UNIVERSITÄT DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

SONDERFORSCHUNGSBEREICH 531

Design und Management komplexer technischer Prozesse
und Systeme mit Methoden der Computational Intelligence

On the Complexity of Overcoming Gaps When Using Elitist Selection and Isotropic Mutations

Jens Jägersküpper

Nr. CI-198/05

Interner Bericht ISSN 1433-3325 November 2005

Sekretariat des SFB 531 · Universität Dortmund · Fachbereich Informatik/XI
44221 Dortmund · Germany

Diese Arbeit ist im Sonderforschungsbereich 531, „Computational Intelligence“, der Universität Dortmund entstanden und wurde auf seine Veranlassung unter Verwendung der ihm von der Deutschen Forschungsgemeinschaft zur Verfügung gestellten Mittel gedruckt.

On the Complexity of Overcoming Gaps When Using Elitist Selection and Isotropic Mutations

Jens Jägersküpper*

Universität Dortmund, Informatik 2, 44221 Dortmund, Germany
`JJ@Ls2.cs.uni-dortmund.de`

Abstract

We consider the (1+$\boldsymbol{\lambda}$) evolution strategy, an evolutionary algorithm for minimization in $\mathbb{R}^{\boldsymbol{n}}$, using isotropic mutations. Thus, for instance, Gaussian mutations adapted by the 1/5-rule or by $\boldsymbol{\sigma}$-self-adaptation are covered. Lower bounds on the (expected) runtime (defined as the number of function evaluations) to overcome a gap in the search space are proved (where the search faces a gap of size $\boldsymbol{\Delta}$ if the distance between the current search point and the set of all better points is at least $\boldsymbol{\Delta}$), showing when the runtime is potentially polynomial and when the runtime is necessarily super-polynomial or even necessarily exponential in $\boldsymbol{n}$, the dimensionality of the search space.

1 Introduction

The (1+λ) evolutionary algorithm ((1+λ) EA) is a simple evolutionary algorithm (EA). The "1" indicates that it uses a population of size 1; thus, crossover is not possible and mutation is the only evolutionary search operator that is used to evolve a search point $\boldsymbol{c}$. The "λ" indicates that, in a single step, λ offspring of the current search point $\boldsymbol{c}$ are generated by λ independent mutations. The best of these λ mutants replaces/becomes the current individual $\boldsymbol{c}$ if (and only if) it is as least as fit as its parent (this is called *elitist selection,* indicated by the "+"). We have just described a single step; this step is repeated in the so-called *evolution loop* until a stopping criterion is met. Fortunately, for the results we are aiming at here, we need not define a reasonable stopping criterion, yet consider an infinite evolution loop.

In the beginning of the analysis of the (1+1) EA for the search space $\{0,1\}^n$ in the early 1990s, symmetric functions $f\colon \{0,1\}^n \to \mathbb{N}$ have been considered, i. e., $|\boldsymbol{x}| = |\boldsymbol{y}| \Rightarrow f(\boldsymbol{x}) = f(\boldsymbol{y})$, where for a bit-string $\boldsymbol{x} \in \{0,1\}^n$, "$|\boldsymbol{x}|$" denotes the L^1-norm, i. e. the number of ones in $\boldsymbol{x}$. Mühlenbein (1992) proved that the expected number of steps the (1+1) EA needs to maximize the L_1-norm (OneMax) is $O(n \log n)$ (the O-notation and other asymptotics are recalled in Section 2). As (1+λ) EAs use a population of size 1, mutation is the only

---

*supported by the German Research Foundation (DFG) as a member of the collaborative research center "Computational Intelligence" (SFB 531)

search operator that is used. The standard mutation operator for the search space $\{0,1\}^n$ independently flips each bit in an individual with the *mutation probability* $p = 1/n$, so that the expected Hamming distance between a mutant and its parent equals 1. It is easy to see that the $O(n \log n)$-bound applies to the minimization of the L^1-norm, too. We will always consider minimization in the following. A simple modification of the proof reveals that the expected number of steps to halve the Hamming distance from the optimum is $O(n)$. As an introductory example/exercise of a lower-bound proof, we will show that it is also $\Omega(n)$:

Let $I = \{i \mid x_i = 1\}$ denote the index set of those bits that are set to one at a specific point in time of a run of the (1+1) EA. For $n \geq 3$, the probability that a fixed bit flips (at least once) within $n/3$ steps equals $1-(1-p)^{n/3} = 1-(1-1/n)^{n/3} \leq 1/3$. Therefore, the expected number of bits in I that flip (at least once) within the next $n/3$ steps is at most $\#I/3$. By Markov's inequality, the probability that at least $\#I/2$ of the bits in I flip (at least once) in $n/3$ steps is at most $(\#I/3)/(\#I/2) = 2/3$. Thus, with a probability of at least 1/3, at least half of the remaining $\#I$ ones never flip within these $n/3$ steps. Consequently, with a probability of at least 1/3, it takes the (1+1) EA at least $n/3$ steps to halve the number of ones, and hence, the expected number of steps to halve the Hamming distance from the optimum is at least $(1/3)(n/3) = n/9 = \Omega(n)$.

Thus, the standard (1+1) EA needs $\Theta(n)$ steps (in expectation) to halve the distance from the optimum when minimizing the number of ones in a bit-string of length n. Jägersküpper (2003) shows that the (1+1) evolution strategy ((1+1) ES) for minimization in $\mathbb{R}^n$ indeed also needs $\Theta(n)$ steps to halve the Euclidian (instead of Hamming) distance from the optimum when minimizing the L^2-norm (instead of the L^1-norm) when using Gaussian mutations adapted by the 1/5-rule (instead of independently flipping each bit with the standard mutation probability $1/n$).

However, for the search space $\{0,1\}^n$, one of the next functions that have been considered in a theoretical analysis of the runtime is $\textsc{Jump}_m : \{0,1\}^n \to \mathbb{N}$ with $2 \leq m \leq n/3$ and $m \in \mathbb{N}$, defined by

$$\textsc{Jump}_m(\boldsymbol{x}) := \begin{cases} 2n & \text{if } 1 \leq |\boldsymbol{x}| \leq m-1, \\ |\boldsymbol{x}| & \text{otherwise.} \end{cases}$$

We call the plateau of worst $\textsc{Jump}$-value $2n$ "the gap" as it separates the global optimum, namely the origin, from the L^1-norm based part of the fitness landscape; all bit-strings with exactly m ones are locally but not globally optimal. Since the (1+1) EA chooses the initial search point $\boldsymbol{x}$ uniformly at random, $\mathsf{E}[|\boldsymbol{x}|] = n/2$ and, by Chernoff bounds, $\mathsf{P}\{|\boldsymbol{x}| \leq n/3\}$ is exponentially small (in n). Consequently, the initial search point is located in the gap only with an exponentially small probability; the probability that the initial search point is the optimum equals 2^{-n}.

