Drift Theory in Continuous Search Spaces: Expected Hitting Time of the (1+1)-ES with 1/5 Success Rule

Youhei Akimoto
Department of Electrical and Electronic Engineering, Shinshu University
4-17-1 Wakasato, Nagano, Japan
y_akimoto@shinshu-u.ac.jp

Anne Auger
RandOpt Team, INRIA Saclay; CMAP, Ecole Polytechnique
Route de Saclay, Ile-de-France, France
anne.auger@inria.fr

Tobias Glasmachers
Institute for Neural Computation, Ruhr-University Bochum
Universitätsstr. 150, Bochum, Germany
tobias.glasmachers@ini.rub.de

Abstract

This paper explores the use of the standard approach for proving runtime bounds in discrete domains—often referred to as drift analysis—in the context of optimization on a continuous domain. Using this framework we analyze the (1+1) Evolution Strategy with one-fifth success rule on the sphere function. To deal with potential functions that are not lower-bounded, we formulate novel drift theorems. We then use the theorems to prove bounds on the expected hitting time to reach a certain target fitness in finite dimension d. The bounds are akin to linear convergence. We then study the dependency of the different terms on d proving a convergence rate dependency of $\Theta(1/d)$. Our results constitute the first non-asymptotic analysis for the algorithm considered as well as the first explicit application of drift analysis to a randomized search heuristic with continuous domain.

1 Introduction

The standard methodology for proving runtime bounds of evolutionary algorithms defined on a discrete search space is often referred to as drift analysis. It consists in proving a drift condition, e.g. expected drift w.r.t. a potential strictly smaller than $c < 0$, that directly translates into a bound on the hitting time to reach the optimum. It allows to decouple generic mathematical arguments, summarized in drift theorems, from arguments specific to the algorithm. With drift analysis, proofs that could take several pages before have been simplified considerably [11, 5, 4, 12, 13].

In this work we explore the utility of such an approach for the analysis of algorithms operating on a continuous domain. For this purpose we focus on the analysis of the $(1+1)$-ES with one-fifth success rule on the sphere function $f : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \|x\|^2$. We are particularly interested in benefits of drift analysis over current tools for analyzing continuous randomized search heuristics, like investigating stability of Markov chains.

The (1+1)-ES We focus here on one of the simplest adaptive algorithms, namely the $(1+1)$ evolution strategy (ES) with one-fifth success rule [14]. It is defined in algorithm 1, where we assume *minimization* of a function $f : \mathbb{R}^d \to \mathbb{R}$. The state of the algorithm at iteration t is $(m_t, \sigma_t) \in \mathbb{R}^d \times \mathbb{R}_{>0}$, where m_t is the mean of the Gaussian sampling distribution and also the best solution found so far, and σ_t is the standard deviation of the distribution or "step-size" that controls the distance at which novel solutions are sampled. This variant of the algorithm, which was fist proposed in [10], implements Rechenberg's idea of maintaining a probability of success of roughly 1/5. This algorithm is not a "toy" algorithm as it features the important flavor of the widely used state-of-the-art CMA-ES [6], namely adaptation of the sampling distribution.

Algorithm 1: (1+1)-ES with 1/5-success rule

input $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $f : \mathbb{R}^d \to \mathbb{R}$, **parameter** $\alpha > 0$
for $t = 1, 2, \ldots$, *until stopping criterion is met* **do**
  sample $x_t \sim m_t + \sigma_t \mathcal{N}(0, I)$
  if $f(x_t) \leq f(m_t)$ **then**
    $m_{t+1} \leftarrow x_t$ ▷ move to the better solution
    $\sigma_{t+1} \leftarrow \sigma_t \cdot e^{\alpha}$ ▷ increase the step size
  else
    $m_{t+1} \leftarrow m_t$ ▷ stay where we are
    $\sigma_{t+1} \leftarrow \sigma_t \cdot e^{-\alpha/4}$ ▷ decrease the step size

Drift Analysis in $\mathbb{R}^d$ Interestingly, although drift theorems are often formulated for finite domains, they naturally generalize to continuous domains [12, 13]. To date however, drift analysis in the style of discrete domains has not been *explicitly* applied to analyze continuous algorithms. Note that drift conditions are also central in other approaches addressing convergence in continuous domains, while they are typically not used for obtaining bounds on the hitting time (see below). At the same time, we will see that some difficulties can arise when dealing with continuous search spaces as it seems natural to use a potential function that converges to minus infinity when approaching the optimum. To overcome those problems we formulate novel drift theorems.

Analyzing state-of-the art continuous evolutionary algorithms means analyzing adaptive algorithms. While this adaptation is the key for the practical success of ES (ensuring linear convergence on wide classes of problems, similar to gradient-based methods on strongly convex functions), in turn it makes the analysis difficult. Indeed, when σ_t is too small compared to $\|m_t\|$, the progress towards the optimum is very small. This complicates the task of finding a suitable potential function and proving a drift condition.

When analyzing algorithms in continuous domains, our goals are (i) to establish how fast the algorithm converges for a fixed dimension d (usually linear convergence), and (ii) to investigate the dependency of the convergence rate on the search space dimension (usually $\Theta(1/d)$)—this is different from discrete domains where the optimum can be located in finite time. In terms of hitting time to reach a certain precision ϵ, property (i) means that for all $\epsilon > 0$ the expected hitting time is finite and proportional to $\log(\|m_0\|) - \log(\epsilon)$, while property (ii) means that it is also proportional to d.

Related work Some of the drift methodology is underlying many results of J. Jägersküpper [7, 8, 9]. Drift is not uncovered explicitly in these works, which makes it arguably difficult to follow the analysis carried out. That might be the reason why nobody built so far on Jägersküpper's impressive work. We also have to point out that Algorithm 1 differs from the variant analyzed by Jägersküpper, where the step-size is kept fixed for several iterations.

For a fixed dimension, the linear convergence of the algorithm on scaling-invariant functions—including in particular the sphere function—has been shown using Markov chain analysis [2]. This analysis is asymptotic in nature and does not provide a dependency of the convergence rate on the dimension. The difficult part in the approach also relies on proving a drift condition, that should however hold only outside a compact set, not on the whole domain.

Drift of a step size adaptive algorithm is also analyzed in [3], the only prior work that uses a potential function in a continuous domain. That approach remains very restricted, applying only to symmetric functions of a single variable.

In this work, we go beyond the state-of-the-art as follows. Other than Jägersküpper's results, our bounds provide (non-asymptotic) constants, and they hold with full probability. In contrast to Markov chain analysis, we obtain a dependency in the dimension and non-asymptotic results. Compared to [3], we go beyond a proof of concept by analyzing a simple yet realistic algorithm.

Outline The rest of the paper is organized as follows. In the next section we introduce novel drift theorems for lower and upper bounds to deal with unbounded potentials, since this is a natural design in our context. In Section 3, we prove technical results needed to derive the drift condition for the upper bound. In Section 4 we define our potential function and show two drift conditions for the lower and the upper bound. By applying the drift theorems to the drift conditions we derive lower and upper bounds on the first hitting time, corresponding to linear convergence with $\Theta(1/d)$ scaling of the convergence rate. For the sake of readability, all proofs are in the appendix.

Notation A multivariate normal distribution is denoted $\mathcal{N}(0, I)$. With Φ_1 we denote the cumulative density function of the standard normal distribution $\mathcal{N}(0, 1)$ on $\mathbb{R}$, and φ_d is the pdf of the standard normal distribution on $\mathbb{R}^d$. The indicator function of a set or condition C is denoted by $\mathbf{1}_{\{C\}}$.

2 Additive Drift on an Unbounded Domain

In the continuous setting considered in this paper, we aim at proving a runtime bound that translates into linear convergence. Linear convergence is typically pictured as the log of the distance to the optimum converging to minus infinity like $-\mathrm{CR} \times t$ with $\mathrm{CR} > 0$. It is thus natural to construct a potential function that involves the log of the distance to the optimum. Yet, this means that the potential function can take values that are arbitrarily negative, while in drift theorems it is typical to assume that the potential function is lower bounded (by zero or one). For this reason we need to adapt existing drift theorems.

We adopt the following formalism. Let $\{X_t : t \geq 0\}$ be a sequence of real-valued random variables adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$. In our typical setting X_t can be homogeneous to the logarithm of the distance to the optimum and thus go to minus infinity when linear convergence occurs. Additionally, from one iteration to the next, X_{t+1} can be arbitrarily much smaller than X_t. This happens if by chance we have made an atypically good step that improves the current solution a lot.

