Robustness of Ant Colony Optimization to Noise

Tobias Friedrich   friedrich@hpi.de
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Timo Kötzing   timo.koetzing@hpi.de
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Martin S. Krejca   martin.krejca@hpi.de
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Andrew M. Sutton   andrew.sutton@hpi.de
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Abstract

Recently, ant colony optimization (ACO) algorithms have proven to be efficient in uncertain environments, such as noisy or dynamically changing fitness functions. Most of these analyses have focused on combinatorial problems such as path finding. We rigorously analyze an ACO algorithm optimizing linear pseudo-Boolean functions under additive posterior noise. We study noise distributions whose tails decay exponentially fast, including the classical case of additive Gaussian noise. Without noise, the classical $(\mu + 1)$ EA outperforms any ACO algorithm, with smaller μ being better; however, in the case of large noise, the $(\mu + 1)$ EA fails, even for high values of μ (which are known to help against small noise). In this article, we show that ACO is able to deal with *arbitrarily large* noise in a graceful manner; that is, as long as the evaporation factor ρ is small enough, dependent on the variance σ^2 of the noise and the dimension n of the search space, optimization will be successful. We also briefly consider the case of *prior noise* and prove that ACO can also efficiently optimize linear functions under this noise model.

Keywords

Ant colony optimization, Noisy Fitness, Theory, Run time analysis.

1   Introduction

Ant colony optimization (ACO) is a metaheuristic for designing randomized general-purpose optimization algorithms inspired by the foraging behavior of ant colonies. ACO has been successfully applied as a heuristic technique for solving combinatorial optimization problems.

In real-world optimization problems, sometimes a large degree of uncertainty is present, due to the complexity of candidate solution generation, noisy measurement processes, and rapidly changing problem environments. Empirically, ACO seems particularly well-suited to uncertain problems, due to its dynamic and distributed nature, and in some cases it can outperform classical state-of-the-art approaches on dynamic network routing problems (Di Caro et al., 2008). We will focus on a version of the *Max–Min Ant System* (MMAS; Stützle and Hoos, 2000) applied to pseudo-Boolean optimization (i.e., optimization where solutions are coded as bit strings).

Jin and Branke (2005) surveyed a number of sources of uncertainty that randomized search heuristics must often deal with in practice: (1) noisy objective functions, (2) dynamically changing problems, (3) approximation errors in the objective function, and (4) a requirement that an optimal solution must be robust to changes in design variables and environmental parameters that occur after optimization is complete. Arguably, the two most important sources of uncertainty are (1) and (2) above, namely, *stochastic* problems and *dynamic* problems (see also Bianchi et al., 2009, for a recent survey). In stochastic problems, the objective function value of a search point follows a random distribution, and that distribution does not change over time. In dynamic problems, the evaluation of fitness is deterministic but changes over time.

In order to address these practical issues, the theoretical analysis of randomized search heuristics under uncertainty has recently gained momentum. For example, a number of recent studies have rigorously analyzed the performance of evolutionary algorithms in stochastic environments (Dang and Lehre, 2014; Gießen and Kötzing, 2014). For ant colony optimization, a series of articles have considered the performance of ACO on single-destination shortest paths (SDSP) problems with stochastic weights. This work was initiated by Sudholt and Thyssen (2012) and later followed up by Doerr et al. (2012a), who showed that by augmenting the ant system with a reevaluation strategy on the best-so-far solution, many of the difficulties with noise discovered by Sudholt and Thyssen could be overcome. Feldmann and Kötzing (2013) showed that an ant system that uses a fitness-proportional update rule (called *MMAS-fp*) can efficiently optimize SDSP on graphs with stochastic weights. MMAS-fp is closer to systems that are used by practitioners (Stützle and Hoos, 2000) and is the ant system variant that we analyze in the present study.

For the optimization of functions over bit strings, analyses of ACO suggest that it often performs worse than evolutionary algorithms (EAs) and simple hill-climbers in a noise-free setting (Kötzing et al., 2011). On the other hand, ACO can outperform EAs on dynamic problems (Kötzing and Molter, 2012; Lissovoi and Witt, in press). So far, the question of how robust ACO is to noisy evaluation on pseudo-Boolean optimization remains unanswered.

The goal of this study is to observe the robustness of ACO to noise on a class of simple objective functions. In particular, we are interested in the *scalability* of the run time of the algorithm as a function of noise intensity (measured by variance). We study the algorithm on linear functions; a linear function maps a bit string x to $\sum_{i \leq n} x_i w_i$, with a predefined sequence of weights w_i. If we set all weights to 1, we recover the definition of the important OneMax test function. Note that we only allow positive weights, as fitness-proportional pheromone updates rely on nonnegative fitness (and weights of 0 do not contribute to the fitness).

The main result of this article is given in Section 3, where we show that robustness can be achieved for additive posterior noise from a Gaussian distribution; for all variances σ^2, there is a parameter setting such that MMAS-fp on $f + \mathcal{N}(0, \sigma^2)$ is successful in time poly(n, σ^2). Thus, we say that MMAS-fp handles Gaussian noise *gracefully*. Such a graceful scaling *cannot* be achieved by the $(\mu + 1)$ EA, as was shown by Friedrich et al. (2015a).

Additionally, we show that a similar statement holds for an ACO algorithm that imposes max–min bounds on the pheromone values. Rather than relying on the standard technique of *pheromone freezing*, employed by many early works on the mathematical analysis of ACO algorithms, we instead achieve this by showing that all pheromones *drift* in the right direction (in particular, they will never be close to any pheromone bound on the wrong side of the spectrum); see Corollary 7.

Algorithm 1 MMAS-fp with parameter ρ for optimizing $f : \{0,1\}^n \to \mathbb{R}$

---

1 **for** $i \in \{1, \dots, n\}$ **do**
2 $\quad \lfloor \tau_i \leftarrow \frac{1}{2}$;
3 **while** optimum not found **do**
4 $\quad$ **for** $i \in \{1, \dots, n\}$ **do**
5 $\quad \quad \lfloor x_i \leftarrow 0$ with probability τ_i, $x_i \leftarrow 1$ with probability $1 - \tau_i$;
6 $\quad$ **for** $i \in \{1, \dots, n\}$ **do**
7 $\quad \quad$ **if** $x_i = 0$ **then**
8 $\quad \quad \quad \lfloor \tau_i \leftarrow \min\left\{\tau_i\left(1 - \rho\frac{f(x)}{W}\right) + \rho\frac{f(x)}{W}, u\right\}$;
9 $\quad \quad$ **else**
10 $\quad \quad \quad \lfloor \tau_i \leftarrow \max\left\{\tau_i\left(1 - \rho\frac{f(x)}{W}\right), \ell\right\}$;

In Section 4 we extend our findings to other noise models and show that we can also achieve the same robustness in the presence of other additive noise distributions (which fulfill a certain restriction), as well as with the model of prior noise from Droste (2004). We discuss our findings and conclude the report in Section 5.

This article extends an earlier conference version (Friedrich et al., 2015b) by generalizing the results on OneMax to arbitrary linear functions.

2 Preliminaries

We consider the optimization of pseudo-Boolean functions—that is, functions $f : \{0,1\}^n \to \mathbb{R}^{\geq 0}$, which we call *fitness functions*. In the following, n always denotes the dimension of the solution space $\{0,1\}^n$. For any bit string $x \in \{0,1\}^n$, the function value $f(x)$ is called the *fitness of* x, and we are interested in finding a solution x^* such that $f(x^*)$ is maximal.

