Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Functions: Validation via Landscapes of Triangles

Xavier F. C. Sánchez-Díaz
Norwegian University of Science and Technology
Trondheim, Norway
xavier.sanchezdz@ntnu.no

Ole Jakob Mengshoel
Norwegian University of Science and Technology
Trondheim, Norway
ole.j.mengshoel@ntnu.no

ABSTRACT

Pseudo-Boolean functions are often multimodal, and it is of interest to find multiple optima. However, the problem of estimating the number of local optima has not been much studied in the combinatorial setting. Since exhaustive enumeration is generally prohibitive, we study an alternative in this paper. Our method, which uses the celebrated Birthday Paradox "in reverse," enables us to estimate the number of local optima in fitness landscapes. We study the method analytically and experimentally, using a new synthetic problem, Triangle. This problem allows us to vary the number of optima and its distribution easily but understandably, which enables analytical validation of our experiments. We conclude by discussing how the approach may be applied and extended in the future.

CCS CONCEPTS

• **Computing methodologies** → **Randomized search**; **Discrete space search**; • **Theory of computation** → *Theory of randomized search heuristics.*

KEYWORDS

Pseudo-Boolean functions, local optima, multimodality

ACM Reference Format:
Xavier F. C. Sánchez-Díaz and Ole Jakob Mengshoel. 2024. Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Functions: Validation via Landscapes of Triangles. In *Genetic and Evolutionary Computation Conference (GECCO '24 Companion), July 14–18, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3638530.3654156

1 INTRODUCTION

Context. Many problems in AI and evolutionary computation (EC) can be formulated using pseudo-Boolean functions (PBFs), e.g., model finding [18], summarization [16], feature selection [8, 9, 19] among others [1, 12, 17]. Synthetic problems include OneMax TwoMax, Hurdle, and CosineTrap, and are motivated by fitness landscapes that include traps normally found in natural problems [5, 14, 19]. Among PBFs, we focus here on the multimodal ones, specially on highly multimodal fitness functions. Previous research has also studied the estimation of the number of local optima for traveling salesman and flow shop scheduling [4], as well as feature selection [10, 11]. To estimate the number of optima, one can use the Birthday Paradox [7]: What is the probability that at least two people (out of a group of n randomly selected) share the same birthday?

Challenges. Despite much progress on niching methods for evolutionary algorithms, the landscapes of many fitness functions, including multimodal fitness functions, are not well understood in general [13]. Li et al. discuss eight open research questions for niching methods [6], one of which is the problem of an unknown number of optima. This is an important challenge, especially when the number of optima is nontrivial. One needs to find adequate problem instances, in contrast to problems with just one or two optima (as in OneMax or TwoMax) while at the same time avoid the other end of the spectrum where problems become too complex (or impossible) to analyze [15].

Contributions. We seek to address the lack of methods for estimating the number of local optima in PBFs. Our main contributions are as follows. (i) We study a method for estimating the number of local optima in PBFs, based upon "reversing" the well-known Birthday Paradox from probability theory. (ii) We propose synthetic functions, Triangle, that can be parameterized to control the landscape and its features—number of optima and regions of attraction—while remaining relatively easy to analyze.

The use of the Birthday Paradox for estimating the number of local optima has been studied previously [2], but empirical studies of several sampling methods have disregarded it as an unreliable method [4]. In this work we highlight synthetic problems where the method works well, in addition to instances where it underperforms.

2 BACKGROUND

We use brackets to delimit a finite range of numbers: define the set of integers $[n] := \{1, 2, \ldots, n\}$. A closed range between two integer numbers is denoted by $[l_b..u_b]$, where l_b and u_b are the lower and upper bounds, respectively. The cardinality of a set B is represented as $|B|$, and the number of 1s in a bitstring $\boldsymbol{b}$ is denoted as $\|\boldsymbol{b}\|$. We define a fitness landscape $(\mathbb{B}^n, f, \mathcal{N})$ [20], where $\mathbb{B}^n$ is the search space, f is a fitness function, and N is a neighborhood function. For the analysis in this paper, we assume a neighborhood $\mathcal{N}(\boldsymbol{b}')$ that encompasses all bitstrings $\boldsymbol{b}_i$ with a Hamming distance of 1 from $\boldsymbol{b}'$, as this is the most likely mutation of a hill-climbing EA with a standard mutation probability of $1/n$.