While arbitrarily good steps should be *helpful* in the sense of making the hitting time only smaller, we face the technical difficulty to distinguish this situation from the following scenario: assume a process X_t with an average decrease of -1, i.e., fulfilling $\mathbb{E}[X_{t+1}|\mathcal{F}_t] - X_t \leq -1$, but where X_{t+1} equals X_t with probability $1-p$, and with probability $p \ll 1$ we jump to $X_{t-1} = X_t - 1/p$, possibly overjumping the target in a single but very improbable step. The time needed to sample this jump is geometrically distributed with expectation $1/p$, resulting in an arbitrarily large hitting time. This small example illustrates that controlling only the expected drift is not enough for bounding the expected hitting time. If the domain is bounded from below then the size of a possible jump is also bounded, avoiding this difficulty. Therefore we have to find a way of controlling extreme events.

To circumvent this problem, instead of controlling directly the drift on X_t we will control the drift of a process with truncated and hence bounded single-step progress. More precisely, for given $\mathrm{A} > 0$ we consider the truncated process defined iteratively as $Y_0^{\mathrm{A}} = X_0$ and

$$Y_{t+1}^{\mathrm{A}} = Y_t^{\mathrm{A}} + \max\Big\{X_{t+1} - X_t, -\mathrm{A}\Big\} \ , \tag{1}$$

where progress (towards minus infinity) larger than $-\mathrm{A}$ is cut. By construction (almost surely[1])

$$Y_{t+1}^{\mathrm{A}} - Y_t^{\mathrm{A}} \geq -\mathrm{A} \ , \tag{2}$$

$$X_t \leq Y_t^{\mathrm{A}} \ , \tag{3}$$

where the latter equation holds as indeed $X_t = X_0 + \sum_{k=0}^{t-1}(X_{k+1} - X_k) \leq Y_0^{\mathrm{A}} + \sum_{k=0}^{t-1} \max\{(X_{k+1} - X_k), -\mathrm{A}\} = Y_0^{\mathrm{A}} + \sum_{k=0}^{t-1}(Y_{k+1}^{\mathrm{A}} - Y_k^{\mathrm{A}}) = Y_t^{\mathrm{A}}$.

As a direct consequence of inequality (3), for $\beta \in \mathbb{R}$, the hitting time $T_\beta^X = \min\{t : X_t \leq \beta\} \in \mathbb{N} \cup \{\infty\}$ of X_t to reach $(-\infty, \beta]$ is upper bounded by the hitting time $T_\beta^{Y^{\mathrm{A}}} = \min\{t : Y_t^{\mathrm{A}} \leq \beta\}$ of Y_t^{A} to reach $(-\infty, \beta]$, i.e., $T_\beta^X \leq T_\beta^{Y^{\mathrm{A}}}$. Hence an upper bound on the hitting time of Y_t^{A} results in an upper bound on the hitting time of X_t. Exploiting this idea, we derive an upper bound on the hitting time T_β^X in the following theorem based on bounding the expected drift of the truncated process $\{Y_t^{\mathrm{A}} : t \in \mathbb{N}\}$.

Theorem 1 (Upper bound via drift on truncated process)**.** *Let $\{X_t : t \geq 0\}$ be a sequence of real-valued random variables adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with $X_0 = x_0 \in \mathbb{R}$. For $\beta < x_0$ let $T_\beta^X = \min\{t : X_t \leq \beta\}$ be the first hitting time of the set $(-\infty, \beta]$. If there exist $\mathrm{A}, \mathrm{B} > 0$ such that Y_t^{A} is integrable, i.e. $\mathbb{E}\left[\left|Y_t^{\mathrm{A}}\right|\right] < \infty$, and*

$$\mathbb{E}\Big[Y_{t+1}^{\mathrm{A}} \,\Big|\, \mathcal{F}_t\Big] - Y_t^{\mathrm{A}} = \mathbb{E}\Big[\max\{X_{t+1} - X_t, -\mathrm{A}\} \,\Big|\, \mathcal{F}_t\Big] \leq -\mathrm{B} \ , \tag{4}$$

then the expectation of T_β^X satisfies

$$\mathbb{E}\left[T_\beta^X\right] \leq \mathbb{E}\left[T_\beta^{Y^{\mathrm{A}}}\right] \leq \frac{x_0 - \beta + \mathrm{A}}{\mathrm{B}} \ . \tag{5}$$

Remark 1. *A drift on the truncated process Y_t^{A} also gives a drift on X_t. Indeed, assume $\mathbb{E}[|X_t|] < +\infty$. Since*

$$X_{t+1} - X_t \leq \max\{X_{t+1} - X_t, -\mathrm{A}\} = Y_{t+1}^{\mathrm{A}} - Y_t^{\mathrm{A}} \ ,$$

[1] We use *almost surely* although the property is deterministic, simply to disambiguate from *in distribution* and *in expectation*.

if inequality (4) *is satisfied then it holds*

$$\mathbb{E}[X_{t+1}|\mathcal{F}_t] - X_t \leq -\mathrm{B} \ . \tag{6}$$

The next proposition ensures that the integrability of the truncated process is implied by the integrability of $\{X_t : t \geq 0\}$.

Proposition 1 (Integrability of the truncated process)**.** *If a process* $\{X_t : t \geq 0\}$ *is integrable, i.e.,* $\mathbb{E}[|X_t|] < \infty$*, then its truncated process* $\{Y_t^{\mathrm{A}} : t \geq 0\}$ *defined in equation* (1) *is integrable as well.*

Our lower bound also relies on an unbounded potential function. Typical drift theorems for establishing lower bounds assume that the potential is bounded and hence cannot be applied directly [9]. Instead we use the following theorem, the proof of which can be seen as a reformulation of the arguments used in [8, Theorem 2] as a drift theorem. It generalizes [9, Lemma 12]. Note that due to the more general setting we lose a (bearable) factor of four in the bound.

Theorem 2. *Let* X_t *be integrable and adapted to* $\mathcal{F}_t$ *such that*

$$X_0 = x_0 \quad \text{and} \quad \mathbb{E}[X_{t+1} \,|\, \mathcal{F}_t] - X_t \geq -\mathrm{C}$$

for $\mathrm{C} > 0$*. For* $\beta < x_0$ *we define* $T_\beta^X = \min\{t : X_t \leq \beta\}$*. Then the expected hitting time is lower bounded by*

$$\mathbb{E}\left[T_\beta^X\right] \geq \frac{x_0 - \beta}{4\mathrm{C}} - \frac{1}{2} \ .$$

3 Probability of Successes with Positive Progress Rate

In this section we derive properties of the success probability that will be central for establishing the drift condition for the upper bound. For an improvement rate $r \in [0,1)$ and $x \sim \mathcal{N}(m, \sigma^2 \mathrm{I})$ in $\mathbb{R}^d$ we define the success probability with rate r given (m, σ) as

$$p_{r,d}^{\mathrm{succ}}(m,\sigma) = \Pr_{x \sim \mathcal{N}(m,\sigma^2 \mathrm{I})}\Big(\|x\| < (1-r)\cdot\|m\|\Big)$$

i.e. as the probability that the norm of the offspring is smaller than $(1-r)\|m\|$. As a consequence of the isotropy of the multivariate normal distribution, this success probability equals

$$p_{r,d}^{\mathrm{succ}}(m,\sigma) = \Pr\left(\left\|e_1 + \frac{\sigma}{\|m\|}\mathcal{N}\right\| < (1-r)\right)$$

where $e_1 = (1, 0, \ldots, 0)$ and $\mathcal{N}$ is a standard normally distributed vector. This latter equation reveals that the probability of success with improvement rate r is a function of $\sigma/\|m\|$. Let us introduce the normalized step size $\bar{\sigma} = d \cdot \sigma/\|m\|$ and define

$$p_{r,d}^{\mathrm{succ}}(\bar{\sigma}) := \Pr\left(\left\|e_1 + \frac{\bar{\sigma}}{d}\mathcal{N}\right\| < (1-r)\right) \ , \tag{7}$$

then $p_{r,d}^{\mathrm{succ}}(\bar{\sigma}) = p_{r,d}^{\mathrm{succ}}(m,\sigma)$. For $r = 0$ we recover the "classic" probability of success

$$p_{0,d}^{\mathrm{succ}}(\bar{\sigma}) := \Pr\left(\left\|e_1 + \frac{\bar{\sigma}}{d}\mathcal{N}\right\| < 1\right) \ .$$

The success probability function is illustrated in Figure 1.

