Refined Runtime Analysis of a Basic Ant Colony Optimization Algorithm*

Benjamin Doerr and Daniel Johannsen

Max-Planck-Institut für Informatik
Campus E1 4, 66123 Saarbrücken, Germany

Abstract

Neumann and Witt (2006) analyzed the runtime of the basic ant colony optimization (ACO) algorithm 1-Ant on pseudo-boolean optimization problems. For the problem OneMax they showed how the runtime depends on the evaporation factor. In particular, they proved a phase transition from exponential to polynomial runtime. In this work, we simplify the view on this problem by an appropriate translation of the pheromone model. This results in a profound simplification of the pheromone update rule and, by that, a refinement of the results of Neumann and Witt. In particular, we show how the exponential runtime bound gradually changes to a polynomial bound inside the phase of transition.

1 Introduction

In 1991 Dorigo, Maniezzo and Colorni [7] introduced the concept of ant colony optimization (ACO). Since then ACO algorithms have proven to be successful heuristics in practice and have been applied to a wide range of combinatorial optimization problems, several of which are NP-hard.

The technique of ACO is based on a natural optimization process, namely the search of an ant colony for a shortest path to a source of food. The ants commute on random paths between their colony and the food source, leaving behind traces of pheromones. The ants preferably choose those paths with high pheromone density. This behavior reinforces existing pheromone trails once they are significantly stronger than the average trail.

Essential for the convergence of this procedure to an optimal path is the fact that pheromones evaporate over time. Since an ant needs more time to traverse a long path then a short one, a larger amount of pheromones evaporates during the traversal of such a long path. Together with the mechanism of reinforcement this leads to a high concentration of pheromones on short paths which eventually are chosen almost exclusively.

Applied to combinatorial optimization problems we interpret the random paths chosen by the ants as random walks on a given *construction graph*. Such a walk represents a solution to the optimization problem. We assign a pheromone value to every edge of the construction graph. This value determines the probability that a random walk uses the edge. In each step of the algorithm a random walk is generated and afterwards the pheromone values are updated according to the quality of the walk as solution to the given problem.

In addition to the problem-specific construction graph we define the *evaporation factor* ρ as the main parameter of the algorithm. During the pheromone update the pheromone values of all edges are decreased by their ρ-th part (evaporation). Then, the pheromone values of the edges in the new walk are increased by ρ (strengthening). Optionally, the pheromone values are normalized afterwards.

For an evaporation factor of zero, this means the new walk does not contribute to the updated pheromone values at all, leaving them at their initial assignment throughout all steps of the algorithm. For an evaporation factor of one, the new walk completely determines the updated pheromone values. In both of these extreme cases the algorithm looses all information obtained in previous steps, thus the evaporation factor is chosen to be strictly between zero and one.

Neumann and Witt showed in [18] that the choice of the evaporation factor ρ can be critical to the optimization[1] time of an ACO algorithm. They investigated the basic ACO algorithm 1-Ant on the pseudo-boolean optimization problem OneMax and proved that with high probability[2] the optimization time of 1-Ant on OneMax is exponential for $\rho \leq 1/n^{1+\epsilon}$ and polynomial for $\rho \geq 1/n^{1-\epsilon}$ where n is the number of variables in OneMax and $\epsilon > 0$.

We continue this study on how the optimization time of 1-Ant on OneMax depends on the evaporation factor. In particular, we focus on the phase transition from exponential to polynomial optimization time, i. e. we investigate the optimization time as ρ assume the critical values between $1/n^{1+\epsilon}$ and $1/n^{1-\epsilon}$.

Our first achievement is that we translate the pheromone model in [18] to a form that is much simpler to analyze but is fully equivalent. This new model was independently proposed in [13], but without showing its equivalence to the existing model.

Building on the techniques of Neumann and Witt, we give bounds on the optimization time of 1-Ant on OneMax withing the critical window $(1/n^{1+\epsilon}, 1/n^{1-\epsilon})$. In particular, in Theorem 4 we show that with high probability the optimization time is already super-polynomial for $\rho = o(1/n \ln n)$ and show how this lower bound can be continuously strengthened as ρ takes values closer to $1/n^{1+\epsilon}$.

*This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive version was published in *Proceedings of the 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, pages 501–507, IEEE. It is available at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4424512.

[1] We use the notions of *runtime* and *optimization time* of an ACO algorithm synonymously. In both cases we mean the number of solutions constructed before an optimal solution is produced.

[2] An event occurs *with high probability* if, for any $\alpha \geq 1$, the event occurs with probability at least $1-c_\alpha/n^\alpha$, where c_α depends only on α.

2 Outline

In the next section we put our work into the context of related research. As indicated already in the title of this paper we focus on theoretical runtime analyses of evolutionary search heuristics in general and ACO algorithms in particular.

In Section 4 we introduce the basic ACO algorithm 1-Ant and discuss general notions like *construction graph*, *random walk generation*, and *pheromone values*. Note that the important concepts of *pheromone updates* and *evaporation factor* are postponed to the following two sections.

The application of 1-Ant to general pseudo-boolean functions is discussed in Section 5 and Section 6. First, we present the model used by Neumann and Witt, in particular, their definitions of the construction graph, pheromone updates and evaporation factor. In Section 6 we translate this model to a simplified version and prove their equivalence. Besides a new construction graph we introduce a new pheromone model which leads to a rescaled evaporation factor and a strongly simplified update rule.

In Section 7 we introduce the pseudo-boolean optimization problem OneMax in the context of the simplified model and show basic properties. Finally, in Section 8, we use the simplified model to prove refined lower bounds on the optimization time of OneMax on 1-Ant.

3 Known Results

Since 1991, ACO algorithms have been applied to a large number of optimization problems, see [8] for an overview of problems in combinatorial optimization. In contrast to the large amount of practical and experimental work on ACO, theoretical results on the runtime behavior of ACO algorithms are still scarce.

Randomized search heuristics in general have been analyzed for twenty years. Examples are the rather general analyses of Randomized Local Search in 1990 [19] and of the Metropolis Algorithm in 1993 [14].

