

Runtime Analysis of a Simple Ant Colony Optimization Algorithm

Frank Neumann · Carsten Witt

Received: 22 January 2007 / Accepted: 20 November 2007 / Published online: 28 November 2007
© Springer Science+Business Media, LLC 2007

Abstract Ant Colony Optimization (ACO) has become quite popular in recent years. In contrast to many successful applications, the theoretical foundation of this randomized search heuristic is rather weak. Building up such a theory is demanded to understand how these heuristics work as well as to come up with better algorithms for certain problems. Up to now, only convergence results have been achieved showing that optimal solutions can be obtained in finite time. We present the first runtime analysis of an ACO algorithm, which transfers many rigorous results with respect to the runtime of a simple evolutionary algorithm to our algorithm. Moreover, we examine the choice of the evaporation factor, a crucial parameter in ACO algorithms, in detail. By deriving new lower bounds on the tails of sums of independent Poisson trials, we determine the effect of the evaporation factor almost completely and prove a phase transition from exponential to polynomial runtime.

Keywords Randomized search heuristics · Ant colony optimization · Runtime analysis

A preliminary version of this paper appeared in the Proceedings of the 17th International Symposium on Algorithms and Computation (ISAAC 2006), volume 4288 of LNCS, pp. 618–627, Springer.

Financial support for C. Witt by the Deutsche Forschungsgemeinschaft (SFB) in terms of the Collaborative Research Center “Computational Intelligence” (SFB 531) is gratefully acknowledged.

F. Neumann (✉)

Algorithms and Complexity, Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany
e-mail: fne@mpi-inf.mpg.de

C. Witt

FB Informatik, LS 2, Universität Dortmund, 44221 Dortmund, Germany
e-mail: carsten.witt@cs.uni-dortmund.de

1 Introduction

The analysis of randomized search heuristics with respect to their runtime is a growing research area where many results have been obtained in recent years. This class of heuristics contains well-known approaches such as Randomized Local Search (RLS), the Metropolis Algorithm (MA), Simulated Annealing (SA), and Evolutionary Algorithms (EAs). Such heuristics are often applied to problems whose structure is not known or if there are not enough resources such as time, money, or knowledge to obtain good specific algorithms. It is widely acknowledged that a solid theoretical foundation for such heuristics is needed.

Some general results on the runtime of RLS can be found in Papadimitriou, Schäfer and Yannakakis [13]. The graph bisection problem has been subject to analysis of MA [11], where MA can be seen as SA with a fixed temperature. For a long time, it was an open question whether there is a natural example where SA outperforms MA for all fixed temperatures. This question has recently been answered positively by Wegener [15] for instances of the minimum spanning tree problem.

In this paper, we focus on another kind of randomized search heuristics, namely Ant Colony Optimization (ACO). Like EAs, these heuristics imitate optimization processes from nature, in this case the search of an ant colony for a common source of food. Solving problems by ACO techniques has become quite popular in recent years. Developed by Dorigo, Maniezzo and Coloni [3], they have shown to be a powerful heuristic approach to solve combinatorial optimization problems such as the TSP (see [2], for an overview on the problems that these heuristics have been applied to). From a theoretical point of view, there are no results that provide estimates of the runtime of ACO algorithms. Despite interesting theoretical investigations of models and dynamics of ACO algorithms [1], convergence results are so far the only results related to their runtimes. Dorigo and Blum [1] explicitly formulate the open problem to determine the runtime of ACO algorithms on simple problems in a similar fashion to what has been done for EAs.

We solve this problem, starting the analysis of ACO algorithms with respect to their expected runtimes and success probability after a specific number of steps. RLS, SA, MA, and simple EAs search more or less locally, and runtime bounds are often obtained by considering the neighborhood structure of the considered problem. Considering ACO algorithms, this is different as search points are obtained by random walks of ants on a so-called construction graph. The traversal of an ant on this graph is determined by values on the edges which are called pheromone values. Larger pheromone values correspond to a higher probability of traversing a certain edge, where the choice of an edge usually fixes a parameter in the current search space. The pheromone values are updated if a good solution has been constructed in this random walk. This update depends on the traversal of the ant and a so-called evaporation factor ρ .

The choice of ρ seems to be a crucial parameter in an ACO algorithm. Using a large value of ρ , the last accepted solution changes the pheromone values by a large amount such that there is a large probability of producing this solution in the next step. In contrast to this, the use of a small evaporation factor leads to a small effect of the last accepted solution such that an improvement may be hard to find in the next step.