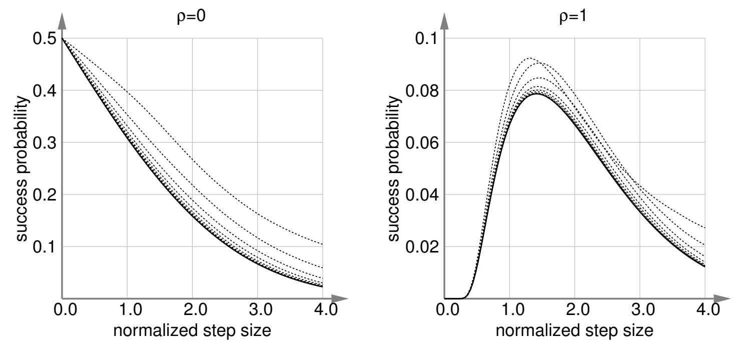

Figure 1: The success probability function $p_{d,r}^{\mathrm{succ}}(\bar{\sigma})$ for $r \cdot d = \rho = 0$ (left) and $r \cdot d = \rho = 1$ (right). The solid curves depict $p_\rho^{\mathrm{succ}}(\bar{\sigma})$, while the dotted curves are $p_{r,d}^{\mathrm{succ}}(\bar{\sigma})$ with $r \cdot d = \rho$ for $d \in \{2, 4, 8, 16, 32, 64, 128, 256\}$. The curves for high dimensions are indistinguishable from the limit curve.

We start by proving that the function $\bar{\sigma} \to p_{r,d}^{\mathrm{succ}}(\bar{\sigma})$ is continuous, and for $r = 0$ it is monotonically decreasing and hence bijective. This is formalized in the following lemma:

Lemma 1.

1. *For all* $d \in \mathbb{N}$ *and* $r \in [0,1)$*,* $\bar{\sigma} \to p_{r,d}^{succ}(\bar{\sigma})$ *is positive and continuous.*
2. *For* $r = 0$ *it is strictly monotonically decreasing and thus bijective.*
3. *For all* $d \in \mathbb{N}$*, the image of* $\bar{\sigma} \to p_{0,d}^{succ}(\bar{\sigma})$ *is* $(0, 1/2)$*.*

We now investigate the asymptotic limit of the function $\bar{\sigma} \to p_{r,d}^{\mathrm{succ}}(\bar{\sigma})$ for d to infinity.

Lemma 2. *For* $r = r(d)$ *fulfilling* $\lim_{d\to\infty} d \cdot r(d) = \rho$ *the limit* $p_\rho^{succ}(\bar{\sigma}) := \lim_{d\to\infty} p_{r,d}^{succ}(\bar{\sigma})$ *exists, and it equals* $p_\rho^{succ}(\bar{\sigma}) = \Phi_1\left(-\frac{\rho}{\bar{\sigma}} - \frac{\bar{\sigma}}{2}\right)$*. For* $\rho = 0$*, the function* p_0^{succ} *is continuous and strictly monotonically decreasing and the image of* p_0^{succ} *is* $(0, 1/2)$*.*

For $\rho = 0$ we recover the known result that the asymptotic limit of the probability of success (for $r = 0$) equals $\Phi_1(-\sigma/2)$ [1]. The above lemma captures the intuition that success is maximized with a small step size (for $\rho = 0$, p_ρ^{succ} is maximal for $\bar{\sigma} \to 0$), while a non-trivial step size ($\bar{\sigma} > 0$) is needed for making significant progress ($\rho > 0$).

4 Potential and Drift

In this section we define a potential function $V(\theta_t)$ that defines the unbounded and untruncated process from Section 2. First we establish that it satisfies the conditions of Theorem 1. Then we prove a drift condition for the lower bound. Finally we apply the drift theorems to obtain lower and upper bounds for the first hitting time of the (1+1)-ES. Our goal is to establish lower and upper bounds on the expected first hitting time of $\log(\|m_t\|)$ to the set $(-\infty, \beta]$, where $\beta = \log(\epsilon)$ is the logarithm of the target distance ϵ to the optimum. Linear or geometric convergence of (1+1)-ES—that is what we observe in simulation and what Jägersküpper found in his analysis with overwhelming probability—is implied if $\log(\|m_t\|)$ decreases at a linear rate towards $-\infty$. The potential function $V(\theta_t)$ will be chosen so that its first hitting time gives an upper bound on the first hitting time of $\log(\|m_t\|)$.

4.1 Potential Function

We fix two probabilities p_u and p_ℓ such that $0 < p_u < 1/5 < p_\ell < 1/2$. Since the probability of success function $\bar{\sigma} \mapsto p^{\text{succ}}_{0,d}(\bar{\sigma})$ with rate $r = 0$ is bijective (see Lemma 1), we know that there exist u and ℓ such that $p^{\text{succ}}_{0,d}(u) = p_u$ and $p^{\text{succ}}_{0,d}(\ell) = p_\ell$. We assume that p_u and p_ℓ are chosen such that $u/\ell \geq \alpha^{5/4}$. Given these parameters, we define the potential function

$$V(\theta) = V(m, \sigma) = \log\left(\|m\|\right) + \max\left\{0\,,\ v \cdot \log\left(\frac{\alpha \cdot \ell \cdot \|m\|}{d \cdot \sigma}\right)\,,\ v \cdot \log\left(\frac{\alpha^{1/4} \cdot \sigma \cdot d}{u \cdot \|m\|}\right)\right\} \tag{8}$$

with coefficient $v > 0$ to be determined later. The potential function consists of three parts. The term $\log(\|m\|)$ measures optimization progress: when approaching the optimum, it decays to $-\infty$. The other terms become positive and hence active only if the step size is not well adapted. The second term in the maximum kicks in if σ is "too small", and the third term turns positive if σ becomes "too large". Hence the potential combines two ways of making progress, namely approaching the optimum and adapting the step size towards a regime where the (1+1)-ES can make significant optimization progress. The parameter v relates these two types of progress by putting them on the same scale.

Lemma 3. *It holds $\mathbb{E}\big[|V(\theta_t)|\big] < \infty$. In other words, $V(\theta_t)$ is integrable for each $t \in \mathbb{N}$. Moreover, for all $\mathrm{A} > 0$ the truncated process Y^{A}_t defined in equation (1) with $X_t = V(\theta_t)$ is integrable for each $t \in \mathbb{N}$.*

4.2 Truncated Drift

In the following, we prove that $V(\theta_t)$ satisfies the prerequisites of Theorem 1. First we prove a proposition with a range of possible choices for the constants A and v. We then show in Proposition 3 how to set those constants to obtain the right scaling with respect to d for the hitting time.

Proposition 2. *Consider optimization of the sphere function $f : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \|x\|^2$ with the (1+1)-ES. If the parameters v and A fulfill $0 < v < \min\{1, 4\mathrm{A}/\log(\alpha)\}$ then the potential function V defined in eq. (8) fulfills*

$$\mathbb{E}\left[\max\{V(\theta_{t+1}) - V(\theta_t)\,,\, -\mathrm{A}\} \mid \theta_t\right] \leq -\mathrm{B}\ , \tag{9}$$

with

$$\mathrm{B} = \min\left\{\mathrm{A} \cdot p^* - \frac{5}{4} \cdot v \cdot \log(\alpha),\ v \cdot \log(\alpha) \cdot \left(\frac{5p_\ell - 1}{4}\right), v \cdot \log(\alpha) \cdot \left(\frac{1 - 5p_u}{4}\right)\right\} \tag{10}$$

and $p^* = \min\limits_{\bar{\sigma} \in [\ell, u]} \left\{p^{succ}_{1 - \exp\left(-\frac{\mathrm{A}}{1-v}\right), d}(\bar{\sigma})\right\}$.

The previous proposition is the core component establishing the drift of the truncated process. The next proposition shows how to arrange the parameters so that the speed of the drift scales as desired in the limit of large dimensions.

Proposition 3. *Consider $d \geq 2$. For $\mathrm{A} = \frac{1}{d}$ and $v = \frac{p'}{2 \cdot d \cdot \log(\alpha)}$ with $p' = \min\limits_{\bar{\sigma} \in [\ell, u]} \left\{p^{succ}_{r', d}(\bar{\sigma})\right\}$ and $r' = 1 - \exp\left(-\frac{\log(\alpha)}{d \cdot \log(\alpha) - 1}\right)$ it holds $\mathrm{B} > 0$ and $\mathrm{B} \in \Theta(1/d)$.*

Proposition 3 implies that the expected truncated drift (9) is of order $\Omega(1/d)$.

