Graceful Scaling on Uniform versus Steep-Tailed Noise

Tobias Friedrich, Timo Kötzing, Martin S. Krejca, and Andrew M. Sutton

Hasso Plattner Institute, Potsdam, Germany

Abstract. Recently, different evolutionary algorithms (EAs) have been analyzed in noisy environments. The most frequently used noise model for this was additive posterior noise (noise added after the fitness evaluation) taken from a Gaussian distribution. In particular, for this setting it was shown that the $(\mu + 1)$-EA on OneMax does not scale gracefully (higher noise cannot efficiently be compensated by higher μ).

In this paper we want to understand whether there is anything special about the Gaussian distribution which makes the $(\mu + 1)$-EA not scale gracefully. We keep the setting of posterior noise, but we look at other distributions. We see that for exponential tails the $(\mu + 1)$-EA on OneMax does also not scale gracefully, for similar reasons as in the case of Gaussian noise. On the other hand, for uniform distributions (as well as other, similar distributions) we see that the $(\mu + 1)$-EA on OneMax does scale gracefully, indicating the importance of the noise model.

Keywords: evolutionary algorithm, noisy fitness, theory

1 Introduction

A major challenge to the theoretical analysis of randomized search heuristics is developing a rigorous understanding of how they behave in the presence of uncertainty. Uncertain problems are pervasive in practice, and practitioners often rely on heuristic techniques in these settings because classical tailored approaches often cannot cope with uncertain environments such as noisy objective functions and dynamically changing problems [1, 10].

It is therefore very important to understand the effect that different properties of uncertainty have on algorithm behavior. In *stochastic* optimization, the fitness of a candidate solution does not have a deterministic value, but instead follows some given (but fixed) *noise distribution*. We are interested in understanding what properties of the noise distribution pose a direct challenge to optimization, and what problems might be overcome by different features of the algorithm. Prior work on stochastic optimization is mostly concerned with the *magnitude* of noise (usually measured by the variance). Our goal in this paper is to also understand how different *kinds* of distributions might affect optimization.

This is in contrast to most recent work on the theoretical analysis of randomized search heuristics in stochastic environments. For ant colony optimization, a series of papers studied the performance of ACO on stochastic path finding problems [13, 3, 5], see also [9, 8] for early work in this area. For evolutionary algorithms, Gießen and Kötzing [7] analyzed the $(\mu+\lambda)$-EA on noisy OneMax and LeadingOnes and found that populations make the EA robust to specific distributions of prior and posterior noise, while [2] considers non-elitist EAs and gives run time bounds in settings of partial information. None of these works aimed at showing difference between noise settings. Posterior noise from a Gaussian was considered in [6, 12] for various algorithms.

We follow [6] in their definition of what counts as a desirable property of an algorithm with respect to noisy optimization: graceful scaling. An algorithm is said to *scale gracefully* if, for any noise strength v, there is a suitable parameter for this algorithm such that optimization is possible in a time polynomial in v (a typical measure of noise strength v is the variance of the noise). We give the formal details in Definition 1.

In this paper we consider the $(\mu + 1)$-EA optimizing the classical OneMax fitness function with additive posterior noise coming from some random variable D. The case of D a Gaussian distribution was considered in [6], where the authors found that the the $(\mu+1)$-EA does not scale gracefully. In this paper we investigate what properties of D lead to graceful scaling, and what properties do not.

In Section 3 we consider the case of exponentially decaying tails in the distribution of the noise. This is similar to the case of Gaussian noise, which decays even faster. In fact, we use a similar proof to show that also in this case the $(\mu + 1)$-EA does not scale gracefully with noise.

After this we turn to another extreme case, uniform distributions. In Section 4 we show that, for noise taken from a uniform distribution, the $(\mu+1)$-EA scales gracefully. Our proof makes use of the fact that the uniform distribution is *truncated* at its lower end: there is a value k such that the noise never takes values below k, but values between k and $k + 1$ are still fairly frequent. Thus, our results generalize to all noise distributions with this property.

