# Concentration of First Hitting Times under Additive Drift

Timo Kötzing
Fiedrich-Schiller-Universität Jena
07743 Jena, Germany

## ABSTRACT

Recent advances in drift analysis have given us better and better tools for understanding random processes, including the run time of randomized search heuristics. In the setting of multiplicative drift we do not only have excellent bounds on the *expected* run time, but also more general results showing the *concentration* of the run time.

In this paper we investigate the setting of additive drift under the assumption of strong concentration of the "step size" of the process. Under sufficiently strong drift towards the goal we show a strong concentration of the hitting time. In contrast to this, we show that in the presence of small drift a Gambler's-Ruin-like behavior of the process overrides the influence of the drift. Finally, in the presence of sufficiently strong negative drift the hitting time is superpolynomial with high probability; this corresponds to the so-called negative drift theorem, for which we give new variants.

## Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

## General Terms

Theory, Algorithms, Performance

## Keywords

Additive Drift, Concentration, Theory, Run Time Analysis

## 1. INTRODUCTION

Suppose we make a random walk on the real line starting with 0. Further suppose that, in each step of the walk, we expect to increase in value by exactly $\varepsilon > 0$; this expected increase is called a (positive) *drift*. The Additive Drift Theorem ([HY04], see Theorem 5) tells us that the *expected* time for the walk to reach some fixed value $n$ (for the first time) is exactly $n/\varepsilon$ (for finite search spaces). The (random) time to reach a given value for the first time is called the *first hitting time* or just *hitting time* (in this paper we will also accept overshooting). This drift theorem is based on a more general result [Haj82] and gave a new and powerful tool for the formal analysis of random processes, such as the progress of randomized search heuristics (like evolutionary algorithms and ant colony optimization).

In fact, after the publication of [HY04], the Additive Drift Theorem became more and more popular as a method to analyze the expected run time of randomized search heuristics. In order to get better bounds from a drift theorem with little effort, new drift theorems were proven, for example for drift proportional to the distance from the target (instead of uniform, as in the Additive Drift Theorem – this is called *multiplicative drift*) [DJW12]. Another very powerful family of drift theorems are the so-called *Variable Drift Theorems* (independently in [Joh10, Theorem 4.6] and [MRC08, Section 8], but see also [RS12] for a discussion and extension).

All these theorems have in common that they can be used for showing *upper* bounds on the run time of randomized algorithms. Aiming for a similarly strong tool for showing lower bounds, [OW11, OW12] derived (again from [Haj82]) a theorem which applies in case that the drift goes *away* from the target (see Theorem 6 for a precise statement and [RS12] for a powerful variant). Just as the drift theorems for upper bounds, the Negative Drift Theorem has proven to be superbly useful for the analysis of randomized search heuristics, providing an easy-to-apply tool for deriving lower bounds.

In addition to bounds on the *expected* hitting time, concentration results are also of interest. These can, for example, be directly used for statements about the concentration of the run time of an algorithm. But sometimes concentration results are necessary for deriving bounds on the expected hitting times as well: imagine, for example, an algorithm which can only be successful when $n$ independent sub-algorithms are successful; in the analysis, one would usually need concentration results for the run time of the sub-algorithms.

For the special case of multiplicative drift, strong concentration results were given in [DG13]. In very recent work [LW13] even more general results are given, providing concentration bounds in a very general setting. In this paper, we take an approach different from that in [LW13] by focusing on the very special case of additive drift and deriving as strong as possible concentration results in this case. The advantage is that, for the theorems in this pa-


*GECCO'14,* July 12–16, Vancouver, BC, Canada.

ACM 978-1-4503-2662-9/14/07 ...$15.00.
http://dx.doi.org/10.1145/2576768.2598364.

