7 plot the relative gap of the best EOV on a log-log scale when three ILS and three RMS algorithms are applied to five TSP instances with over 10,000 cities from TSPLIB (see Appendix B.3 for details on running the algorithms to generate these figures). Neither the RMS nor the ILS algorithms reach the optimal solution in these results. The figures show that the relative gaps realized by the RMS algorithms remain above 1% even after $10^6$ iterations, while those realized by the ILS algorithms drop below 0.5% in all trials and even below 0.1% in some cases. Thus, ILS is superior to RMS. Furthermore, the figures show that even with the ILS algorithms, the relative gap decreases in an approximately sigmoidal fashion and eventually decays (at best) at a power-law rate, indicating that ILS cannot overcome the curse of scale-freeness in our experiments. Note that this result does not necessarily imply poor performance of ILS but rather highlights the intractability of overcoming the curse of scale-freeness.

We consider why ILS only reduces the relative gap of the best EOV in an almost sigmoidal manner and why it cannot overcome the curse of scale-freeness in large-scale optimization. From Figures 5–7, we observe a characteristic of ILS: in the early stages of the search, the relative gap of the best EOVs does not decrease significantly, but then decays rapidly (though at best following a power-law) in the middle stage before its decay rate gradually slows down. This behavior can be attributed to the fact that ILS is more depth-first exploratory than the RMS method. Specifically: (i) the depth-first nature of ILS delays finding deep and large ”valleys” where good solutions accumulate; (ii) once such valleys are reached, the best empirical solution is continuously updated for a while; (iii) however, as unreached good
solutions in these valleys eventually decrease, the best empirical solution is updated less frequently. For these reasons, overcoming the curse of scale-freeness in large-scale optimization may remain difficult, even with ILS.

Finally, we summarize the insights gained from our theoretical and numerical results. Suppose that we develop the most powerful LS algorithm possible for a large-scale optimization problem, but it generates a virtually infinite number of locally optimal solutions when starting from different initial solutions. In this scenario, the curse of scale-freeness cannot be overcome without integrating a diversification mechanism that is exponentially more effective than RMS (as implied by (4.14b)). However, such an exponential improvement in performance would not be achievable unless the target problem has highly tractable properties, leading to an exponential reduction of the search space. Although ILS is widely recognized as a powerful metaheuristic, it does not seem to achieve such an exponential reduction in search space by its very nature. In fact, as shown in Figures 5–7, our ILS outperforms our RMS, but its diversification mechanism is not effective enough to overcome the curse of scale-freeness, at least within the scope of our numerical experiments.

5 Concluding Remarks

We have derived power-law formulas for the expected improvement rate and the expected relative gap of the best empirical objective value (EOV) generated by the RMS method. The formulas show that both expectations vary at most at power-law rates as the number of iterations increases.

Based on these power-law formulas, we introduced the new perspective of the curse of scale-freeness, interpreted as a Zeno’s paradox-like phenomenon encapsulated by the metaphor “Reaching for the goal makes it slip away.” This perspective highlights the inherent intractability of large-scale optimization with multi-start methods.

The curse of scale-freeness brings us back to the basics of developing metaheuristics for large-scale optimization: adopting a practical strategy to generate the best possible solution within a reasonable computational time. In other words, the focus should be on “when to stop computing” rather than “how to close the gap.” Moreover, this decision should be guided by a cost-benefit tradeoff.

However, this practical strategy cannot be fully realized by simply measuring the gap from the optimal value. As our theoretical and numerical results show, even when the gap between the optimal value and the best EOV is small, the computational cost of closing this gap can often exceed expectations. Therefore, the decision to stop an algorithm should be based on the expected improvement rate from additional trials. From this perspective, a promising future task would be to develop a method for estimating the expected improvement rate.

A Proof of Theorem 3.1

The proof of Theorem 3.1 needs the following proposition.
Figure 4: Good-solution-ratio function $r(\varepsilon)$ calculated from the data used to generate Figure 3.
Figure 5: Evolution of the relative gap of the best EOV when performing 100 random sets of $10^6$ iterations of the RA+LK RMS and ILS algorithms, respectively, for the five TSPLIB instances listed in Table I.
Figure 6: Evolution of the relative gap of the best EOV when performing 100 random sets of $10^6$ iterations of the NN+LK RMS and ILS algorithms, respectively, for the five TSPLIB instances listed in Table 1.
Figure 7: Evolution of the relative gap of the best EOV when performing 100 random sets of $10^6$ iterations of the GR+LK RMS and ILS algorithms, respectively, for the five TSPLIB instances listed in Table I.
Proposition A.1 Suppose that Assumption 2.1 holds, and that $E[X] = \int_0^\infty x dF(x) < \infty$ and $\xi < 1$. For $x \in (0, x^*)$, we then have

$$\lim_{n \to \infty} E \left[ \frac{1}{a(n)} \{(Z_n - x)_+ - (U(n) - x)\} \right] = \overline{m}_\xi,$$