Pseudo-Boolean functions. Let $\mathbb{B}^n$ be the set of bitstrings of length n. We study PBFs of the form $f : \mathbb{B}^n \to \mathbb{R}_{\geq 0}$. We maximize the fitness function f:

$$\boldsymbol{b}^* = \arg\max_{\boldsymbol{b} \in \mathbb{B}^n} f(\boldsymbol{b}). \quad (1)$$

GECCO '24 Companion, July 14–18, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0495-6/24/07.
https://doi.org/10.1145/3638530.3654156

We are interested in multiple optima $B^* = \{\boldsymbol{b}_1^*, \boldsymbol{b}_2^*, \dots\}$ such that $f(\boldsymbol{b}_1^*) \geq f(\boldsymbol{b}_2^*) \geq f(\boldsymbol{b}_3^*), \dots$, where each $\boldsymbol{b}_i^*$ is a local optimum. In particular, we are interested in estimating the number of local optima, i.e., $|B^*|$, which is often unknown [6].

Local optima in Pseudo-Boolean Functions. We identify two types of local optima for PBFs: Peaked Local Optima (PLOP) and Flat Local Optima (FLOP). We study PLOP functions, where an optimum $f(\boldsymbol{b}_i^*)$ with $\boldsymbol{b}_i^* \in B$ is strictly greater than its neighborhood. We refer to a PLOP with k local optima as k-PLOP function.

Symmetry. A k-PLOP function f is said to be *symmetric* if each of the k optima has an equal probability $p = 1/k$ of being found by an algorithm under uniformly at random initialization.

Region of Attraction. To formalize the concept of probability of convergence in PLOP functions, we introduce the concept of *region of attraction* (also called basin of attraction [13]). Let f be a PLOP function. The *region of attraction* (RoA) of a state $\boldsymbol{b}$ is the set of states from which the state $\boldsymbol{b}$ can be reached by a given algorithm A with a probability $p > \epsilon$.[1]

Overlap of RoAs. Starting from certain states, for a given f, an algorithm may reach multiple optima. We call this an *overlap of RoAs* (OoRoA). Intuitively, this occurs at all states $\boldsymbol{b}$ from which multiple optima can be reached, by a given algorithm, with probability $p > \epsilon$. At these overlapping states, we consider that a portion of the RoA is *shared*. Let $f(\|\boldsymbol{b}\|^*)$ be the fitness value (or *peak*, $\|\boldsymbol{b}\|^*$) of a k-PLOP function. The *share of RoA* (or SoRoA for short) of a peak $\|\boldsymbol{b}\|_i^*$, $\text{SoRoA}(\|\boldsymbol{b}\|_i^*)$, is the subset of $\text{RoA}(\|\boldsymbol{b}\|_i^*)$ from which we expect convergence to $\|\boldsymbol{b}\|_i^*$.

3 ALGORITHMS, METHOD, AND MODELS

3.1 Hill-Climbing and Sampling Algorithms

For our empirical study, we use a hill-climbing EA (HCEA) also known as the 1+1 EA [3], which explores the landscape using stochastic hill climbing. The HCEA employs one individual for searching. Then, it creates a clone which is mutated (in this work, flipping each bit with an independent probability $p = 1/n$), and keeps the best solution. In case of a tie, the offspring is kept. This process is repeated until the stopping criterion has been met. In this paper, the HCEA stops when it has reached a known local optimum. The following result can be shown for HCEA:

Theorem 3.1. *When using HCEA, the probability of convergence to an optimum in peak $\|\boldsymbol{b}\|_i^*$, or $P(c(\|\boldsymbol{b}\|_i^*))$, is approximately equal to the size of the share of its RoA, divided by the sum of the sizes of SoRoAs of all peaks*[2] $\|\boldsymbol{b}\|_j^* \in |B|$:

$$P(c(\|\boldsymbol{b}\|_i^*)) \approx \frac{|\,\text{SoRoA}(\|\boldsymbol{b}\|_i^*)|}{\sum\limits_{j \in |B|} |\,\text{SoRoA}(\|\boldsymbol{b}\|_j^*)|}$$