Our results have some interesting implications. First, if it is possible to truncate the noise distribution artificially, then this can potentially improve the run time of an EA. This is an attractive option since no noise-specific modifications need to be made to the algorithm (such as performing re-evaluations to sample the distribution and thereby reduce the variance). Second, there are settings where even very large populations do not sufficiently reduce the effect of the noise, so that other techniques are required. This serves as a cautionary tale to practitioners that increasing the population size does not always improve an EA's robustness to noise.

Before we discuss our results on exponential tails and uniform distributions in Sections 3 and 4, respectively, we introduce the algorithm and noise model in Section 2. Finally in Section 5 we summarize our findings and conclude the paper.

2 Preliminaries

In this paper we study the optimization of pseudo-Boolean functions (mapping $\{0,1\}^n$ for fixed n to real numbers). As our main test function we use OneMax, where

$$\text{OneMax}: \{0,1\}^n \to \mathbb{R}, \quad x \mapsto \|x\|_1 := |\{i : x_i = 1\}| \ .$$

As algorithm for the optimization, we consider the $(\mu+1)$-EA, defined in Algorithm 1. The $(\mu+1)$-EA is a simple mutation-only evolutionary algorithm that maintains a population of μ solutions and uses elitist survival selection.

Algorithm 1: The $(\mu+1)$-EA.

$t \leftarrow 1$;
$P_t \leftarrow \mu$ elements of $\{0,1\}^n$ uniformly at random;
while termination criterion not met **do**
    Select $x \in P_t$ uniformly at random;
    Create y by flipping each bit of x with probability $1/n$;
    $P_{t+1} \leftarrow P_t \cup \{y\} \setminus \{z\}$ where $\forall v \in P_t : f(z) \leq f(v)$;
    $t \leftarrow t+1$;

2.1 Noise Model

We consider *additive posterior noise*, meaning that the noisy fitness value is given by the actual fitness value plus some term sampled (independently for each sample) from some fixed random variable D. For OneMax and a fixed distribution, we call this noisy fitness function OneMax_D.

Let F be a family of pseudo-Boolean functions $(F_n)_{n\in\mathbb{N}}$ where each F_n is a set of functions $f : \{0,1\}^n \to \mathbb{R}$. Let D be a family of distributions $(D_v)_v$ such that for all $D_v \in D$, $\mathrm{E}(D_v) = 0$. We define F with additive posterior D-noise as the set $F[D] := \{f_n + D_v : f_n \in F_n, D_v \in D\}$.

Definition 1. *An algorithm A* scales gracefully with noise *on $F[D]$ if there is a polynomial q such that, for all $g_{n,v} = f_n + D_v \in F[D]$, there exists a parameter setting p such that $A(p)$ finds the optimum of f_n using at most $q(n,v)$ calls to $g_{n,v}$.*

We will need the following result regarding noisy optimization from [6] for our negative results.

Theorem 2 ([6]). *Let $\mu \geq 1$ and D a distribution on $\mathbb{R}$. Let Y be the random variable describing the minimum over μ independent copies of D. Suppose*

$$\Pr(Y > D + n) \geq \frac{1}{2(\mu+1)} \ .$$

Consider optimization of OneMax_D *by* $(\mu+1)$*-EA. Then, for* μ *bounded from above by a polynomial, the optimum will* not *be evaluated after polynomially many iterations w.h.p.*

Intuitively, whenever the selection pressure is so weak that selection is almost uniform (which would mean a probability of $1/(\mu+1)$ for choosing any particular individual), optimization will not succeed.

2.2 Drift Analysis

For the theoretical analysis we will use the following *drift theorem*.