We show that a simple ACO algorithm behaves for very large values of ρ (namely $\rho \geq 1/3$) as the simplest EA called $(1 + 1)$ EA. This algorithm has been studied extensively with respect to its runtime on pseudo-boolean functions $f: \{0, 1\}^n \rightarrow \mathbb{R}$ (see, e.g. [4]) as well as on combinatorial optimization problems. The list of problems where runtime bounds have been obtained include some of the best-known polynomially solvable problems such as maximum matchings [7] and minimum spanning trees [12]. It should be clear that we cannot expect such general heuristics to outperform the best-known algorithms for these mentioned problems. The main aim of such analyses is to get an understanding how these heuristics work. In the case of NP-hard problems, one is usually interested in good approximations of optimal solutions. Witt [16] has presented a worst-case and average-case analysis of the $(1 + 1)$ EA for the partition problem, which is one of the first results on NP-hard problems. All these results immediately transfer to our ACO algorithm with very large ρ .

After these general results, we consider the effect of the evaporation factor ρ on the runtime of our ACO algorithm in detail. As proposed in the open problem stated by Dorigo and Blum [1], we examine the simplest non-trivial pseudo-boolean function called ONEMAX and analyze for the first time for which choices of ρ the runtime with high probability is upper bounded by a polynomial and for which choices it is exponential. We observe a phase transition from exponential to small polynomial runtime when ρ crosses the threshold value $1/n$. Larger values of ρ imply that the expected function value of a new solution is determined by the function value of the best seen solution. Then an improvement will be achieved after an expected polynomial number of steps. In the case of smaller ρ , an improvement does not increase the expected function value sufficiently. Here exponential lower bounds are obtained by showing that there is a large gap between the expected value and the best-so-far function value. Both the proof of the upper and the lower runtime bound contain new analytical tools to lower bound the tail of a sum of independent trials with different success probabilities. The new tools may be of independent interest in other probabilistic analyses.

In Sect. 2, we introduce the simple ACO algorithm which we will consider. We investigate its relation to the $(1 + 1)$ EA in Sect. 3 and transfer the results on this EA to our algorithm. In Sect. 4, we investigate the choice of the evaporation factor ρ for the function ONEMAX in greater detail. We finish with some conclusions.

2 The Algorithm

Gutjahr [9] has considered a graph-based ant system and investigated under which conditions such an algorithm converges to an optimal solution. We consider a simple graph-based ant system metaheuristic that has been inspired by this algorithm. Such a heuristic produces solutions by random walks on a construction graph (see, e.g., [2]). Let $C = (V, E)$ be the construction graph with a designated start vertex s and pheromone values τ on the edges. Starting at s , an ant traverses the construction graph depending on the pheromone value using Algorithm 1. Assuming that the ant is at vertex v , the ant moves to a successor w of v , where w is chosen proportionally to the pheromone values of all non-visited successors of v . The process is iterated until a situation is reached where all successors of the current vertex v have been visited.

Algorithm 1 (Construct(C, τ))

1. $v := s$, mark v as visited.
2. While there is a successor of v in C that has not been visited:
 - (a) Let N_v be the set of non-visited successors of v and $T := \sum_{(v,w) | w \in N_v} \tau_{(v,w)}$.
 - (b) Choose one successor w of v where the probability of selection of any fixed $u \in N_v$ is $\tau_{(v,u)} / T$.
 - (c) Mark w as visited, set $v := w$ and go to 2.
3. Return the solution x and the path $P(x)$ constructed by this procedure.

Based on this construction procedure, solutions of our simple ACO algorithm (see Algorithm 2) called 1-ANT are constructed. In the initialization step, each edge gets a pheromone value of $1/|E|$ such that the pheromone values sum up to 1. After that, an initial solution x^* is produced by a random walk on the construction graph and the pheromone values are updated with respect to this walk. In each iteration, a new solution x is constructed and the pheromone values are updated if this solution is not inferior to the currently best solution x^* . We formulate our algorithm for maximization problems although it can be easily adapted to minimization.

Algorithm 2 (1-ANT)

1. Set $\tau_{(u,v)} = 1/|E|$ for all $(u, v) \in E$.
2. Compute x (and $P(x)$) using Construct(C, τ).
3. Update($\tau, P(x)$) and set $x^* := x$.
4. Compute x (and $P(x)$) using Construct(C, τ).
5. If $f(x) \geq f(x^*)$, Update($\tau, P(x)$) and set $x^* := x$.
6. Go to 4.