The Sampling algorithm is shown in Algorithm 1. Using this algorithm, we run the HCEA until an optimum $\boldsymbol{b}^*$ has been found and is added to a set of found optima, B. Then, HCEA restarts and continues sampling. This process continues until a *revisit* occurs, i.e., an optimum $\boldsymbol{b}^* \in B$ is found a second time. We record the number of restarts of HCEA it took to do a revisit, $r = |B| + 1$, and continue sampling. The process is repeated until a sample with N measurements has been collected, $\boldsymbol{R} = (r_1, \dots, r_N)$. Using the Birthday Paradox analogy, a repetition or revisit by Sampling represents the number of people needed in a room to get the same birthday for two people. We denote this number as r. We study the median $\tilde{r}$[3] as computed from the collected sample $\boldsymbol{R}$ of size N as an empirical estimate of r. We denote this estimate as $\tilde{r}_N$.[4]

Algorithm 1 Pseudocode for the sampling procedure

```
Require: A fitness function f and an HCEA
procedure SAMPLING(f, HCEA)
    R ← ∅                                  ▷ Initialize set of samples
    for j = 1, ..., N do
        B ← ∅                              ▷ Initialize set of local optima
        b* ← HCEA(f)                       ▷ Find local optimum b*
        Push b* into B
        while true do
            b* ← HCEA(f)                   ▷ Find another local optimum b*
            if b* ∈ B then
                break                      ▷ "Repeated birthday"
            else
                Push b* into B             ▷ "New birthday"
        r ← |B| + 1                        ▷ Number of repetitions r
        Push r into R
    return R
```

We use $\tilde{r}$ to approximate the number of optima d as:

$$\tilde{d} = \frac{(\tilde{r} - 0.5)^2 - 0.25}{2\log(2)}. \tag{2}$$

Eq. 2 is derived from a formula to calculate the number of people needed for the Birthday Paradox to hold [7]:

$$r = \left\lceil 0.5 + \sqrt{0.25 + 2 \cdot \log(2) \cdot d} \right\rceil. \tag{3}$$

We use Eq. 3 to analytically approximate the repetition number r, and hence get an idea of the true number of optima d.

3.2 Synthetic Triangle Functions

To analyze multimodal functions with an arbitrary number of optima, we propose a number of synthetic PBFs for testing. These functions receive a bitstring $\boldsymbol{b}$ and return an integer fitness value, depending on the number of 1s in $\boldsymbol{b}$. Different bitstrings with the same number of 1s in their genotype have the same phenotypic values $\|\boldsymbol{b}\|$ and hence the same fitness.

Definition 3.2. Consider $\|\boldsymbol{b}\|$ as the number of 1s in a given bit string $\boldsymbol{b}$. We define the **triangular positive wave function** as:

$$\text{Triangle}(\boldsymbol{b}, m, s) = \begin{cases} g(\boldsymbol{b}), & \text{if } \left\lceil \frac{\|\boldsymbol{b}\|}{s} \right\rceil \mod 2 = 1 \\ m\left(\left\lceil \frac{\|\boldsymbol{b}\|}{s} \right\rceil \cdot s - \|\boldsymbol{b}\|\right) & \text{otherwise} \end{cases} \tag{4}$$

where

$$g(\boldsymbol{b}) = \begin{cases} m \cdot s, & \text{if } \|\boldsymbol{b}\| \mod s = 0 \\ m(\|\boldsymbol{b}\| \mod s) & \text{otherwise.} \end{cases} \tag{5}$$

[1] We omit highly unlikely mutations for simplicity.

[2] It can also be approximated by dividing over the size of the search space, 2^n, assuming that RoAs were to partition the search space.

[3] We use the median as the sampling distribution is not always normal.