The fitness functions we are going to investigate are linear, following the notation of Droste et al. (2002)—that is, given a vector w of length n of positive weights, $f(x) = \sum_i w_i x_i$, where $i = \{1, \dots, n\}$; thus, the all-ones string is the global optimum. In this study, index variables i refer to the set $\{1, \dots, n\}$ if not stated otherwise.

For a given weight vector w, we let $w_{\min}$ denote the minimum weight, $w_{\max}$ the maximum weight, and $W = \sum_i w_i$ the sum of all weights. We let $\|x\|_0$ denote the number of zeros in x and $\|w(x)\|_0$ denote the sum of weights w_i with $x_i = 0$. Analogously, we define $\|x\|_1$ and $\|w(x)\|_1$. Note that $W = \|w(x)\|_0 + \|w(x)\|_1$ for all x.

Throughout the article, we consider a noisy version of f. The noise added to the actual fitness will be, at first, a normally distributed random variable D with mean 0 and variance σ^2. More formally, for a given deterministic fitness function f, $f_{\sigma^2}(x) := f(x) + D$ with $D \sim \mathcal{N}(0, \sigma^2)$.

We say that an event A occurs *with high probability* if $\overline{A}$ only occurs with a probability less than the inverse of any polynomial—that is, $\Pr(\overline{A}) = o(1/\mathrm{poly}(n))$ for any polynomial $\mathrm{poly}(n)$.

2.1 Algorithms

Our main algorithm of interest is MMAS-fp (Alg. 1), an ACO algorithm with a so-called *fitness-proportional* update rule.
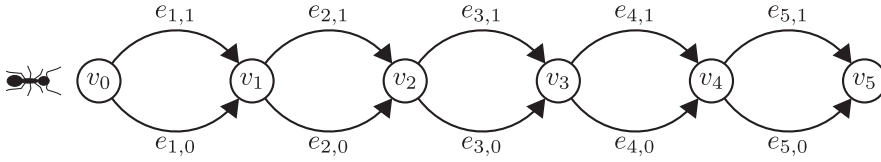
Figure 1: Construction graph for pseudo-Boolean optimization with $n = 5$ bits.

To find an optimal solution, MMAS-fp maintains a vector $\tau \in [\ell, u]^n$, the components of which are called *pheromones*, where $0 \leq \ell < 1/2$ and $1 \geq u > 1/2$ are the lower and upper bounds for each pheromone, respectively; we will assume $\ell + u = 1$. Further, we let $\tau^{(t)}$ with $t \in \mathbb{N}$ denote the pheromone vector in the t-th iteration of the algorithm.

MMAS-fp starts off with every pheromone $\tau_i^{(0)} = 1/2$ and iteratively generates solutions $x^{(t)} \in \{0, 1\}^n$ until $x^{(t)}$ is optimal. These solutions $x^{(t)}$ are sampled according to $\tau^{(t)}$ in the manner $\forall i : \Pr(x_i^{(t)} = 0) = \tau_i^{(t)} \wedge \Pr(x_i^{(t)} = 1) = 1 - \tau_i^{(t)}$. Thus, our pheromone vector corresponds to the probabilities of sampling 0s, and we desire the pheromones to go *down*. This is equivalent to a definition of using a simulated ant walk on the construction graph shown in Figure 1, where choosing an edge $e_{i,1}$ means that $x_i^{(t)} = 1$ and choosing an edge $e_{i,0}$ means that $x_i^{(t)} = 0$. Our pheromones $\tau_i^{(t)}$ correspond to the pheromones on the edges $e_{i,0}$.

Afterward, the pheromone vector for the next iteration is updated according to the fitness of the sampled solution:

$$
\tau_i^{(t+1)} := \begin{cases} \min\left\{\tau_i^{(t)}\left(1 - \rho\dfrac{f(x^{(t)})}{W}\right) + \rho\dfrac{f(x^{(t)})}{W}, u\right\} & : \ x_i^{(t)} = 0, \\[2em] \max\left\{\tau_i^{(t)}\left(1 - \rho\dfrac{f(x^{(t)})}{W}\right), \ell\right\} & : \ x_i^{(t)} = 1. \end{cases}
$$

Note that increasing $\tau_i^{(t)}$ is the same as decreasing $1 - \tau_i^{(t)}$, and vice versa for decreasing $\tau_i^{(t)}$. Hence, MMAS-fp does not prefer 1s over 0s, or the other way around: It is *unbiased* in the sense of Lehre and Witt (2010).

The variable $\rho \in [0, 1]$ is the so-called *evaporation factor*; ρ is a parameter of the algorithm. Intuitively, ρ regulates the impact of the fitness of a sampled solution on the corresponding pheromone update. To maintain meaningful values of $\tau_i^{(t)}$ and sensible updates, it must hold that $\rho f(x)/W = o(1)$ for all x.

We assume that MMAS-fp knows W in its update rule for theoretical purposes. In reality, one either has to have an upper bound on W or choose ρ accordingly. We further assume that W is a polynomial in n and that $w_{\min} = \Omega(1)$, so that a *polynomial* run time is possible with high probability. The proofs given could be extended to arbitrary values of W, but this would need case distinctions with respect to the growth of the weights and would make the proofs overly technical.

Note that MMAS-fp does not keep a best-so-far solution but, instead, always updates its pheromone vector regardless of the quality of the solution sampled. The update is proportional to the fitness of the sampled solution, hence the suffix *fp*. If an updated pheromone is not capped due to the pheromone bounds ℓ or u, we speak of a *normal update*.

In the following, we will denote a pheromone vector $\tau^{(t)}$ simply as τ if we are not interested in t. Likewise, x denotes the solution sampled by τ, and τ' denotes the pheromone vector after the update according to x.

The other algorithm we consider is ACO-fp. It is a special case of MMAS-fp where $\ell = 0$ and $u = 1$; that is, only trivial pheromone bounds are enforced.

2.2 Tools Used

The proofs in this article rely heavily on *drift theory*, a tool that bounds the hitting times of stochastic processes with a bias (called *drift*).

We first state one drift theorem, as well as two concentration bounds regarding drift, that we will use in our proofs.

THEOREM 1 (MULTIPLICATIVE DRIFT) (Doerr et al., 2012b): *Let $(X_t)_{t \in \mathbb{N}}$ be nonnegative random variables over $\mathbb{R}$, each with finite expectation, and let $T = \min\{t \mid X_t < 1\}$.*

Suppose there exists an $\varepsilon > 0$ such that, for all t,

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

Then

$$E(T \mid X_0) \leq \frac{1 + \ln X_0}{\varepsilon}.$$

THEOREM 2 (TAIL BOUNDS FOR MULTIPLICATIVE DRIFT) (Doerr and Goldberg, 2013; Lehre and Witt, 2014): *Let $(X_t)_{t \in \mathbb{N}}$ be nonnegative random variables over $\mathbb{R}$, each with finite expectation, and let $T = \min\{t \mid X_t < 1\}$.*

Suppose there exists an $\varepsilon > 0$ such that, for all t,

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

Then, for all $k > 0$,

$$Pr\left(T > \frac{k + \ln X_0}{\varepsilon} \mid X_0\right) \leq e^{-k}.$$

THEOREM 3 (NEGATIVE DRIFT) (Kötzing, 2014): *Let $(X_t)_{t \in \mathbb{N}}$ be random variables over $\mathbb{R}$, each with finite expectation, let $d > 0$, and let $T = \min\{t : X_t \geq d \mid X_0 \leq 0\}$. Suppose there are $s, 0 < s < d$, and $\varepsilon < 0$ such that, for all t,*

1. $E(X_{t+1} - X_t \mid X_t, t < T) \leq \varepsilon$, *and*

2. $|X_t - X_{t+1}| < s$.