(A.1a)
or equivalently,

$$E[(Z_n - x)_+] = (\overline{m}_\xi + o(1)) a(n) + U(n) - x, \quad \text{as } n \to \infty,$$

(A.1b)

where $\overline{m}_\xi$ is given in (2.4) and $a(\cdot)$ is a positive function satisfying (2.5).

We prove only (A.1a) because it is equivalent to (A.1b). Dividing the expectation on the left-hand side of equation (A.1a) based on the events $\{Z_n > x\}$ and $\{Z_n \leq x\}$, we obtain

$$E \left[ \frac{1}{a(n)} \{(Z_n - x)_+ - (U(n) - x)\} \right]
= E \left\{ \mathbb{1}(Z_n > x) \frac{Z_n - U(n)}{a(n)} \right\} + E \left\{ \mathbb{1}(Z_n \leq x) \frac{x - U(n)}{a(n)} \right\}
= E \left[ \frac{Z_n - U(n)}{a(n)} \right] - E \left\{ \mathbb{1}(Z_n \leq x) \frac{Z_n - U(n)}{a(n)} \right\} + E \left\{ \mathbb{1}(Z_n \leq x) \frac{x - U(n)}{a(n)} \right\}
= E \left[ \frac{Z_n - U(n)}{a(n)} \right] + E \left\{ \mathbb{1}(Z_n \leq x) \frac{x - Z_n}{a(n)} \right\}, \quad 0 < x < x^*.

(A.2)

Letting $n \to \infty$ in the second term on the left-hand side of (A.2), we obtain the following: For $x \in (0, x^*)$,

$$0 \leq \lim_{n \to \infty} E \left\{ \mathbb{1}(Z_n \leq x) \frac{x - Z_n}{a(n)} \right\} \leq \lim_{n \to \infty} \frac{x E[\mathbb{1}(Z_n \leq x)]}{a(n)} = \lim_{n \to \infty} \frac{x P(Z_n \leq x)}{a(n)} = \lim_{n \to \infty} \frac{x [F(x)]^n}{a(n)} = 0,$$

(A.3)

where the last equality holds because function $a(\cdot)$ is of at most polynomial growth for any $\xi < 1$ (see Propositions 2.6 and 2.8). Applying (A.3) and (de Haan and Ferreira 2006 Theorem 5.3.1) to (A.2) yields

$$\lim_{n \to \infty} E \left[ \frac{1}{a(n)} \{(Z_n - x)_+ - (U(n) - x)\} \right]
= \lim_{n \to \infty} E \left[ \frac{Z_n - U(n)}{a(n)} \right] = \int_{-\infty}^\infty z dG_\xi(z) = \overline{m}_\xi, \quad 0 < x < x^*,$$

which shows that (A.1a) holds.

Using Proposition A.1, we prove the statements (i)–(iv) of Theorem 3.1. To achieve this, we need (A.4), which is derived from (1.3), (2.4), and (A.1b):

$$E[R_n(x)] = \begin{cases} \frac{1}{x} \left[ -\Gamma(-\xi) - \frac{1}{\xi} + o(1) \right] a(n) + U(n) - x, & \xi \neq 0, \xi < 1, \\
\frac{1}{x} \left[ (\gamma + o(1)) a(n) + U(n) - x \right], & \xi = 0,
\end{cases}
\quad \text{(A.4a)}$$

\text{(A.4b)}
as $n \to \infty$ for any $x \in (0,x^*)$.

**Proof of the statement (i).** Suppose that $0 < \xi < 1$. It then follows from Proposition 2.8 (ii) that

$$\lim_{n \to \infty} \frac{U(n) - x}{a(n)} = \frac{1}{\xi}, \quad x > 0. \quad \text{(A.5)}$$

Using (A.5) and (A.4a), we obtain

$$\lim_{n \to \infty} \frac{1}{a(n)} E[R_n(x)] = \frac{1}{x} \lim_{n \to \infty} \left[ -\Gamma(-\xi) - \frac{1}{\xi} + o(1) + \frac{U(n) - x}{a(n)} \right]$$

$$= \frac{1}{x} \left[ -\Gamma(-\xi) - \frac{1}{\xi} + \frac{1}{\xi} \right] = \frac{\Gamma(-\xi)}{x}, \quad x > 0,$$

which shows that (3.1) holds.