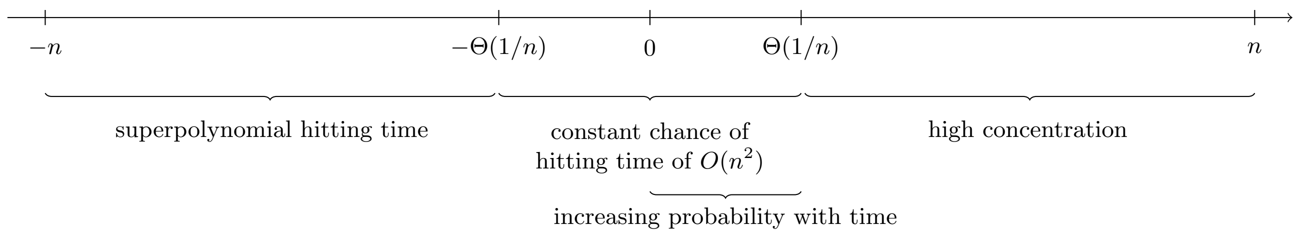


**Figure 1: Intuitive regimes of additive drift. Depicted are possible values of the additive drift $\varepsilon$; the important change points are (up to constant factors) at $1/n$ and $-1/n$. Note that for large (superconstant) bounds on the possible jump size $c$ of the process, the results get worse.**

per, checking whether they apply is easy, and so is using the conclusion; the downside is the restricted scope.

Outside of the evolutionary computation community, a number of results also regarding positive drift are known. In particular, most of the work of this paper is based on the technique of bounded differences (basically all proofs are applications of the Azuma-Hoeffding Inequality), which is wide spread, and the applications given here are straightforward instances of these methods (see [DP09] for an introduction). A notable example where similar results regarding positive drift are shown is [Wor99] (a version of Theorem 1 can be found in its Section 4). The main purpose of the present paper is to introduce these methods to the evolutionary computation community in an easily accessible way (the Azuma-Hoeffding Inequality itself was already used in the community to derive concentration, see, for example, [DK13]).

## 1.1 Discussion of Results

Recall that, if we start with 0 and drift an expected amount of at most $\varepsilon$ towards $n$, we have an expected time of at least $n/\varepsilon$ to reach $n$. However, it is possible that $n$ is already reached after one round with constant probability: the process might, in the first iteration, jump to the goal ($= n$) with probability $1/2$ and with the remaining $1/2$ probability it jumps to $-n$, giving an expected progress of $0 \leq \varepsilon$.[1] Similarly, one can give examples where the drift is high, but the probability to reach the goal within the expected number of steps is low.

We would like to give sufficient conditions (which hold in many cases for analyses of randomized search heuristics) under which the hitting time is concentrated around the expectation. To that end we will assume that no large jumps occur: we require that the largest possible jump is $c$, for some value $c > 0$. Under this condition we can derive strong results; Figure 1 gives an overview. Note the three regimes of additive drift: if it is strong, we get high concentration; if it is between about $-1/n$ and $1/n$, we get a behavior similarly to the Gambler's Ruin problem, with a constant probability of reaching $n$ regardless of the strength of the drift due to a (sufficiently unbiased) random walk on the real line; note that this result requires a constant variance. This constant probability can be significantly boosted by allowing more time, in case of non-negative drift. Finally, for low values of drift, we get an exponential hitting time, (this is the regime of negative drift theorems). In the following we discuss these statements in more detail.

Our first theorem informs about an exponentially small probability of arriving at $n$ significantly before the expected $n/\varepsilon$ iterations. Note that a version of this bound was shown in [Wor99, Corollary 4.1].

THEOREM 1. *Let $(X_t)_{t \geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ we denote the random variable describing the earliest point that the random process exceeds $n$, given a starting value of at most 0. Suppose there are $\varepsilon, c > 0$ such that, for all $t$,*

1. *$E(X_{t+1} \mid X_0, \ldots X_t, T > t) \leq X_t + \varepsilon$, and*
2. *$|X_t - X_{t+1}| < c$.*

*Then, for all $s \leq n/(2\varepsilon)$,*

$$P(T < s) \leq \exp\left(-\frac{n^2}{16c^2 s}\right).$$

Note that the first condition in the previous theorem formalizes an *additive drift of at most $\varepsilon$*, the second condition formalizes the bounded step width.

We see that, for example for constant $c$ and $\varepsilon = \Omega(1/n)$, we have a superpolynomially small probability of hitting $n$ in less than $n^2/\omega(\log n)$ iterations. Note that the bound is no longer useful (i.e. greater than 1) when $s \geq n^2$. This means that after more than $n^2$ steps we cannot exclude having exceeded $n$ (at least not with this theorem). If $\varepsilon \geq 1/n$, we expected to hit $n$ after $n^2$ steps anyway (due to the drift), and the bound of $s \leq n/(2\varepsilon)$ makes the bound inapplicable for values of $s \geq n^2$. As soon as we have drift of $\varepsilon < 1/n$, the drift process is intuitively more and more drowned by the random walk due to the variance (which we will consider later).