In 2002 the runtime of basic evolutionary algorithms (EAs) on pseudo-boolean optimization problems [9] was studied. The pseudo-boolean functions analyzed were example functions such as OneMax, LeadingOnes, and BinVal. Due to the well-known structure of these function it was possible to develop stochastic methods for the analysis of EAs, some of which were highly non-trivial.

In the following years, the runtime of EAs on several problems in combinatorial optimization was studied. Some examples are the partition problem [20] and the problems of finding maximum matchings [10], minimum spanning trees [16], and Euler tours [15, 2, 4, 3] in graphs.

Although models and dynamics of ACO algorithms were studied before, until recently only convergence results on the runtime of such algorithms existed. In [6] Dorigo and Blum explicitly formulate the demand for the theoretic investigation of the runtime of simple optimization problems similar to which was done for general EAs.

In 2006 Neumann and Witt proved the first rigorous bounds on the runtime of an ACO algorithm [18]. They analyzed the optimization time of the basic ACO algorithm 1-Ant on the pseudo-boolean optimization problem OneMax. In [5] other pseudo-boolean functions were investigated and in [17] the combinatorial optimization problem of finding a minimum spanning tree was addressed. We continue this work of theoretic runtime analysis and refine the results presented in [18].

4 The 1–Ant Algorithm

We revisit the basic ACO algorithm 1-Ant analyzed by Neumann and Witt in [18]. This algorithm is derived from the graph-based ant system proposed in [11]. Like other randomized search heuristics, 1-Ant successively generates potential solutions to the given optimization problem until an optimal solution is found.

A solution generated by 1-Ant is the outcome of a random walk of a single ant on a given *construction graph* G. More precisely, let $G = (V, E)$ be a directed graph with edge weights $\tau\colon E \to [0, 1]$. The weights τ are called the *pheromone values* of the edges. Given a starting vertex $s \in V$, the random walk of a single ant is constructed as follows. In every step of the walk the ant randomly chooses an edge to an unvisited neighbor. This is done with probability proportional to the pheromone value of that edge. The walk ends if all neighbors of the current vertex are visited.

In the procedure AntWalk the random walks are generates. The input of AntWalk consists of the construction graph G, the pheromone values τ, and the starting vertex s. AntWalk then returns a walk $W = (V_W, E_W)$ constructed randomly as described above.

```
AntWalk($G, s, \tau$)
$V_W := s$
$E_W := \emptyset$
$v := s$
while ($N_G(v) \setminus V_W \neq \emptyset$)
    do Choose $w \in N_G(v) \setminus V_W$ randomly
         with probability $\frac{\tau(v,w)}{\sum_{w' \in N_G(v)\setminus V_W} \tau(v,w')}$
       $V_W := V_W \cup \{w\}$
       $E_W := E_W \cup \{(v, w)\}$
       $v := w$
return $W = (V_W, E_W)$
```

Let $\mathcal{W}$ be the set of maximal non-self-intersecting walks on G starting in s, i. e., the set of walks that can be returned by AntWalk. The algorithm 1-Ant tries to optimize an *objective function* $f\colon \mathcal{W} \to \mathbb{R}$ on all such walks. This is done by successively generating such walks and storing the currently best walk W^*. Whenever W^* is changed, 1-Ant also updates the pheromone values τ according to the edges used in W^*.

The input of 1-Ant is the construction graph G, the starting vertex s and the initial pheromone values τ_{INIT}. Initially, all edges are assigned the same pheromone value.

First, 1-Ant generates an initial solution W^* using the procedure AntWalk. Then, the pheromone values τ are updated according to W^*. This update is performed by the procedure Update which is discussed in the next two sections.

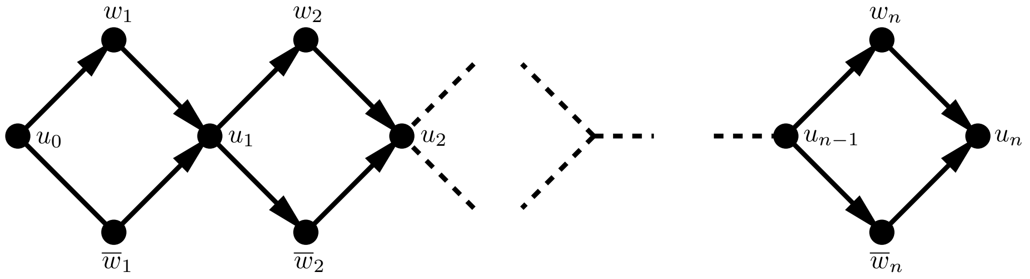

Figure 1: The classical chain G_{bool}.

Next, 1-Ant repeatedly generates solutions W. If the value of the new solution $f(W)$ is at least as large as the objective value of the currently best solution $f(W^*)$ then W^* is replaced by W and the pheromone values τ are again updated.

```
1-Ant($G, s, \tau_{\text{INIT}}$)
$\tau := \tau_{\text{INIT}}$
$W^* :=$ AntWalk($G, s, \tau$)
$\tau :=$ Update($G, \tau, W^*$)
repeat
        $W :=$ AntWalk($G, s, \tau$)
        if ( $f(W) \geq f(W^*)$ )
           then $W^* := W$
                $\tau :=$ Update($G, \tau, W^*$)
forever
```

The *optimization time* of an ACO is the number of solutions generated until an optimal solution is found. Clearly, the optimization time is a random variable depending on the choice of G, s, τ_{INIT} and the definition of the procedure Update. For the theoretical investigation of the optimization time it is common to assume that the algorithm never terminates.

5 The Classical Model for Ant Colony Optimization on Pseudo-Boolean Functions

In this section we present the ACO model for pseudo-boolean optimization problems as used in [12] and [18]. We discuss a suitable construction graph and pheromone values for such problems and define a corresponding Update procedure.

Neumann and Witt use the *chain graph* for the representation of pseudo-boolean functions $f\colon \{0,1\}^n \to \mathbb{R}$. Let $n \in \mathbb{N}$. The *chain graph* G_{bool} as depicted in Figure 1 is the directed graph (V, E) with distinct vertices $u_0, \ldots, u_n, w_1, \ldots, w_n, \overline{w}_1, \ldots, \overline{w}_n$ and edges (u_{i-1}, w_i), (w_i, u_i), $(u_{i-1}, \overline{w}_i)$, and $(\overline{w}_i, u_i)$ for all $i \in \{1, \ldots, n\}$. The vertex u_0 serves as *starting vertex* s.