**Proof of the statement (ii).** Suppose that $\xi = 0$ and $x^* = \infty$. It then follows from (2.2) and (2.8) (see Proposition 2.8 (iv)) that

$$\lim_{n \to \infty} U(n) = x^* = \infty, \quad \lim_{n \to \infty} \frac{a(n)}{U(n)} = 0. \quad \text{(A.6)}$$

Using (A.6) and (A.4b), we obtain

$$\lim_{n \to \infty} \frac{1}{U(n)} E[R_n(x)] = \frac{1}{x} \lim_{n \to \infty} \left[ (\xi + o(1)) \frac{a(n)}{U(n)} + 1 - \frac{x}{U(n)} \right] = \frac{1}{x}, \quad x > 0,$$

which shows that (3.2) holds.

**Proof of the statement (iii).** Suppose that $\xi = 0$ and $x^* < \infty$. It then follows from (2.2) and (2.9) (see Proposition 2.8 (iv)) that

$$\lim_{n \to \infty} \frac{a(n)}{x^* - U(n)} = 0. \quad \text{(A.7)}$$

Using (A.7) and (A.4b) that

$$\lim_{n \to \infty} \frac{1}{x^* - U(n)} \left[ \frac{x^* - x}{x} - E[R_n(x)] \right]$$

$$= \frac{1}{x} \lim_{n \to \infty} \frac{1}{x^* - U(n)} \left[ x^* - x - (\gamma + o(1))a(n) - U(n) + x \right]$$

$$= \frac{1}{x} \lim_{n \to \infty} \frac{1}{x^* - U(n)} \left[ x^* - U(n) - \gamma a(n) \right]$$

$$= \frac{1}{x} \lim_{n \to \infty} \left[ 1 - \frac{\gamma a(n)}{x^* - U(n)} \right] = \frac{1}{x}, \quad 0 < x < x^*,$$

which shows that (3.3a) holds. Note here (1.5) yields

$$E[\Delta_n(x)] = \frac{x}{x^*} \left[ \frac{x^* - x}{x} - E[R_n(x)] \right], \quad 0 < x < x^*. \quad \text{(A.8)}$$
Combining (A.8) and (3.3a) leads to (3.3b).

**Proof of the statement (iv).** Suppose that $\xi < 0$. It then follows from Proposition 2.8 (iii) that

$$\lim_{n \to \infty} U(n) = x^* < \infty, \quad \lim_{n \to \infty} \frac{x^* - U(n)}{a(n)} = -\frac{1}{\xi},$$

(A.9)

Using (A.9) and (A.4a), we obtain

$$\lim_{n \to \infty} \frac{1}{a(n)} \left[ \frac{x^* - x}{x} - \mathbb{E}[R_n(x)] \right] = \lim_{x \to \infty} \frac{1}{a(n)} \left[ x^* - x - \left( -\Gamma(-\xi) - \frac{1}{\xi} + o(1) \right) a(n) - U(n) + x \right] = \lim_{x \to \infty} \frac{1}{a(n)} \left[ \frac{x^* - U(n)}{a(n)} + \Gamma(-\xi) + \frac{1}{\xi} \right] = \frac{1}{x} \left[ -\frac{1}{\xi} + \Gamma(-\xi) + \frac{1}{\xi} \right] = \frac{\Gamma(-\xi)}{x}, \quad 0 < x < x^*,$$

which shows that (3.4a) holds. Furthermore, Combining (A.8) and (3.4a) leads to (3.4b). The proof of Theorem 3.1 is completed.

### B Setup of Numerical Experiments

This section describes the setup of our numerical experiments aimed at observing whether the power-law phenomenon, referred to as the *curse of scale-freeness*, occurs when improving the best EOV. First, we describe the Traveling Salesman Problem (TSP) instances used to evaluate the practical implications of our mathematical results. Next, we present the RMS and ILS algorithms implemented in the Concorde TSP Solver [Applegate et al., 2003], a software package that includes an efficient branch-and-cut algorithm along with several heuristic algorithms for TSP. Finally, we outline the procedure for running our RMS and ILS algorithms. Note that these algorithms are run to obtain empirical solutions that verify the phenomena suggested by our mathematical results.