[4] We use r as the number of *repetitions*, $\tilde{r}$ for its analytical computation, and $\tilde{r}_N$ for its empirical estimate in a sample of size N.

Here, m and s are the function's slope and step, respectively.

An example of Triangle with $n = 16$, $m = 1$ and $s = 4$ is shown in Figure 1. By adjusting the Triangle parameters, different PLOP landscapes can be generated. When $n = s$, the problem is identical to OneMax, and when $n = 4s$, two peaks with equal RoAs are present.

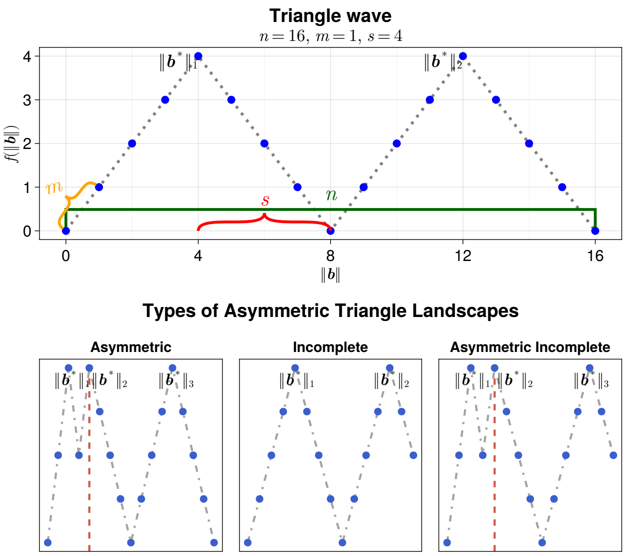

Figure 1: The top panel shows a Triangle function with $n = 16$, $m = 1$, and $s = 4$. Panels at the bottom show asymmetric Triangle landscapes, with a red line separating the α- and β-regions.

Generalized Triangle. A generalized version of Triangle can be constructed using piecewise definitions. For simplicity, we focus on at most two intervals here. Consider a function, f, with 2 intervals (or **regions**) of Triangle, a 2-range Triangle:

$$f(\boldsymbol{b}) = \begin{cases} \text{Triangle}(\boldsymbol{b}, m_1, s_1), & 0 \leq \|\boldsymbol{b}\| \leq \alpha \\ \text{Triangle}(\boldsymbol{b}, m_2, s_2), & \alpha \leq \|\boldsymbol{b}\| \leq \beta = n \end{cases} \tag{6}$$

where each definition is bounded by an upper limit and creates a region such that $0 \leq \alpha \leq \beta = n$.

Symmetric Triangle Function. Let f be a 2-range Triangle function with equal parameters m_i and s_i, where $n = 4s_i$. Using $\alpha = n/2$ and $\beta = n$ divides f into two regions of equal size. We denote this special case as a **symmetric triangle function** (STF).

Figure 1 illustrates an STF. An OoRoA occurs at $\|\boldsymbol{b}\| = 8$ but since both regions have equal size, the probability of converging to any of the peaks is the same. It can be shown that any STF, when searched by the HCEA, is a symmetric k-PLOP function with two regions of equal size 2^{n-1}. This assumption of equal probability holds for a symmetric k-PLOP and can be applied for HCEA when searching an STF (similar to birthdays spread uniformly).

In contrast to the typical use of the Birthday Paradox, we do not start with a fixed number of days d. Instead, we seek to estimate d from sampling the fitness function using *repetitions*, r, found by HCEA (number of people in a room).

4 EXPERIMENTAL RESULTS

4.1 Experimental Settings

We run HCEA to sample the landscape of f. When a sample of size N has been collected, we use $\tilde{r}_N$ to estimate the number of optima (see Eq. 2). We use four different sample sizes $N \in \{50, 100, 200, 300\}$, and report $\tilde{r}_N$ to compare against the ideal $\tilde{r}$.

4.2 Symmetric Fitness Functions

Goal. We estimate the number of repetitions, r, as a way to estimate the number of optima, d, for different symmetric PLOP functions with many optima. We know the number of optima d, and hence can assess the quality of our estimate $\tilde{r}$.