Droste, Jansen, and Wegener (2002) prove that the expected runtime of the (1+1) EA using the static mutation probability $p = 1/n$ on $\textsc{Jump}_m$ is $\Theta(n^m)$; in fact, a slightly different function which is to be maximized is investigated, yet the proof carries over. Roughly speaking, the (1+1) EA minimizes $\textsc{Jump}_m$ as it minimizes the L^1-norm up to the point when a locally but not globally optimum point with Hamming distance m from the origin is created. Then a mutation must flip exactly the remaining m ones, for the (1+1) EA to overcome the gap,

i. e., to obtain a search point with better fitness/smaller Jump-value, namely the optimum. The probability of this event (called "success" in the following) equals $p^m(1-p)^{n-m}$, where p denotes the mutation probability (recall that a mutation consists in flipping each bit independently with probability p). Since $\frac{\mathrm{d}}{\mathrm{d}p}\, p^m(1-p)^{n-m} = 0$ for $p = m/n$, the success probability is maximum when using the mutation probability $p = m/n$, and hence, even if the (1+1) EA could adapt p optimally, the success probability is upper bounded by $(m/n)^m(1-m/n)^{n-m}$. Since the number of trials until a mutation actually creates a better point is geometrically distributed, the expected runtime of the (1+1) EA on $\mathrm{J{\scriptsize UMP}}_m$ is lower bounded by the reciprocal of the success probability. Thus, we expect a superpolynomial number of steps if $(n/m)^m$ is superpolynomial or if $(1-m/n)^{n-m}$ is superpolynomially small.

For $m \in [n^\varepsilon, n/3]$ with $\varepsilon \in (0,1)$ we have

$$(1-m/n)^{n-m} \le (1-n^\varepsilon/n)^{n\cdot 2/3} < \exp(-n^\varepsilon 2/3)$$

since $(1-1/k)^k < 1/\mathrm{e}$, and hence, the success probability is exponentially small (so that the expected runtime is exponential). For $m \in [\log n, n^\varepsilon]$ with $\varepsilon \in (0,1)$

$$(n/m)^m \ge (n^{1-\varepsilon})^{\log n} = n^{(1-\varepsilon)\cdot \log n},$$

and thus, the expected runtime is superpolynomial. Finally, we consider the case $m \le \log n$. Then $(n/m)^m = 2^{m(\log n - \log m)} \ge 2^{m(\log n - \log\log n)} = 2^{m\cdot\Omega(\log n)} = n^{m\cdot\Omega(1)}$, and hence, the expected runtime is superpolynomial unless $m = O(1)$.

All in all, the expected runtime of the (1+1) EA on $\mathrm{J{\scriptsize UMP}}_m$ is polynomial (in n) if $m = O(1)$ when using the standard mutation probability $1/n$, and—as we we have just shown—it is superpolynomial if m is not $O(1)$ *even when the mutation probability could be adapted optimally,* i. e., our lower bound applies also, for instance, to the dynamic (1+1) EA introduced by Droste et al. (2001), which varies the mutation probability according to a static periodic schedule. In other words, an efficient optimization (i. e. a polynomial runtime) is possible only for a gap corresponding to a constant number of bits which have to be flipped simultaneously by a single mutation.

The aim of this paper is to prove a similar result for minimization using "isotropic-mutation hill-climbing" in the search space $\mathbb{R}^n$. We will also start off with a symmetric function—spherically symmetric, of course.

Evolutionary algorithms for optimization in $\mathbb{R}^n$ are commonly subsumed under the term *evolution(ary) strategies (ESs)* coined by Rechenberg and Schwefel; cf. Rechenberg (1973) and Schwefel (1995). Rigorous analyses of the (expected) runtime of ESs like those by Jägersküpper (2005, 2003) for the (1+1) ES using Gaussian mutations adapted by the 1/5-rule succeeded only recently. Therein, unimodal functions, namely the well-known Sphere-function and positive definite quadratic forms, are considered. For multimodal functions, no rigorous analysis has been published at all by now (up to the author's knowledge, of course). However, an almost uncountable number of experimental results on the optimization of multimodal functions exist. In attempting to find an explanation for the experimental findings obtained for search space dimensions usually ranging from 10 to 30, some works additionally present some calculations for the 1-dimensional search space $\mathbb{R}$ (e. g. Yao et al. (1999)). Unfortunately, such calculations cannot help us with an analysis of how the runtime grows with the dimensionality of the search space, i. e., with explaining the dramatic effect which an increasing search space dimensionality has on the performance of an ES on a multimodal function.

In the considerably developed theory on local performance measures (*progress rate, quality gain;* cf. Beyer (2001)), the progress which a single step yields (after the single-step gain has become steady-state) is analyzed, i. e., the optimization process is (implicitly) assumed to stabilize w. r. t. to the single-step progress. As this is hardly the case for multimodal optimization, it is not apparent how to apply this approach to the scenario considered here.

In the following section, some notions and results that will be used are recapitulated. Subsequently, the function scenario that we are going to investigate is introduced, and in Section 4, the method by which the lower bounds on the expected runtime will be obtained is summarized. In Section 5, we are going to have a closer look at the probability with which a mutation succeeds, and afterwards, exponential lower bounds on the expected runtime will be obtained in Section 6. The crucial size of the gap to overcome making the expected runtime to change from potentially polynomial to necessarily superpolynomial is determined in Section 7. Finally, we discuss those results and finish with some concluding remarks in Section 8.

2 Preliminaries

First of all, recall the following **asymptotics**, where $g, h\colon \mathbb{N} \to \mathbb{R}$,

- $g(n) = O(h(n))$ iff there exists a positive constant κ such that $g(n) \leq \kappa \cdot h(n)$ for n large enough, i. e., for $n \geq n_0$ where n_0 is a constant,
- $g(n) = \Omega(h(n))$ iff $h(n) = O(g(n))$,
- $g(n) = \Theta(h(n))$ iff $g(n)$ is $O(h(n))$ and $\Omega(h(n))$,
- $g(n) = \mathrm{poly}(n)$ iff $g(n) = O(n^c)$ for a constant c;

furthermore, for $g, h > 0$,

- $g(n) = o(h(n))$ iff $g(n)/h(n) \to 0$ as $n \to \infty$,
- $g(n) = \omega(h(n))$ iff $h(n) = o(g(n))$.

As we are interested in how the runtime depends on n, the dimensionality of the search space, all asymptotics are w. r. t. to this parameter (unless stated differently).

A probability is **exponentially small** (in n) if it is upper bounded by $\exp(-\Omega(n^{\varepsilon})$ for a constant $\varepsilon > 0$.

As mentioned above, we will consider isotropic mutations for $\mathbb{R}^n$. For $\boldsymbol{m} \in \mathbb{R}^n$ let $|\boldsymbol{m}|$ denote $\boldsymbol{m}$'s Euclidean length, i. e. its L^2-norm. A random mutation vector $\boldsymbol{m} \in \mathbb{R}^n$ is **isotropically distributed** if

- the normalized vector $\boldsymbol{m}/|\boldsymbol{m}|$ is uniformly distributed upon the unit hypersphere $\{\boldsymbol{x} \in \mathbb{R}^n \mid |\boldsymbol{x}| = 1\}$ and
- the random length $|\boldsymbol{m}|$ is independent of the random direction $\boldsymbol{m}/|\boldsymbol{m}|$.