#### B.1 TSP Instances

TSP is one of the most well-known combinatorial optimization problems [Lawler et al., 1985; Applegate et al., 2007]. Its difficulty is easily adjustable, and extensive collections of benchmark problems with known optimal solutions are readily available. Therefore, we chose TSP to validate our mathematical results from a practical perspective. Note that while TSP is originally a minimization problem, EVT, our mathematical tool, is designed for maximum value data. To fit TSP into our EVT framework, we multiplied the objective function by $-1$.

To generate Figures 2, 3, and 4 we generated 100 random TSP instances as follows. Each instance consists of 1000 cities with two-dimensional coordinates $(x^{(1)}, x^{(2)})$ that are uniformly
distributed random integers chosen from $[0, 10^5 - 1]$. The travel cost $c_{i,j}$ between cities $i$ and $j$, with coordinates $(x_i^{(1)}, x_i^{(2)})$ and $(x_j^{(1)}, x_j^{(2)})$, respectively, is defined as

$$c_{i,j} = \left\lfloor \sqrt{(x_i^{(1)} - x_j^{(1)})^2 + (x_i^{(2)} - x_j^{(2)})^2 + 0.5} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the floor function. We computed the optimal values for all 100 instances using the branch-and-cut algorithm of the Concorde TSP Solver (Lawler et al., 1985; Applegate et al., 2007).

To generate Figures 5–7, we used the five TSP instances listed in Table 1 from TSPLIB (Reinelt, 1991), a well-known and widely used benchmark set of TSP instances. These instances contain more than 10,000 cities, and their optimal values are already known.

Table 1: TSPLIB instances

| Instance Name | Number of Cities |
|---------------|------------------|
| brd14051      | 14,051           |
| d15112        | 15,112           |
| d18512        | 18,512           |
| rl11849       | 11,849           |
| usa13509      | 13,509           |

B.2 Algorithms

This subsection describes the six versions of RMS and ILS, respectively, that were used to solve the TSP instances introduced in the previous subsection. For both RMS and ILS, these versions were generated by combining three initial solution generators with two local search algorithms (see Table 2). We implemented these versions by modifying the heuristic algorithms included in the Concorde TSP Solver. The three initial solution generators and the two local search algorithms are described in detail below (see Johnson and McGeoch 2003 for more information).

The three initial solution generators are (i) the RAndom (RA) algorithm, (ii) the randomized Nearest Neighbor (NN) algorithm, and (iii) the randomized GReedy (GR) algorithm. The RA algorithm is originally provided by the Concorde TSP Solver, while the NN and GR algorithms have been modified to incorporate a random mode. The characteristics of these three algorithms are as follows:

(i) The RA algorithm generates an initial solution by visiting each city exactly once in a random order.

(ii) The NN algorithm starts from a randomly selected city and moves to a randomly selected city of the three nearest unvisited ones.
Table 2: The options for initial solution generation and local search in RMS and ILS

|                      | Initial solution generation | Local search               |
|----------------------|----------------------------|-----------------------------|
| RA + 3-opt           | RAndom                     | 3-opt                       |
| NN + 3-opt           | randomized Nearest Neighbor| 3-opt                       |
| GR + 3-opt           | randomized GReedy           | 3-opt                       |
| RA + LK              | RAndom                     | Lin-Kernighan               |
| NN + LK              | randomized Nearest Neighbor| Lin-Kernighan               |
| GR + LK              | randomized GReedy           | Lin-Kernighan               |

(iii) The GR algorithm starts with the shortest edge and adds a randomly selected edge of the three shortest remaining ones to the current path, avoiding the creation of subtours and vertices of degree three.

B.3 How to Run RMS and ILS Algorithms

This subsection describes how the RMS and ILS algorithms were run to obtain the data shown in Figures 3–7, with a focus on the use of random numbers generated by the Mersenne Twister.

We replaced the default random number generator (Knuth 1997, Section 3.2.2, Algorithm A) in the Concorde TSP Solver with the Mersenne Twister (Nishimura and Matsumoto 2004), which is specifically designed for 64-bit machines and has a much longer period length of $2^{19937} - 1$ (compared to the default generator’s period length of $2^{55} - 1$). This change was necessary because our experiment verifying the curse of scale-freeness required running the RMS and ILS algorithms for an extremely long time.

We obtained the results shown in Figures 3 and 4 by running $10^6$ trials for each of the six RMS algorithms (see Table 2) on 100 random TSP instances (introduced in Section B.1). For each set of $10^6$ trials, a unique random seed was assigned to the Mersenne Twister. The Concorde TSP Solver then used the sequence of random numbers generated by the Mersenne Twister based on the specified seed.