Then, for all $t \in \mathbb{N}$,

$$Pr(T \leq t) \leq t \exp\left(-\frac{d|\varepsilon|}{16s^2}\right).$$

The following proposition gives tail bounds for our noise $D \sim \mathcal{N}(0, \sigma^2)$ by using standard estimates of the Q-function (Madhow, 2008).

PROPOSITION 4: *Let D be a zero-mean Gaussian random variable with variance σ^2. For all $t > 0$, we have*

$$Pr(D < -t) \leq \frac{1}{2} \exp\left(\frac{-t^2}{2\sigma^2}\right).$$

3   Gaussian Noise

In this section we are going to bound the run time of MMAS-fp for optimizing f_{σ^2}, which will be denoted as f, for convenience, in the proofs but not in the statements of the lemmas and so forth.

First off, we bound the noise and thus the difference of two consecutive pheromones with high probability. Since the noise is unbounded, an update can exceed the pheromone bounds by a large amount. By bounding the noise with high probability, we make sure that such updates happen only very rarely. We then set ρ such that it can compensate for likely noise levels.

LEMMA 5: *Consider MMAS-fp optimizing f_{σ^2}. Then, for any $m = \omega(\sigma \log n)$, we can bound $|\tau_i - \tau_i'|$ for any pheromone τ_i with high probability as follows:*

$$|\tau_i - \tau_i'| \leq \frac{\rho}{W}(W + m).$$

PROOF: We make a case distinction for τ_i' and assume $\ell = 0$ and $u = 1$, since capping a pheromone would only make the difference smaller.

1. $\tau_i' = \tau_i\left(1 - \rho\frac{f(x)}{W}\right)$:

$$|\tau_i - \tau_i'| = \left|\tau_i - \tau_i\left(1 - \rho\frac{f(x)}{W}\right)\right| = \frac{\rho}{W}|\overset{\leq 1}{\overbrace{\tau_i}}f(x)| \leq \frac{\rho}{W}|f(x)|.$$

2. $\tau_i' = \tau_i\left(1 - \rho\frac{f(x)}{W}\right) + \rho\frac{f(x)}{W}$:

$$|\tau_i - \tau_i'| = \left|\tau_i - \left(\tau_i\left(1 - \rho\frac{f(x)}{W}\right) + \rho\frac{f(x)}{W}\right)\right| = \frac{\rho}{W}|\overset{\geq -1}{\overbrace{(\tau_i - 1)}}f(x)| \leq \frac{\rho}{W}|f(x)|.$$

We bound $|f(x)| \leq |W + D|$. By Proposition 4 we get

$$\Pr(D > t) = \Pr(D < -t) \leq \frac{1}{2}\exp\left(-\frac{t^2}{2\sigma^2}\right).$$

For $t = \omega(\sigma \log n) = m$, this probability is superpolynomially small. So we can bound $|f(x)| \leq W + m$ and get $|\tau_i - \tau_i'| \leq \frac{\rho}{W}(W + m)$.   □

In the following we always assume, for any $m = \omega(\sigma \log n)$, that we choose $\rho = o(W/(W + m))$ such that $\rho f(x)/W = o(1)$ holds for all x with high probability, and we let $s = \rho(W + m)/W$. Note that s is the greatest difference between two consecutive pheromones, as stated by Lemma 5.

For pheromones in $[\ell + s, u - s]$, every update is normal with high probability. However, since the update is done relative to τ_i and s is an upper bound, $\ell + s$ can be quite pessimistic as a bound, because τ_i will be small when it is close to ℓ. Thus, we shortly want to discuss when a *decreasing* update is normal—that is, when $\tau_i' = \tau_i(1 - \rho f(x)/W) > \ell$. Because of our previous assumption regarding ρ, we have $\tau_i' = \tau_i(1 - o(1)) \geq \tau_i/2$. So if $\tau_i = \omega(\ell)$, then $\tau_i' = \omega(\ell)$. Hence, we only have to consider a nonnormal decreasing update if $\tau_i = \mathcal{O}(\ell)$.

From now on, we always implicitly condition on the event that the noise does not exceed the bounds mentioned in the proof of Lemma 5, which holds with high probability in any polynomial number of steps of MMAS-fp for any of the n pheromone values via a simple union bound. Other runs are referred to as *fails*.

We now show that, in expectation, the result of a normal update of any pheromone $\tau_i \leq u - s$ will be farther away from the upper bound u than before, thus meaning that the pheromone drifts *away* from u.

LEMMA 6: *Consider MMAS-fp optimizing f_{σ^2} and let a pheromone vector τ in some iteration of the algorithm be given. Let $i \leq n$ be given, and suppose $b := \tau_i \leq u - s$ and $b = \omega(\ell)$. Then the drift of τ_i toward 1 is at most $-b(1-b)w_{min}\rho/W$.*

PROOF: From the conditions of the lemma we see that the pheromone τ_i is updated normally.

Let A denote the event that x_i was sampled as 1, and consider the expected value of τ_i'. Note that we can express a (normal) pheromone update as in terms of an unconditional decrease and an increase in the case of $\overline{A}$.

$$\mathrm{E}(\tau_i') = \mathrm{E}\left(\tau_i\left(1 - \rho\frac{f(x)}{W}\right)\right) + \mathrm{E}\left(\rho\frac{f(x)}{W}\,\middle|\,\overline{A}\right)\overbrace{\mathrm{Pr}(\overline{A})}^{=\tau_i=b}$$

$$= b + b\frac{\rho}{W}\Big(\mathrm{E}(f(x)\mid\overline{A}) - \mathrm{E}(f(x))\Big).$$

We get the negative drift:

$$\mathrm{E}(\tau_i' - \tau_i) = b + b\frac{\rho}{W}\Big(\mathrm{E}(f(x)\mid\overline{A}) - \mathrm{E}(f(x))\Big) - b$$

$$= b\frac{\rho}{W}(\mathrm{E}(W - \|w(x)\|_0 + D\mid\overline{A}) - \mathrm{E}(W - \|w(x)\|_0 + D))$$

$$= b\frac{\rho}{W}\left(W - \left(w_i + \sum_{\substack{j\\j\neq i}}w_j\tau_j\right) - \left(W - \left(w_i b + \sum_{\substack{j\\j\neq i}}w_j\tau_j\right)\right)\right)$$

$$= b(b-1)w_i\frac{\rho}{W}$$

$$\leq -b(1-b)w_{min}\frac{\rho}{W}. \qquad \square$$

We proceed by showing that each pheromone considered in Lemma 6 cannot reach a constant value greater than $1/2$ within polynomial time during a nonfailing run of MMAS-fp with high probability; that is, they go in the wrong direction for this to occur. This is due to the negative drift we just proved—that is, the considered pheromones move, in expectation, down, and are thus unlikely to cover large distances going up.