It is well known that a Gaussian mutation vector $\widetilde{\boldsymbol{m}}$, each component of which is independently standard normal distributed, is isotropically distributed. It is easy to see that a scaled Gaussian mutation $\sigma \cdot \widetilde{\boldsymbol{m}}$ with $\sigma > 0$ is also isotropically distributed. Scaled Gaussian mutations are commonly used, for instance within Rechenberg's 1/5-rule or Schwefel's σ-self-adaptation. Within the *covariance matrix adaptation (CMA)* due to Hansen and Ostermeier (1996), however, $\sigma \cdot \mathbf{B} \cdot \widetilde{\boldsymbol{m}}$ makes up the mutation vector with a matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ which is also adapted, and thus, unless $\mathbf{B} = s \cdot \mathbf{I}$ for some scalar s, the mutation vector is not isotropically distributed.

However, the crucial property of an isotropic mutation is that we can first choose a random direction and subsequently sample a point on the already chosen half line according to the distribution of $|\boldsymbol{m}|$; or we can (and will) do it vice versa: First we sample the mutation's length ℓ according to the random variable $|\boldsymbol{m}|$; subsequently, the mutant is uniformly distributed upon the hypersphere with radius ℓ centered at the mutated search point.

Let G_ℓ denote the spatial gain of an isotropically distributed mutation vector $\boldsymbol{m}$ parallel to an arbitrarily fixed direction, given that (the random variable) $|\boldsymbol{m}|$ takes the value $\ell > 0$; note that $G_\ell \in [-\ell, \ell]$. In other words, when $\boldsymbol{c} \in \mathbb{R}^n$ is mutated and $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{m}$ is the mutant, then the absolute value of G_ℓ equals the distance between the two hyper-planes containing $\boldsymbol{c}$ resp. $\boldsymbol{c}'$ that are perpendicular to the predefined direction, respectively. Jägersküpper (2003) shows that for $n \geq 4$ the density of the random variable G_ℓ at $g \in [-\ell, \ell]$ equals $(1/\ell) \cdot \left(1 - (g/\ell)^2\right)^{(n-3)/2} / \Psi_{n-2}$, where $\sqrt{2\pi/(k+1)} < \Psi_k < \sqrt{2\pi/k}$ for $k \geq 2$, and thus, $1/\Psi_{n-2} = \Theta(\sqrt{n})$.

With this density function, the probability of $\boldsymbol{c}'$ being closer to some fixed point in the search space than $\boldsymbol{c}$ has been estimated. For notational convenience, we assume w. l. o. g. that the unique optimum point coincides with the origin $\boldsymbol{o}$ so that we can use $|\boldsymbol{c}|$ for the distance from the optimum. Obviously, if $\ell > 2|\boldsymbol{c}|$ then $|\boldsymbol{c}'| > |\boldsymbol{c}|$; if $\ell \leq 2|\boldsymbol{c}|$, however, $|\boldsymbol{c}'| \leq |\boldsymbol{c}|$ if and only if the spatial gain parallel to $\overrightarrow{\boldsymbol{c}\boldsymbol{o}}$ is at least $\ell^2/(2|\boldsymbol{c}|)$ as has been shown. Thus, for instance, the (conditional) probability that the mutant is at least as close to the optimum/origin as its parent equals

$$\begin{aligned} & \mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}| \text{ given that } |\boldsymbol{m}| = \ell\} \\ = \; & \Theta(\sqrt{n}) \cdot \int_{\min\{1\,,\,\ell/(2|\boldsymbol{c}|)\}}^{1} \left(1 - y^2\right)^{(n-3)/2} \mathrm{d}y \end{aligned}$$

(since $|\boldsymbol{c}'| \leq |\boldsymbol{c}|$ is impossible if $\ell > 2|\boldsymbol{c}|$).

3 Function Scenario/Fitness Landscape

As mentioned before, we want to investigate symmetric functions for $\mathbb{R}^n$ that correspond to $\textsc{Jump}_m$ for $\{0,1\}^n$. Therefore, let $\Delta\colon \mathbb{N} \to (0, 1/3]$ denote a function. The sequence of functions $\textsc{Gap}_n^\Delta\colon \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, is defined by

$$\textsc{Gap}_n^\Delta(\boldsymbol{x}) \quad := \quad \begin{cases} 2 & \text{for } |\boldsymbol{x}| \in [1{-}\Delta(n)\,, 1) \text{ and} \\ |\boldsymbol{x}| & \text{otherwise.} \end{cases}$$

Due to Δ's range/codomain, all $\boldsymbol{x}$ in the unit hypersphere $U := \{\boldsymbol{x} \mid |\boldsymbol{x}| = 1\}$ are locally but not globally optimal, and the origin is the unique global optimum.

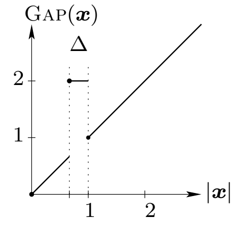
Let $B := \{\boldsymbol{x} \mid |\boldsymbol{x}| < 1 - \Delta\}$ denote the open hyper-ball separated from the rest $R := \{\boldsymbol{x} \mid |\boldsymbol{x}| \geq 1\}$ of the search space by "the gap" $P := \{\boldsymbol{x} \mid 1 - \Delta \leq |\boldsymbol{x}| < 1\}$, a plateau of constant value 2. Hence, for a mutation of $\boldsymbol{c} \in R$ to succeed in conquering the gap, i. e. $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{m} \in B$, the (isotropic) mutation $\boldsymbol{m}$ must reduce the distance from the global optimum by at least $\Delta(n)$. Thus, the parameter Δ corresponds to the parameter m in $\textsc{Jump}_m$ discussed in the introduction.

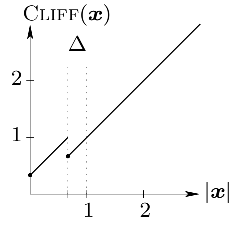
A similar function, which has been discussed already by Jägersküpper and Witt (2005), is

$$\textsc{Cliff}_n^{\Delta}(\boldsymbol{x}) := \begin{cases} |\boldsymbol{x}| + \Delta(n) & \text{for } |\boldsymbol{x}| < 1 - \Delta(n), \\ |\boldsymbol{x}| & \text{otherwise.} \end{cases}$$

When $\Delta(n)$ is $o(1)$, then Cliff converges uniformly to the L^2-norm, which is simple to minimize[1] for the (1+1) ES. Since in the limit of infinite search space dimension "$\textsc{Cliff}_{\infty}^{\Delta} = L^2$-norm" (because $\lim_{n\to\infty} \Delta(n) = 0$), Cliff serves as a perfect example for how the assumption "$n \to \infty$" can potentially lead to results that reveal nothing but ill-founded conclusions for finite n.