For theoretical investigations, it is common to have no termination condition in such an algorithm. One is interested in the random optimization time which equals the number of constructed solutions until the algorithm has produced an optimal search point. Usually, we try to bound the expected value of this time.

We take a general view and consider optimization for pseudo-boolean goal functions $f: \{0, 1\}^n \rightarrow \mathbb{R}$ for $n \geq 3$ using the canonical construction graph in our setting, $C_{\text{bool}} = (V, E)$ (see Fig. 1) with $s = v_0$. In the literature, this graph is also known as *Chain* [10]. Optimizing bitstrings of length n , the graph has $3n + 1$ vertices and $4n$ edges. The decision whether a bit $x_i, 1 \leq i \leq n$, is set to 1 is made

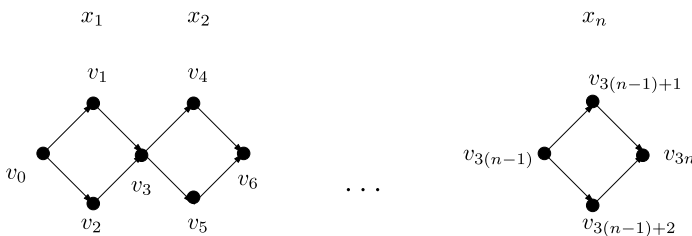


Fig. 1 Construction graph for pseudo-boolean optimization

at node $v_{3(i-1)}$. In case that the edge $(v_{3(i-1)}, v_{3(i-1)+1})$ is chosen, x_i is set to 1 in the constructed solution. Otherwise $x_i = 0$ holds. After this decision has been made, there is only one single edge which can be traversed in the next step. In case that $(v_{3(i-1)}, v_{3(i-1)+1})$ has been chosen, the next edge is $(v_{3(i-1)+1}, v_{3i})$, and otherwise the edge $(v_{3(i-1)+2}, v_{3i})$ will be traversed. Hence, these edges have no influence on the constructed solution and we can assume $\tau_{(v_{3(i-1)}, v_{3(i-1)+1})} = \tau_{(v_{3(i-1)+1}, v_{3i})}$ and $\tau_{(v_{3(i-1)}, v_{3(i-1)+2})} = \tau_{(v_{3(i-1)+2}, v_{3i})}$ for $1 \leq i \leq n$. We call the edges $(v_{3(i-1)}, v_{3(i-1)+1})$ and $(v_{3(i-1)+1}, v_{3i})$ *1-edges* and the other edges *0-edges*. We define the *1-potential* as the sum of pheromone values on 1-edges and call the edges $(v_{3(i-1)}, v_{3(i-1)+1})$ and $(v_{3(i-1)}, v_{3(i-1)+2})$ as well as $(v_{3(i-1)+1}, v_{3i})$ and $(v_{3(i-1)+2}, v_{3i})$ *complementary* to each other.

The pheromone values are chosen such that at each time $\sum_{(u,v) \in E} \tau_{(u,v)} = 1$ holds. In addition, it seems to be useful to have bounds on the pheromone values (see, e.g., [1]) to ensure that each search point has a positive probability of being chosen in the next step. We restrict each $\tau_{(u,v)}$ to the interval $[\frac{1}{2n^2}, \frac{n-1}{2n^2}]$ and ensure $\sum_{(u,\cdot) \in E} \tau_{(u,\cdot)} = \frac{1}{2n}$ for $u = v_{3i}$, $0 \leq i \leq n - 1$, and $\sum_{(\cdot,v)} \tau_{(\cdot,v)} = \frac{1}{2n}$ for $v = v_{3i}$, $1 \leq i \leq n$. This can be achieved by normalizing the pheromone values after an update and replacing the current value by $\frac{1}{2n^2}$ if $\tau_{(u,v)} < \frac{1}{2n^2}$ and by $\frac{n-1}{2n^2}$ if $\tau_{(u,v)} > \frac{n-1}{2n^2}$ holds. As a consequence, the probability that a certain bit is set to 1 by the construction procedure is the pheromone value on the corresponding 1-edges multiplied by $2n$.