COROLLARY 7: *Consider MMAS-fp optimizing f_{σ^2} and let $d > 1/2$ be a constant. If $\rho = o\big(Ww_{min}/\big((W+m)^2\log n\big)\big)$, for any $m = \omega(\sigma\log n)$, then each pheromone, during a nonfailing run of MMAS-fp, reaches values of at least d in polynomial time with only superpolynomially low probability.*

PROOF: Consider $b \in [c, d]$, with $1/2 - s > c > \ell$ being a constant. We apply Theorem 3 by using the bound given in Lemma 5 and the negative drift theorem of Lemma 6. Note that we have to translate our values by $-1/2$ to make use of Lemma 6. This does not affect the drift.

$$\mathrm{Pr}(T \leq t) \leq t\exp\left(-\frac{\overbrace{\left(d - \frac{1}{2}\right)}^{>0}b(1-b)w_{min}\frac{\rho}{W}}{16\frac{\rho^2}{W^2}(W+m)^2}\right) \leq t\exp\left(-\frac{\Omega(1)Ww_{min}}{\rho(W+m)^2}\right).$$

This means that the probability for a single pheromone starting from $1/2$ to reach d in any polynomial number of steps is superpolynomially small for $\rho = o(W w_{\min}/((W + m)^2 \log n))$, as given, if that pheromone only takes values in $[c, d]$.

Now consider a pheromone to be in $[\ell, c)$. Even the best update could not get that pheromone above $1/2$. So we can apply the above argument.

Note that $s = o(1) < d - 1/2 = \Omega(1)$, as this is what we need to apply Theorem 3.

$\qquad\square$

Using a union bound, we can now say that none of the n pheromones reaches d as in Corollary 7 in polynomially many steps with high probability. So we now always assume pheromones to be in $[\ell, d)$ if not stated otherwise.

We now consider the sum of all pheromone values and prove that it drifts toward 0; that is, all pheromone values together decrease in expectation. This does, however, not hold when *all* pheromones are very close to ℓ because, in this case, an increase of any pheromone would have a far higher impact than a decrease. The following lemma thus only holds if there is at least one pheromone with a sufficiently large value $(1/n)$.

LEMMA 8: *Consider MMAS-fp optimizing f_{σ^2} with $\rho = o\big(W w_{\min}/\big((W + m)^2 \log n\big)\big)$, for any $m = \omega(\sigma \log n)$, and $\ell = o\big(w_{\min}/(w_{\max} n^3)\big)$. Consider further that $k \in [0, n] \cap \mathbb{N}$ pheromones dropped below $1/n$. If at least one pheromone is at least $1/n$, then the drift of the sum of all pheromones τ_{total} toward 0 is $\Omega(\tau_{total} \rho w_{\min}/W)$.*

PROOF: In this proof, for the drift calculations, we pessimistically assume pheromones below $1/n$ as having reached ℓ. The drift per pheromone is relative to the pheromones value, according to Lemma 6, but it also is positive (i.e., toward 0) as long as the pheromone is updated normally. So there can only be a negative drift (i.e., drift away from 0) for pheromone values in $\mathcal{O}(\ell)$.

In this proof, we write ℓ and ignore constant factors, because we make sure that the overall drift will be positive toward 0 in asymptotics. Note that there cannot be any positive drift toward 0 if $k = n$ since none of the pheromones could drop any lower.

We split up the overall drift of the sum of all pheromones into the (positive) drift (toward 0) of the $n - k$ pheromones that are at least $1/n$ and into the (negative) drift of the k pheromones that are below $1/n$.

Let Y denote the index set of those former $n - k$ pheromones, Z the index set of the latter k, and let K be the event that exactly k pheromones are below $1/n$. The string x^y denotes the bit string consisting only of those elements of x whose index is in Y, and x^z is defined analogously with respect to Z. The sum of pheromones with index in Y is denoted by τ_{total}^y, and the sum of those with index in Z by τ_{total}^z. Further, let T_{total} denote $E(\|w(x)\|_0 \mid K, \tau_{total}^y = b)$; let T_{total}^y and T_{total}^z be defined analogously.

First, we have a look at the $n - k$ pheromones. Note that, for each pheromone τ_i with index in Y, both the increase and decrease during an update scale τ_i with $1 - \rho f(x)/W$, regardless of the corresponding sampled bit x_i. However, if a bit x_i was sampled as 0, the corresponding pheromone τ_i additionally gets an increase of $\rho f(x)/W$.

$$
E(\tau_{total}^y - \tau_{total}^{y\,\prime} \mid \overbrace{K, \tau_{total}^y = b}^{:=B}) \geq E\left(\tau_{total}^y - \left(\tau_{total}^y \left(1 - \rho \frac{f(x)}{W}\right) + \|x^y\|_0 \rho \frac{f(x)}{W} \right) \,\Big|\, B \right)
$$

$$
= \frac{\rho}{W} E\left(\tau_{total}^y f(x) - \|x^y\|_0 f(x) \mid B \right)
$$

$$
= \frac{\rho}{W} \big(b E(f(x) \mid B) - E(\|x^y\|_0 f(x) \mid B) \big).
$$

We give bounds for the expected values needed.

$$E(f(x) \mid B) = E(W - \|w(x)\|_0 + D \mid B) = W - T_{\text{total}},$$

and

$$\begin{aligned}
E(\|x^y\|_0 f(x) \mid B) &= E(\|x^y\|_0(W - \|w(x)\|_0 + D) \mid B) \\
&= W E(\|x^y\|_0 \mid B) - E(\|x^y\|_0 \cdot \|w(x)\|_0 \mid B) \\
&= Wb - E(\|x^y\|_0 \cdot \|w(x)\|_0 \mid B).
\end{aligned}$$

Let a be a variable ranging over Y; j ranges over all pheromone indices. Let $[\cdot]$ denote the Iverson bracket (indicator function). We make use of Corollary 7.

$$\begin{aligned}
E(\|x^y\|_0 \cdot \|w(x)\|_0 \mid B) &= E\left(\sum_a \sum_j [x_a = 0] w_j [x_j = 0] \,\middle|\, B \right) \\
&= \sum_a \sum_j w_j E\left([x_a = 0][x_j = 0] \mid B\right) \\
&= \sum_a \sum_{\substack{j \\ j \neq a}} w_j E\left([x_a = 0] \mid B\right) E\left([x_j = 0] \mid B\right) + \sum_a w_a E\left([x_a = 0][x_a = 0] \mid B\right) \\
&= \sum_a \sum_{\substack{j \\ j \neq a}} w_j \tau_a \tau_j + \sum_a w_a \tau_a \\
&= \sum_a \tau_a \sum_j w_j \tau_j - \sum_a w_a \tau_a^2 + \sum_a w_a \tau_a \\
&= \sum_a \tau_a \sum_j w_j \tau_j + \sum_a w_a \tau_a (1 - \overbrace{\tau_a}^{\leq d}) \\
&\geq b T_{\text{total}} + (1 - d) T_{\text{total}}^y.
\end{aligned}$$

Now we can calculate the desired drift.

$$\begin{aligned}
E(\tau_{\text{total}}^y - {\tau_{\text{total}}^y}' \mid B) &\geq \frac{\rho}{W}\left(b\left(W - T_{\text{total}}\right) - \left(Wb - b T_{\text{total}} - (1 - d) T_{\text{total}}^y\right)\right) \\
&= \frac{\rho}{W}(1 - d) T_{\text{total}}^y.
\end{aligned}$$