Design. We sample different symmetric Triangle functions using different values for parameters n and s to obtain landscapes with one and two peaks over the phenotypic values. The first function is an STF with $n = 16$ and $s = 4$—with two peaks and $d = 3640$ optima. For the second test, we use Triangle$(\boldsymbol{b}, m = 1, s = 8)$ with $n = 16$, with $d = 12870$ optima in $\mathbb{B}^n$ concentrated on a single peak at $\|\boldsymbol{b}\| = 8$.

Results and Discussion. Results are presented in the top of Figure 2. In all cases, the true number of repetitions $\tilde{r}$ (see Eq. 3) is quite close to the median $\tilde{r}_N$ of each sample. See for example $\tilde{r}_{200} = 71.5$ for the first symmetric landscape. This allows for very accurate estimates of the number of optima, d.

4.3 Asymmetric Fitness Functions

Goal. Symmetry, as discussed in Section 4.2, might be too strong of an assumption for many fitness functions, so we study other Triangle landscapes as well (see the bottom row of Figure 1).

Design. We consider three asymmetric cases (see Figure 1). First, a 2-range Triangle with $n = 16$, where the α- and β-regions have different sizes:

$$\text{Triangle}_{asym}(\boldsymbol{b}) = \begin{cases} \text{Triangle}(\boldsymbol{b}, m = 2, s = 2), & 0 \leq x \leq 4 \\ \text{Triangle}(\boldsymbol{b}, m = 1, s = 4), & 4 \leq x \leq n. \end{cases} \tag{7}$$

We also estimate the number of optima on an *incomplete* Triangle function, Triangle$(\boldsymbol{b}, m = 1, s = 4)$, but using $n = 14$. By setting n to be non-divisible by s, the last slope of Triangle is *shorter* than the rest, affecting the probability of convergence to the second peak. Finally, we combine the two concepts and test on a 2-range *asymmetric incomplete* Triangle, using Eq. 7 but setting $n = 14$.

Results and Discussion. Results are presented in the bottom part of Figure 2. In the asymmetrical landscape (cf. left panel of Figure 1), the empirical medians, $\tilde{r}_N$ are still close to the actual $\tilde{r}$ value. See for example $\tilde{r}_{200}$ and $\tilde{r}_{300}$ with 66.5 and 68 repetitions respectively, compared to the *ideal* computed median of $\tilde{r} = 72.7$. This occurs because the probability of converging to any peak of the β-region is about the same: the peak $\|\boldsymbol{b}\|_3^*$ at $\|\boldsymbol{b}\| = 12$ can be reached by any state in the β-region with $\|\boldsymbol{b}\| \geq 8$, i.e., $P(c(\|\boldsymbol{b}\|_3^*)) \approx 0.5$; while for the second peak in β-region $P(c(\|\boldsymbol{b}\|_2^*)) \approx 0.4936$.

On the *incomplete* Triangle landscape, the estimates deviate more from the ideal $\tilde{r}$. Looking at the RoA for both peaks for the incomplete landscape (c.f. middle panel of Figure 1) reveals a big difference: For $\|\boldsymbol{b}\|_1^*$ at $\|\boldsymbol{b}\| = 4$, $|\,\text{RoA}(\|\boldsymbol{b}\|_1^*)| = 12911$. For $\|\boldsymbol{b}\|_2^*$ at