Depending on whether edge (u, v) is contained in the path $P(x)$ of the accepted solution x , the pheromone values are updated to τ' in the procedure $\text{Update}(\tau, P(x))$ as follows:

$$\tau'_{(u,v)} = \min \left\{ \frac{(1 - \rho) \cdot \tau_{(u,v)} + \rho}{1 - \rho + 2n\rho}, \frac{n - 1}{2n^2} \right\} \quad \text{if } (u, v) \in P(x),$$

and

$$\tau'_{(u,v)} = \max \left\{ \frac{(1 - \rho) \cdot \tau_{(u,v)}}{1 - \rho + 2n\rho}, \frac{1}{2n^2} \right\} \quad \text{if } (u, v) \notin P(x).$$

Due to the bounds on the pheromone values, the probability of fixing x_i as in an optimal solution is at least $1/n$. Hence, the 1-ANT finds an optimum for each pseudo-boolean function f regardless of ρ in expected time at most n^n .

3 1-ANT and (1 + 1) EA

We consider the relation between the 1-ANT and a simple evolutionary algorithm called (1 + 1) EA, which has extensively been studied with respect to its runtime distribution. The (1 + 1) EA starts with a solution x^* that is chosen uniformly at random and produces in each iteration a new solution x from a currently best solution x^* by flipping each bit of x^* with probability $1/n$. Hence, the probability of producing a certain solution x with Hamming distance $H(x, x^*)$ to x^* is $(1/n)^{H(x, x^*)} \cdot (1 - 1/n)^{n-H(x, x^*)}$.

Algorithm 3 ((1 + 1) EA)

1. Choose $x^* \in \{0, 1\}^n$ uniformly at random.
2. Construct x by flipping each bit of x^* independently with probability $1/n$.
3. Replace x^* by x if $f(x) \geq f(x^*)$.
4. Go to 2.

In the following, we consider the 1-ANT with values of ρ at least $\frac{n-2}{3n-2}$, which is for large n approximately $1/3$. In this case, we show that the 1-ANT behaves as the (1 + 1) EA on each function. This also means that the 1-ANT has the same expected optimization time as the (1 + 1) EA on each function.

Theorem 1 *Choosing $\rho \geq (n - 2)/(3n - 2)$, the 1-ANT has the same runtime distribution as the (1 + 1) EA on each function.*

Proof In the initialization step of the (1 + 1) EA, a bitstring is chosen uniformly at random, which means that $\text{Prob}(x_i = 1) = \text{Prob}(x_i = 0) = 1/2$ for all $i, 1 \leq i \leq n$. As $\tau_{(u,v)} = 1/(4n)$ holds for each edge $(u, v) \in E$, the probability to choose the edge (v_{3i}, v_{3i+1}) equals the probability of choosing the edge (v_{3i}, v_{3i+2}) at vertex $v_{3i}, 0 \leq i \leq n - 1$, and is $1/2$. Hence, the 1-ANT chooses the first solution uniformly at random from the search space $\{0, 1\}^n$ as the (1 + 1) EA.

Assume that the best solution so far constructed by the 1-ANT is x^* . This implies that the edges of the construction graph corresponding to this solution have been updated in the last update operation. Before the update, the value $\tau_{(u,v)}$ of each edge $(u, v) \in P(x^*)$ was at least $\frac{1}{2n^2}$ and the value $\tau_{(u,v)}$ of edges $(u, v) \notin P(x^*)$ was at most $\frac{n-1}{2n^2}$.

We inspect the case of an edge $(u, v) \in P(x^*)$ in greater detail and consider the function

$$h(\rho) := \frac{(1 - \rho) \cdot \tau_{(u,v)} + \rho}{1 - \rho + 2n\rho} \geq \frac{(1 - \rho) \cdot \frac{1}{2n^2} + \rho}{1 - \rho + 2n\rho} = \frac{1}{2n^2} \cdot \frac{1 + (2n^2 - 1)\rho}{1 + (2n - 1)\rho} =: h'(\rho).$$

For each fixed $n \geq 1$, $h'(\rho)$ is a non-decreasing function. Using $\rho \geq (n - 2)/(3n - 2)$, we get

$$h(\rho) \geq \frac{1 + (2n^2 - 1) \frac{n-2}{3n-2}}{2n^2 + (4n^3 - 2n^2) \frac{n-2}{3n-2}} = \frac{2n^3 - 4n^2 + 2n}{4n^4 - 4n^3} = \frac{n - 1}{2n^2}.$$