But what is now the probability of arriving significantly after the expected time? For that we need a *lower* bound on the expected progress (drift).

THEOREM 2. *Let $(X_t)_{t \geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$ we denote the random variable describing the earliest point that the random process exceeds $n$, given a starting value of at least 0. Suppose there are $\varepsilon, c > 0$ such that, for all $t$,*

[1] Note that iterating this idea leads to an example where, under arbitrary additive drift, the expected number of iterations until $n$ is reached is 2, seemingly contradicting the Additive Drift Theorem; however, this iterated example requires an unbounded search space, which is ruled out by the requirements of the Additive Drift Theorem.

1. $E(X_{t+1} \mid X_0, \dots X_t, T > t) \geq X_t + \varepsilon$, and
2. $|X_t - X_{t+1}| < c$.

*Then, for all $s \geq 2n/\varepsilon$,*

$$P(T \geq s) \leq \exp\left(-\frac{s\varepsilon^2}{16c^2}\right).$$

Thus, unless the drift is small, $n$ will be exceeded with high probability after twice the expected number of steps. For small drift ($O(1/n)$), the bound is only meaningful for larger numbers of iterations, so that Markov's Inequality will give better bounds in this case for $s$ close to $n/\varepsilon$.

If the drift is significantly negative, then we cannot hope to reach the goal in polynomial time with reasonable probability; this is the statement of the Negative Drift Theorem ([OW11, OW12], see Theorem 6). However, this theorem requires *constant* negative drift, a restriction which we replace to negative drift of essentially $\omega(1/n)$ with the following theorem. Note that a further version of such drift theorems is available in [OW13].

THEOREM 3. *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ we denote the random variable describing the earliest point that the random process exceeds $n$, given a starting value of at most 0. Suppose there are $c$, $0 < c < n$ and $\varepsilon < 0$ such that, for all $t$,*

1. $E(X_{t+1} \mid X_0, \dots X_t, T > t) \leq X_t + \varepsilon$, and
2. $|X_t - X_{t+1}| < c$.

*Then, for all $s \geq 0$,*

$$P(T \leq s) \leq s \exp\left(-\frac{n|\varepsilon|}{16c^2}\right).$$

For example, for $c$ constant and $\varepsilon = -\omega(\log(n)/n)$, this gives a superpolynomially small hitting probability for any polynomial number of steps.

Finally, we consider the case where there is only small drift $\varepsilon \in [0, 1/n]$.

THEOREM 4. *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$ we denote the random variable describing the earliest point that the random process exceeds $n$, given a starting value of at least 0. Suppose there are $\varepsilon, c \geq 0$ with $c < n$ such that, for all $t$,*

1. $E(X_{t+1} \mid X_0, \dots X_t, T > t) \geq X_t + \varepsilon$,
2. $Var(X_{t+1} - X_t \mid X_0, \dots X_t, T > t) \geq 1$, and
3. $|X_t - X_{t+1}| < c$.

*Then there is a constant $\ell$ (independent of $n, c$ and $\varepsilon$) such that, for all $\delta > 0$,*

$$P(T \leq n^2/\delta^{\ell \log(c)}) \geq 1 - \delta.$$

For example, if we have $c$ constant and want any constant hitting probability $\delta$, then a quadratic number of steps suffices (just as in the Gambler's Ruin problem).

## 1.2 Application and Open Problems

Note that all results are applicable not only for going from 0 to some $n$, but for drifting from any start $a$ to any $b$, regardless of whether $a < b$ or $b < a$ – a simple transformation of the $(X_t)_{t\geq 0}$ shows this. Also, whenever there is no sufficiently small bound $c$ to be used, one can condition the process on never jumping larger than some $c$ and compute the failure probability of jumping more than this $c$; for example, many mutation operators can jump arbitrarily large distances (so-called *global operators*), but usually stay within some small (logarithmic) neighborhood with large probability.