For the negative drift, we have to look at τ_{total}^z and ${\tau_{\text{total}}^z}'$. All of the k pheromones where the corresponding bit position was sampled as 1 cannot drop any lower due to our assumption. The remainder of the k pheromones get a normal update in the best case:

$$\begin{aligned}
{\tau_{\text{total}}^z}' &\leq \|x^z\|_1 \ell + \|x^z\|_0 \left(\ell\left(1 - \rho \frac{f(x)}{W}\right) + \rho \frac{f(x)}{W} \right) \\
&= \overbrace{\|x^z\|_1 \ell + \|x^z\|_0 \ell}^{=k\ell} - \|x^z\|_0 \ell \rho \frac{f(x)}{W} + \|x^z\|_0 \rho \frac{f(x)}{W} \\
&= k\ell + \frac{\rho}{W}(1 - \ell)\|x^z\|_0 f(x).
\end{aligned}$$

Thus, we get:

$$E(\tau_{\text{total}}^{z} - \tau_{\text{total}}^{z}{}' \mid B) \geq E\left(k\ell - k\ell - \frac{\rho}{W}(1-\ell)\|x^{z}\|_0 f(x) \,\middle|\, B\right)$$

$$= -\frac{\rho}{W}(1-\ell)E(\|x^{z}\|_0 f(x) \mid B).$$

The remaining estimations follow analogously to the ones beforehand:

$$E(\|x^{z}\|_0 f(x) \mid B) = E\left(\|x^{z}\|_0(W - \|w(x)\|_0 + D) \mid B\right)$$

$$= WE(\|x^{z}\|_0 \mid B) - E(\|x^{z}\|_0 \cdot \|w(x)\|_0 \mid B)$$

$$= Wk\ell - E(\|x^{z}\|_0 \cdot \|w(x)\|_0 \mid B).$$

Note that a now refers to indices in Z.

$$E\left(\|x^{z}\|_0 \cdot \|w(x)\|_0 \mid B\right) = \sum_a \tau_a \sum_j w_j \tau_j + \sum_a w_a \tau_a (1 - \tau_a)$$

$$\geq k\ell T_{\text{total}} + (1-d)\, T_{\text{total}}^{z}.$$

So we get the negative drift:

$$E\left(\tau_{\text{total}}^{z} - \tau_{\text{total}}^{z}{}' \mid B\right) \geq -\frac{\rho}{W}(1-\ell)\left(Wk\ell - k\ell T_{\text{total}} - (1-d)T_{\text{total}}^{z}\right)$$

$$\geq -\frac{\rho}{W}(1-\ell)Wk\ell.$$

Now we can look at the general drift:

$$E(\tau_{\text{total}} - \tau_{\text{total}}' \mid B) \geq \frac{\rho}{W}(1-d)T_{\text{total}}^{y} - \frac{\rho}{W}(1-\ell)Wk\ell$$

$$= \frac{\rho}{W}\left((1-d)T_{\text{total}}^{y} - (1-\ell)Wk\right),$$

which we want to be positive.

Recall that we only consider $k \in [0, n-1] \cap \mathbb{N}$. For the drift to be positive, it suffices to show that $T_{\text{total}}^{y} = \omega(Wk\ell)$. Again, a ranges over indices in Y, and j ranges over all indices.

$$\sum_a \overbrace{w_a}^{\geq w_{\min}\frac{1}{n}} \tau_a = \omega\left(\overbrace{k}^{\leq n-1}\, \ell \overbrace{\sum_j w_j}^{\leq n w_{\max}}\right),$$

so we just check that the following holds:

$$\frac{w_{\min}}{n} = \omega(n^2 \ell w_{\max}) \;\Leftrightarrow\; 1 = \omega\left(n^3 \frac{w_{\max}}{w_{\min}}\ell\right).$$

This holds, because we assume $\ell = o(w_{\min}/(w_{\max}n^3))$. So we can write the drift as $\Omega(T_{\text{total}}^{y}\rho/W)$.

Since we want the drift to be expressed in terms of τ_{total}, we need $T_{\text{total}}^{y}\rho/W \geq \varepsilon\tau_{\text{total}}$ to hold for an $\varepsilon > 0$. We bound $T_{\text{total}}^{y} \geq w_{\min}b$ and $\tau_{\text{total}} = b + k\ell < b + k/n$, and get $w_{\min}b\rho/W \geq \varepsilon(b + k/n)$. This leaves us with

$$\varepsilon \leq \frac{w_{\min}b\rho}{W(b + \frac{k}{n})} = \frac{w_{\min}\rho}{W(1 + \frac{k}{bn})} = \mathcal{O}\left(\frac{w_{\min}\rho}{W}\right).$$

Overall, we finally get a drift of $\Omega(\tau_{\text{total}} \rho w_{\min}/W)$. □

We are now ready to give the main theorem, which shows efficient optimization given additive posterior noise. It follows from an application of the Multiplicative Drift Theorem (Theorem 1) with its respective concentration bound (Theorem 2).

THEOREM 9: *Consider MMAS-fp with parameters $\rho = o\big(W w_{\min}/\big((W+m)^2 \log n\big)\big)$, for any $m = \omega(\sigma \log n)$, and $\ell = o\big(w_{\min}/(w_{\max}n^3)\big)$ optimizing f_{σ^2}. Then, for any $k = \omega(\log(n))$, the algorithm finds the optimum after $\mathcal{O}(Wk/(\rho w_{\min}))$ steps with high probability.*

PROOF: Let $T = \min\{t \in \mathbb{N} \mid \forall i : \tau_i^{(t)} < 1/n\}$ be the time until all pheromones are below $1/n$. That means that τ_{total} has dropped below n.

According to Lemma 8, there is a positive drift of the pheromone sum toward 0 of at least $\tau_{\text{total}} \rho w_{\min}/W$.

With this multiplicative drift, we get an upper bound on the expectation of T using Theorem 1. Note that we have to scale all values by n to do so. This, however, does not affect the drift.

$$
\mathrm{E}\left(T \,\middle|\, \forall i : \tau_i^{(0)} = \frac{1}{2}\right) = \mathcal{O}\left(\frac{1 + \log \frac{n}{2}}{\frac{\rho w_{\min}}{W}}\right) = \mathcal{O}\left(\frac{W \log n}{\rho w_{\min}}\right).
$$

Now assume that all pheromones have dropped below $1/n$. The probability to sample the optimum now is $(1 - 1/n)^n \geq (1 - 1/n)/e$. Thus, the optimum will be sampled in expectation in $en/(n-1) = \mathcal{O}(1)$ many tries.

The overall expected run time is bounded above by the following worst case scenario: The algorithm needs $\mathcal{O}(W \log(n)/(\rho w_{\min}))$ steps until all pheromones drop below $1/n$. Then the algorithm does not sample the optimum, and all pheromones are set back to $1/2$ (that does not actually happen, but we are looking at a worst case). As we mentioned before, in this scenario, the algorithm would need $\mathcal{O}(1)$ tries to sample the optimum, once the pheromones are low enough. Hence, the algorithm has to be restarted $\mathcal{O}(1)$ times. This results in an expected run time in $\mathcal{O}(W \log(n)/(\rho w_{\min}))$.

To lower-bound the probability of success, we upper-bound the probability of failure (conditioned on a nonfailing run) by using Theorem 2. We choose $k = \omega(\log n)$ to get a superpolynomially small failing probability.

$$
\Pr\left(T > \frac{k + \ln \frac{n}{2}}{\frac{\rho w_{\min}}{W}} \,\middle|\, \forall i : \tau_i^{(0)} = \frac{1}{2}\right) = \Pr\left(T > \frac{W\omega(\log n)}{\rho w_{\min}} \,\middle|\, \forall i : \tau_i^{(0)} = \frac{1}{2}\right)
$$

$$
\leq e^{-\omega(\log n)}.
$$

Since we now have an $\omega(\log n)$ in the numerator, we have to make a slight change to our run time; it is now $\mathcal{O}(W\omega(\log n)/(\rho w_{\min}))$.