Hence, the pheromone value of each edge $(u, v) \in P(x^*)$ is $\frac{n-1}{2n^2}$ after the update. The pheromone value of each edge $(u, v) \notin P(x^*)$ is $\frac{1}{2n^2}$ as the sum of the pheromone values of two complementary edges is $\frac{1}{2n}$. After this update, the probability to choose in the next solution x the bit $x_i = x_i^*$ is $\frac{2n(n-1)}{2n^2} = 1 - \frac{1}{n}$ and the probability to choose $x_i = 1 - x_i^*$ is $\frac{2n}{2n^2} = 1/n$. Hence the probability to produce a specific solution x that has Hamming distance $H(x, x^*)$ to x^* is $(1/n)^{H(x,x^*)} \cdot (1 - 1/n)^{n-H(x,x^*)}$ as in the case of the (1 + 1) EA. □

4 1-ANT on OneMax

In the following, we inspect the choice of ρ in detail for a simple pseudo-boolean function called ONEMAX defined by $\text{ONEMAX}(x) = \sum_{i=1}^n x_i$. This is the simplest non-trivial function that can be considered and analyses of ACO algorithms for such simple functions are explicitly demanded by Dorigo and Blum [1]. Note that due to results on the $(1 + 1)$ EA by Droste, Jansen and Wegener [4], the expected optimization time of the 1-ANT is $O(n \log n)$ on each linear function if $\rho \geq (n - 2)/(3n - 2)$ holds.

We prepare ourselves by considering the effects of pheromone updates for a solution x^* in greater detail. Let $\tau(e)$ and $\tau'(e)$ be the pheromone values on edge e before respectively after the update. If $e \in P(x^*)$, $\tau'(e) \geq \tau(e)$ and $\tau'(e) \leq \tau(e)$ otherwise. The amount by which the pheromone value is increased on a 1-edge equals the amount the pheromone value is decreased on the complementary 0-edge. However, the change of a pheromone value depends on the previous value on the edge. In the following lemma, we bound the relative change of pheromone values. We call an edge *saturated* if and only if its pheromone value is either $\frac{1}{2n^2}$ or $\frac{n-1}{2n^2}$.

Lemma 2 *Let e_1 and e_2 be two edges of C_{bool} and let τ_1 respectively τ_2 be their current pheromone values in the 1-ANT. Let τ'_1 respectively τ'_2 be their updated pheromone values for the next accepted solution x . If $e_1, e_2 \in P(x^*)$ and none of the edges is saturated before or after the update, then $|(\tau'_1 - \tau_1) - (\tau'_2 - \tau_2)| \leq \rho|\tau_1 - \tau_2|$.*

Proof W.l.o.g., $\tau_2 \geq \tau_1$. Since $e_1, e_2 \in P(x^*)$ and no edge is saturated,

$$\tau'_1 = \frac{(1 - \rho)\tau_1 + \rho}{1 - \rho + 2n\rho} \quad \text{and} \quad \tau'_2 = \frac{(1 - \rho)\tau_2 + \rho}{1 - \rho + 2n\rho}.$$

This implies

$$(\tau'_1 - \tau_1) - (\tau'_2 - \tau_2) = \frac{\rho - \tau_1 2n\rho - (\rho - \tau_2 2n\rho)}{1 - \rho + 2n\rho} \geq 0.$$

Second, since the denominator is at least 1, we obtain

$$\tau'_1 - \tau'_2 \leq \rho(\tau_2 - \tau_1) + (\tau_1 - \tau_2), \quad \text{implying} \quad (\tau'_1 - \tau_1) - (\tau'_2 - \tau_2) \leq \rho|\tau_1 - \tau_2|.$$

Taking the absolute value of $(\tau'_1 - \tau_1) - (\tau'_2 - \tau_2)$, the claim follows. □

In the following, we will figure out which values of ρ lead to efficient runtimes of the 1-ANT and which do not. Intuitively, $1/n$ is a threshold value for ρ since the denominator $1 - \rho + 2n\rho$ of the normalization factor diverges for $\rho = \omega(1/n)$ and is $1 - \rho - o(1)$ for $\rho = o(1/n)$. We will make precise that there is a phase transition in the behavior of the 1-ANT on ONEMAX when ρ is asymptotically smaller respectively larger than $1/n$.