4.3 Hit-and-Run

A very general lower bound on the expected first hitting time was established by Jägersküpper. His argumentation in [8, Theorem 2] is based on the hit-and-run algorithm. Here we use a similar approach for proving the lower bound. In iteration t, given a mutation direction $\delta_t = x_t - m_t$ (with the notation of algorithm 1), the hit-and-run algorithm selects the optimal length of δ_t maintaining its direction and produces the offspring $x^*_t = m_t + \gamma^* \delta_t$ with $\gamma^* = \arg\min_\gamma f(m_t + \gamma \cdot \delta_t)$. By construction, the progress of the hit-and-run method upper bounds the progress of the (1+1)-ES. Using the same realization for the Gaussian vector creating the offspring x_t (see Algorithm 1), we indeed have:

$$\log(\|m_t\|) - \log(\|m_{t+1}\|) \leq \log(\|m_t\|) - \log(\|x^*_t\|)\ . \tag{11}$$

The log-progress of the hit-and-run on the sphere is

$$\log\left(\|m_t\|\right) - \log\left(\|x^*_t\|\right) = -\log\left(\Big\|e_1 + (\gamma^*/\|m_t\|)\delta_t\Big\|\right)\ .$$

In the next lemma, we bound the expectation of its progress.

Lemma 4. *For $d \geq 2$, the expected log progress of the hit-and-run algorithm is upper bounded by $1/d$.*

Using inequality (11) we find that the expected log progress of the (1+1)-ES is upper-bounded by $1/d$.

$$\mathbb{E}[\log(\|m_t\|) - \log(\|m_{t+1}\|)|\mathcal{F}_t] \leq \frac{1}{d}. \tag{12}$$

4.4 Bounds on the First Hitting Time

Finally, all preparations are in place and we can reap the fruit of our labor, which are formulated in the following theorem. To this end, let $T_\epsilon = \min\{t : \|m_t\| \leq \epsilon\}$ be the first hitting time of $(-\infty, \log(\epsilon)]$ by $\log(\|m_t\|)$, where m_t is defined in Algorithm 1.

Theorem 3. *The expected first hitting time of the (1+1)-ES starting from $\theta_0 = (m_0, \sigma_0)$ on the sphere function $f(x) = \|x\|^2$ fulfills*

$$\frac{\big(\log(\|m_0\|) - \log(\epsilon)\big) \cdot d}{4} - \frac{1}{2} \leq \mathbb{E}[T_\epsilon] \leq \frac{V(\theta_0) - \log(\epsilon) + \frac{1}{d}}{\mathrm{B}}$$

with $V(\theta)$ defined in eq. (8) *and* B *given in eq.* (10)*. With the choice of constants A and v given in Proposition 3, it is hence of the form*

$$\mathbb{E}[T_\epsilon] \in \Theta\Big(\big(\log(\|m_0\|) + \log(1/\epsilon)\big) \cdot d\Big) \ . \tag{13}$$

The asymptotic form (13) of the expected first hitting time implies (i) that the process is akin to linear convergence due to the term $\log(1/\epsilon)$, and (ii) a convergence rate of the form $\Theta(1/d)$ due to the factor d in the expected hitting time.

5 Discussion and Conclusion

We have established the first non-asymptotic runtime bound for the first hitting time of the (1+1)-ES with one-fifth success rule (Algorithm 1) on the sphere function. Our proof is based on a global drift condition, a generic approach that has proven invaluable for the analysis of discrete algorithms. Our work shows that such approaches are a promising tool also for continuous domains. As usual in drift analysis, constructing the potential function and establishing drift conditions makes up the lion's share of the efforts. In this sense, our drift theorems merely add convenience.

Establishing a drift condition is simplified in the stability analysis of the underlying Markov chain, since drift is needed only outside a compact set, i.e., for very small and very large normalized step size $\bar{\sigma}$. On the other hand, the current analysis is non-asymptotic and provides estimates of the convergence rate as a function of the problem dimension.

Jägersküpper established similar results already more than a decade ago, when drift analysis was only in its infancy. His results are hard to follow from a modern perspective. We improve on his work by proving non-asymptotic bounds for finite dimensions.

Acknowledgement We gratefully acknowledge support by Dagstuhl seminar 17191 "Theory of Randomized Search Heuristics". We would like to thank Per Kristian Lehre, Carsten Witt, and Johannes Lengler for valuable discussions and advice on drift theory.

References

[1] Anne Auger, Dimo Brockhoff, and Nikolaus Hansen. Analyzing the impact of mirrored sampling and sequential selection in elitist evolution strategies. In *Proceedings of the 11th Workshop Proceedings on Foundations of Genetic Algorithms*, FOGA '11, pages 127–138, New York, NY, USA, 2011. ACM.

[2] Anne Auger and Nikolaus Hansen. Linear convergence on positively homogeneous functions of a comparison based step-size adaptive randomized search: the (1+1) ES with generalized one-fifth success rule. *CoRR*, abs/1310.8397, 2013.

[3] Claudia R. Correa, Elizabeth F. Wanner, and Carlos M. Fonseca. Lyapunov design of a simple step-size adaptation strategy based on success. In *Parallel Problem Solving from Nature (PPSN)*, pages 101–110, 2016.

[4] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative drift analysis. *Algorithmica*, 64(4):673–697, December 2012.

[5] Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276(1):51 – 81, 2002.

[6] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[7] Jens Jägersküpper. Analysis of a simple evolutionary algorithm for minimization in Euclidean spaces. *Automata, Languages and Programming*, pages 188–188, 2003.

[8] Jens Jägersküpper. How the (1+1)-ES using isotropic mutations minimizes positive definite quadratic forms. *Theoretical Computer Science*, 361(1):38–56, 2006.

[9] Jens Jägersküpper. Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. *Theoretical Computer Science*, 379(3):329–347, 2007.

[10] S. Kern, S. D. Müller, N. Hansen, D. Büche, J. Ocenasek, and P. Koumoutsakos. Learning probability distributions in continuous evolutionary algorithms–a comparative review. *Natural Computing*, 3(1):77–112, 2004.

[11] Per Kristian Lehre. Drift analysis. In *Genetic and Evolutionary Computation Conference, GECCO '12, Philadelphia, PA, USA, July 7-11, 2012, Companion Material Proceedings*, pages 1239–1258, 2012.

[12] Per Kristian Lehre and Carsten Witt. General drift analysis with tail bounds. Technical Report arXiv:1307.2559, 2013.

[13] Johannes Lengler and Angelika Steger. Drift analysis and evolutionary algorithms revisited. Technical Report arXiv:1608.03226, 2016.

[14] Ingo Rechenberg. *Evolutionsstrategie: Optimierung technisher Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog, 1973.

Appendix

proof of theorem 1. We consider the truncated process defined above and the stopped truncated process as $Z_0 = Y_0^{\text{A}}$ and $Z_t = Y^{\text{A}}_{\min\{t, T_\beta^{Y^{\text{A}}}\}}$. By construction it holds $X_t \leq Y_t^{\text{A}} \leq Z_t$ and $T_\beta^X \leq T_\beta^{Y^{\text{A}}}$. We will prove that

$$E[Z_{t+1} \,|\, \mathcal{F}_t] \leq Z_t - \mathrm{B} \cdot \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} \ . \tag{14}$$

We start from

$$E[Z_{t+1} \,|\, \mathcal{F}_t] = E[Z_{t+1} \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} \leq t\}} \,|\, \mathcal{F}_t] + E[Z_{t+1} \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} \,|\, \mathcal{F}_t] \tag{15}$$

and estimate the different terms:

$$E[Z_{t+1} \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} \leq t\}} \,|\, \mathcal{F}_t] = E[Z_t \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} \leq t\}} \,|\, \mathcal{F}_t] = Z_t \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} \leq t\}} \tag{16}$$

where we have used that $\mathbf{1}_{\{T_\beta^{Y^{\text{A}}} \leq t\}}$ is $\mathcal{F}_t$-measurable, and this also implies that Y_t^{A} and Z_t, being functions of X_t and $T^{Y^{\text{A}}}$, are $\mathcal{F}_t$-measurable. Also

$$\begin{aligned} E[Z_{t+1} \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} \,|\, \mathcal{F}_t] &= E[Y_{t+1} \,|\, \mathcal{F}_t] \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} \\ &\leq (Y_t - \mathrm{B}) \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} = (Z_t - \mathrm{B}) \mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}} \end{aligned} \tag{17}$$

where we have also used that $\mathbf{1}_{\{T_\beta^{Y^{\text{A}}} > t\}}$ is $\mathcal{F}_t$ measurable. Hence injecting (16) and (17) into (15), we end up with (14). From (14), by taking the expectation we deduce

$$\mathbb{E}[Z_{t+1}] \leq \mathbb{E}[Z_t] - \mathrm{B} \cdot \Pr[T_{\text{A}} > t] \ . \tag{18}$$