The reason that we require such a bound $c$ is that we base our results on the Azuma-Hoeffding Inequality (see Theorem 7) which requires just that. This entails also that our bounds are best when the steps of the process have a high variance (while many randomized search heuristics have step sizes of rather small variance, as compared with the range that they cover with reasonable probability). Thus, it is an *open problem* to improve the above bounds, for example by replacing the frequently occurring term $c^2$ by a bound on the variance of the steps of the process, or something similar. This would require finding a suitable replacement for the Azuma-Hoeffding Inequality.

# 2. KNOWN BOUNDS

The literature knows a large number of drift theorems; we give the two most important with respect to our setting of additive drift.

The simplest drift theorem concerns the expected hitting time under additive drift.

THEOREM 5 (ADDITIVE DRIFT [HY04]). *Let $(X_t)_{t\geq 0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}$. Let $T$ be the random variable that denotes the earliest point in time $t \geq 0$ such that $X_t \geq n$. If there exist $\varepsilon > 0$ such that*

$$E(X_{t+1} - X_t \mid T > t) \leq \varepsilon,$$

*then*

$$E(T \mid X_0) \geq \frac{X_0}{\varepsilon}.$$

*If there exist $\varepsilon > 0$ such that*

$$E(X_{t+1} - X_t \mid T > t) \geq \varepsilon,$$

*then*

$$E(T \mid X_0) \leq \frac{X_0}{\varepsilon}.$$

Second, the *Negative Drift Theorem* concerns adverse drift and shows a high hitting time, which can be used to derive lower bounds on the run time of algorithms.

THEOREM 6 (NEGATIVE DRIFT [OW11, OW12]). *Let $(X_t)_{t\geq 0}$ be real-valued random variables describing a stochastic process over some state space. Suppose there is an interval $[a, b] \subseteq \mathbb{R}$, two constants $\delta, \varepsilon > 0$ and, possibly depending on $\ell = b - a$, a function $r(\ell)$ satisfying $1 \leq r(\ell) = o(\ell/\log \ell)$ such that, for all $t \geq 0$, the following conditions hold.*

1. $E(X_{t+1} - X_t \mid a < X_t < b) \geq \varepsilon$;
2. For all $j \geq 0$, $P(|X_{t+1} - X_t| \geq j \mid a < X_t) \leq \frac{r(\ell)}{(1+\delta)^j}$.

*Then there is a constant c such that, for $T = \min\{t \geq 0 : X_t \leq a \mid X_0 \geq b\}$, we have*

$$P(T \leq 2^{c\ell/r(\ell)}) = 2^{-\Omega(\ell/r(\ell))}.$$

A crucial requirement of the theorem is a restriction on the jump size of the random process: the larger the step, the less likely it must be. A further important requirement is that of *constant* drift away from the target. See Corollary 15 for a comparison with the results of this paper.

For our results, we make use of the Azuma-Hoeffding Inequality for supermartingales, see [Azu67]. We give a version from [FGL12, Corollary 2.1], with scaling incorporated. For this we mention that $(X_t)_{t\geq 0}$ is a *supermartingale* if each random variable has finite expectation and, for all $t \geq 0$,

$$E(X_{t+1} \mid X_0, \ldots X_t, T > t) \leq X_t.$$

THEOREM 7 (AZUMA-HOEFFDING INEQUALITY). *Let $(X_t)_{t\geq 0}$ be a supermartingale such that there is c such that, for all t, $|X_t - X_{t+1}| < c$. Then, for all t and all x,*

$$P\left(\max_{0\leq j\leq t}(X_j - X_0) \geq x\right) \leq \exp\left(-\frac{x^2}{4c^2 t}\right).$$

## 3. DETAILED PROOFS

In this section we give the proofs of the theorems from Section 1. The proofs for Theorems 1-3 are straightforward applications of the Azuma-Hoeffding Inequality (Theorem 7); we will discuss these in Section 3.1 on large drift (they mostly apply when the drift is $\Omega(1/n)$ in either direction). After that we will consider small drift in Section 3.2.