A random walk of the ant has to pass the vertices $u_0, \ldots, u_n$ in their given order, having the choice between w_i and $\overline{w}_i$ when passing from u_{i-1} to u_i. Thus, a walk $W = (V_W, E_W)$ of the ant defines a vector $x \in \{0,1\}^n$ such that $x_i = 1$ if $w_i \in V_W$ and $x_i = 0$ if $\overline{w}_i \in V_W$.

The pheromone values $\tau\colon E \to [0,1]$ are chosen such that in every step of the algorithm $1/2n^2 \leq \tau(e) \leq (n-1)/2n^2$, $\sum_{e \in E} \tau(e) = 1$, and $\tau(u_{i-1}, w_i) + \tau(u_{i-1}, \overline{w}_i) = 1/2n$ hold for all $i \in \{1, \ldots, n\}$. Initially, all pheromone values are set to $\tau_{\text{INIT}}(e) = 1/4n$ for all $e \in E$. The reason to restrict the values of τ to the interval $[1/2n^2, (n-1)/2n^2]$ is to maintain the flexibility of the algorithm.

The main idea behind ACO is to strengthen the pheromone values of successful solutions. More precisely, if the algorithm generates a walk that has an objective value of at least the current optimum, then the pheromone values of all edges used by the walk are increased (*strengthening*). At the same time the pheromone values of unused edges are decreased (*evaporation*). These changes are determined by the *evaporation factor* $\rho \in [0,1]$.

In 1-Ant, the pheromone values are changed by the procedure Update. In the model proposed in [18] this procedure is defined as follows.

```
Update($G, \tau, E_W$)
for ($e \in E$)
     do if ($e \in E_W$)
          then $\tau(e) := \min\{\frac{(1-\rho)\tau(e)+\rho}{1-\rho+2n\rho}, \frac{n-1}{2n^2}\}$
          else  $\tau(e) := \max\{\frac{(1-\rho)\tau(e)}{1-\rho+2n\rho}, \frac{1}{2n^2}\}$
return $\tau$
```

Note that whenever an ant uses the edge (u_{i-1}, w_i), then it also uses the edge (w_i, u_i). Respectively, the edge $(u_{i-1}, \overline{w}_i)$ is alway used simultaneously with the edge $(\overline{w}_i, u_i)$. Thus, $\tau(u_{i-1}, w_i) = \tau(w_i, u_i)$ and $\tau(u_{i-1}, \overline{w}_i) = \tau(\overline{w}_i, u_i)$.

In order to maintain the two properties $\sum_{e \in E} \tau(e) = 1$ and $\tau(u_{i-1}, w_i) + \tau(u_{i-1}, \overline{w}_i) = 1/2n$ for all $i \in \{1, \ldots, n\}$ the normalization of τ by $1 - \rho + 2n\rho$ is necessary. Because of the second property the probability for an ant to choose the edge $e \in E$ is $p_e = 2n\tau(e)$.

6 The Simplified Model

In order to ease the study of the random optimization time of 1-Ant for pseudo-boolean functions we propose a simplified model. We introduce three modifications to the model of the previous section: We simplify the construction graph by removing unnecessary vertices and edges. We identify the pheromone value of an edge with the probability that the edge is chosen by the ant. Most important, as a consequence of the previous modification we find an appropriate scaling

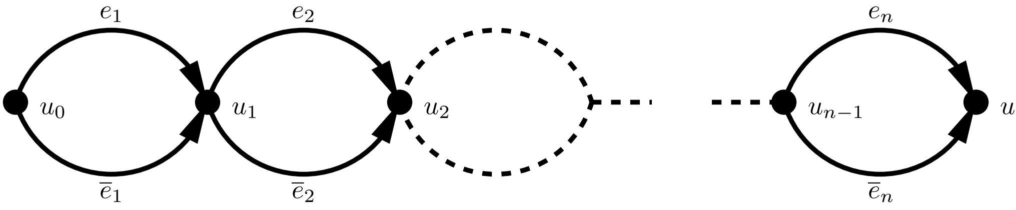

Figure 2: The simplified chain H_{bool}.

of the evaporation factor which supports the understanding of the model and eases an improvement of the given runtime bounds.

First, we adapt the graph G_{bool} from the previous section to our specific needs. At the end of the previous section we observed that for all $i \in \{1, \ldots, n\}$, whenever the ant uses the edge (u_{i-1}, w_i) it also uses the edge (w_i, u_i) and that the same holds for the edges $(u_{i-1}, \overline{w}_i)$ and $(\overline{w}_i, u_i)$. Thus, as new construction graph we use a multi-graph instead, i. e., a directed graph with multiple edges. We replace the paths u_{i-1}, w_i, u_i and $u_{i-1}, \overline{w}_i, u_i$ by two parallel edges $e_i = (u_{i-1}, u_i)$ and $\overline{e_i} = (\overline{u_{i-1}, u_i})$, see Figure 2. By this, the graph we consider becomes a sequence of vertices such that successive vertices are joined by two parallel edges.

For $n \in \mathbb{N}$ let $H_{\text{bool}} = (U, E)$ be a multi-graph with vertices $u_0, \ldots, u_n$ and distinct edges $e_1, \ldots, e_n, \overline{e}_1, \ldots, \overline{e}_n$, such that for all $i \in \{1, \ldots, n\}$, both e_i and $\overline{e}_i$ start in u_{i-1} and end in u_i. A walk $W = (V_W, E_W)$ on H now defines the boolean vector $x \in \{0, 1\}^n$ by setting x_i to one if $e_i \in E_W$ and x_i to zero if $\overline{e}_i \in E_W$.