For Cliff, let $R := \mathbb{R}^n \setminus B = \{\boldsymbol{x} \mid |\boldsymbol{x}| \geq 1 - \Delta\}$ denote the "rest of the search space". Each $\boldsymbol{x}$ with $|\boldsymbol{x}| = 1 - \Delta$ is locally but not globally optimal; however, again $|\boldsymbol{c}| = 1$ results in maximum probability of conquering the cliff: For $|\boldsymbol{c}| \in [1 - \Delta, 1]$ a spatial gain of Δ is necessary for $\boldsymbol{c}' \in B \wedge \textsc{Cliff}(\boldsymbol{c}') \leq \textsc{Cliff}(\boldsymbol{c})$, whereas for $|\boldsymbol{c}| = 1 + \varepsilon$ with $\varepsilon > 0$, a spatial gain of $\Delta + \varepsilon$ is necessary. Thus, the ratio of distance from the optimum to necessary gain is maximum for $|\boldsymbol{c}| = 1$—and with it the success probability as we will see.





4 Lower-Bound Argumentation

In this section the argumentation by which we obtain lower bounds on the expected runtime is described. Therefore, recall the lower bound of $\Omega(n^m)$ for $\textsc{Jump}_m$ presented in the introduction, which is valid for any (1+1) EA that uses a symmetric mutation operator[2].

1. We show that the search starts far away from the optimum point so that the gap has to be overcome in oder to find the optimum. (In fact, for Jump we showed that the probability of starting with the optimum or a search point in the gap is exponentially small.) For the search space $\mathbb{R}^n$

---

[1] In fact, the approximation of its minimum is simple, not the minimization itself, when using the 1/5-rule for mutation adaptation, for instance.

[2] Here "symmetric" means that a mutation of $\boldsymbol{x} \in \{0,1\}^n$ consists in flipping each of the n bits independently with the same probability p, independent of whether this bit is set to one or to zero; p may be arbitrarily varied/adapted from one mutation to the next.

we cannot sample a uniformly distributed starting point, and hence, we will assume that the search starts outside the gap.

2. We optimistically assume that the search makes it to a point at which the probability that an optimal adapted mutation overcomes the gap is maximum, and furthermore, we disregard the steps/mutations up to this point. (For JUMP and GAP such a point is a local but non-global minimum; for CLIFF, however, such a point is not a local minimum as already mentioned.)

3. Due to elitist selection, the algorithm repeats mutating such a point until the gap is conquered. Thus, the expected number of steps necessary to overcome the gap is bounded below by the expectation value of a geometrically distributed random variable over $\{1, 2, \dots\}$, namely by $1/p_{\text{opt}}$, where p_{opt} is the best-case (w. r. t. the mutation adaptation as well as the location in the search space) success probability of a mutation to bridge the gap.

To obtain a lower bound by this argumentation, we need to find out at which location in the search space the chance of bridging the gap is maximum, and moreover, we need a good estimate of a mutation's hitting probability. Before we consider such best-case hitting probabilities, however, we will have a closer look at (1+λ) EAs.

Obviously, creating λ mutants rather than one does not affect argument 1. Since we do not count the number of steps in argument 2 anyway, there is also no effect there. In argument 3, the algorithm repeats to mutate the same[3] point again and again, and hence, the expected number of steps is lower bounded by $\lceil (1/p_{\text{opt}})/\lambda \rceil$. Apparently, a larger λ results in a larger lower bound on the number of expected steps: If the gap is overcome already in the first of the λ mutations in a step, the (1+λ) EA may perform up to $\lambda - 1$ vain mutations. So the total expected number of mutations/function evaluations performed in argument 3 may be larger[4]; however, it cannot be smaller—that is the point.

5 Hitting Probabilities

As mentioned in the preceding section, we need a good upper bound on the best-case probability of hitting a point lying in the global-optimum region B, i. e., the goal is an estimation of the *hitting probability* $\mathsf{P}\{\boldsymbol{c}' \in B\}$, the mass of B with respect to the probability measure over $\mathbb{R}^n$ induced by $\boldsymbol{c}$ and the distribution of $\boldsymbol{m}$. As $\boldsymbol{m}$ is isotropically distributed, i. e., $\boldsymbol{m}/|\boldsymbol{m}|$ is uniformly distributed upon the unit hyper-sphere,

$$\begin{aligned} & \mathsf{P}\{\boldsymbol{c}' \in B \mid |\boldsymbol{m}| = \ell\} \\ = \; & \frac{\text{hypersurface area of } B \cap S_{\boldsymbol{c},\ell}}{\text{hypersurface area of } S_{\boldsymbol{c},\ell}} \\ =: \; & \beta_{\boldsymbol{c},B}(\ell), \end{aligned}$$

---

[3] in fact, equivalent points w. r. t. the best-case probability of an accepted mutation to overcome the gap

[4] up to a factor close to λ if p_{opt} is close to 1

where $S_{\boldsymbol{c},\ell}$ denotes the hyper-sphere with radius ℓ centered at $\boldsymbol{c}$, and furthermore,

$$\begin{aligned} \mathsf{P}\{\boldsymbol{c}' \in B\} &= \int_0^\infty \beta_{\boldsymbol{c},B}(\ell) \cdot \text{density}_{|\boldsymbol{m}|}(\ell)\, \mathrm{d}\ell \\ &\leq \sup_{\ell>0} \beta_{\boldsymbol{c},B}(\ell). \end{aligned}$$

As we are aiming at an upper bound on the hitting probability in the best case w. r. t. to $|\boldsymbol{m}|$, we must assume that the mutation adaptation chooses $|\boldsymbol{m}|$ such that the hitting probability is maximum, and hence, we must find $\sup_{\ell>0} \beta_{\boldsymbol{c},B}(\ell)$—or at least a good (i. e. as tight as possible) upper bound.

6 Exponential Lower Bound

We are interested in the *success probability* that a mutation conquers the cliff, i. e., in $\mathsf{P}\{\boldsymbol{c}' \in B \cap A_{\boldsymbol{c}} \mid \boldsymbol{c} \in R\}$, where $A_{\boldsymbol{c}} := \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq f(\boldsymbol{c})\} \subseteq \mathbb{R}^n$ denotes the set of all acceptable (by elitist selection) points with respect to $\boldsymbol{c}$. $A_{\boldsymbol{c}}$ is also called the *lower level set* of $\boldsymbol{c}$ (w. r. t. the function f to be optimized). Obviously, this probability is upper bounded by $\mathsf{P}\{\boldsymbol{c}' \in B \mid \boldsymbol{c} \in R\}$ as well as by $\mathsf{P}\{\boldsymbol{c}' \in A_{\boldsymbol{c}} \mid \boldsymbol{c} \in R\}$. Furthermore, note that these two probabilities are again upper bounded by $\mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}|\}$, respectively.