For this section, let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation. Furthermore, we let $n \in \mathbb{N}$ and let $T_{\leq} = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ be the random variable that denotes hitting time of $n$ (similarly, $T_{\geq} = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$). For a given $\varepsilon$, we say that $(X_t)_{t\geq 0}$ *has drift of at most $\varepsilon$* iff

$$E(X_{t+1} \mid X_0, \ldots X_t, T > t) \leq X_t + \varepsilon. \quad (1)$$

Symmetrically, we say, for a given $\varepsilon$, that $(X_t)_{t\geq 0}$ *has drift of at least $\varepsilon$* iff

$$E(X_{t+1} \mid X_0, \ldots X_t, T > t) \geq X_t + \varepsilon. \quad (2)$$

We will assume that there is $c$ such that, for all $t \geq 0$,

$$|X_{t+1} - X_t| < c. \quad (3)$$

We call a process where Equation (3) holds *c-bounded* or *of step width at most c*.

Note that (1) is almost exactly the definition of a supermartingale (similarly, (2) corresponds to submartingales), except for the additional $\varepsilon$. In fact, we can derive a supermartingale in case that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon$ by letting, for all $t \geq 0$,

$$Y_t = X_t - t\varepsilon.$$

Since $(Y_t)_{t\geq 0}$ is now a supermartingale, we can now apply the Azuma-Hoeffding Inequality (see Theorem 7). Similarly, if $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon$, then $(-Y_t)_{t\geq 0}$ is a supermartingale.

Note that if $(X_t)_{t\geq 0}$ is $c$-bounded, then $(Y_t)_{t\geq 0}$ is $2c$-bounded, thanks to $|\varepsilon| \leq c$.

### 3.1 Large Drift

We will now use our observations to get proofs for Theorems 1-3.

THEOREM 8 (EQUIVALENT TO THEOREM 1). *Suppose that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon > 0$. Then, for all $s \leq n/(2\varepsilon)$,*

$$P(T_{\leq} < s) \leq \exp\left(-\frac{n^2}{16c^2 s}\right).$$

PROOF. Let $s \leq n/(2\varepsilon)$. We apply the Azuma-Hoeffding Inequality (Theorem 7) on the supermartingale $(Y_t)_{t\geq 0}$ (which is $2c$-bounded) to see that the probability of any of the $Y_t - Y_0$ for $t \leq s$ exceeding $n/2$ is bounded from above by

$$\exp\left(-\frac{n^2}{64c^2 s}\right).$$

Intuitively, we bound the probability of going half the way by random variance; as, for $s \leq n/(2\varepsilon)$, we have made at most (the other) half in expectation, this translates to the desired bound as follows. We observe that, for all $t \leq s$, $Y_t < n/2$ implies $X_t < n$ (as $s \leq n/(2\varepsilon)$). □

Similarly, we get the bound that the hitting time is reached with high probability after sufficient iterations.

THEOREM 9 (EQUIVALENT TO THEOREM 2). *Suppose that $(X_t)_{t\geq 0}$ has drift of at least $\varepsilon > 0$. Then, for all $s \geq 2n/\varepsilon$,*

$$P(T_{\geq} \geq s) \leq \exp\left(-\frac{s\varepsilon^2}{64c^2}\right).$$

PROOF. Let $s \geq 2n/\varepsilon$. We apply the Azuma-Hoeffding Inequality (Theorem 7) on the supermartingale $(-Y_t)_{t\geq 0}$ to see that the probability of $-Y_s + Y_0$ exceeding $s\varepsilon/2$ is bounded from above by

$$\exp\left(-\frac{(s\varepsilon)^2}{64c^2 s}\right) = \exp\left(-\frac{s\varepsilon^2}{64c^2}\right).$$

Intuitively, we bound the probability of making only half the steps that we should have made; this is meaningful once we should have overshot by a factor of 2, i.e. for $s \geq 2n/\varepsilon$ as desired. The proof is completed by the observation that $-Y_s < s\varepsilon/2$ implies $X_s > n$ (as $s \geq 2n/\varepsilon$). □

We now use this approach to prove the theorem concerning negative drift.