Following the same approach as [13, Theorem 1], since $T_\beta^{Y^{\text{A}}}$ is a random variable taking values in $\mathbb{N}$, it can be rewritten as $\mathbb{E}[T_\beta^{Y^{\text{A}}}] = \sum_{t=0}^{+\infty} \Pr[T_\beta^{Y^{\text{A}}} > t]$ and thus it holds

$$\begin{aligned} \mathrm{B} \cdot \mathbb{E}\left[T_\beta^{Y^{\text{A}}}\right] \overset{\tilde{t} \to \infty}{\longleftarrow} \sum_{t=0}^{\tilde{t}} \mathrm{B} \cdot \Pr\left[T_\beta^{Y^{\text{A}}} > t\right] &\leq \sum_{t=0}^{\tilde{t}} \Big(\mathbb{E}[Z_t] - \mathbb{E}[Z_{t+1}]\Big) \\ &\leq \mathbb{E}[Z_0] - \mathbb{E}[Z_{\tilde{t}}] = x_0 - \mathbb{E}[Z_{\tilde{t}}] \ . \end{aligned} \tag{19}$$

Since $Y_{t+1} \geq Y_t - \mathrm{A}$, then $Y_{T_\beta^{Y^{\text{A}}}} \geq \beta - \mathrm{A}$ and given that $Z_t \geq Y_{T_\beta^{Y^{\text{A}}}}$, we deduce that $E[Z_{\tilde{t}}] \geq \beta - \mathrm{A}$ for all $\tilde{t}$, which implies

$$\mathbb{E}\left[T_\beta^{Y^{\text{A}}}\right] \leq \frac{x_0 - \beta + \mathrm{A}}{\mathrm{B}} \ .$$

With $\mathbb{E}[T_\beta^X] \leq \mathbb{E}[T_\beta^{Y^{\text{A}}}]$ this proves the upper bound. □

---

proof of proposition 1. From the definition of the truncated process (1) we obtain $|Y_{t+1}^{\text{A}}| \leq |Y_t^{\text{A}}| + |X_{t+1} - X_t| + \mathrm{A}$ which implies

$$\begin{aligned} \mathbb{E}[|Y_{t+1}^{\text{A}}|] &\leq \mathbb{E}[|Y_t^{\text{A}}|] + \mathbb{E}[|X_{t+1} - X_t|] + \mathrm{A} \\ &\leq \mathbb{E}[|Y_t^{\text{A}}|] + \mathbb{E}[|X_{t+1}|] + E[|X_t|] + \mathrm{A} \ , \end{aligned}$$

where the second to fourth terms are finite. Since $Y_0^{\text{A}} = X_0$ is integrable, Y_t^{A} is integrable by induction. □

---

proof of theorem 2. After $T = \lfloor (x_0 - \beta)/(2\mathrm{C}) \rfloor$ iterations it holds $\mathbb{E}[x_0 - X_T] \leq \mathrm{C} \cdot T \leq (x_0 - \beta)/2$. From Markov's inequality we conclude $\Pr(x_0 - X_T \geq x_0 - \beta) \leq \frac{1}{2}$ and thus $\Pr(x_0 - X_T \leq x_0 - \beta) \geq \frac{1}{2}$, which is equivalent to $\Pr(T_\beta^X \geq T) \geq \frac{1}{2}$. Applying the Markov inequality once more we obtain

$$\mathbb{E}[T_\beta^X] \geq \Pr\left(T_\beta^X \geq T\right) \cdot T \geq T/2 \geq \frac{x_0 - \beta}{4\mathrm{C}} - \frac{1}{2} \ .$$

□

---

proof of lemma 1. We introduce the sample $z \sim \mathcal{N}(0, \mathrm{I})$ through $z = (x - m)/\sigma$, or equivalently, $x = m + \sigma \cdot z$. We write the success rate in the form

$$p_{r,d}^{\text{succ}}(\bar{\sigma}) = \int\limits_{A_{r,d}(\bar{\sigma}) := \mathrm{B}\left(\frac{d}{\bar{\sigma}} \frac{-m}{\|m\|}, \frac{d}{\bar{\sigma}}(1-r)\right)} \varphi_d(z)\, dz.$$

For increasing values of $\bar{\sigma}$ the ball-shaped integration area shrinks, and in case of $r > 0$ it also moves away from the origin. Together with the monotonicity of φ_d w.r.t. $\|z\|$ this proves that $p_{r,d}^{\text{succ}}$ is monotonically decreasing. Continuity of $p_{r,d}^{\text{succ}}$ follows from the boundedness of φ_d, and positivity from the fact that $A_{r,d}(\bar{\sigma})$ is non-empty and φ_d is positive. This proves the first claim. For $r = 0$ the balls are nested. This immediately proves the second claim. From

$$\bigcap_{\bar{\sigma} > 0} A_{0,d}(\bar{\sigma}) = \emptyset \quad \text{and} \quad \bigcup_{\bar{\sigma} > 0} A_{0,d}(\bar{\sigma}) = \left\{ z \in \mathbb{R}^d \,\middle|\, m^T z < 0 \right\}$$

we conclude $\lim_{\bar{\sigma} \to 0} p_{0,d}^{\text{succ}}(\bar{\sigma}) = 1/2$ and $\lim_{\bar{\sigma} \to \infty} p_{0,d}^{\text{succ}}(\bar{\sigma}) = 0$, which proves the last claim. □

---

proof of lemma 2. We consider the sequence of random variables

$$\begin{aligned} J_d = \mathbf{1}_{\left\{\left\|e_1 + \frac{\bar{\sigma}}{d}\mathcal{N}\right\|^2 < (1-r)^2\right\}} &= \mathbf{1}_{\left\{1 + 2\frac{\bar{\sigma}}{d}\mathcal{N}_1 + \frac{\bar{\sigma}^2}{d^2}\|\mathcal{N}\|^2 < 1 - 2r + r^2\right\}} \\ &= \mathbf{1}_{\left\{2\bar{\sigma}\mathcal{N}_1 + \frac{\bar{\sigma}^2}{d}\|\mathcal{N}\|^2 < -2rd + r^2 d\right\}} \end{aligned}$$

indexed by d. Here $\mathcal{N}$ denotes a standard normally distributed vector in $\mathbb{R}^d$, and $\mathcal{N}_1$ is its first component. Almost surely by the Law of Large Numbers, $\|\mathcal{N}\|^2/d$ converges to 1 such that when d goes to infinity then it holds

$$\lim_{d \to \infty} \mathbf{1}_{\left\{2\bar{\sigma}\mathcal{N}_1 + \frac{\bar{\sigma}^2}{d}\|\mathcal{N}\|^2 < -2rd + r^2 d\right\}} = \mathbf{1}_{\{2\bar{\sigma}\mathcal{N}_1 + \bar{\sigma}^2 < -2\rho\}} = \mathbf{1}_{\left\{\mathcal{N}_1 < -\frac{\rho}{\bar{\sigma}} - \frac{\bar{\sigma}}{2}\right\}}$$

almost surely. Since $p_{d,r}^{\text{succ}}(\bar{\sigma}) = \mathbb{E}\left[J_d\right]$ and J_d converges almost surely to $\mathbf{1}_{\left\{\mathcal{N}_1 < -\frac{\rho}{\bar{\sigma}} - \frac{\bar{\sigma}}{2}\right\}}$ we need to prove the uniform integrability to ensure that the limit also holds in expectation. However the uniform integrability is here obvious since $\mathbb{E}\left[|J_d|\right] \leq 1$ for all d. Hence we have proven

$$\lim_{d\to\infty} \mathbb{E}\left[J_d\right] = \mathbb{E}\left[\mathbf{1}_{\{\mathcal{N}_1 < -\frac{\rho}{\bar\sigma} - \frac{\bar\sigma}{2}\}}\right] = \Pr\left(\mathcal{N}_1 < -\frac{\rho}{\bar\sigma} - \frac{1}{2}\bar\sigma\right) = \Phi_1\left[-\frac{\rho}{\bar\sigma} - \frac{1}{2}\bar\sigma\right] \quad . \quad \square$$