First, we investigate $\Delta(n) \geq \delta > 0$ for a constant δ. As we have just noted, for $\ell \leq 2|\boldsymbol{c}|$

$$\mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}|\} = \Theta(\sqrt{n}) \int_{\ell/(2|\boldsymbol{c}|)}^{1} \left(1 - y^2\right)^{(n-3)/2} \mathrm{d}y$$

(cf. Section 2 or/and Jägersküpper (2003)) is an upper bound on the success probability, and obviously,

$$\int_{\ell/(2|\boldsymbol{c}|)}^{1} \left(1 - y^2\right)^{(n-3)/2} \mathrm{d}y \leq \left(1 - \left(\tfrac{\ell}{2|\boldsymbol{c}|}\right)^2\right)^{(n-3)/2}.$$

Hence, if $\ell/(2|\boldsymbol{c}|) \geq \varepsilon > 0$ for a constant ε, the success probability is at most $\Theta(\sqrt{n}) \cdot (1 - \varepsilon^2)^{(n-3)/2} = \exp(-\Omega(n))$. Since a success is impossible if $|\boldsymbol{m}| < |\boldsymbol{c}| - (1 - \Delta)$, i. e., if the mutation is too short, we have for $\boldsymbol{c} \in R$ such that $|\boldsymbol{c}| \geq 1$

$$\frac{\ell}{2|\boldsymbol{c}|} \geq \frac{|\boldsymbol{c}| - (1 - \Delta)}{2|\boldsymbol{c}|} = \frac{1}{2} - \frac{1}{2|\boldsymbol{c}|} + \frac{\Delta}{2|\boldsymbol{c}|} \geq \frac{\Delta}{2}$$

for Cliff as well as for Gap. If we have $\boldsymbol{c} \in R$ such that $|\boldsymbol{c}| \in [1 - \Delta, 1)$ for Cliff, then $|\boldsymbol{m}| \geq \Delta$ is necessary for a mutation to hit $B \cap A_{\boldsymbol{c}}$, and hence, also in such situations $\ell/(2|\boldsymbol{c}|) \geq \Delta/2$. Since Δ is bounden below by the constant $\delta > 0$, we have just shown that—whatever the length of $\boldsymbol{m}$ may be—either a mutation is to short (so that a success is precluded) or the (conditional) probability (given that $|\boldsymbol{m}| = \ell$; yet ℓ is arbitrary) that a mutant is at least as close to the optimum is already exponentially small (choose $\varepsilon := \delta/2$). This implies that the success probability of overcoming the gap/the cliff is also exponentially small.

A by-product of this result is that for Gap we have $\mathsf{P}\{\boldsymbol{c}' \in P \cap A_{\boldsymbol{c}} \mid \boldsymbol{c} \notin P\} = \exp(-\Omega(n))$ since a mutation hitting the plateau $P = \{\boldsymbol{x} \mid 1-\Delta \leq |\boldsymbol{x}| < 1\}$ is accepted iff a point $\boldsymbol{c}$ is mutated for which $|\boldsymbol{c}| \geq 2$, i. e., the search is facing a Cliff-like situation.

In the argumentation above we used that necessarily $\ell/(2|\boldsymbol{c}|) = \Omega(1)$ because $\Delta = \Omega(1)$. However, the term $(1 - \ell^2/(2|\boldsymbol{c}|)^2)^{(n-3)/2}$ is upper bounded by $\exp(-\Omega(n^{\varepsilon})$ for a constant $\varepsilon > 0$ also when $\Delta(n) = o(1)$ such that $\Delta(n) = \Omega(n^{\phi}/n)$ for a constant $\phi > 0$ (and yet $0 < \Delta \leq 1/3$).

Recall that for $\boldsymbol{c} \in R$ a spatial gain towards the optimum $\boldsymbol{o}$ of at least $\Delta = \Omega(n^{\phi}/n)$ is necessary for the mutant to hit B *and* $A_{\boldsymbol{c}}$, and that the spatial gain parallel to $\overrightarrow{\boldsymbol{c}\boldsymbol{o}}$ has to be at least as large as the one towards the optimum. As we have seen, the success probability is smaller than $\mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}|\} = O\big(\sqrt{n} \cdot (1-\alpha^2)^{(n-3)/2}\big)$ with $\alpha := \ell/(2|\boldsymbol{c}|)$. Since $(1 - 1/k)^k < 1/\mathrm{e}$ for $k > 1$, in fact $(1-\alpha^2)^{(n-3)/2}$ is exponentially small if $\alpha^2 = \Omega(n^{\varepsilon}/n)$ for some constant $\varepsilon > 0$ such that $0 \leq \alpha \leq 1$. Consequently, $\mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}|\}$ is exponentially small if $\ell = \alpha 2|\boldsymbol{c}| = \Omega(|\boldsymbol{c}| n^{\varepsilon/2}/\sqrt{n})$. In other words, the success probability is upper bounded by

$$\mathsf{P}\{G_{\ell} \geq \alpha \cdot \ell\} \quad = \quad O\Big(\sqrt{n} \cdot (1-\alpha^2)^{(n-3)/2}\Big)\,,$$

where $\alpha \in [0, 1]$ denotes an arbitrarily fixed relative (to ℓ) spatial gain parallel to some fixed direction. As a consequence, $\mathsf{P}\big\{G_{\ell} = \Omega(\ell\, n^{\delta}/\sqrt{n}\big\}$ is exponentially small for any constant $\delta > 0$.

We consider the case $\boldsymbol{c} \in R$ such that $|\boldsymbol{c}| \leq 2$. As shown above, if $|\boldsymbol{m}| = \Omega(|\boldsymbol{c}|\, n^{\varepsilon}/\sqrt{n})$ for a constant $\varepsilon > 0$, then $\mathsf{P}\{|\boldsymbol{c}'| \leq |\boldsymbol{c}|\}$ is exponentially small. Consequently, $\mathsf{P}\{\boldsymbol{c}' \in B \cap A_{\boldsymbol{c}} \mid \boldsymbol{c} \in R\}$ is exponentially small if $|\boldsymbol{m}| = \Omega(n^{\varepsilon}/\sqrt{n})$ for a constant $\varepsilon > 0$ since $|\boldsymbol{c}| = \Theta(1)$. On the other hand, the spatial gain (parallel to some fixed direction) of an isotropic mutation $\boldsymbol{m}$ is $\Omega(|\boldsymbol{m}| \cdot n^{\delta}/\sqrt{n})$ only with an exponentially small probability for any constant $\delta > 0$ as also shown above. Thus, if on the other hand $|\boldsymbol{m}| = O(n^{\varepsilon}/\sqrt{n})$, a spatial gain of $\Omega(n^{\varepsilon+\delta}/n)$ is realized only with exponentially small probability. By choosing $\varepsilon := \delta := \phi/2$, we obtain that $\mathsf{P}\{\boldsymbol{c}' \in B \cap A_{\boldsymbol{c}} \mid 1 - \Delta \leq |\boldsymbol{c}| \leq 2\}$ is exponentially small either way, i. e., whatever the distribution of $|\boldsymbol{m}|$ may be.