Theorem 3 (Multiplicative Drift [4]). *Let* $(X_t)_{t\geq 0}$ *be a sequence of random variables over* $\mathbb{R}_{\geq 0}$*. Let* T *be the random variable that denotes the earliest point in time* $t \leq 0$ *such that* $X_t < 1$*. If there exist* $c > 0$ *such that, for all* a*,*

$$E[X_t - X_{t+1} \mid T > t,\, X_t = a] \geq c\, a\ ,$$

then, for all a*,*

$$E[T \mid X_0 = a] \leq \frac{1 + \ln(a)}{c}\ .$$

3 Exponential Tails

In this section we consider noise taken from a random variable D that decays exponentially fast, i.e., we assume

$$F(t) := \Pr(D < k) = \frac{1}{2}\mathrm{e}^{ct} \quad \text{if } t \leq 0 \text{ and}$$
$$F(t) := 1 - \frac{1}{2}\mathrm{e}^{-ct} \quad \text{if } t > 0\,,$$

for some constant c. By taking the derivative of F, we get the probability mass function p of D, i.e.,

$$p(t) = F'(t) = \frac{c}{2}\mathrm{e}^{ct} \quad \text{if } t \leq 0 \text{ and}$$
$$p(t) = \frac{c}{2}\mathrm{e}^{-ct} \quad \text{if } t > 0\,.$$

This is basically a symmetric variant of the exponential distribution. Note that D is a distribution, since F is non-negative and monotonically increasing, $\lim_{t\to-\infty} F(t) = 0$, and $\lim_{t\to\infty} F(t) = 1$. Because p is symmetric around 0, it follows that D has mean 0.

We calculate the variance of D:

$$\begin{aligned}\mathrm{Var}(D) &= \int_{-\infty}^{\infty} t^2 p(t)\mathrm{d}t = \frac{c}{2}\left(\int_{-\infty}^{0} t^2\mathrm{e}^{ct}\mathrm{d}t + \int_{0}^{\infty} t^2\mathrm{e}^{-ct}\mathrm{d}t\right)\\ &= c\int_{-\infty}^{0} t^2\mathrm{e}^{ct}\mathrm{d}t = c\left[\frac{\left(2 - 2ct + t^2c^2\right)\mathrm{e}^{ct}}{c^3}\right]_{-\infty}^{0}\\ &= \frac{2}{c^2} =: \sigma^2\ ,\end{aligned}$$

where the integral can be computed by integrating by parts twice. This leads to

$$F(t) = \frac{1}{2}\mathrm{e}^{\sqrt{2}\frac{t}{\sigma}} \qquad \text{if } t \le 0\,;\text{ and}$$
$$F(t) = 1 - \frac{1}{2}\mathrm{e}^{-\sqrt{2}\frac{t}{\sigma}} \quad \text{if } t > 0\,.$$

We now want to show that, in this setting and for sufficiently large variance, the $(\mu+1)$-EA is not successful. We will start with the case of $\mu \in \{1,2\}$, as this case is not covered by our main theorem of this section (Theorem 5 below). The proof is instructive, since its structure is similar to the proof of Theorem 5, while it is a bit simpler in the details.

Proposition 4. *Consider optimization of* OneMax_D *by the* $(\mu+1)$*-EA with* $\mu \in \{1,2\}$. *Suppose* $\sigma^2 = \omega(n^2)$. *Then the optimum will* not *be evaluated after polynomially many iterations w.h.p.*

Proof. We set up to use Theorem 2. Thus, let $t^- < 0$ and $t^+ > 0$ be such that $\Pr(D < t^-) = \Pr(D \ge t^+) = 1/4$. Hence, $\Pr(t^- \le D < 0) = \Pr(0 \le D < t^+) = 1/4$ because D is symmetric around 0.

We consider D and μ copies of it: D_i^*, for $i = 1,\ldots,\mu$. We want to bound

$$\Pr\left(\min_{i=1,\ldots,\mu}\{D_i^*\} > D + n\right) = \Pr\left(\bigwedge_{i=1}^{\mu} D_i^* > D + n\right) .$$

We lower-bound the above probability as follows:

$$\begin{aligned}\Pr\left(\bigwedge_{i=1}^{\mu} D_i^* > D+n\right) \ge{}& \Pr(D < t^-)\prod_{i=1}^{\mu}\Pr(D_i^* \ge t^- + n) + \\ & \Pr(t^- < D < 0)\prod_{i=1}^{\mu}\Pr(D_i^* \ge n) + \\ & \Pr(0 < D < t^+)\prod_{i=1}^{\mu}\Pr(D_i^* \ge t^+ + n) \\ ={}& \frac{1}{4}\left(\prod_{i=1}^{\mu}\Pr(D_i^* \ge t^- + n) + \prod_{i=1}^{\mu}\Pr(D_i^* \ge n) + \right. \\ & \left.\prod_{i=1}^{\mu}\Pr(D_i^* \ge t^+ + n)\right) .\end{aligned}$$