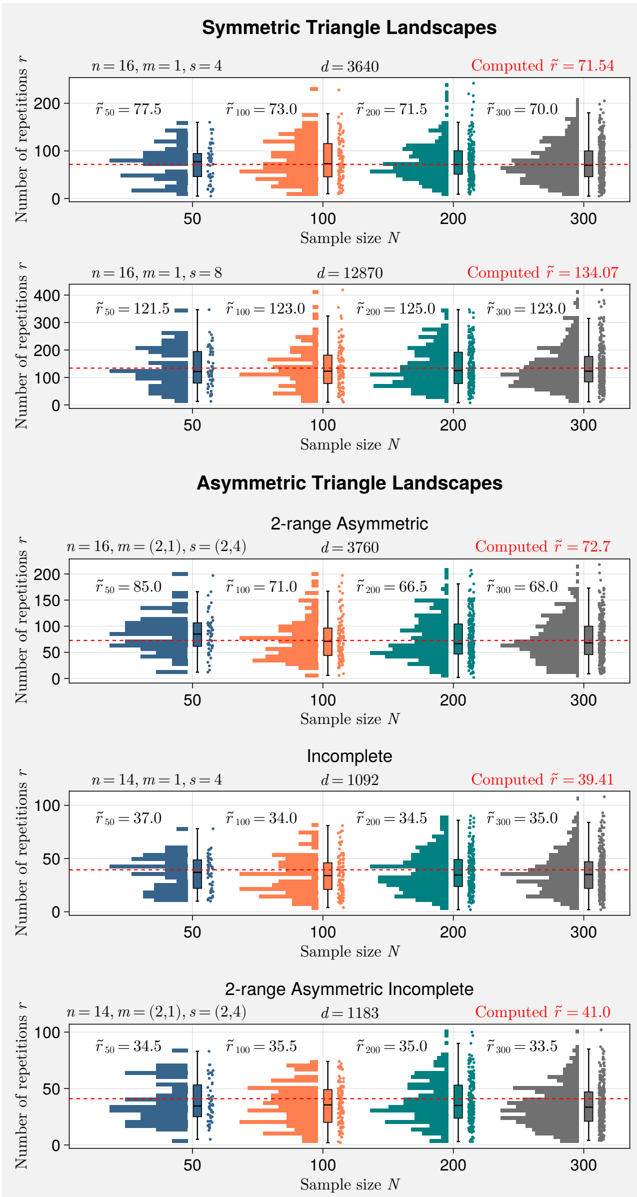

Figure 2: Sampling results on TRIANGLE landscapes. For each trial, we present a histogram, a box plot and a scatter plot of *R*. The median $\tilde{r}$ is shown with a red dashed line.

$\|\boldsymbol{b}\| = 12$, $|\operatorname{RoA}(\|\boldsymbol{b}\|_1^*)| = 6476$. Hence, the probability of convergence for both peaks is altered.

The method performs similarly on the 2-range *asymmetric incomplete* TRIANGLE. The distribution of optima is not even. Examining the RoA of each peak (c.f. right panel of Figure 1) we see that it is much different between peaks: $|\operatorname{RoA}(\|\boldsymbol{b}\|_1^*)| = 470$, $|\operatorname{RoA}(\|\boldsymbol{b}\|_2^*)| = 12805$ and $|\operatorname{RoA}(\|\boldsymbol{b}\|_3^*)| = 6476$. This highlights the limitations of the estimation method on PLOP landscapes with multiple peaks where some of them have much larger RoAs than others. Nevertheless, the values for $\tilde{r}$ are still inside the interquartile range of the distribution in the box plots in Figure 2, serving as an estimate of the number of optima in the landscape.

5 CONCLUSION AND FUTURE WORK

Much research has focused on one or a few local optima, while there is evidence that many real world problems are multimodal [6, 9, 13]. In this work, we focus on estimating the number of optima in highly multimodal fitness landscapes using a new synthetic function, TRIANGLE. We study methods for estimating the region of attraction and the probability of convergence in multimodal (PLOP) functions to estimate the number of local optima of pseudo-Boolean functions. We test our estimation procedure, which is based on the reverse Birthday Paradox, on various pseudo-Boolean function landscapes created from our synthetic function. The method gives very accurate estimates on functions where optima are spread evenly through the fitness landscape. When facing landscapes with big differences in the region of attraction of their peaks, the accuracy declines but still gives an approximation of the number of optima. Several opportunities exist for building on this work: parameter setting in multimodal optimization or improving the estimation procedure by trying out other sampling methods.