This adaptation, while minor, seems natural and obviously simplifies the graph. First, the artificial coupling of the pheromone values of the edges (u_{i-1}, w_i) and (w_i, u_i) in G_{bool} disappears and reduces the number of edges and pheromone values by one half. Second, the entries of the boolean vector are defined by the edges of the corresponding random walk rather then by its vertices. This fact better reflects the paradigm of ACO algorithms to have an edge-based representation of the solutions.

The second modification of the previous model is again minor in detail but major with respect to understanding the behavior of the algorithm. In the previous section, the pheromone values were requested to sum up to one. We now drop this constraint. Instead, we directly assign the probability to chose an edge in the random walks as its pheromone value. We therefore denote the pheromone value of the edge e_i by p_i and that of the edge $\overline{e}_i$ by $\overline{p}_i = 1 - p_i$. The value p_i now marks both, the pheromone value of the edge e_i and the probability of its occurrence in a walk of the ant. As in the previous model, we restrict the p_i to the interval $[1/n, 1 - 1/n]$ and set the initial pheromone values to $\tau_{\text{INIT}} := p^{(0)} = (1/2, \ldots, 1/2)$.

In consequence to the modification of the pheromone values we propose the third, most important simplification. In the old model, the probability to choose an edge e is $p_e = 2n\tau(e)$. Thus, after an update

$$p_e := \begin{cases} \min\{\frac{1-\rho}{1-\rho+2n\rho}\, p_e + \frac{2n\rho}{1-\rho+2n\rho}, \frac{n-1}{n}\} & \text{if } e \in E_W, \\ \max\{\frac{1-\rho}{1-\rho+2n\rho}\, p_e, \frac{1}{n}\} & \text{else.} \end{cases}$$

We redefine the evaporation factor by $\tilde{\rho} := \frac{2n\rho}{1-\rho+2n\rho}$. By this, the procedure UPDATE changes to

```
UPDATE($H, p, E_W$)
for ($i \in \{1, \ldots, n\}$)
    do if ($e_i \in E_W$)
        then $p_i := \min\{(1-\tilde{\rho})p_i + \tilde{\rho}, 1 - \frac{1}{n}\}$
        else  $p_i := \max\{(1-\tilde{\rho})p_i, \frac{1}{n}\}$
return $p$
```

Comparing this new UPDATE procedure to the previous one clearly demonstrates the simplification introduced by the new model. Since $\rho := \frac{\tilde{\rho}}{2n-(2n-1)\tilde{\rho}}$ all results in the new model can be translated to the previous model and vice versa. To ease reading, we computed some useful values in Table 1.

The influence of the evaporation factor $\tilde{\rho}$ on the pheromone values now becomes directly clear. For pheromone values not close to $1/n$ or $1 - 1/n$ we can interpret an update as the evaporation of a $\tilde{\rho}$-th of the existing pheromones and the subsequent strengthening of the used edges by $\tilde{\rho}$.

Alternatively, we can rewrite $(1-\tilde{\rho})p_i + \tilde{\rho}$ as $p_i + \tilde{\rho}(1 - p_i)$ and $(1-\tilde{\rho})p_i$ as $p_i - \tilde{\rho}p_i$. Thus, if e_i is used by the ant, then its pheromone value is raised by $\tilde{\rho}$ times the probability that the edge is not chosen. Respectively, if e_i is not used then p_i is decreased by $\tilde{\rho}$ times its current value. Because of this, pheromone values are less increased if close to one and less decreased if close to zero.

ρ	0	$1/n^{1+\epsilon}$	$\sim 1/(2n \ln n)$	$1/n^{1-\epsilon}$	1
$\tilde{\rho}$	0	$\sim 2/n^{\epsilon}$	$1/\ln n$	$\sim 1 - 1/2n^{\epsilon}$	1

Table 1: The relation between ρ and $\tilde{\rho}$.

7 Optimization of OneMax

The main advantage of the simplified model becomes clear if we investigate optimization problems with linear objective functions like ONEMAX. The objective function to be maximized in ONEMAX is the *size* $|x| = \sum_{i=1}^{n} x_i$ of the boolean vector $x \in \{0, 1\}^n$. Obviously, this maximum is unique and attained by the vector $(1, \ldots, 1)$ of size n.

We consider the vector $x \in \{0,1\}^n$ corresponding to a random walk of an ant to be a sample of a random 0-1-vector $X = (X_1, \ldots, X_n)$. The distribution of X is depends on the new pheromone values $p_1, \ldots, p_n$, since p_i is the probability that $X_i = 1$ for all $i \in \{1, \ldots, n\}$. Note that all X_i are mutually independent.

The value of the objective function is the size $|x|$ of x and represented by the random variable $S_X = |X|$. Observe that by linearity of expectation $\mathrm{E}\big[S_X\big] = \sum_{i=1}^n p_i$.

A typical run of 1-Ant on OneMax first produces a sample $x \in \{0,1\}^n$ of the random variable $X^{(0)}$ distributed according to $p^{(0)} = (1/2, \ldots, 1/2)$. Let $Y^{(0)} = x$, then $Y^{(0)}$ is a random vector distributed like $X^{(0)}$. Next, 1-Ant updates $p^{(0)}$ to $p^{(1)}$. Suppose $\tilde{\rho} \le 1 - 2/n$, then

$$p_i^{(1)} = \begin{cases} \frac{1}{2} + \frac{\tilde{\rho}}{2} & \text{if } Y_i^{(0)} = 1\,, \\ \frac{1}{2} - \frac{\tilde{\rho}}{2} & \text{if } Y_i^{(0)} = 0\,. \end{cases}$$

We can now identify *phases* $t \in \mathbb{N}$ in the run of 1-Ant. The initial triple $p^{(0)}$, $X^{(0)}$, and $Y^{(0)}$ marks the 0-th phase. In general, a new phase starts after the pheromone vector has been updated. Thus, there is a new pheromone vector $p^{(t)}$ for each phase t.

This pheromone vector defines the random vector $X^{(t)}$ and its size $S_X^{(t)} = \sum_{i=1}^n X_i^{(t)}$. During phase t 1-Ant repeatedly generates samples $x \in \{0,1\}$ of $X^{(t)}$. The phase ends when a sample x^* with size at least the size of all the x from previous phases has been generated.