We can easily make a union bound over the run failing via too large noise (as bounded in the proof of Lemma 5) and the run failing because of the drift concentrations (per pheromone or for the sum). Since all of these probabilities are superpolynomially small, the overall failing probability of the algorithm is superpolynomially small—that is, the algorithm succeeds with high probability. □

As a direct corollary, we get that we can disregard pheromone bounds (i.e., setting these bounds to 0 and 1) and still get the same result. Intuitively, this holds because we have drift in the right direction in each pheromone at any time (while still not close to the target). Recall that ACO-fp is a special case of MMAS-fp (Algorithm 1) with trivial pheromone bounds.

COROLLARY 10: *Consider ACO-fp optimizing f_{σ^2}. If $\rho = o\big(W w_{min} / \big((W + m)^2 \log n\big)\big)$, for any $m = \omega(\sigma \log n)$, then, for any $k = \omega(\log(n))$, the algorithm finds the optimum after $\mathcal{O}(Wk/(\rho w_{min}))$ steps with high probability.*

PROOF: Recall that all previous proofs are actually proofs for ACO-fp as well. We can thus argue analogously as in the proof of Theorem 9. The upper bound d is at most 1 because $d \leq u \leq 1$. The lower bound of $\ell = 0$ for ACO-fp satisfies $\ell = o(w_{min}/(w_{max}n^3))$ as well. $\square$

4 Other Noise Models

In this section we consider the optimization of linear functions perturbed by other noise than additive posterior Gaussian noise.

4.1 Posterior Noise

We start with a generalization of Corollary 10 by taking non-Gaussian noise into account; that is, we optimize f_D, with $f_D(x) := f(x) + D$ and D being a random variable (possibly not Gaussian).

The idea behind why the following noise models still do not harm optimization is that we can, again, bound the noise with high probability as we already did in the proof of Lemma 5.

COROLLARY 11: *Consider MMAS-fp optimizing f_D with $\ell = o(w_{min}/(w_{max}n^3))$, where D is a random variable such that, for some $t^*(n) > 0$ and all polynomials p, the following holds:*

$$Pr\left(|D| > t^*(n)\right) \leq 1 - o\left(\frac{1}{p(n)}\right). \tag{1}$$

If $\rho = o\big(W w_{min}/((W + t^*(n))^2 \log n)\big)$, then, for any $k = \omega(\log(n))$, the algorithm finds the optimum after $\mathcal{O}(Wk/(\rho w_{min}))$ steps with high probability.

PROOF: The drifts of Lemmas 6 and 8 hold here as well, since the noise is additive and posterior, and thus cancels out in expectation.

To be able to use Corollary 7, we need to bound $|f(x)|$ with high probability. Because of Equation (1) we can do so, and it follows that any pheromone τ_i reaches values of at least a constant $d > 1/2$ with superpolynomially low probability for $\rho = o\big(W w_{min}/((W + t^*(n))^2 \log n)\big)$.

We can now use an argumentation analogous to that in Theorem 9, proving the corollary. $\square$

A similar corollary holds for ACO-fp.

COROLLARY 12: *Consider ACO-fp optimizing f_D, where D is a random variable such that, for some $t^*(n) > 0$ and all polynomials p, the following holds:*

$$Pr\left(|D| > t^*(n)\right) \leq 1 - o\left(\frac{1}{p(n)}\right).$$

If $\rho = o\big(W w_{min}/((W + t^*(n))^2 \log n)\big)$, then, for any $k = \omega(\log(n))$, the algorithm finds the optimum after $\mathcal{O}(Wk/(\rho w_{min}))$ steps with high probability.

PROOF: We can use the same proof as for Corollary 11 and argue as in the proof of Corollary 10 that we do not need the special pheromone bounds. $\square$

4.2 Prior Noise

In this section, we have a look at the noise model from Droste (2004), where $f_p(x) :=$ $f(\widetilde{x})$, whereas $\widetilde{x} = x$ with probability $(1 - p)$, and $\widetilde{x} = x'$ with $d_H(x, x') = 1$ for each such x' with probability p/n. That means that, with probability p, a single bit in x gets flipped chosen uniformly at random.

Note that we can bound $|\tau_i - \tau_i'| \leq \rho$ in the same way as we did in the proof of Lemma 5, since $|f(x)| \leq W$. Since the noise is not additive, we do *not* have to bound ρ with respect to the noise.

Because we assume again that $\rho = o(1)$, we have normal updates for pheromones at least in $\omega(\ell)$ and at most $u - \rho$, this time with probability 1.

The following proofs are all similar to the ones from Section 3, and the order of argumentation is the same. We first prove that individual pheromones, when updated normally, drift away from u. This results in an upper bound on each pheromone that holds with high probability. We then, again, look at the sum of all pheromones and prove that it drifts (under sufficient conditions) toward 0. Applying Theorem 1 finally yields the desired run time.

LEMMA 13: *Consider MMAS-fp optimizing f_p and let a pheromone vector τ in some iteration of the algorithm be given. Let $i \leq n$ be given and suppose $b := \tau_i \leq u - \rho$ and $b = \omega(\ell)$. Then the drift of τ_i toward 1 is at most $-b(1 - b)w_{min}\rho/W$.*

PROOF: As the noise does not have any influence on the update of the pheromones and we have a normal update, we get the same basic equation for the desired drift as in Lemma 6. Let A denote the event that x_i was sampled as 1. Note that τ_i is updated normally.

$$E(\tau_i' - \tau_i) = b\frac{\rho}{W}(E(f(x) \mid \overline{A}) - E(f(x))).$$

Let P denote the event that $\widetilde{x} = x'$ during the evaluation of f—that is, a bit flipped—and let j and k range from 1 to n.

For $E(f(x) \mid \overline{A})$ we have

$$E(f(x) \mid \overline{A}) = E(f(x) \mid \overline{A}, \overline{P})(1 - p) + E(f(x) \mid \overline{A}, P)p$$

$$= \left(W - \left(w_i + \sum_{\substack{j \\ j \neq i}} w_j\tau_j\right)\right)(1 - p)$$

$$+ \left(W - \sum_{\substack{k \\ k \neq i}} \frac{1}{n}\left(w_i + w_k(1 - \tau_k) + \sum_{\substack{j \\ j \neq k \\ j \neq i}} w_j\tau_j\right) - \frac{1}{n}\sum_{\substack{j \\ j \neq i}} w_j\tau_j\right)p.$$

And for $E(f(x))$ we get

$$E(f(x)) = E(f(x) \mid \overline{P})(1 - p) + E(f(x) \mid P)p$$

$$= \left(W - \left(w_ib + \sum_{\substack{j \\ j \neq i}} w_j\tau_j\right)\right)(1 - p)$$

$$+ \left(W - \sum_{\substack{k \\ k \neq i}} \frac{1}{n}\left(w_ib + w_k(1 - \tau_k) + \sum_{\substack{j \\ j \neq k \\ j \neq i}} w_j\tau_j\right) - \frac{1}{n}\left(w_i(1 - b) + \sum_{\substack{j \\ j \neq i}} w_j\tau_j\right)\right)p.$$