We also obtained the results shown in Figures 5, 6, and 7 by applying a similar procedure as described above to each of the five TSPLIB instances listed in Table 1.

(i) 100 runs of $10^6$ trials for each of the three RMS algorithms (RA + LK, NN + LK, and GR + LK); and

(ii) 100 runs of “one initial start and $(10^6 - 1)$ random-walk kicks” for each of the three ILS algorithms (RA + LK, NN + LK, and GR + LK).

The introduction of randomness into these RMS and ILS algorithms followed the same methodology used to generate Figures 3 and 4.
References

Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J., 2003. Concorde TSP Solver. URL: https://www.math.uwaterloo.ca/tsp/concorde/index.html.

Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J., 2007. The Traveling Salesman Problem: A Computational Study. Princeton University Press, New Jersey.

Bingham, N.H., Goldie, C.M., Teugels, J.L., 1989. Regular Variation. Paperback ed., Cambridge University Press, New York.

Carling, K., Meng, X., 2015. Confidence in heuristic solutions? Journal of Global Optimization 63, 381–399.

Carling, K., Meng, X., 2016. On statistical bounds of heuristic solutions to location problems. Journal of Combinatorial Optimization 31, 1518–1549.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Fisher, M.L., Northup, W.D., Shapiro, J.F., 1975. Using duality to solve discrete optimization problems: Theory and computational experience. Mathematical Programming Study 3, 56–94.

Giddings, A.P., Rardin, R.L., Uzsoy, R., 2014. Statistical optimum estimation techniques for combinatorial optimization problems: A review and critique. Journal of Heuristics 20, 329–358.

Golden, B.L., 1977. A statistical approach to the TSP. Networks 7, 209–225.

Golden, B.L., 1978. Point estimation of a global optimum for large combinatorial problems. Communications in Statistics - Simulation and Computation 7, 361–367.

Golden, B.L., Alt, F.B., 1979. Interval estimation of a global optimum for large combinatorial problems. Naval Research Logistics Quarterly 26, 69–77.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer, New York.

Hansen, P., Mladenović, N., 2001. Variable neighborhood search: Principles and applications. European Journal of Operational Research 130, 449–467.

Held, M., Karp, R.M., 1970. The traveling-salesman problem and minimum spanning trees. Operations Research 18, 1138–1162.

Johnson, D.S., McGeoch, L.A., 2003. The traveling salesman problem: A case study, in: Aarts, E., Lenstra, J.K. (Eds.), Local Search in Combinatorial Optimization. Princeton University Press, New Jersey. chapter 8, pp. 215–310.
Knuth, D.E., 1997. The Art of Computer Programming, Volume 2: Seminumerical Algorithms, 3rd Edition. Addison-Wesley, Massachusetts.

Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G., Shmoys, D.B. (Eds.), 1985. The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization. Wiley, Chichester.

Lourenço, H.R., Martin, O.C., Stützle, T., 2019. Iterated local search: Framework and applications, in: Gendreau, M., Potvin, J.Y. (Eds.), Handbook of Metaheuristics (3rd edition). Springer, Cham. chapter 5, pp. 129–168.

Martí, R., Lozano, J.A., Mendiburu, A., Hernando, L., 2018. Multi-start methods, in: Martí, R., Pardalos, P.M., Resende, M.G.C. (Eds.), Handbook of Heuristics. Springer, Cham. chapter 6, pp. 155–175.

Martí, R., Resende, M.G.C., Ribeiro, C.C., 2013. Multi-start methods for combinatorial optimization. European Journal of Operational Research 226, 1–8.

Neri, F., Cotta, C., Moscato, P. (Eds.), 2012. Handbook of Memetic Algorithms. Springer, Berlin.

Nishimura, T., Matsumoto, M., 2004. Mersenne twister 64bit version. URL: http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt64.html

Reinelt, G., 1991. TSPLIB—A traveling salesman problem library. ORSA Journal on Computing 3, 376–384. URL: http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95.

Resende, M.G.C., Rebeiro, C.C., 2016. Optimization by GRASP: Greedy Randomized Adaptive Search Procedures. Springer, New York.

Velasco, L., Guerrero, H., Hospitaler, A., 2022. Can the global optimum of a combinatorial optimization problem be reliably estimated through extreme value theory? Swarm and Evolutionary Computation 75, 101172.