The random vector $Y^{(t)}$ is now defined to be this last x^* of phase t. Accordingly, $S_Y^{(t)} = \sum_{i=1}^n Y^{(t)}$ is defined to be its size. Since $Y^{(t)}$ has size at least as the size of all its predecessors, $S_Y^{(t)} \ge S_Y^{(t-1)}$ holds for all $t > 0$. In particular, phase t ends exactly when a generated x has size at least $S_Y^{(t-1)}$.

Between phase t and phase $t+1$ the pheromone vector $p^{(t)}$ is updated:

$$p_i^{(t+1)} = \begin{cases} \min\{(1-\tilde{\rho})p_i^{(t)} + \tilde{\rho}, 1 - \frac{1}{n}\} & \text{if } Y_i^{(t)} = 1\,, \\ \max\{(1-\tilde{\rho})p_i^{(t)}, \frac{1}{n}\} & \text{if } Y_i^{(t)} = 0\,. \end{cases}$$

For early phases t, the values of $p^{(t)}$ are bounded as follows.

Lemma 1. *Let $\tilde{\rho} \le 1/2$, $t \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$. Then*

$$\tfrac{1}{2} e^{-2\tilde{\rho}t} \le p_i^{(t)} \le 1 - \tfrac{1}{2} e^{-2\tilde{\rho}t}\,.$$

Proof. By induction $\frac{1}{2}(1-\tilde{\rho})^t \le p_i^{(t)} \le 1 - \frac{1}{2}(1-\tilde{\rho})^t$ holds for all $i \in \{1, \ldots, n\}$. The claim then follows as $1 - \tilde{\rho} \ge e^{-2\tilde{\rho}}$ holds for all $0 \le \tilde{\rho} \le 1/2$. □

As a corollary we obtain for $t \le 1/\tilde{\rho}$ and $n \ge 2e^2$

$$p_i^{(t)} = \begin{cases} (1-\tilde{\rho})p_i^{(t-1)} + \tilde{\rho} & \text{if } Y_i^{(t-1)} = 1\,, \\ (1-\tilde{\rho})p_i^{(t-1)} & \text{if } Y_i^{(t-1)} = 0\,. \end{cases}$$

Note that for $t \ge 1$ the distributions of $Y^{(t)}$ and $X^{(t)}$ differ. This can be easily seen, for example $S_X^{(t)}$ can be arbitrarily small while $S_Y^{(t)}$ is at least $S_Y^{(t)}$. However, $Y^{(t)}$ and $X^{(t)}$ have the same distribution under the condition $X^{(t)} \ge Y^{(t-1)}$. Formally, for $t > 0$ and $x \in \{0,1\}^n$

$$\Pr\big[Y^{(t)} = x\big] = \Pr\big[X^{(t)} = x \mid S_X^{(t)} \ge S_Y^{(t-1)}\big]\,.$$

As we see, the distribution of $Y^{(t)}$ depends not only on $p^{(t)}$ like that of $X^{(t)}$ but also on $Y^{(t-1)}$.

The two random variables $S_X^{(t)}$ and $S_Y^{(t)}$ turn out to be the key to the understanding of how the optimization time of 1-Ant behaves. While $S_X^{(t)}$ describes the single solutions generated in the current phase, $S_Y^{(t)}$ denotes the best solution 1-Ant has generated at the end of phase t. The distributions of $S_X^{(t)}$ and $S_Y^{(t)}$ are related like those of $X^{(t)}$ and $Y^{(t)}$. For $t > 0$ and $k \in \{0, \ldots, n\}$

$$\Pr\big[S_Y^{(t)} = k\big] = \Pr\big[S_X^{(t)} = k \mid S_X^{(t)} \ge S_Y^{(t-1)}\big]\,.$$

Due to the choice of our model, the values $\mu_X^{(t)} = \mathrm{E}\big[S_X^{(t)}\big]$ and $\sigma_X^{(t)} = \mathrm{Var}[S_X^{(t)}]^{\frac{1}{2}}$ can now be interpreted meaningfully. Namely, $\mu_X^{(t)}$ denotes the average value of a single solution in phase t with derivation $\sigma_X^{(t)}$. For example, the size of first vector almost surely differs by at most $\sigma_X^{(0)} = \sqrt{n}/2$ from $\mu_X^{(0)} = n/2$. We will see that we can keep track on the behavior of $\mu_X^{(t)}$ and $\sigma_X^{(t)}$.

To monitor $\mu_Y^{(t)} = \mathrm{E}\big[S_Y^{(t)}\big]$ is more difficult. Obviously, $\mu_Y^{(0)} = n/2$ since $Y^{(0)}$ is distributed like $X^{(0)}$. But for $t > 0$ we can express $\mu_Y^{(t)}$ only as $\mu_Y^{(t)} = \mathrm{E}\big[S_X^{(t)} \mid S_X^{(t)} \ge S_Y^{(t-1)}\big]$.

We use the same stochastic techniques as [18] to approximate the probability that $S_X^{(t)}$ exceeds $S_Y^{(t-1)}$. Nevertheless, our scaling of the pheromone values allow us to directly translate the results to an understanding of the algorithm's behavior and results in more concise proofs.

The following lemma is central in the study of 1-Ant on OneMax and a good example of this claim. It states that the average size of a vector generated in early phases depends only on the size of the so far accepted solutions, but not on how these solutions actually look like.

Lemma 2. *Let $\tilde{\rho} \le 1/2$ and $0 < t \le 1/\tilde{\rho}$, then*

$$\mu_X^{(t)} = \mu_X^{(t-1)} + \tilde{\rho}(S_Y^{(t-1)} - \mu_X^{(t-1)})\,.$$

Proof.

$$\begin{aligned} \mu_X^{(t)} &= \sum_{i\colon Y_i^{(t-1)}=1} p_i^{(t)} + \sum_{i\colon Y_i^{(t-1)}=0} p_i^{(t)} \\ &= \left(\sum_{i=1}^n (1-\tilde{\rho})p_i^{(t-1)}\right) + \tilde{\rho} S_Y^{(t-1)}\,. \end{aligned}$$

and the statement follows from the definition of $\mu_X^{(t-1)}$. □

We conclude this section with a general lower and upper bounds on the standard deviation $\sigma_X^{(t)} = \mathrm{Var}[S_X^{(t)}]^{\frac{1}{2}}$.