Finally, we consider the case $\boldsymbol{c} \in R$ such that $|\boldsymbol{c}| > 2$. In this case, a spatial gain towards the optimum greater than $|\boldsymbol{c}| - 1 > |\boldsymbol{c}|/2 = \Omega(|\boldsymbol{c}|)$ is necessary for $\boldsymbol{c}'$ to hit B. In other words, $\boldsymbol{c}' \in B$ implies $|\boldsymbol{m}| = \Theta(|\boldsymbol{c}|)$, and since a spatial gain of $\Omega(|\boldsymbol{m}|)$ is realized only with an exponentially small probability, $\mathsf{P}\{\boldsymbol{c}' \in B \mid |\boldsymbol{c}| > 2\}$ is exponentially small.

All in all, for $\Delta(n) = \Omega(n^{\phi}/n)$ with ϕ a positive constant (which also covers $\Delta(n) = \Omega(1)$), we have proved that the probability that an accepted mutation conquers the cliff is exponentially small, whatever the length of the mutation vector might be. Since the number of mutations that are necessary until a success is observed is geometrically distributed, the expected number of mutations/function evaluations a (1+λ) ES performs until $\boldsymbol{c} \in B$ when starting in R is exponential in n (if it is finite at all; depending on the adaptation mechanism, we might expect that B is never reached in a finite number of steps; this will be discussed later).

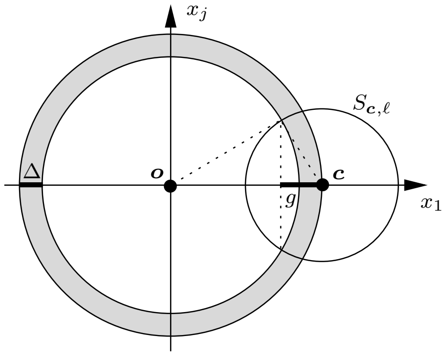
Theorem 1 *Let the (1+λ) ES minimize* $\textsc{Gap}^{\Delta}$ *or* $\textsc{Cliff}^{\Delta}$ *in* $\mathbb{R}^n$ *using isotropic mutations; let the initial search point be in* R*. Then the expected number of mutations/function evaluations until a point in the global-optimum region* B *is selected to become the current search point is exponentially large in* n *if* $\Delta(n) = \Omega(n^{\phi}/n)$ *for a positive constant* ϕ*; moreover, the probability that* $n^{O(1)}$ *mutations/function evaluations suffice is exponentially small in* n*.*

(That $n^{O(1)}$ mutations suffice only with an exponentially small probability is a direct consequence of a single mutation being successful only with an exponentially small probability.)

7 Polynomial vs. Superpolynomial

The estimations made in the preceding section seem too rough to find out when the best-case probability of overcoming the gap/cliff changes from polynomially small to superpolynomially small. In particular, we assumed that the spatial gain parallel to $\overrightarrow{\boldsymbol{co}}$ had to be at least Δ. The "parallel spatial gain" (on which G_ℓ bases) corresponding to a spatial gain of Δ towards the optimum, however, can be much larger than Δ. This may result in a poor bound, and therefore, the estimates must be made more precise.

We exemplarily consider the situation $|\boldsymbol{c}| = 1$. Let H denote the hyper-plane containing the $(n{-}1)$-sphere $\{\boldsymbol{x} \in \mathbb{R}^n \mid \operatorname{dist}(\boldsymbol{x}, \boldsymbol{c}) = \ell \wedge |\boldsymbol{x}| = |\boldsymbol{c}| - \Delta\}$, which is the intersection of the mutation-sphere $S_{\boldsymbol{c},\ell} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \operatorname{dist}(\boldsymbol{x}, \boldsymbol{c}) = \ell\}$ and the hyper-sphere $\{\boldsymbol{x} \in \mathbb{R}^n \mid |\boldsymbol{x}| = |\boldsymbol{c}| - \Delta\}$ consisting of all points that are exactly Δ closer to the minimizer than $\boldsymbol{c}$. Let g denote the distance between $\boldsymbol{c}$ and H, i. e. the "parallel spatial gain" corresponding to Δ. Then $\mathsf{P}\{G_\ell \geq g\}$ is an upper bound on the probability of conquering the cliff and equality holds if $\ell \leq 2|\boldsymbol{c}| - \Delta$. Assume $\boldsymbol{c}$ lies on the (positive halve of the) first axis; then the following figure depicts the situation (where $j \in \{2, \ldots, n\}$).



As our objective is an estimation of the success probability that is independent of the adaptation of the length of an isotropic mutation, we must assume that the success probability is maximal with respect to the relation of g to ℓ. Solving

$$(|\boldsymbol{c}| - \Delta)^2 - (|\boldsymbol{c}| - g)^2 \quad = \quad \ell^2 - g^2$$

for g, the "parallel gain" corresponding to Δ (cf. the figure), yields

$$g \;=\; \Delta + \frac{\ell^2 - \Delta^2}{2|\boldsymbol{c}|}$$

(if $|\boldsymbol{m}| < \Delta$, conquering the cliff is precluded, and thus, we can assume $\ell \geq \Delta$). Since $\mathsf{P}\{G_\ell \geq g\}$ is maximal if g/ℓ is minimal, we are interested in ℓ^* minimizing this ratio. As

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\ell}\frac{g}{\ell} &= \frac{\mathrm{d}}{\mathrm{d}\ell}\left(\frac{\Delta}{\ell} + \frac{\ell}{2|\boldsymbol{c}|} - \frac{\Delta^2}{\ell 2|\boldsymbol{c}|}\right) \\ &= \frac{-\Delta}{\ell^2} + \frac{1}{2|\boldsymbol{c}|} + \frac{\Delta^2}{\ell^2 2|\boldsymbol{c}|} \\ &= \frac{1}{2|\boldsymbol{c}|} - \frac{\Delta(2|\boldsymbol{c}| - \Delta)}{\ell^2 2|\boldsymbol{c}|}, \end{aligned}$$

solving $\frac{\mathrm{d}}{\mathrm{d}\ell}\, g/\ell = 0$ for ℓ yields that $|\boldsymbol{m}| = \ell^* := \sqrt{\Delta(2|\boldsymbol{c}| - \Delta)}$ results in maximum success probability. Since $\Delta < 2|\boldsymbol{c}| - \Delta$, we have $\ell^* \leq 2|\boldsymbol{c}| - \Delta$, and thus,

$$\max_\ell \mathsf{P}\{\boldsymbol{c}' \in B \cap A_{\boldsymbol{c}} \mid |\boldsymbol{c}| = 1, |\boldsymbol{m}| = \ell\} \;=\; \mathsf{P}\{G_{\ell^*} \geq g\}\,.$$