REFERENCES

[1] C. Bian, C. Feng, C. Qian, and Y. Yu. 2020. An Efficient Evolutionary Algorithm for Subset Selection with General Cost Constraints. In *Proc. AAAI.* 3267–3274.
[2] R. Caruana and M. Mullin. 2000. Estimating the Number of Local Minima in Big, Nasty Search Spaces. (March 2000).
[3] S. Droste, T. Jansen, and I. Wegener. 1998. A Rigorous Complexity Analysis of the (1+1) Evolutionary Algorithm for Linear Functions with Boolean Inputs. In *Proc. CEC.* 499–504. https://doi.org/10.1109/ICEC.1998.700079
[4] L. Hernando, A. Mendiburu, and J. A. Lozano. 2013. An Evaluation of Methods for Estimating the Number of Local Optima in Combinatorial Optimization Problems. *Evolutionary Computation* 21, 4 (2013), 625–658.
[5] H. H. Hoos. 2002. A mixture-model for the behaviour of SLS algorithms for SAT. In *Proc. AAAI.* 661–667.
[6] X. Li, M. G. Epitropakis, K. Deb, and A. Engelbrecht. 2017. Seeking Multiple Solutions: An Updated Survey on Niching Methods and Their Applications. *IEEE Transactions on Evolutionary Computation* 21, 4 (2017), 518–538.
[7] F. H. Mathis. 1991. A Generalized Birthday Problem. *SIAM Rev.* 33, 2 (1991), 265–270. https://doi.org/10.1137/1033051
[8] O. J. Mengshoel, Y. Ahres, and T. Yu. 2016. Markov Chain Analysis of Noise and Restart in Stochastic Local Search. In *Proc. IJCAI.* 639–646.
[9] O. J. Mengshoel, E. Flogard, J. Riege, and T. Yu. 2021. Stochastic Local Search Heuristics for Efficient Feature Selection: An Experimental Study. In *Proc. NIKT.* 58–71.
[10] W. Mostert, K. Malan, and A. Engelbrecht. 2018. Filter Versus Wrapper Feature Selection Based on Problem Landscape Features. In *Proc. GECCO Companion.* 1489–1496.
[11] W. Mostert, K. Malan, G. Ochoa, and A. P. Engelbrecht. 2019. Insights into the Feature Selection Problem using Local Optima Networks. In *Proc. EvoCOP 2019.* 147–162.
[12] D. K. Pal and O. J. Mengshoel. 2016. Stochastic CoSaMP: Randomizing Greedy Pursuit for Sparse Signal Recovery.. In *Proc. ECML-PKDD.* 761–776.
[13] M. Preuss. 2015. *Multimodal Optimization by Means of Evolutionary Algorithms* (1st ed.). Springer.
[14] A. Prügel-Bennett. 2004. When a Genetic Algorithm Outperforms Hill-Climbing. *Theoretical Computer Science* 320, 1 (2004), 135 – 153.
[15] B. Y. Qu, J. J. Liang, Z. Y. Wang, Q. Chen, and P. N. Suganthan. 2016. Novel Benchmark Functions for Continuous Multimodal Optimization with Comparative Results. *Swarm and Evolutionary Computation* 26 (Feb. 2016), 23–34.
[16] R. Schumann, L. Mou, Y. Lu, O. Vechtomova, and K. Markert. 2020. Discrete Optimization for Unsupervised Sentence Summarization with Word-Level Extraction. In *Proc. ACL.* 5032–5042.
[17] K. Shang, H. Ishibuchi, and W. Chen. 2021. Greedy approximated hypervolume subset selection for many-objective optimization. In *Proc. GECCO.* 448–456.
[18] E. Yolcu and B. Poczos. 2019. Learning Local Search Heuristics for Boolean Satisfiability. In *Proc. NeurIPS.* 7992–8003.
[19] T. Yu, B. Kveton, and O. J. Mengshoel. 2017. Thompson Sampling for Optimizing Stochastic Local Search. In *Proc. ECML-PKDD.* 493–510.
[20] F. Zou, D. Chen, H. Liu, S. Cao, X. Ji, and Y. Zhang. 2022. A Survey of Fitness Landscape Analysis for Optimization. *Neurocomputing* 503 (2022), 129–139.