Thus we get

$$E\left(f(x) \mid \overline{A}\right) - E(f(x)) = -(1-b)w_i(1-p) + \left(-\frac{n-1}{n}(1-b)w_i + \frac{1}{n}(1-b)w_i\right)p$$

$$= -(1-b)w_i(1-p) - (1-b)w_i\frac{n-2}{n}p$$

$$= -(1-b)w_i\left(1 - \frac{2}{n}p\right).$$

Finally, we have a negative drift as desired:

$$E(\tau_i' - \tau_i) \le -b(1-b)w_{\min}\left(1 - \frac{2}{n}p\right)\frac{\rho}{W} \le -b(1-b)w_{\min}\frac{\rho}{W}. \qquad \Box$$

COROLLARY 14: *Consider MMAS-fp optimizing f_p, and let $d > 1/2$ be a constant. If $\rho = o(w_{min}/(W \log n))$, then each pheromone, during a nonfailing run of MMAS-fp, reaches values of at least d in polynomial time with only superpolynomially low probability.*

PROOF: The proof is basically the same as the one of Corollary 7. Again using Theorem 3, for all $t \ge 0$, the following proposition holds for T as the hitting time of one pheromone reaching a value of at least d:

$$\Pr(T \le t) \le t \exp\left(-\frac{\left(d - \frac{1}{2}\right)b(1-b)w_{\min}\frac{\rho}{W}}{16\rho^2}\right)$$

$$= t \exp\left(-\frac{\Omega(1)w_{\min}}{\rho W}\right).$$

The probability for a single pheromone starting from $1/2$ to reach d in any polynomial number of steps is superpolynomially small for $\rho = o(w_{\min}/(W \log n))$. $\qquad \Box$

Again, a union bound argument gives us the guarantees of Corollary 14 for all pheromones.

LEMMA 15: *Consider MMAS-fp optimizing f_p with $\rho = o(w_{min}/(W \log n))$ and $\ell = o(w_{min}/(w_{max}n^3))$. Consider further that $k \in [0, n] \cap \mathbb{N}$ pheromones dropped below $1/n$. If at least one pheromone is at least $1/n$, then the drift of the sum of all pheromones τ_{total} toward 0 is $\Omega(\tau_{total}\rho w_{min}/W)$.*

PROOF: We are going to use the same notation as in the proof of Lemma 8. P shall, again, denote the event that the bit flip occurred.

As we mentioned before, the kind of noise has no impact on the general form of the drifts. So we are going to use the general forms as in the proof of Lemma 8, and for the $n - k$ pheromones we have

$$E(\tau_{\text{total}}^y - \tau_{\text{total}}^{y\,'} \mid B) \ge \frac{\rho}{W}(bE(f(x) \mid B) - E(\|x^y\|_0 f(x) \mid B)).$$

We can now calculate the required expected values.

$$E(f(x) \mid B) = E(f(x) \mid B, \overline{P})(1-p) + E(f(x) \mid B, P)p$$

$$= (W - T_{\text{total}})(1-p) + (W - E(\|w(x)\|_0 \mid B, P))p.$$

Furthermore, we get

$$E(\|w(x)\|_0 \mid B, P) = \sum_i \frac{1}{n}\left(w_i(1 - \tau_i) + \sum_{\substack{j \\ j \neq i}} w_j \tau_j\right)$$

$$= \frac{1}{n}\sum_i (w_i(1 - \tau_i) + T_{\text{total}} - w_i \tau_i)$$

$$= \frac{1}{n}\sum_i (w_i + T_{\text{total}}) - \frac{2}{n}\sum_i w_i \tau_i$$

$$= \frac{W}{n} + T_{\text{total}} - \frac{2}{n}T_{\text{total}}.$$

This results in

$$E(f(x) \mid B) = W - T_{\text{total}} + p\left(\frac{2}{n}T_{\text{total}} - \frac{W}{n}\right).$$

For $E(\|x^y\|_0 f(x) \mid B)$ we get

$$E(\|x^y\|_0 f(x) \mid B) = Wb - E(\|x^y\|_0 \cdot \|w(\widetilde{x})\|_0 \mid B),$$

and by using Corollary 14,

$$E(\|x^y\|_0 \cdot \|w(\widetilde{x})\|_0 \mid B) = \sum_a \sum_{\substack{j \\ j \neq a}} w_j E\left([x_a = 0] \mid B\right) E\left([\widetilde{x}_j = 0] \mid B\right)$$

$$+ \sum_a w_a E\left([x_a = 0][\widetilde{x}_a = 0] \mid B\right)$$

$$= \sum_a \sum_{\substack{j \\ j \neq a}} w_j \tau_a \left(\tau_j(1 - p) + \tau_j p\frac{n-1}{n} + (1 - \tau_j)\frac{p}{n}\right)$$

$$+ \sum_a w_a \tau_a \left(1 - \frac{p}{n}\right)$$

$$= \sum_a \sum_{\substack{j \\ j \neq a}} w_j \tau_a \left(\tau_j\frac{n-2p}{n} + \frac{p}{n}\right) + \sum_a w_a \tau_a \left(1 - \frac{p}{n}\right)$$

$$= \frac{n-2p}{n}\sum_a \sum_{\substack{j \\ j \neq a}} w_j \tau_a \tau_j + \frac{p}{n}\sum_a \sum_{\substack{j \\ j \neq a}} w_j \tau_a + \sum_a w_a \tau_a \left(1 - \frac{p}{n}\right)$$

$$= \frac{n-2p}{n}\sum_a \sum_j w_j \tau_a \tau_j - \frac{n-2p}{n}\sum_a w_a \tau_a^2$$

$$+ \frac{n-p}{n}\sum_a w_a \tau_a + \frac{p}{n}\sum_a \tau_a \left(\sum_j w_j - w_a\right)$$

$$\overset{\tau_a \leq d}{\geq} \frac{n-2p}{n}bT_{\text{total}} + \sum_a w_a \tau_a \left(\frac{n-p}{n} - \frac{n-2p}{n}d\right) + \frac{p}{n}\left(Wb - T_{\text{total}}^y\right)$$

$$= \frac{n-2p}{n}bT_{\text{total}} + T_{\text{total}}^y \frac{n(1-d) - p(1-2d)}{n} + \frac{p}{n}(Wb - T_{\text{total}}^y).$$

Thus we get

$$\mathrm{E}(\|x^y\|_0 f(x) \mid B) \le Wb - \frac{n-2p}{n} b T_{\text{total}} - T_{\text{total}}^y \frac{n(1-d) - p(1-2d)}{n} - \frac{p}{n}(bW - T_{\text{total}}^y).$$

Now we can calculate the desired drift as follows.

$$\mathrm{E}(\tau_{\text{total}}^y - {\tau_{\text{total}}^y}' \mid B)$$

$$\ge \frac{\rho}{W}\left(b\left(W - T_{\text{total}} + p\left(\frac{2}{n}T_{\text{total}} - \frac{W}{n}\right)\right) - Wb + \frac{n-2p}{n}bT_{\text{total}}\right.$$

$$\left. + T_{\text{total}}^y \frac{n(1-d) - p(1-2d)}{n} + \frac{p}{n}\left(bW - T_{\text{total}}^y\right)\right)$$

$$= \frac{\rho}{W}\left(Wb - bT_{\text{total}} - Wb + bT_{\text{total}} + \frac{1}{n}\left(-Wbp\right.\right.$$

$$\left.\left. + T_{\text{total}}^y(n(1-d) - p(1-2d)) + Wbp - T_{\text{total}}^y p\right)\right)$$

$$= \frac{\rho T_{\text{total}}^y}{Wn}(n(1-d) - p(2-2d))$$

$$= \frac{\rho}{W}T_{\text{total}}^y(1-d)\frac{n-2p}{n}$$

$$= \frac{\rho}{W}T_{\text{total}}^y(1-d)(1-o(1)).$$

Regarding the negative drift, we have

$$\mathrm{E}(\tau_{\text{total}}^z - {\tau_{\text{total}}^z}' \mid B) = -\frac{\rho}{W}(1-\ell)\mathrm{E}(\|x^z\|_0 f(x) \mid B).$$