THEOREM 10 (EQUIVALENT TO THEOREM 3). *Suppose that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon < 0$ and assume $c \leq n/2$. Then, for all $s \geq 0$,*

$$P(T_{\leq} \leq s) \leq s \exp\left(-\frac{n|\varepsilon|}{64c^2}\right).$$

PROOF. We make an analysis with phases. A phase begins when, for some $t$, $X_t \geq 0$ and ends when either $X_{t'} \geq n$ or $X_{t'} < 0$; in the first case we call the phase *successful*, in the second case *unsuccessful*. We will show that a phase is successful with probability at most $\exp\left(-\frac{n|\varepsilon|}{64c^2}\right)$, as then a union bound (or an application of Bernoulli's Inequality) will give the desired result, lower bounding the length of each phase with the trivial bound of 1. In order to bound the probability of a phase being successful, we use the following reasoning. Any phase starts $\leq c$. If the process does not overshoot its expectation by $n - c$ ever within $n/|\varepsilon|$ iterations, it not only did not reach $n$ (starting from $\leq c$) but also drops below 0 (as, after $n/|\varepsilon|$ iterations, the expectation is $\leq -n$). For this we apply again the Azuma-Hoeffding Inequality (Theorem 7) and see that the probability of a phase being successful is at most

$$\exp\left(-\frac{(n-c)^2|\varepsilon|}{16c^2 n}\right) \leq \exp\left(-\frac{n|\varepsilon|}{64c^2}\right),$$

where the inequality comes from $c \leq n/2$. □

## 3.2 Small Drift

We start with a lemma which is interesting in its own right, showing the theorem concerning negative drift (Theorem 3) to be reasonably tight. The proof of the lemma makes use of one of the result concerning the concentration of the hitting time under positive drift (Theorems 1).

LEMMA 11. *Let $b = 3072$. Then, for all $n$ and $c$, the following holds. Let $k = bc^2 n$. Suppose that $(X_t)_{t\geq 0}$ has a drift of at least $\varepsilon \geq -1/(4k)$, a variance in each step of at least 1 and $1 \leq c < n/2$. Let $s = 24bc^2n^2$. Then we have that, within $s$ steps, the process does not drop below $-k$ with probability $\geq 1/2$ and*

$$P(T_{\geq} \leq s) \geq \frac{1}{2}.$$

PROOF. We give the proof for $\varepsilon \leq 0$; the case of $\varepsilon > 0$ is analogous, but easier. We let $A$ be the event that the process does not drop below $-k$ within $s$ steps. We first show $P(A) \geq 3/4$, after that we show that, conditional on $A$, the process reaches $n$ with probability 3/4, which will imply the claim.

For all $t$, we let

$$Y_t = (X_t)^2$$

and

$$\Delta_t = X_{t+1} - X_t.$$

In all of the following computations of expectation and variance the conditioning on all relevant (previous) random variables is implicitly understood but not made explicit for clarity (and brevity) of the exposition. We note that, for all $t$, $Var(\Delta_t) \leq 4c^2$. It suffices to show that $Y_t$ does not reach $k^2$ within $s$ steps with probability $\geq 3/4$. We want to apply Theorem 1 to $(Y_t)_{t\geq 0}$, so we compute the expected drift.

$$\begin{aligned}
E(Y_{t+1}) &= E((\Delta_t + X_t)^2) \\
&= E((\Delta_t)^2 + 2\Delta_t X_t + X_t^2) \\
&= E((\Delta_t)^2) + 2E(\Delta_t)X_t + X_t^2 \\
&= Var(\Delta_t) + E(\Delta_t)^2 + 2E(\Delta_t)X_t + Y_t \\
&\leq 4c^2 + 1 + Y_t \\
&\leq 5c^2 + Y_t.
\end{aligned}$$

The last inequality follows from our bound on the variance, together with our bounds on the expected drift and the value of $X_t$.

In order to estimate the number of steps until $(X_t)_{t\geq 0}$ reaches $-k$, we wait until the process drops below 0, and bound the time that the process $(Y_t)_{t\geq 0}$ takes to get from 0 to $k^2$, which, using the Additive drift theorem, has an expectation of

$$\geq k^2/(5c^2) \geq 2s$$

steps (using $b \geq 240$). It is easy to see that the process $(Y_t)_{t\geq 0}$ is $2kc$-bounded. Thus, Theorem 1 gives that $(Y_t)_{t\geq 0}$ *does* exceed $k^2$ within $s$ steps starting from 0 with probability

$$\leq \exp\left(-\frac{k^4}{16 \cdot 4k^2c^2 s}\right) = \exp\left(-\frac{b}{64 \cdot 24}\right) \leq 1/4$$

as desired.