Lemma 3. *Let $\tilde{\rho} \le 1/2$ and $t \in \mathbb{N}$. Then*

$$\sigma_X^{(t)} \ge \tfrac{1}{2} e^{-\tilde{\rho}t} n^{1/2}\,.$$

Proof. By Lemma 1 and $(\sigma_X^{(t)})^2 = \sum_{i=1}^n p_i^{(t)}(1 - p_i^{(t)})$. □

8 A New Lower Bound for the Optimization Time of 1–Ant on OneMax

In this section we study how the optimization time of 1-Ant on OneMax depends on the evaporation factor $\tilde{\rho}$. In [18] it was shown that for all fixed $\epsilon > 0$ and $\tilde{\rho} = 1/n^{\epsilon}$ (equivalently, $\rho = 1/n^{1+\epsilon}$) and $\epsilon > 0$ with high probability the optimization time is exponential in n.

The following theorem refines this result and shows that with high probability already for $\tilde{\rho} = o(1/\ln n)$ the optimization time is super-polynomial. It particular, we derive an explicit correspondence between a bound on the optimization time and the probability that this bound holds.

Theorem 4. *There exists a $c \in (0,1)$ and an $N \in \mathbb{N}$ such that for $n \geq N$ and $\tilde{\rho} \leq c$ the optimization time of* 1-Ant *on* OneMax *with n variables and evaporation factor $\tilde{\rho}$ is at most $e^{\frac{c}{\tilde{\rho}}}$ with probability less then $1/e^{\frac{c}{\tilde{\rho}}}$.*

We first motivate this theorem, the full proof is given at the end of this section. Suppose that $\tilde{\rho} \leq 1/\ln n$ and let us consider a typical run of 1-Ant. We are particularly interested in the first $T = 1/\tilde{\rho} \geq \ln n$ phases. Throughout these T phases, the solutions $S_X^{(t)}$ are more or less binomially distributed. In particular, the deviation $\sigma_X^{(t)}$ stays approximately $\sigma = \sqrt{n}/2$.

With high probability, the first solution $S_Y^{(0)}$ is close to its expected value, hence we suppose $S_Y^{(0)} = n/2$.

Whenever a solution is accepted as new $S_Y^{(t)}$, then with constant probability the increase $S_Y^{(t)} - S_Y^{(t-1)}$ is at least $\sigma/\sqrt{\ln n}$. Thus, by ap=plying the Chernoff bounds, we see that with high probability $S_Y^{(T)} \geq n/2 + \sqrt{n/\tilde{\rho}}$.

On the other hand, Lemma 2 tells us, that after T phases

$$\mu_X^{(T+1)} \leq n/2 + 1/2\sqrt{n/\tilde{\rho}}\,.$$

Applying the Chernoff bounds again, we can bound the probability that the next solution $S_X^{(T+1)}$ can bridge the gap of $\sigma\sqrt{1/\tilde{\rho}}$ between its expected value $\mu_X^{(T+1)}$ and the current optimal solution $S_Y^{(T)}$. For $\tilde{\rho} = 1/\ln n$ this probability is roughly polynomial in $1/n$, for $\tilde{\rho} = o(1/\ln n)$ it decreases rapidly.

The remainder of this section is devoted to the proof of Theorem 4.

First, note that while $S_Y^{(t)}$ is clearly increasing in t, this is not necessarily true for $\mu_Y^{(t)}$. This is due to the fact that we may accept solutions of size less then $n/2$ in early phases. We do, however, have the following lower bound on $\mu_X^{(t)}$.

Lemma 5. *For all $t \geq 0$,*

$$\mu_X^{(t)} \geq \min\{\tfrac{n}{2}, S_Y^{(0)}\}\,.$$

Proof. Since $p_i^{(0)} = 1/2$ for all $i \in \{1,\ldots,n\}$, we have that $\mu_X^{(0)} = \sum_{i=0} p_i^{(0)} = n/2$. Furthermore, $S_Y^{(t)} \geq S_Y^{(0)}$ and by Lemma 2 also $\mu_X^{(t)} = (1-\tilde{\rho})\mu_X^{(t-1)} + \tilde{\rho}S_Y^{(t-1)}$. Thus, the inequality $\mu_X^{(t)} \geq \min\{\mu_X^{(t-1)}, S_Y^{(t-1)}\}$ holds by a simple induction on t. □

From this lower bound on $\mu_X^{(t)}$ we obtain the following upper bound of $\mu_X^{(t)}$.

Lemma 6. *For all $t \geq 0$,*

$$\mu_X^{(t)} \leq \tfrac{n}{2} + \tilde{\rho}t\max\{S_Y^{(t-1)} - \tfrac{n}{2}, S_Y^{(t-1)} - S_Y^{(0)}\}\,.$$

Proof. Again invoking Lemma 2, we compute

$$\begin{aligned}
\mu_X^{(t)} &= \mu_X^{(0)} + \sum_{s=1}^{t}(\mu_X^{(s)} - \mu_X^{(s-1)}) \\
&= \tfrac{n}{2} + \sum_{s=1}^{t}\tilde{\rho}(S_Y^{(s-1)} - \mu_X^{(s-1)}) \\
&\leq \tfrac{n}{2} + \tilde{\rho}\sum_{s=0}^{t}(S_Y^{(t-1)} - \min\{\tfrac{n}{2}, S_Y^{(0)}\}) \\
&= \tfrac{n}{2} + \tilde{\rho}t\max\{S_Y^{(t-1)} - \tfrac{n}{2}, S_Y^{(t-1)} - S_Y^{(0)}\}\,.
\end{aligned}$$

□

We need two tools from probability theory. The following lemma was proven by Neumann and Witt ([18] , Lemma 3).