Thus, we have to bound the probabilities of the form $\Pr(D_i^* \ge a + n)$. We do so by showing $\Pr(D_i^* \ge a+n) \ge \big(1 - \mathrm{o}(1)\big)\Pr(D_i^* \ge a)$. First, consider $a \ge 0$.

$$\begin{aligned}\Pr(D_i^* \ge a+n) &= \frac{1}{2}\mathrm{e}^{-\sqrt{2}\frac{a+n}{\sigma}} = \frac{1}{2}\mathrm{e}^{-\sqrt{2}\left(\frac{a}{\sigma}+\frac{n}{\sigma}\right)} = \frac{1}{2}\mathrm{e}^{-\sqrt{2}\left(\frac{a}{\sigma}+\mathrm{o}(1)\right)} \\ &= \big(1-\mathrm{o}(1)\big)\Pr(D_i^* \ge a) .\end{aligned}$$

Note that if $a = t^-$, we get $t^- + n < 0$, because we assume that $\Pr(D < t^-) = 1/4$, which means that $t^- = -\Theta(\sigma) = -\omega(n^2)$.

$$\begin{aligned}\Pr(D_i^* \geq t^- + n) &= 1 - \Pr(D_i^* < t^- + n) = 1 - \frac{1}{2}\mathrm{e}^{\sqrt{2}\frac{t^- + n}{\sigma}} \\ &= 1 - \bigl(1 + \mathrm{o}(1)\bigr)\mathrm{e}^{\sqrt{2}\frac{t^-}{\sigma}} = \bigl(1 - \mathrm{o}(1)\bigr)\Pr(D_i^* \geq t^-) \ .\end{aligned}$$

This results in

$$\Pr\left(\bigwedge_{i=1}^{\mu} D_i^* > D + n\right) \geq \bigl(1 - \mathrm{o}(1)\bigr)\frac{1}{4}\left(\left(\frac{3}{4}\right)^{\mu} + \left(\frac{2}{4}\right)^{\mu} + \left(\frac{1}{4}\right)^{\mu}\right) \ ,$$

which is at least $1/\bigl(2(\mu+1)\bigr)$ for $\mu \in \{1, 2\}$. Applying Theorem 2 completes the proof. □

We now turn to the more general case.

Theorem 5. *Consider optimization of* OneMax_D *by the* $(\mu+1)$*-EA with* $\mu \geq 3$ *and* μ *bounded from above by a polynomial in* n*. Suppose* $\sigma^2 = \omega(n^2)$*. Then the optimum will* not *be evaluated after polynomially many iterations w.h.p.*

Proof. This proof follows the ideas of the one of Corollary 6 from [6]. Let $a = \omega(1)$ be such that $\sigma^2 \geq (na)^2$.

Again, we want to use Theorem 2, thus, let Y be the minimum of μ independent copies of D, whereas D is a distribution as defined above. We want to bound $\Pr(D + n < Y)$. Hence, we choose two points $t_0 < t_1 < 0$ such that $\Pr(D < t_0) = 0.7/\mu$ and $\Pr(D < t_1) = 1.4/\mu$. Note that $\Pr(D < t_0) < \Pr(D < t_1) < 1/2$, since $\mu \geq 3$. Thus, t_0 and t_1 actually exist.

We define the following two disjoint events that are a subset of the event $D + n < Y$:

A: The event that $D < t_0 - n$ and $t_0 < Y$.
B: The event that $t_0 - n < D < t_1 - n$ and $t_1 < Y$.

We first focus on bounding $\Pr(D < t_0 - n)$ and do so showing that $t_0 \leq -na/32$ holds via contraposition.