proof of lemma 3. The statement holds trivially for $t = 0$, since the initial condition is a constant. The following elementary calculation shows that the pole of the logarithm in the definition of V is not problematic. Let $B(0,1)$ denote the open ball of radius one around the origin, then he have:

$$\begin{aligned}
\int_{B(0,1)} \log(\|z\|)\,dz &= \int_0^1 \int_{S(0,r)} \log(\|z\|)\,dz\,dr \\
=\int_0^1 \left(\int_{S(0,r)} dz\right) \log(r)\,dr &= \frac{2\cdot\pi^{d/2}}{\Gamma(d/2)} \int_0^1 r^{d-1}\log(r)\,dr \\
=\frac{2\cdot\pi^{d/2}}{\Gamma(d/2)} \cdot \left[\frac{r^d(d\log(r)-1)}{d^2}\right]_0^1 &= -\frac{2\cdot\pi^{d/2}}{\Gamma(d/2)\cdot d^2},
\end{aligned}$$

where Γ denotes the Gamma function. Therefore $\mathbb{E}\big[|V(\theta_t)| \,\big|\, \theta_{t-1}\big] < \infty$ for all t, and the statement follows by induction. The integrability of the truncated process is straightforward from the above statement and Proposition 1. $\square$

proof of proposition 2. For the sake of simplicity we introduce $\log^+(x) = \log(x)\cdot\mathbf{1}_{\{x\geq 1\}}$. We rewrite the potential function as

$$\begin{aligned}
V(m_t,\ \sigma_t) =& \log\left(\|m_t\|\right) \\
&+ v\cdot\log^+\left(\frac{\alpha\cdot\ell\cdot\|m_t\|}{\sigma_t\cdot d}\right) \qquad (20)\\
&+ v\cdot\log^+\left(\frac{\sigma_t\cdot d}{\alpha^{-1/4}\cdot u\cdot\|m_t\|}\right) \quad . \qquad (21)
\end{aligned}$$

We want to estimate the conditional expectation

$$\mathbb{E}\left[\max\{V(\theta_{t+1}) - V(\theta_t)\,,\, -\mathrm{A}\} \mid \theta_t\right]. \qquad (22)$$

We partition the possible values of θ_t into three sets. First the set of θ_t such that $\sigma_t < \ell\cdot\|m_t\|/d$ (σ_t is small), second the set of θ_t such that $\sigma_t > u\cdot\|m_t\|/d$ (σ_t is large), and last the set of θ_t such that $\ell\cdot\|m_t\|/d \leq \sigma_t \leq u\cdot\|m_t\|/d$ (reasonable σ_t). In the following, we bound eq. (22) for each of the three cases and in the end our bound B will equal the minimum of the three bounds obtained for each case.

Reasonable σ_t case: $\frac{\|m_t\|}{d\sigma_t} \in \left[\frac{1}{u}, \frac{1}{\ell}\right]$. The potential function at time $t+1$ can be written as

$$\begin{aligned}
V(\theta_{t+1}) =& \log\left(\|m_{t+1}\|\right) \\
&+ v\cdot\log\left(\frac{\alpha\cdot\ell\cdot\|m_{t+1}\|}{d\cdot\sigma_{t+1}}\right) 1_{\{\alpha\ell\|m_{t+1}\|>d\cdot\sigma_{t+1}\}} 1_{\{\sigma_{t+1}>\sigma_t\}} \\
&+ v\cdot\log\left(\frac{\alpha\cdot\ell\cdot\|m_{t+1}\|}{d\cdot\sigma_{t+1}}\right) 1_{\{\alpha\ell\|m_{t+1}\|>d\cdot\sigma_{t+1}\}} 1_{\{\sigma_{t+1}<\sigma_t\}} \\
&+ v\cdot\log\left(\frac{d\cdot\alpha^{1/4}\cdot\sigma_{t+1}}{u\cdot\|m_{t+1}\|}\right) 1_{\left\{\alpha^{-1/4}u\|m_{t+1}\|<d\cdot\sigma_{t+1}\right\}} 1_{\{\sigma_{t+1}>\sigma_t\}} \\
&+ v\cdot\log\left(\frac{d\cdot\alpha^{1/4}\cdot\sigma_{t+1}}{u\cdot\|m_{t+1}\|}\right) 1_{\left\{\alpha^{-1/4}u\|m_{t+1}\|<d\cdot\sigma_{t+1}\right\}} 1_{\{\sigma_{t+1}<\sigma_t\}}.
\end{aligned}$$

In case of success, where thus $\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}\}} = 1$, we have $\|m_{t+1}\|/\sigma_{t+1} < \|m_t\|/(\alpha\sigma_t) \leq d/(\alpha\ell)$, implying that the conditions in the second term never hold at the same time and thus the second term is always 0. Similarly, in case of failure, $\|m_{t+1}\|/\sigma_t = \|m_t\|/(\alpha^{-1/4}\sigma) \leq d/(\alpha^{-1/4}u)$ and we find that the fifth term is always zero. We rearrange the third and fourth term into

$$\begin{aligned}
(3^{\mathrm{rd}}) &= v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot\ell\cdot\|m_t\|}{d\cdot\sigma_t}\right)\cdot 1_{\{\sigma_{t+1}<\sigma_t\}} \quad , \\
(4^{\mathrm{th}}) &= -v\cdot\left[\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right) - \log\left(\frac{d\cdot\sigma_t}{\alpha^{-5/4}\cdot u\cdot\|m_t\|}\right)\right] \\
&\quad\times 1_{\left\{\alpha^{-5/4}u\|m_{t+1}\|<d\cdot\sigma_t\right\}} 1_{\{\sigma_{t+1}>\sigma_t\}} \quad .
\end{aligned}$$

Then, the drift $\Delta_t = V(\theta_{t+1}) - V(\theta_t)$ is upper bounded by

$$\begin{aligned}
\Delta_t \leq& \left(1 - v\cdot 1_{\left\{\alpha^{-5/4}u\|m_t\|<d\cdot\sigma_t\right\}}\cdot 1_{\{\sigma_{t+1}>\sigma_t\}}\right)\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right) \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot\ell\cdot\|m_t\|}{d\cdot\sigma_t}\right)\cdot 1_{\{\sigma_{t+1}<\sigma_t\}} \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot d\cdot\sigma_t}{u\cdot\|m_t\|}\right)\cdot 1_{\{\sigma_{t+1}>\sigma_t\}} \\
\leq& (1-v)\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right) \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot\ell\cdot\|m_t\|}{d\cdot\sigma_t}\right)\cdot 1_{\{\sigma_{t+1}<\sigma_t\}} \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot d\cdot\sigma_t}{u\cdot\|m_t\|}\right)\cdot 1_{\{\sigma_{t+1}>\sigma_t\}} \quad .
\end{aligned}$$

The truncated drift $\max\{\Delta_t\,,\, -\mathrm{A}\}$ is upper bounded by

$$\begin{aligned}
\max\{\Delta_t\,,\,-\mathrm{A}\} \leq& (1-v)\max\left\{\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right)\ ,\ -\frac{\mathrm{A}}{1-v}\right\} \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot\ell\cdot\|m_t\|}{d\cdot\sigma_t}\right)\cdot 1_{\{\sigma_{t+1}<\sigma_t\}} \\
&+ v\cdot\log^+\left(\frac{\alpha^{5/4}\cdot d\cdot\sigma_t}{u\cdot\|m_t\|}\right)\cdot 1_{\{\sigma_{t+1}>\sigma_t\}} \quad .
\end{aligned}$$

To consider the expectation of the above upper bound on the truncated drift, we need to compute the expectation of the maximum of $\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right)$ and $-\frac{\mathrm{A}}{1-v}$. Let $a \leq 0$ and $b \in \mathbb{R}$ then

$$\max(a,b) = a\cdot\mathbf{1}_{\{a>b\}\}} + b\cdot\mathbf{1}_{\{a\leq b\}\}} \leq b\cdot\mathbf{1}_{\{a\leq b\}\}} \quad .$$