Now we want to bound the probability for reaching $n$. To this end we let, for all $t \geq 0$,

$$Z_t = (X_t + k)^2 - k^2$$

and we condition on $A$. In a computation analogous to that for $(Y_t)_{t\geq 0}$ we see that

$$E(Z_{t+1}) \geq 1/2 + Z_t.$$

From the Additive Drift Theorem (Theorem 5) we now know that $(Z_t)_{t\geq 0}$ reaches $(n+k)^2 - k^2$ in an *expected* number of

$$2((n+k)^2 - k^2) = 2(2nk + n^2) \leq 6nk$$

steps. Thus, using Markov's Inequality, we get that $(Z_t)_{t\geq 0}$ reaches $(n+k)^2 - k^2$ within $s$ steps with probability $\geq 3/4$ as desired, as $s = 4(6nk)$. □

Now we come to the last theorem of this paper.

THEOREM 12 (EQUIVALENT TO THEOREM 4). *Let $b = 3072$ just as in the preceding lemma. Suppose that $(X_t)_{t\geq 0}$ has drift of at least $\varepsilon \geq 0$, a variance in each step of at least 1 and $1 \leq c < n/2$. Then there is a constant $\ell$ (independent of $n, c$ and $\varepsilon$) such that, for all $\delta > 0$,*

$$P(T_{\geq} \leq n^2/\delta^{\ell \log(c)}) \geq 1 - \delta.$$

PROOF. We let $k_0 = 0$ and, for all $i$,

$$k_{i+1} = bc^2(k_i + n) + n + k_i$$

and

$$s_i = 24bc^2(k_i + n)^2.$$

We analyze the process in an infinite series of phases, starting with phase 0. For each $i$, Phase $i+1$ starts as soon as Phase $i$ ends (Phase 0 starts at time $t = 0$). Phase $i$ ends when either the goal is reached (the process is $\geq n$), the process is $\leq -k_{i+1}$, or $s_i$ steps passed in Phase $i$, whichever happens first. We call a phase *successful* if it ends with reaching the goal.

Trivially, just before the beginning of Phase $i$, the process is $\geq -k_i$. We want to apply Lemma 11, where the $n$ of the Lemma corresponds to $k_i + n$ (for the application of the Lemma, we shift the process by $k_i$). Thus, we see that each phase is successful with probability $\geq 1/2$.

Let $\delta > 0$ and let $a > 0$ be such that $2^{a-1} \geq 1/\delta \geq 2^a$. Thus, after $a$ phases, we have a success probability of at least $1 - \delta$ as desired. As, for all $i$, Phase $i$ takes at most $s_i$ steps, we get the desired result. □

# 4. COROLLARIES

In this section we will derive some useful corollaries to our theorems. We will use the terminology of the preceding section.

The first corollary is derived from Theorems 1 and 2 and gives an interval in which the hitting time is with high probability.

COROLLARY 13 (CONCENTRATION). *Suppose that $(X_t)_{t\geq 0}$ has drift of $\varepsilon = \Omega(1/n)$ and each step is bounded by $c$. Also suppose $X_0 = 0$ and let $T$ be the hitting time of $n$. Then, for each $k$, there is a $k'$ (independent of $\varepsilon$) such that*

$$P\left((n/\varepsilon)/(k'c^2 \log n) \leq T \leq k'(n/\varepsilon)c^2 \log n\right) \geq 1 - n^{-k}.$$

*Furthermore, for all $r = \omega(c^2 \log n)$,*

$$P\left((n/\varepsilon)/r \leq T \leq (n/\varepsilon)r\right) \geq 1 - n^{-\omega(1)}.$$

The next corollary is derived from Theorem 3. It shows that sufficiently negative drift gives strong (lower) bounds on the hitting time.

COROLLARY 14 (NEGATIVE DRIFT). *Suppose that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon = -\omega(c^2 \log n/n)$ and assume $c \leq n/2$. Then, for all polynomials $p$ and all $n$ large enough,*

$$P(T_{\leq} \leq p(n)) \leq \frac{1}{p(n)}.$$

As another corollary, we can get a statement similar to the Negative Drift Theorem ([OW11, OW12], see Theorem 6). This is not directly a corollary to Theorem 3, but a corollary to its proof (where it is easy to see that negative drift is only required in some bounded interval). Note that, in order to follow the notation of Theorem 6, the process attempts to go *down* (from $b$ to $a$), so what is called a "negative drift" is a positive value (away from the goal). The constraint of boundedness is lifted by conditioning on the (very likely) event of never making a large jump.