Assume that $t_0 > -na/32$ (still $t_0 < 0$). Because we assume $\sigma \geq na$ as well, we get that $s := -t_0/\sigma < 1/32$. Due to the definition of D we get

$$\Pr(D < t_0) = F(t_0) = \frac{1}{2}\mathrm{e}^{-\sqrt{2}s} > \frac{1}{2}\mathrm{e}^{-\frac{\sqrt{2}}{32}} > \frac{0.7}{3} \ .$$

Since we assume $\mu \geq 3$, this contradicts the definition of t_0. Hence, the bound $t_0 \leq -na/32$ holds which is equivalent to $t_0(1+32/a) \leq t_0 - n$. Thus, we estimate

$$\begin{aligned}\Pr(D < t_0 - n) &\geq \Pr\left(t_0\left(1 + \frac{32}{a}\right)\right) = \frac{1}{2}\mathrm{e}^{\sqrt{2}\frac{t_0\left(1+\frac{32}{a}\right)}{\sigma}} = \frac{\mathrm{e}^{-\mathrm{o}(1)}}{2}\mathrm{e}^{\sqrt{2}\frac{t_0}{\sigma}} \\ &= \bigl(1 - \mathrm{o}(1)\bigr)\frac{1}{2}\mathrm{e}^{\sqrt{2}\frac{t_0}{\sigma}} = \bigl(1 - \mathrm{o}(1)\bigr)\Pr(D < t_0) \ .\end{aligned}$$

Because $\Pr(D < t_0 - n) \leq \Pr(D < t_0)$ holds trivially, we have $\Pr(D < t_0 - n) = \big(1 - \mathrm{o}(1)\big)\Pr(D < t_0)$.

We now bound $\Pr(t_0 - n < D < t_1 - n) = \Pr(D < t_1 - n) - \Pr(t_0 - n < D)$, where we are left with bounding $\Pr(D < t_1 - n)$. We do so analogously to the calculations before. This time, we bound $t_1 \leq -na/64$, since assuming $t_1 > -na/64$ leads to

$$\Pr(D < t_1) > \frac{1}{2}\mathrm{e}^{-\frac{\sqrt{2}}{64}} > \frac{1.4}{3} \ .$$

All the remaining calculations can be done as before, and we get $\Pr(t_0 - n < D < t_1 - n) = \big(1 - \mathrm{o}(1)\big)1.4/\mu - \big(1 - \mathrm{o}(1)\big)0.7/\mu = \big(1 - \mathrm{o}(1)\big)0.7/\mu$. Overall, we have

$$\begin{aligned}\Pr(Y > D + n) \geq \Pr(A) + \Pr(B) &= \big(1 - \mathrm{o}(1)\big)\frac{0.7}{\mu}\left(\left(1 - \frac{0.7}{\mu}\right)^{\mu} + \left(1 - \frac{1.4}{\mu}\right)^{\mu}\right) \\ &\geq \big(1 - \mathrm{o}(1)\big)^2\frac{0.7}{\mu}\left(\mathrm{e}^{-0.7} + \mathrm{e}^{-1.4}\right) \geq \frac{1}{2\mu} \geq \frac{1}{2(\mu + 1)} \ .\end{aligned}$$

Applying Theorem 2 completes the proof. □

The statement of Theorem 5 is basically that if the standard deviation of the noise is asymptotically larger than the greatest OneMax value (n), the noise will dominate and optimization will fail. The proof idea can be expanded to the case when $\Pr(D < t_0) = \Omega(1)$ if $t_0 = -\Omega(\sigma)$, i.e., there is at least a constant probability to deviate by at least one standard deviation from the mean. In such a case the OneMax value is irrelevant, since it will easily be dominated by the noise.

Overall, we get the following statement regarding graceful scaling.

Corollary 6. *The $(\mu + 1)$-EA does* not *scale gracefully on OneMax with additive posterior noise from a distribution with exponential tails as given above (parametrized in the variance).*

4 Truncated Distributions

In this section we consider *truncated* distributions; these distributions are a generalization of uniform distributions, which capture the essence of what our proofs need to show that the $(\mu + 1)$-EA can scale gracefully. Truncated distributions are distributions whose density functions vanish above (respectively, below) some point k and whose mass near that point is bounded from below by some value q.