Applying this and taking the conditional expectation, a trivial upper bound for the conditional expectation of

$$\max\left\{\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right), -\frac{\mathrm{A}}{1-v}\right\}$$

is $-\frac{\mathrm{A}}{1-v}$ times the probability of $\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right)$ being no greater than $-\frac{\mathrm{A}}{1-v}$. The latter condition is equivalent to $\|m_{t+1}\| \leq (1-r)\cdot\|m_t\|$ corresponding to successes with rate $r = 1 - \exp\left(-\frac{\mathrm{A}}{1-v}\right)$ or better. That is,

$$(1-v)\cdot\mathbb{E}\left[\max\left\{\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right), -\frac{\mathrm{A}}{1-v}\right\}\right] \leq -\mathrm{A}\cdot p^{\mathrm{succ}}_{r,d}\left(\frac{d\cdot\sigma_t}{\|m_t\|}\right) \quad (23)$$

Note also that the expected value of $\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}$ is the success probability $p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)$. We obtain an upper bound for the conditional expectation of the truncated drift in the case of reasonable σ_t as

$$\begin{aligned}\mathbb{E}\left[\max\{\Delta_t, -\mathrm{A}\}|\theta_t\right] &\leq -\mathrm{A}\cdot p^{\mathrm{succ}}_{r,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)\\ &+\left(\frac{5}{4}\log(\alpha) + \underbrace{\log\left(\frac{\ell\|m_t\|}{d\sigma_t}\right)}_{\leq 0}\right)\cdot v\cdot\left(1 - p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)\right)\\ &+\left(\frac{5}{4}\log(\alpha) + \underbrace{\log\left(\frac{d\sigma_t}{u\|m_t\|}\right)}_{\leq 0}\right)\cdot v\cdot p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)\\ &\leq -\mathrm{A}\cdot p^* + \frac{5}{4}\log(\alpha)\cdot v\ ,\end{aligned} \quad (24)$$

where $r = 1 - \exp\left(-\frac{\mathrm{A}}{1-v}\right)$.

Small σ_t case: $\frac{\|m_t\|}{d\sigma_t} > \frac{1}{\ell}$. If $\ell\|m_t\| > d\sigma_t$, the summand (20) is positive. Moreover, if $\sigma_{t+1} < \sigma_t$, we have $\ell\|m_{t+1}\|/d = \ell\|m_t\|/d > \sigma_t > \sigma_{t+1}$ and hence the summand (20) is positive for $V(\theta_{t+1})$ as well. If $\sigma_{t+1} > \sigma_t$, any regime can happen. Then,

$$\begin{aligned}&V(\theta_{t+1}) - V(\theta_t)\\ &= \log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right) - v\cdot\log\left(\frac{\alpha\cdot\ell\cdot\|m_t\|}{d\cdot\sigma_t}\right)\\ &+ v\cdot\log\left(\frac{\alpha\cdot\ell\cdot\|m_{t+1}\|}{d\sigma_{t+1}}\right)\mathbf{1}_{\{\alpha\ell\|m_{t+1}\|>d\cdot\sigma_{t+1}\}}\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}\\ &+ v\cdot\log\left(\frac{\alpha\cdot\ell\cdot\|m_{t+1}\|}{d\sigma_{t+1}}\right)\mathbf{1}_{\{\alpha\ell\|m_{t+1}\|>d\cdot\sigma_{t+1}\}}\mathbf{1}_{\{\sigma_{t+1}<\sigma_t\}}\\ &+ v\cdot\log\left(\frac{\alpha^{1/4}\cdot\sigma_{t+1}d}{u\cdot\|m_{t+1}\|}\right)\mathbf{1}_{\left\{\alpha^{-1/4}u\|m_{t+1}\|<d\cdot\sigma_{t+1}\right\}}\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}\\ &= \Big[1 + \Big(v\cdot\mathbf{1}_{\{\ell\|m_{t+1}\|>d\sigma_t\}} - v\cdot\mathbf{1}_{\left\{\alpha^{-5/4}u\|m_{t+1}\|<d\cdot\sigma_t\right\}}\Big)\\ &\quad\cdot\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}\Big]\cdot\log\left(\frac{\|m_{t+1}\|}{\|m_t\|}\right)\\ &- v\cdot\log\left(\frac{u\|m_t\|}{\alpha^{5/4}\cdot d\cdot\sigma_t}\right)\mathbf{1}_{\left\{\alpha^{-5/4}u\|m_{t+1}\|<d\cdot\sigma_t\right\}}\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}\\ &- v\cdot\log\left(\frac{\ell\|m_t\|}{d\cdot\sigma_t}\right)\cdot\Big(1 - \mathbf{1}_{\{\ell\|m_{t+1}\|>d\cdot\sigma_t\}}\mathbf{1}_{\{\sigma_{t+1}>\sigma_t\}}\\ &\quad - \mathbf{1}_{\left\{\alpha^{5/4}\ell\|m_{t+1}\|>d\cdot\sigma_t\right\}}\mathbf{1}_{\{\sigma_{t+1}<\sigma_t\}}\Big)\\ &- v\cdot\log(\alpha)\cdot\left(1 - \frac{5}{4}\mathbf{1}_{\left\{\alpha^{5/4}\ell\|m_{t+1}\|>d\cdot\sigma_t\right\}}\mathbf{1}_{\{\sigma_{t+1}<\sigma_t\}}\right)\end{aligned}$$

On the RHS of the above equality, the first term is guaranteed to be non-positive since $v \in (0,1)$. The second and third terms are non-positive as well since $\frac{u\|m_t\|}{d\alpha^{5/4}\sigma_t} > \frac{u}{\alpha^{5/4}\ell} > 1$ and $\frac{\ell\|m_t\|}{d\sigma_t} > 1$. Since $v\cdot\log(\alpha)$ is positive, replacing the indicator $\mathbf{1}_{\left\{\alpha^{5/4}\ell\|m_{t+1}\|>d\sigma_t\right\}}$ with 1 provides an upper bound. Altogether, we obtain

$$V(\theta_{t+1}) - V(\theta_t) \leq -v\cdot\log(\alpha)\cdot\left(1 - \frac{5}{4}\mathbf{1}_{\{\sigma_{t+1}<\sigma_t\}}\right)\ .$$

Note that the RHS is larger than $-\mathrm{A}$. Then, the conditional expectation of the truncated drift is

$$\begin{aligned}\mathbb{E}\left[\max\{\Delta_t, -\mathrm{A}\}|\mathcal{F}_t\right] &\leq -v\cdot\log(\alpha)\cdot\left(\frac{5}{4}p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right) - \frac{1}{4}\right)\\ &\leq -v\cdot\log(\alpha)\cdot\left(\frac{5p_\ell - 1}{4}\right) < 0\ .\end{aligned} \quad (25)$$

Here we used $p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right) > p_\ell > 1/5$.

Large σ_t case: $\frac{\|m_t\|}{d\sigma_t} < \frac{1}{u}$. Since $\frac{\|m_{t+1}\|}{\sigma_{t+1}} \leq \frac{\|m_t\|}{\alpha^{-1/4}\sigma_t} < \frac{d}{\alpha^{-1/4}u}$, the summand (21) is positive in both $V(\theta_t)$ and $V(\theta_{t+1})$. For the summand (20), recall that $\alpha\ell\|m_t\|/d\sigma_t < \alpha\ell/u \leq \alpha\cdot\alpha^{-5/4} = \alpha^{-1/4} < 1$ since we have assumed that $u/\ell \geq \alpha^{5/4}$. Hence, for $V(\theta_t)$ the summand (20) is zero. Also, $\alpha\ell\|m_{t+1}\|/d\sigma_{t+1} \leq \alpha\ell/(\alpha^{-1/4}u) = \alpha^{5/4}\ell/u \geq 1$ and thus for $V(\theta_{t+1})$ the summand (20) also equals 0. We obtain

$$\begin{aligned}V(\theta_{t+1}) - V(\theta_t) =& (1-v)\Big(\log\left(\|m_{t+1}\|\right) - \log\left(\|m_t\|\right)\Big)\\ &+ v\cdot\log\left(\sigma_{t+1}/\sigma_t\right),\end{aligned}$$

where $\log\left(\sigma_{t+1}/\sigma_t\right)$ equals $\log(\alpha)$ with probability $p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)$, and $-\frac{1}{4}\log(\alpha)$ with probability $1 - p^{\mathrm{succ}}_{0,d}\left(\frac{d\sigma_t}{\|m_t\|}\right)$. The first term on the RHS is guaranteed to be non-positive since $v < 1$, yielding $\Delta_t \leq v\cdot\log(\sigma_{t+1}/\sigma_t)$. On the other hand,