Lemma 7. *Let $S_n = \sum_{i=0}^{n} X_i$ be the sum of mutually independent random 0-1-variables $X_1,\ldots,X_n$, $\mu_n = E[S_n]$ and $\sigma_n^2 = Var[S_n]$. Let $0 \leq a \leq n - \sigma$, $\gamma_n = \max\{\frac{a-\mu_n}{\sigma_n}, 2\}$ and $\sigma_n \to \infty$. Then there exists a $p_0 \in (0,1)$ and an $N \in \mathbb{N}$ such that $Pr[S_n \geq a + \frac{\sigma_n}{\gamma_n} \mid S_n \geq a] \geq p_0$ for all $n \geq N$.*

The second tool we use are the well-known Chernoff bounds (see for example[1]).

Theorem 8 (Chernoff)**.** *Let $S_n = \sum_{i=0}^{n} X_i$ be the sum of mutually independent random 0-1-variables $X_1,\ldots,X_n$ and let $\mu = E[S_n]$. Then for all $a > 0$ and $\delta \in [0,1]$,*

(i) $Pr[S_n \leq \mu + a] \leq e^{-\frac{2a^2}{n}}$,

(ii) $Pr[S_n \geq \mu + a] \leq e^{-\frac{2a^2}{n}}$, and

(iii) $Pr[S_n \leq (1-\delta)\mu] \leq e^{-\frac{\delta^2\mu}{2}}$.

We are now ready to prove Theorem 4.

Proof. Let $N \in \mathbb{N}$ and $p_0 \in (0,1)$ as in Lemma 7. Also, let $c = \frac{p_0}{5120}$, $\tilde{\rho} \leq c$, $\Delta = \sqrt{8cn/\tilde{\rho}}$, and $T = \lfloor \frac{1}{4\tilde{\rho}} \rfloor$. Then, it holds that $c \in (0,1)$, $\Delta \geq \sqrt{n}$, and $\frac{1}{5\tilde{\rho}} \leq T \leq \frac{1}{4\tilde{\rho}}$.

Let E_1 be the event that $\frac{n}{2} - \Delta < S_Y^{(0)} < \frac{n}{2} + \Delta$ and E_2 the event that $S_Y^{(T)} > \frac{n}{2} + \Delta$.

We first show that $\Pr[E_1] \geq 1 - 2e^{-\frac{16c}{\tilde{\rho}}}$. Since $S_Y^{(0)}$ has the same distribution as $S_X^{(0)}$, we apply the first two Chernoff bounds and obtain

$$\Pr[S_Y^{(0)} \geq \frac{n}{2} + \Delta] \leq e^{-\frac{2\Delta^2}{n}}$$

and

$$\Pr[S_Y^{(0)} \leq \frac{n}{2} - \Delta] \leq e^{-\frac{2\Delta^2}{n}}\,.$$

Hence, E_1 occurs at least with probability

$$1-2e^{-\frac{2\Delta^2}{n}} = 1-2e^{-16c/\tilde{\rho}}\,.$$

Next, we show that $\Pr\bigl[E_2 \mid E_1\bigr] \ge 1-e^{-\frac{128c}{\tilde{\rho}}}$. Suppose, that E_1 holds, in particular $S_Y^{(0)} \ge \frac{n}{2}-\Delta$. Suppose further, that $S_Y^{(T)} \le \frac{n}{2}+\Delta$. We show that for this to happen, a number of very unlikely events have to occur.

For $t \le T$, let

$$\gamma_t = \max\{\frac{S_Y^{(t-1)}-\mu_X^{(t)}}{\sigma_X^{(t)}}, 2\}\,.$$

Since $T \le \frac{1}{4\tilde{\rho}}$, we have $S_Y^{(t-1)} - \mu_X^{(t)} \le 2\Delta$ by Lemma 6 and $\sigma_X^{(t)} \ge \frac{1}{4}n^{\frac{1}{2}}$ by Lemma 3. It also holds that $8\Delta n^{-\frac{1}{2}} \ge 2$, and thus we get $\gamma_t \le 8\Delta n^{-\frac{1}{2}}$. Hence, by Lemma 7 we obtain

$$\Pr\bigl[S_X^{(t)} \ge S_Y^{(t-1)} + \tfrac{n}{32\Delta} \mid S_X^{(t)} \ge S_Y^{(t-1)} \wedge E_1\bigr] \ge p_0\,.$$

Thus,

$$\Pr\bigl[S_Y^{(t)} - S_Y^{(t-1)} \ge \tfrac{n}{32\Delta} \mid E_1\bigr] \ge p_0\,.$$

For $t \ge 0$ we define the auxiliary random 0-1-variables

$$Z^{(t)} = \begin{cases} 1 & \text{if } S_Y^{(t)} - S_Y^{(t-1)} \ge \frac{n}{32\Delta}\,, \\ 0 & \text{else.} \end{cases}$$

Let $S_Z = \sum_{t=1}^{T} Z^{(t)}$ and $\mu_Z = \mathrm{E}\bigl[S_Z \mid E_1\bigr] \ge p_0 T$. Then $\mu_Z \ge \frac{1024c}{\tilde{\rho}}$, as $T \ge \frac{1}{5\tilde{\rho}}$ and $p_0 = 5120c$. By the third Chernoff bound

$$\Pr\bigl[S_Z \le \tfrac{1}{2}\mu_Z \mid E_1\bigr] \le e^{-\frac{128c}{\tilde{\rho}}}\,.$$

Hence,

$$\Pr\bigl[S_Z > \tfrac{64\Delta^2}{n} \mid E_1\bigr] \ge 1 - e^{-\frac{128c}{\tilde{\rho}}}\,,$$

since $\frac{\Delta^2}{n} = \frac{8c}{\tilde{\rho}}$. Now,

$$S_Y^{(T)} = S_Y^{(0)} + \sum_{t=1}^{T} S_Y^{(t)} - S_Y^{(t-1)} \ge \tfrac{n}{2} - \Delta + \tfrac{n}{32\Delta} S_Z\,.$$

Thus, $\Pr\bigl[S_Y^{(T)} > \frac{n}{2}+\Delta \mid E_1\bigr] \ge 1-e^{-128c/\tilde{\rho}}$.