We also get

$$\mathrm{E}(\|x^z\|_0 f(x) \mid B) = Wk\ell - \mathrm{E}(\|x^z\|_0 \cdot \|w(\tilde{x})\|_0 \mid B)$$

$$\le Wk\ell - \frac{n-2p}{n}k\ell T_{\text{total}} - T_{\text{total}}^z\left(\frac{n-p}{n} - \frac{n-2p}{n}d\right)$$

$$- \frac{p}{n}\left(Wk\ell - T_{\text{total}}^z\right)$$

$$\le Wk\ell.$$

Now we have the negative drift as

$$\mathrm{E}(\tau_{\text{total}}^z - {\tau_{\text{total}}^z}' \mid B) \ge -\frac{\rho}{W}(1-\ell)Wk\ell.$$

Taking all the parts together, we get the total drift

$$\mathrm{E}(\tau_{\text{total}} - \tau_{\text{total}}' \mid B) \ge \frac{\rho}{W}T_{\text{total}}^y(1-d)(1-o(1)) - \frac{\rho}{W}(1-\ell)Wk\ell$$

$$= \frac{\rho}{W}(T_{\text{total}}^y(1-d)(1-o(1)) - (1-\ell)Wk\ell),$$

which we want to be positive.

This is basically the same expression as in the proof of Lemma 8, and the drift is again $\Omega(T_{\text{total}}^y \rho / W)$. So we can conclude analogously. $\square$

We can now state that MMAS-fp and ACO-fp are able to optimize efficiently, given prior noise.

THEOREM 16: *Consider MMAS-fp or ACO-fp with parameters $\rho = o\big(w_{min}/(W \log n)\big)$ and $\ell = o(w_{min}/(w_{max}n^3))$ optimizing f_p. Then, for any $k = \omega(log(n))$, both algorithms find the optimum after $\mathcal{O}(Wk/(\rho w_{min}))$ steps with high probability.*

PROOF: The argumentation is exactly as in the proof of Theorem 9, but now we use Lemmas 13 and 15, and Corollary 14. The value of ρ has to be changed accordingly using Corollary 14. □

5   Discussion and Summary

In this work we saw that two simple ACO algorithms on linear pseudo-Boolean functions scale gracefully with noise for Gaussian distributions—that is, the run time depends only linearly on the variance of the noise. We get similar results for many other noise models and for different ACO algorithms, suggesting that ACO algorithms are generally good for dealing with noise (at least in settings where the underlying fitness function is simple enough, as in the case of linear functions). Many of these settings are not solvable by simple hill climbers (Droste, 2004).

The analysis of metaheuristics such as ACO on noisy fitness functions is of particular interest because this is a specific area where only very few tailored approaches exist for designing efficient algorithms. Our proofs give insight into why ACO is robust to various noise models: By choosing the evaporation factor ρ to be small enough, the noise cannot harm an update with high probability. In the end, better solutions turn out to be better in expectation, and thus, optimization succeeds.

One drawback of the ACO algorithms analyzed in this report is that the variance of the noise must be known in order to correctly set the evaporation factor ρ. This problem can be bypassed as follows: Guess a variance of 1 and run the algorithm until it has a constant success probability if the guess was correct. If the optimum was not found so far, double the guess and repeat. This standard doubling scheme leads to a *noise-oblivious* ACO algorithm with an expected run time of at most a constant factor away from the ACO that knows the noise in advance.

Acknowledgments

References

Bianchi, L., Dorigo, M., Gambardella, L., and Gutjahr, W. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8:239–287.

Dang, D.-C., and Lehre, P. K. (2014). Evolution under partial information. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1359–1366.

Di Caro, G. A., Ducatelle, F., and Gambardella, L. M. (2008). Theory and practice of Ant Colony Optimization for routing in dynamic telecommunications networks. In F. Orsucci and N. Sala (Eds.), *Reflexing interfaces: The complex coevolution of information technology ecosystems*, pp. 185–216. Hershey, PA: Information Science Reference.

Doerr, B., and Goldberg, L. A. (2013). Adaptive drift analysis. *Algorithmica*, 65(1): 224–250.

Doerr, B., Hota, A., and Kötzing, T. (2012a). Ants easily solve stochastic shortest path problems. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 17–24.

Doerr, B., Johannsen, D., and Winzen, C. (2012b). Multiplicative drift analysis. *Algorithmica*, 64(4): 673–697.

Droste, S. (2004). Analysis of the (1+1) EA for a Noisy OneMax. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1088–1099.

Droste, S., Jansen, T., and Wegener, I. (2002). On the analysis of the (1+1) Evolutionary Algorithm. *Theoretical Computer Science*, 276:51–81.

Feldmann, M., and Kötzing, T. (2013). Optimizing expected path lengths with ant colony optimization using fitness proportional update. In *Proceedings of Foundations of Genetic Algorithms*, pp. 65–74.

Friedrich, T., Kötzing, T., Krejca, M. S., and Sutton, A. M. (2015a). The benefit of recombination in noisy evolutionary search. In *Proceedings of the International Symposium on Algorithms and Computation (ISAAC)*, pp. 140–150.

Friedrich, T., Kötzing, T., Krejca, M. S., and Sutton, A. M. (2015b). Robustness of ant colony optimization to noise. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 17–24.

Gießen, C., and Kötzing, T. (2014). Robustness of populations in stochastic environments. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1383–1390.

Jin, Y., and Branke, J. (2005). Evolutionary optimization in uncertain environments—A survey. *IEEE Transactions on Evolutionary Computation*, 9:303–317.

Kötzing, T. (2014). Concentration of first hitting times under additive drift. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1391–1397.

Kötzing, T., and Molter, H. (2012). ACO beats EA on a dynamic pseudo-Boolean function. In *Proceedings of the Parallel Problem Solving from Nature (PPSN)*, pp. 113–122.

Kötzing, T., Neumann, F., Sudholt, D., and Wagner, M. (2011). Simple max–min ant systems and the optimization of linear pseudo-Boolean functions. In *Proceedings of Foundations of Genetic Algorithms*, pp. 209–218.

Lehre, P. K., and Witt, C. (2010). Black box search by unbiased variation. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1441–1448.

Lehre, P. K., and Witt, C. (2014). Concentrated hitting times of randomized search heuristics with variable drift. In *Proceedings of the International Symposium on Algorithms and Computation (ISAAC)*, pp. 686–697.

Lissovoi, A., and Witt, C. (in press). MMAS versus population-based EA on a family of dynamic fitness functions. *Algorithmica*.

Madhow, U. (2008). *Fundamentals of digital communication*. Cambridge: Cambridge University Press.

Stützle, T., and Hoos, H. H. (2000). MAX–MIN Ant System. *Future Generation Computer Systems*, 16:889–914.

Sudholt, D., and Thyssen, C. (2012). A simple ant colony optimizer for stochastic shortest path problems. *Algorithmica*, 64:643–672.