Definition 7. *Let D be a random variable. If there are $k, q \in \mathbb{R}$ such that*

$$\Pr(D > k) = 0 \wedge \Pr\big(D \in (k - 1, k]\big) \geq q \ ,$$

then we call D upper q-truncated. *Analogously, we call D* lower q-truncated *if there is a $k \in \mathbb{R}$ with*

$$\Pr(D < k) = 0 \wedge \Pr\big(D \in [k, k + 1)\big) \geq q \ .$$

From [14] we know that the run time of the $(\mu+1)$-EA on OneMax is $\mathcal{O}(\mu n + n \log n)$ when no noise is present. The following theorem looks at the optimization behavior for truncated noise and gives a slightly weaker run time bound in the presence of noise which is suitably truncated.

Theorem 8. *Let $\mu \geq 1$ and let D be lower $2\log(n\mu)/\mu$-truncated. Consider optimization of OneMax_D by the $(\mu+1)$-EA. Then the optimum will be evaluated, in expectation, after $\mathcal{O}(\mu n \log n)$ iterations.*

Proof. We argue with drift on the number of 1s in the search point with the most number of 1s. If the search point with the most number of 1s is never removed within the first $\mathcal{O}(\mu n \log n)$ iterations, then multiplicative drift (Theorem 3) will give us the result. If any other search point evaluates in the minimal bracket $[k, k+1)$, then the best search point is safe (if there are multiple best, then it is safe anyway). The probability that none evaluates in the minimal bracket is at most

$$\begin{aligned}\Pr\left(D \notin [k, k+1)\right)^{\mu} &\leq \left(1 - \frac{2\log(n\mu)}{\mu}\right)^{\mu} \\ &\leq \exp\bigl(-2\log(n\mu)\bigr) \\ &\leq \mathcal{O}\left(\frac{1}{(n\mu)^2}\right) .\end{aligned}$$

Thus, the expected number of iterations until the best search point decreases in number of 1s is $\omega(\mu n \log n)$ iterations. Since this holds from any starting configuration, the result follows from the fact that the optimum can be found by iteratively increasing the best individual in $\mathcal{O}(\mu n \log n)$ iterations. □

As a corollary to Theorem 8, we turn the statement of the previous theorem around and show how large a population is required for efficient optimization in the presence of truncated noise.

Corollary 9. *Let D be lower q-truncated. Then, for all $\mu \geq 3q^{-1}\log(nq^{-1})$, in the optimization of OneMax_D by $(\mu+1)$-EA, the optimum will be evaluated, in expectation, after $\mathcal{O}(\mu n \log n)$ iterations.*

Corollary 10. *Let $\mu, r \geq 1$. Consider optimization of OneMax with reevaluated additive posterior noise uniformly from $[-r, r]$ by $(\mu+1)$-EA without crossover. Then the optimum*

1. *will be evaluated within $\mathcal{O}(\mu n \log n)$ iterations in expectation if $r \leq \mu/\bigl(4\log(n\mu)\bigr)$;*
2. *will* not *be evaluated within polynomially many iterations w.h.p. if $r \geq n(\mu+1)$.*

Proof. Let D be the uniform distribution on $[-r, r]$.

Regarding the first claim, we note that D is lower $1/(2r)$-truncated, so the result follows from Theorem 8.

For the second claim we want to use Theorem 2. Define

$$f\colon [-r,r] \to [-r,r]+n, \quad x \mapsto \begin{cases} x & \text{if } x \geq -r+n\,; \\ r+n-x & \text{otherwise.} \end{cases}$$

Then we have that $D+n$ and $f(D)$ have the same distribution. Let Y be the minimum over μ independent copies of D. Due to symmetry, we have $\Pr(Y > D) = 1/(\mu+1)$. Thus, we have

$$\begin{aligned}
\Pr(Y > D+n) &= \Pr\big(Y > f(D)\big) \\
&\geq \Pr(Y > f(D) \wedge D > -r+n) \\
&= \Pr(Y > D \wedge D > -r+n) \\
&\geq \Pr(Y > D) - \Pr(D \leq -r+n) \\
&= \frac{1}{\mu+1} - \frac{n}{2r} \\
&\geq \frac{1}{2(\mu+1)}\ .
\end{aligned}$$

The result now follows with Theorem 2. □

From this we get the result regarding graceful scaling on uniform noise.