COROLLARY 15 (NEGATIVE DRIFT II). *Suppose there is an interval $[a, b] \subseteq \mathbb{R}$, two constants $\delta, \varepsilon > 0$ and, possibly depending on $\ell = b - a$, a function $r(\ell)$ satisfying $1 \leq r(\ell) = \exp(o(\sqrt[4]{\ell}))$ such that, for all $t \geq 0$, the following conditions hold.*

1. *$E(X_{t+1} - X_t \mid a < X_t < b) \geq \varepsilon$;*
2. *For all $j \geq 0$, $P(|X_{t+1} - X_t| \geq j \mid a < X_t) \leq \frac{r(\ell)}{(1+\delta)^j}$.*

*Then there is a constant $c$ such that, for $T = \min\{t \geq 0 : X_t \leq a \mid X_0 \geq b\}$, we have*

$$P(T \leq 2^{c\sqrt{\ell}}) = 2^{-\Omega(\sqrt[4]{\ell})}.$$

In comparison with the Negative Drift Theorem (Theorem 6) we see that we get only lower bounds on the hitting time ($2^{\sqrt{\ell}}$ instead of $2^{\ell}$), for the case of small $r(\ell)$, also with higher probability. However, $r(\ell)$ can increase much higher without these bounds degrading. It is not surprising that Corollary 15 works for high values of $r(\ell)$, as these high values only increase the variance logarithmically.


## Acknowledgments

I would like to thank Tobias Friedrich, Benjamin Doerr and Anton Krohmer for many useful discussions on the topic of this paper; all of them also provided valuable pointers to the literature. A further source of pointers to the literature as well as many helpful comments came from the reviewers of this paper, who have proved to be very knowledgeable in this subject.


# 5. REFERENCES


[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19:357–367, 1967.

[DG13] Benjamin Doerr and Leslie Ann Goldberg. Adaptive drift analysis. *Algorithmica*, 65:224–250, 2013.

[DJW12] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative drift analysis. *Algorithmica*, 64:673–697, 2012.

[DK13] Benjamin Doerr and Marvin Künnemann. How the (1+λ) evolutionary algorithm optimizes linear functions. In *Proc. of GECCO'13*, pages 1589–1596, 2013.

[DP09] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[FGL12] Xiequan Fan, Ion Grama, and Quansheng Liu. Hoeffding's inequality for supermartingales. *Stochastic Processes and their Applications*, 122:3545–3559, 2012.

[Haj82] Bruce Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 13:502–525, 1982.

[HY04] Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3:21–35, 2004.

[Joh10] Daniel Johannsen. *Random Combinatorial Structures and Randomized Search Heuristics*. PhD thesis, Universität des Saarlandes, 2010. Available online at http://scidok.sulb.uni-saarland.de/volltexte/2011/3529/pdf/Dissertation_3166_Joha_Dani_2010.pdf.

[LW13] Per Kristian Lehre and Carsten Witt. General drift analysis with tail bounds. *CoRR*, abs/1307.2559, 2013.

[MRC08] Boris Mitavskiy, Jonathan E. Rowe, and Chris Cannings. Preliminary theoretical analysis of a local search algorithm to optimize network communication subject to preserving the total number of links. In *Proc. of CEC'08*, pages 1484–1491, 2008.

[OW11] Pietro Simone Oliveto and Carsten Witt. Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica*, 59:369–386, 2011.

[OW12] Pietro Simone Oliveto and Carsten Witt. Erratum: Simplified drift analysis for proving lower bounds in evolutionary computation. *arXiv*, abs/1211.7184, 2012.

[OW13] Pietro Simone Oliveto and Carsten Witt. On the runtime analysis of the simple genetic algorithm. *Theoretical Computer Science*, 2013. In press.

[RS12] Jonathan E. Rowe and Dirk Sudholt. The choice of the offspring population size in the (1, $\lambda$) ea. In *Proc. of GECCO'12*, pages 1349–1356, 2012.

[Wor99] Nicholas C. Wormald. The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms*, pages 73–155, 1999.