Since

$$\begin{aligned} \min_\ell \frac{g}{\ell} \;=\; \frac{g}{\ell^*} &= \frac{\Delta + \dfrac{(\ell^*)^2 - \Delta^2}{2|\boldsymbol{c}|}}{\ell^*} \\ &= \frac{\sqrt{\Delta(2|\boldsymbol{c}| - \Delta)}}{|\boldsymbol{c}|} \end{aligned}$$

and $0 < \Delta < |\boldsymbol{c}|$, in fact

$$\begin{aligned} \frac{\sqrt{\Delta|\boldsymbol{c}|}}{|\boldsymbol{c}|} &= \frac{\sqrt{\Delta(2|\boldsymbol{c}| - |\boldsymbol{c}|)}}{|\boldsymbol{c}|} \\ &< \frac{g}{\ell^*} \\ &< \frac{\sqrt{\Delta(2|\boldsymbol{c}| - 0)}}{|\boldsymbol{c}|} \\ &= \frac{\sqrt{\Delta 2|\boldsymbol{c}|}}{|\boldsymbol{c}|}, \end{aligned}$$

i. e., $g/\ell^* = \sqrt{\xi\Delta/|\boldsymbol{c}|}$ for some $\xi \in (1, 2)$. Hence,

$$\begin{aligned} \mathsf{P}\{G_{\ell^*} \geq g\} &= \frac{1}{\Psi_{n-2}}\int_{g/\ell^*}^{1} \left(1 - y^2\right)^{(n-3)/2} \mathrm{d}y \\ &= \Theta(\sqrt{n}) \int_{\sqrt{\xi\Delta/|\boldsymbol{c}|}}^{1} \left(1 - y^2\right)^{(n-3)/2} \mathrm{d}y. \end{aligned}$$

Since $\left(1 - y^2\right)^{(n-3)/2}$ is decreasing (for $0 < y < 1$), the right-most integral is in fact bounded from above by $(1 - \xi\Delta/|\boldsymbol{c}|)^{(n-3)/2}$ and becomes superpolynomially small if $\xi\Delta/|\boldsymbol{c}|$ is not $O(\log n/n)$ because $(1-1/k)^k < 1/\mathrm{e}$ for $k > 1$. Since $\xi \in (1,2)$ and $|\boldsymbol{c}| = 1$, this is the case if $\Delta(n)$ is not $O(\log n/n)$. Note that the (negative) degree of the polynomial depends on the disguised constant in the O-notation. On the other hand, for any $a \in (0, 1/2)$,

$$\int_a^{2a} \left(1-y^2\right)^{(n-3)/2} \mathrm{d}y \quad \geq \quad a \cdot \left(1-(2a)^2\right)^{(n-3)/2},$$

and hence, $\int_{\sqrt{\xi\Delta/|\boldsymbol{c}|}}^1 \left(1-y^2\right)^{(n-3)/2} \mathrm{d}y$ is bounded also below by a polynomial (of negative degree) for $\xi\Delta/|\boldsymbol{c}| = O(\log n/n)$. In shorter words, we have proved

$$1/\mathsf{P}\{G_{\ell^*} \geq g\} = \mathrm{poly}(n) \quad \Longleftrightarrow \quad \Delta(n) = O(\log n/n).$$

In the preceding argumentation we assumed $|\boldsymbol{c}| = 1$; for GAP this means that the search is located at a local but non-global minimum; for CLIFF this is neither a global nor a local minimum, yet it is a point where the probability of overcoming the cliff is maximum—given that $|\boldsymbol{m}|$ is optimally adapted/chosen: Obviously, the same argumentation holds for $1-\Delta \leq |\boldsymbol{c}| < 1$ since a candidate search point is in $B \cap A_{\boldsymbol{c}}$ only if the spatial gain towards the optimum is at least Δ; yet the ratio $\ell/|\boldsymbol{c}|$ is greater, resulting in a smaller success probability. For $|\boldsymbol{c}| = 1+\varepsilon$, $\varepsilon > 0$, the spatial gain towards the optimum has to be at least $\Delta+\varepsilon$ for the mutant to hit $B \cap A_{\boldsymbol{c}}$, and going through the same calculations again yields that, as expected, $g/\ell^*_{\Delta+\varepsilon}$ increases with ε. All in all, we have proved the following theorem.

Theorem 2 *Let the (1+λ) ES minimize* GAP^{Δ} *or* CLIFF^{Δ} *in* $\mathbb{R}^n$ *using isotropic mutations; let the initial search point be in* R*. Then the expected number of mutations/function evaluations until a point in the global-optimum region* B *is selected to become the current search point is polynomial in* n *only if* $\Delta(n) = O(\log n/n)$*, i. e. there exists a constant* κ *such that* $\Delta(n) \leq \kappa \cdot \log n/n$ *for* n *large enough; where the degree of the polynomial increases with* κ*.*

8 Discussion and Conclusion

Note that, as long as $\lambda = \mathrm{poly}(n)$, even if all λ mutations in a step of a (1+λ) ES could be performed in parallel, the expected number of steps is polynomial only if $\Delta(n) = O(\log n/n)$. However, remember that $\mathsf{P}\{G_{\ell^*} \geq g\}$ equals the *best-case* probability of overcoming the gap/cliff w. r. t. to $|\boldsymbol{m}|$, i. e., if $|\boldsymbol{m}|$ is adapted optimally. Consequently, the expected runtime of the (1+1) ES on $\mathrm{CLIFF}^{\Delta}_n$ is necessarily super-polynomial if $\Delta(n)$ is not $O(\log n/n)$, yet it is not necessarily polynomial if $\Delta(n) = O(\log n/n)$—in contrast to the (1+1) EA on JUMP_m in the search space $\{0,1\}^n$ when $m = O(1)$. This crucially depends on the mutation adaptation, and in particular, the 1/5-rule for mutation adaptation is supposed to fail even if Δ is chosen such that we obtain a lower bound that is $\Theta(n^2)$. The 1/5-rule shortens the (expected) length of the mutation vector (by a constant factor) after $\Theta(n)$ steps (unless 1/5 of the these steps have been successful). After the search approaches the gap/cliff, the success probability drops below 1/10, and hence, we expect at most 10% successful mutations in the following phase. By Chernoff bounds, with very high probability less than 1/5 of the steps in this phase are actually successful so that the step length is reduced by a constant factor. This is fatal because the success probability is also reduced by reducing the (expected) step length. Since the probability of overcoming the gap in this phase is only $1 - (1 - \Theta(1/n^2))^n = O(1/n)$, with high probability the gap is not overcome in this phase. In addition, in the following phase, the shortened step length results is a success probability that is even smaller, and so on. Already Rudolph (2001) proved this unsurprising "premature convergence" for a one-dimensional bimodal function. The modified 1/5-rule proposed by Greenwood and Zhu (2001) is supposed to overcome this deficiency. In these two works (and also in other works), however, merely convergence properties are investigated, i. e., the question is addressed whether the number of steps is finite or not (depending on the mode of stochastic convergence considered).