$$v\cdot\log(\sigma_{t+1}/\sigma_t)$$

$$= v \cdot \left(\log(\alpha) \mathbf{1}_{\{\|m_{t+1}\| < \|m_t\|\}\}} - \frac{1}{4} \log(\alpha) \mathbf{1}_{\{\|m_{t+1}\| = \|m_t\|\}\}} \right)$$
$$= v \cdot \left(\frac{5}{4} \log(\alpha) \mathbf{1}_{\{\|m_{t+1}\| < \|m_t\|\}\}} - \frac{1}{4} \log(\alpha) \right)$$
$$\geq -\frac{1}{4} \log(\alpha) v \geq -\mathrm{A}$$

where the last inequality comes from the prerequisite $v \leq 4\mathrm{A}/\log(\alpha)$. Hence, $\max\{v \cdot \log(\sigma_{t+1}/\sigma_t), -\mathrm{A}\} = v \log(\sigma_{t+1}/\sigma_t)$ such that $\max\{\Delta_t\,, -\mathrm{A}\} \leq v \cdot \log(\sigma_{t+1}/\sigma_t)$. Then, the conditional expectation of the truncated drift is

$$\mathbb{E}\left[\max\{\Delta_t\,, -\mathrm{A}\} | \theta_t\right] \leq -\frac{1}{4} v \log(\alpha) \left(1 - 5 p_{0,d}^{\mathrm{succ}} \left(\frac{d\sigma_t}{\|m_t\|} \right) \right)$$
$$\leq -v \log(\alpha) \left(\frac{1 - 5p_u}{4} \right) < 0 \ . \qquad (26)$$

Here we used $p_{0,d}^{\mathrm{succ}} \left(\frac{d\sigma_t}{\|m_t\|} \right) \leq p_u < 1/5$.

Inequalities (25), (26), and (24) together cover all possible cases and hence imply the bound (10). ☐

proof of proposition 3. We rewrite $r' = 1 - \exp\left(-\frac{\mathrm{A}}{1 - \frac{1}{d \cdot \log(\alpha)}} \right)$. It holds $v < \frac{1}{d \cdot \log(\alpha)}$ and hence $r' > r$, from which we obtain $p' < p^*$. Now we consider the terms in equation (10) one by one. We start with

$$\mathrm{A} \cdot p^* - \frac{5}{4} \cdot \log(\alpha) \cdot v = \frac{p^*}{d} - \frac{5}{8} \frac{p'}{d}$$

which is lower bounded by $-\frac{3}{8} \frac{p'}{d}$ and upper bounded by $-\frac{3}{8} \frac{p^*}{d}$. Furthermore, we obtain

$$\log(\alpha) \cdot v \cdot \frac{5p_\ell - 1}{4} = \frac{p'}{d} \cdot \frac{5p_\ell - 1}{8}$$
$$\log(\alpha) \cdot v \cdot \frac{1 - 5p_u}{4} = \frac{p'}{d} \cdot \frac{1 - 5p_u}{8} \ .$$

We collect these results into the definition of lower and upper bounds

$$L = \frac{p'}{d} \cdot \min\left\{ \frac{3}{8}, \frac{5p_\ell - 1}{8}, \frac{1 - 5p_u}{8} \right\}$$
$$U = \frac{p^*}{d} \cdot \max\left\{ \frac{3}{8}, \frac{5p_\ell - 1}{8}, \frac{1 - 5p_u}{8} \right\}$$

for $L \leq \mathrm{B} \leq U$. From $L > 0$ we immediately obtain $\mathrm{B} > 0$. We have $\lim_{d \to \infty} d \cdot r = 1$ and hence according to Lemma 2

$$\lim_{d \to \infty} p^* = \lim_{d \to \infty} \left(\min_{\bar{\sigma} \in [\ell, u]} \left\{ p_{r,d}^{\mathrm{succ}}(\bar{\sigma}) \right\} \right)$$
$$\overset{(*)}{=} \min_{\bar{\sigma} \in [\ell, u]} \left\{ \lim_{d \to \infty} \left(p_{r,d}^{\mathrm{succ}}(\bar{\sigma}) \right) \right\} = \min_{\bar{\sigma} \in [\ell, u]} \left\{ \Phi_1 \left(-\frac{1}{\bar{\sigma}} - \frac{\bar{\sigma}}{2} \right) \right\}$$
$$= \min\left\{ \Phi_1 \left(-\frac{1}{\ell} - \frac{\ell}{2} \right), \Phi_1 \left(-\frac{1}{u} - \frac{u}{2} \right) \right\} > 0 \ .$$

The equality (*) holds as follows: Let $(\bar{\sigma}_d)_{d \in \mathbb{N}}$ be a sequence of points where the minimum is attained, then the Bolzano-Weierstraß property provides a convergent sub-sequence with limit point $\bar{\sigma} \in [\ell, u]$. Since the success probability functions and its limit are continuous, the minimum of the limit function is attained at $\bar{\sigma}$. We obtain $U \in \Theta(1/d)$. Analogously, with $\lim_{d \to \infty} d \cdot r' = 1$ and

$$\lim_{d \to \infty} p' = \min_{\bar{\sigma} \in [\ell, u]} \left\{ \Phi_1 \left(-\frac{1}{\bar{\sigma}} - \frac{\bar{\sigma}}{2} \right) \right\} > 0$$

we also obtain $L \in \Theta(1/d)$. Combining the results for L and U proves $\mathrm{B} \in \Theta(1/d)$. ☐

proof of lemma 4. The log progress of the hit-and-run algorithm amounts to $-\log(\sin(\theta)) \cdot 1_{\{\theta \leq \pi/2\}}$, where $\theta \in [0, \pi)$ is the angle between δ_t and e_1. This follows from a geometric interpretation of the algorithm. Let $W_d = \int_0^{\pi/2} \sin^d(\theta) d\theta$ denote the Wallis integral. Then the density of θ is $(2W_{d-2})^{-1} |\sin(\theta)|^{d-2}$. The expected log progress of the hit-and-run algorithm is written as

$$\frac{-1}{2W_{d-2}} \int_0^{\pi/2} \log\left(\sin(\theta)\right) \sin^{d-2}(\theta) d\theta$$
$$= \frac{-1}{2(d-1)^2 W_{d-2}} \int_0^1 \frac{\log(r)}{\sqrt{1 - r^{\frac{2}{d-1}}}} dr \ .$$

Here we applied the change of variables $\sin(\theta)^{d-1} = r$. Note that $\log(r)$ and $1/\sqrt{1 - r^{\frac{2}{d-1}}}$ are positively correlated. Therefore, the integral on the RHS is lower bounded by the product of the integrals of the two terms, which reads

$$\int_0^1 \frac{\log(r)}{\sqrt{1 - r^{\frac{2}{d-1}}}} dr \geq \underbrace{\int_0^1 \log(r)\, dr}_{=-1} \underbrace{\int_0^1 \sqrt{\frac{1}{1 - r^{\frac{2}{d-1}}}} dr}_{=(d-1)W_{d-2}} \ .$$

using $d \leq 2(d-1)$ for all $d \geq 2$ concludes the proof. ☐

proof of theorem 3. Since $\log(\|m_t\|) \leq V(\theta_t)$, then the hitting time of $(-\infty, \log(\epsilon)]$ by $V(\theta_t)$, denoted T_ϵ^V, is not less than T_ϵ. In Proposition 2 we have shown that for $\mathrm{A} = 1/d$ and v as set in Proposition 3, the drift

$$\mathbb{E}\left[\max\{V(\theta_{t+1}) - V(\theta_t)\,, -1/\mathrm{d}\} \mid \theta_t\right] \leq -\mathrm{B}$$

holds. By applying Theorem 1 we obtain

$$\mathbb{E}\left[T_\epsilon^V\right] \leq \frac{V(\theta_0) - \log(\epsilon) + \frac{1}{d}}{\mathrm{B}} \in \Theta\Big((V(\theta_0) - \log(\epsilon)) \cdot d \Big) \ .$$

Together with $\mathbb{E}\left[T_\epsilon\right] \leq \mathbb{E}\left[T_\epsilon^V\right]$ this shows the upper bound. Lemma 4 bounds the drift of $X_t = \log(\|m_t\|)$, see eq. (12). With the bound $\mathrm{C} = \frac{1}{d}$ and $\beta = \log(\epsilon)$, Theorem 2 yields the lower bound $\mathbb{E}[T_\epsilon] \geq \frac{(x_0 - \beta) \cdot d}{4} - \frac{1}{2} \in \Theta\Big(\big(x_0 - \log(\epsilon)\big) \cdot d \Big)$. ☐