Suppose,

$$\frac{n}{2} - \Delta \le S_Y^{(0)} < \frac{n}{2} + \Delta < S_Y^{(T)}\,,$$

that is, E_1 and E_2 hold. Then there is a $t^* \le T$ such that $S_Y^{(t^*-1)} \le \frac{n}{2}+\Delta$, but $S_Y^{(t^*)} > \frac{n}{2}+\Delta$.

Let $\mu^* = \mathrm{E}\bigl[S_X^{(t^*)} \mid E_1 \wedge E_2\bigr]$. Then by Lemma 6 it holds that

$$\mu^* \le \tfrac{n}{2} + 2\tilde{\rho}T\Delta \le \tfrac{n}{2} + \tfrac{\Delta}{2}\,.$$

Thus, by the Chernoff bound

$$\Pr\bigl[S_X^{(t^*)} \ge \mu^* + \tfrac{\Delta}{2} \mid E_1 \wedge E_2\bigr] \le e^{-\frac{\Delta^2}{2n}}\,.$$

Hence, $\Pr\bigl[S_X^{(t^*)} > \frac{n}{2}+\Delta \mid E_1 \wedge E_2\bigr] \le e^{-4c/\tilde{\rho}}$.

Under the condition $E_1 \wedge E_2$, the probability that the optimization time of 1-Ant on OneMax is at most $e^{\frac{c}{\tilde{\rho}}}$ is bounded by $e^{\frac{c}{\tilde{\rho}}}\Pr\bigl[S_X^{(t^*)} > \frac{n}{2}+\Delta \mid E_1 \wedge E_2\bigr] \le e^{-3c/\tilde{\rho}}$. But then the unconditional probability that the optimization time of 1-Ant on OneMax is at most $e^{\frac{c}{\tilde{\rho}}}$ is bounded by $2e^{-16c/\tilde{\rho}} + e^{-128c/\tilde{\rho}} + e^{-3c/\tilde{\rho}} \le e^{-c/\tilde{\rho}}$. □

9 Conclusion

In this work we proposed a novel pheromone model for pseudo-boolean optimization problems. We identified the pheromone values with their corresponding probabilities. This lead to a redefined evaporation factor that supports a better understanding of pheromone updates and links the mathematical notions of expectancy and variance to conceptual invariants of the optimization process.

We were able to refine the runtime bounds in [18] and to extended them to values of ρ inside the critical window of the phase transition. We are confident that our refined model will allow further insight in the runtime behavior within the phase transition and prove superior for the future analysis of ACO algorithms on other pseudo-boolean optimization problems.

References

[1] N. Alon and J. Spencer. *The Probabilistic Method.* Interscience Series in Discrete Mathematics and Optimization. Wiley, 2000.

[2] B. Doerr, N. Hebbinghaus, and F. Neumann. Speeding up evolutionary algorithms through restricted mutation operators. In *Proceedings of the 9th International Conference on Parallel Problem Solving From Nature (PPSN)*, volume 4193 of *Lecture Notes in Computer Science*, pages 978–987. Springer, 2006.

[3] B. Doerr and D. Johannsen. Adjacency list matchings — an ideal genotype for cycle covers. In *Genetic and Evolutionary Computation Conference (GECCO-2007)*, pages 1203–1210. ACM, 2007.

[4] B. Doerr, C. Klein, and T. Storch. Faster evolutionary algorithms by superior graph representation. In *Proceedings of the First IEEE Symposium on Foundations of Computational Intelligence (FOCI)*, pages 245–250. IEEE Press, 2007.

[5] B. Doerr, F. Neumann, D. Sudholt, and C. Witt. On the runtime analysis of the 1-ANT ACO algorithm. In D. Thierens, editor, *Genetic and Evolutionary Computation Conference (GECCO-2007)*, pages 33–40. ACM, 2007.

[6] M. Dorigo and C. Blum. Ant colony optimization theory: A survey. *Theor. Comput. Sci.*, 344:243–278, 2005.

[7] M. Dorigo, V. Maniezzo, and A. Colorni. The ant system: An autocatalytic optimizing process. Technical Report 91-016 Revised, Politecnico di Milano, 1991.

[8] M. Dorigo and T. Stützle. *Ant Colony Optimization.* MIT Press, 2004.

[9] S. Droste, T. Jansen, and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theor. Comput. Sci.*, 276:51–81, 2002.

[10] O. Giel and I. Wegener. Evolutionary algorithms and the maximum matching problem. In *Proc. of STACS '03*, volume 2607 of *LNCS*, pages 415–426, 2003.

[11] W. J. Gutjahr. A generalized convergence result for the graph-based ant system metaheuristic. *Probab. Eng. Inform. Sc.*, 17:545–569, 2003.

[12] W. J. Gutjahr. On the finite-time dynamics of ant colony optimization. *Methodology and Computing in Applied Probability*, 8:105–133, 2006.

[13] W. J. Gutjahr and G. Sebastiani. Runtime analysis of ant colony optimization. Technical report, Mathematics department, Sapienza University of Rome, 2007/03, 2007.

[14] M. Jerrum and G. B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Appl. Math.*, 82:155–175, 1998.

[15] F. Neumann. Expected runtimes of evolutionary algorithms for the Eulerian cycle problem. In *Proceedings of the 2004 IEEE Congress on Evolutionary Computation (CEC)*, pages 904–910. IEEE Press, 2004.

[16] F. Neumann and I. Wegener. Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. In *Proc. of GECCO '04*, volume 3102 of *LNCS*, pages 713–724, 2004.

[17] F. Neumann and C. Witt. Ant colony optimization and the minimum spanning tree problem. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2006. Report No. 143.

[18] F. Neumann and C. Witt. Runtime analysis of a simple ant colony optimization algorithm. In *Proc. of ISAAC '06*, volume 4288 of *LNCS*, pages 618–627. Springer, 2006.

[19] C. H. Papadimitriou, A. A. Schäffer, and M. Yannakakis. On the complexity of local search. In *Proc. of STOC '90*, pages 438–445. ACM Press, 1990.

[20] C. Witt. Worst-case and average-case approximations by simple randomized search heuristics. In *Proc. of STACS '05*, volume 3404 of *LNCS*, pages 44–56, 2005.