If one aims at an upper bound on the runtime w. r. t. the dimensionality of the search space, things become more complicated or, say, challenging—the lower bounds presented here may be taken as an indicator. In particular, the additional randomness introduced by σ-self-adaptation (σ-SA) may raise difficulties in a stochastic analysis. Besides, also σ-SA was reported to suffer from a "loss of step-length control" (the step length becomes too short too fast); Liang et al. (2001) propose two kinds of a so-called *dynamic lower bound scheme* to prevent the σ-SA from this loss of control.

Cliff might also serve as an example function for showing that a Metropolis-like selection might reduce the expected runtime significantly. Also a (2+1) ES using steady state intermediate recombination, or a (2+2) ES using 1-point cross-over might result in a significantly smaller number of Cliff-evaluations. As discussed in Jägersküpper and Witt (2005), also a (μ+1) ES using Rechenberg's 1/5-rule may manage to optimize Cliff: If the best individual in the population is very close to B and the worst one is at a distance greater than 1 from the optimum, a mutation of the best one hitting B makes it into the population.

For all these algorithms, showing that a search point in B is created and, moreover, accepted (to become the current individual/part of the current population) is no problem. The problem of a rigorous analysis is, however, to show that the B-individual remains in the population so that it keeps on evolving. In particular, the search could leave the global optimum region B again yet in the next step so that, hypothetically, the search could go back and forth between R and B forever.

All in all, these examples tell us that there is a variety of upper-bound problems to tackle.

Finally note that the method presented here is not restricted to Cliff and Gap but can be applied more generally: A vector $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{o}\}$ and a scalar $s \in \mathbb{R}$ define a hyper-plane $H := \{\boldsymbol{x} \mid \boldsymbol{x}^\top \boldsymbol{v} = s\} \subset \mathbb{R}^n$. Thus, $\boldsymbol{x} \in H \iff \boldsymbol{x}^\top \boldsymbol{v} = s$, and we let "$\boldsymbol{x} \le H$" and "$\boldsymbol{x} \ge H$" abbreviate $\boldsymbol{x}^\top \boldsymbol{v} \le s$ resp. $\boldsymbol{x}^\top \boldsymbol{v} \ge s$. (Note that switching from $\boldsymbol{v}$ to $-\boldsymbol{v}$ doesn't affect H, but merely turns things upside down.) So, for the probability of generating an acceptable mutant we obtain: If we can pick a separating hyper-plane H such that $A_{\boldsymbol{c}} \le H$ but $\boldsymbol{c} \ge H$, then the (conditional) probability of hitting the lower level set of $\boldsymbol{c}$ in a step in which $|\boldsymbol{m}| = \ell$ is upper bounded by $\mathsf{P}\{G_\ell \ge \operatorname{dist}(\boldsymbol{c}, H) \mid |\boldsymbol{m}| = \ell\}$. As this probability increases with ℓ, we additionally need an upper bound on ℓ. Without further assumptions, $\ell_{\sup} := \sup_{\boldsymbol{x} \in A_{\boldsymbol{c}}} \operatorname{dist}(\boldsymbol{x}, \boldsymbol{c})$ is the best upper bound on ℓ, and if $\ell_{\sup} < \infty$, then (by the law of total probability) $\mathsf{P}\big\{G_{\ell_{\sup}} \ge \operatorname{dist}(\boldsymbol{c}, H)\big\}$

is an upper bound on $\mathsf{P}\{\boldsymbol{c}+\boldsymbol{m} \in A_{\boldsymbol{c}}\}$. However, if we can additionally pick a hyper-ball that contains $A_{\boldsymbol{c}}$, then we can use the result on the optimal length ℓ^* presented above—resulting in a much better bound if $\ell^* \ll \ell_{\text{sup}}$.

Naturally, we can replace $A_{\boldsymbol{c}}$, the lower level set of the search point $\boldsymbol{c}$ w. r. t. the function to be minimized, by any other (measurable) set M. Then we obtain an upper bound on the "hitting probability" $\mathsf{P}\{\boldsymbol{c}' \in M\}$ rather than on the probability of generating a mutant lying in the lower level set. The investigation of CLIFF and GAP given above follows the approach just described by choosing $M := B \cap A_{\boldsymbol{c}}$, and thus, CLIFF and GAP are just two examples in the class of functions that are covered.

References

Beyer, H.-G. (2001): *The Theory of Evolution Strategies.* Springer.

Droste, S., Jansen, T., Wegener, I. (2001): *Dynamic parameter control in simple evolutionary algorithms.* In *Foundations of Genetic Algorithms 6 (FOGA 2000)*, 275–294, Morgan Kaufmann, San Francisco.

Droste, S., Jansen, T., Wegener, I. (2002): *On the analysis of the (1+1) evolutionary algorithm.* Theoretical Computer Science, 276:51–82.

Greenwood, G. W., Zhu, Q. J. (2001): *Convergence in evolutionary programs with self-adaptation.* Evolutionary Computation, 9(2):147–157.

Hansen, N., Ostermeier, A. (1996): *Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation.* In *Proceedings of the IEEE Int'l Conference on Evolutionary Computation (ICEC)*, 312–317.

Jägersküpper, J. (2003): *Analysis of a simple evolutionary algorithm for minimization in Euclidean spaces.* In *Proceedings of the 30th Int'l Colloquium on Automata, Languages and Programming (ICALP)*, volume 2719 of *LNCS*, 1068–1079, Springer.

Jägersküpper, J. (2005): *Rigorous runtime analysis of the (1+1) ES:* 1/5*-rule and ellipsoidal fitness landscapes.* In *Proceedings of Foundations of Genetic Algorithms 8 (FOGA 2005)*, volume 3469 of *LNCS*, Springer, to appear.

Jägersküpper, J., Witt, C. (2005): *Rigorous runtime analysis of a (μ+1) ES for the sphere function.* In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, ACM, to appear.

Liang, K., Yao, X., Newton, C. S. (2001): *Adapting self-adaptive parameters in evolutionary algorithms.* Applied Intelligence, 15(3):171–180.

Mühlenbein, H. (1992): *How genetic algorithmis really work: Mutation and hillclimbing.* In *Parallel Problem Solving from Nature 2 (PPSN)*, 15–25, North-Holland, Amsterdam.

Rechenberg, I. (1973): *Evolutionsstrategie*. Frommann-Holzboog, Stuttgart, Germany.

Rudolph, G. (2001): *Self-adaptive mutations may lead to premature convergence*. IEEE Transactions on Evolutionary Computation, 5(4):410–414.

Schwefel, H.-P. (1995): *Evolution and Optimum Seeking*. Wiley, New York.

Yao, X., Liu, Y., Lin, G. (1999): *Evolutionary programming made faster*. IEEE Transactions on Evolutionary Computation, 3(2):82–102.