Corollary 11. *The* $(\mu+1)$*-EA scales gracefully on OneMax with additive posterior noise from the uniform distribution on* $[-r,r]$*.*

5 Summary

In this work we saw indications that the shape of the distributions plays an important role in settings with noisy fitness functions. For the case of the uniform distributions, we can give bounds for when optimization is successful and for when it is not. The analysis is significantly more complicated for other distributions, but our results still suggest that more even distributions make optimization easier.

It seems that further results are hard to come by and probably require a new way of dealing with diversity of populations (a long standing open problem). In particular, the only theorem for lower bounds we have is Theorem 2, which makes significant worst-case assumptions about the diversity of the population. Similarly, all upper bounds usually make worst-case assumptions on the diversity, but this time in the other direction (namely that the population is clustered, while Theorem 2 is based on the assumption that a single good individual works against many bad individuals). This also explains the gap in the bounds for the uniform distribution (Corollary 10). An alternative route could be in adapting drift theorems specifically for populations [11].

It is open whether we need better tools for showing lower or upper bounds; a useful first step could thus be to conjecture run time bounds based on empirical evidence and analyzing also the spread of the population carefully in dependence on the distribution.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 618091 (SAGE) and the German Research Foundation (DFG) under grant agreement no. FR 2988 (TOSU).

References

1. L. Bianchi, M. Dorigo, L. Gambardella, and W. Gutjahr. A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8:239–287, 2009.
2. D.-C. Dang and P. K. Lehre. Evolution under partial information. In *Proc. of GECCO'14*, pages 1359–1366, 2014.
3. B. Doerr, A. Hota, and T. Kötzing. Ants easily solve stochastic shortest path problems. In *Proc. of GECCO'12*, pages 17–24, 2012.
4. B. Doerr, D. Johannsen, and C. Winzen. Multiplicative drift analysis. *Algorithmica*, 64(4):673–697, 2012.
5. M. Feldmann and T. Kötzing. Optimizing expected path lengths with ant colony optimization using fitness proportional update. In *Proc. of FOGA'13*, pages 65–74, 2013.
6. T. Friedrich, T. Kötzing, M. S. Krejca, and A. M. Sutton. The benefit of recombination in noisy evolutionary search. In *Proc. of ISAAC'15*, pages 140–150, 2015.
7. C. Gießen and T. Kötzing. Robustness of populations in stochastic environments. In *Proc. of GECCO'14*, pages 1383–1390, 2014.
8. W. Gutjahr. A converging ACO algorithm for stochastic combinatorial optimization. In *Proc. of SAGA'03*, pages 10–25, 2003.
9. W. Gutjahr and G. Pflug. Simulated annealing for noisy cost functions. *Journal of Global Optimization*, 8:1–13, 1996.
10. Y. Jin and J. Branke. Evolutionary optimization in uncertain environments—a survey. *IEEE Trans. on Evol. Comp.*, 9:303–317, 2005.
11. P. K. Lehre. Negative drift in populations. In *Proc. of PPSN'10*, pages 244–253, 2010.
12. A. Prügel-Bennett, J. E. Rowe, and J. Shapiro. Run-time analysis of population-based evolutionary algorithm in noisy environments. In *Proc. of FOGA'15*, pages 69–75, 2015.
13. D. Sudholt and C. Thyssen. A simple ant colony optimizer for stochastic shortest path problems. *Algorithmica*, 64:643–672, 2012.
14. C. Witt. Runtime analysis of the $(\mu + 1)$ EA on simple pseudo-boolean functions. *Evolutionary Computation*, 14:65–86, 2006.