4.1 Exponential Lower Bounds

Choosing $\rho = 0$, the pheromone value on each edge is $1/(4n)$ at each time step. This implies that the expected optimization time of the 1-ANT on ONEMAX is 2^n as each solution is chosen uniformly at random from $\{0, 1\}^n$. In the following, we show that the optimization time with overwhelming probability still is exponential if ρ is convergent to 0 only polynomially fast.

Assume that the currently best solution x^* has value k . Then the following lemma gives a lower bound on the probability of overshooting k by a certain amount in the next accepted step. All asymptotics are with respect to n , the number of variables.

Lemma 3 *Let $X_1, \dots, X_n \in \{0, 1\}$ be independent Poisson trials with success probabilities $p_i, 1 \leq i \leq n$. Let $X := X_1 + \dots + X_n, \mu := E(X) = p_1 + \dots + p_n$ and $\sigma := \sqrt{\text{Var}(X)}$. For any $0 \leq k \leq n - \sigma$, let $\gamma_k = \max\{2, (k - \mu)/\sigma\}$. If $\sigma \rightarrow \infty$ then $\text{Prob}(X \geq k + \sigma/\gamma_k \mid X \geq k) = \Omega(1)$.*

Proof Since the X_i are bounded and σ diverges, Lindeberg’s generalization of the Central Limit Theorem [6, Chap. VIII.4] holds s. t. the distribution of X converges to a Normal distribution with expectation μ and variance σ^2 . We use approximations of the Normal distribution (with the common notion $\Phi(x)$ for its cumulative distribution function) and distinguish two cases.

If 2 maximizes γ_k , we even show $\tilde{p}_k := \text{Prob}(X \geq k + \sigma/\gamma_k) = \Omega(1)$. Let $\tilde{d}_k := (k + \sigma/\gamma_k - \mu)/\sigma$ be the normalized deviation from the expectation. Since by our assumptions $(k - \mu)/\sigma \leq 2$, we obtain $\tilde{d}_k = O(1)$. The Central Limit Theorem implies $\tilde{p}_k = (1 \pm o(1))(1 - \Phi(\tilde{d}_k)) = \Omega(1)$.

Now let $\gamma_k > 2$. Let $p_k := \text{Prob}(X \geq k), d_k := (k - \mu)/\sigma$, and let \tilde{p}_k and \tilde{d}_k as above. By our assumptions, $2 \leq d_k \leq \tilde{d}_k \leq d_k + 1/d_k$. We have to bound \tilde{p}_k/p_k from below. We reuse the Central Limit Theorem and employ the inequalities

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} < 1 - \Phi(x) < \frac{1}{x} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$$

(see [5, Chap. VII.1]). Hence,

$$\frac{\tilde{p}_k}{p_k} \geq \frac{1 - o(1)}{1 + o(1)} \cdot \left(\frac{d_k}{\tilde{d}_k} - \frac{d_k}{(\tilde{d}_k)^3}\right) \cdot e^{-(1/2)((\tilde{d}_k)^2 - (d_k)^2)}.$$

The first fraction and the term in brackets are $\Omega(1)$. Finally, $e^{-(1/2)((\tilde{d}_k)^2 - (d_k)^2)} = \Omega(1)$ since $(\tilde{d}_k)^2 - (d_k)^2 \leq (d_k + 1/d_k)^2 - (d_k)^2 \leq 2 + 1/(d_k)^2 \leq 3$. □

Using this lemma, we are able to prove an exponential lower bound on the runtime of the 1-ANT on ONEMAX. In order to show that the success probability in an exponential number of steps is still exponentially small, we assume that $\rho = O(n^{-1-\epsilon})$ for some constant $\epsilon > 0$.

Theorem 4 *Let $\rho = O(n^{-1-\epsilon})$ for some constant $\epsilon > 0$. Then the optimization time of the 1-ANT on ONEMAX is $2^{\Omega(n^{\epsilon/3})}$ with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$.*

Proof The main idea is to keep track of the 1-potential defined in Sect. 2. Note that the 1-potential multiplied by n equals the expected ONEMAX-value of the next constructed solution x . If the 1-potential is bounded by $1/2 + O(1/\sqrt{n})$, Chernoff bounds yield that the probability of $\text{ONEMAX}(x) \geq n/2 + n^{1/2+\epsilon/3}$ is bounded above by $2^{-\Omega(n^{\epsilon/3})}$. We will show that with overwhelming probability, the 1-potential is bounded as suggested.

Starting with initialization, we consider a phase of length $\lfloor 2^{cn^{\epsilon/3}} \rfloor$ for some constant c to be chosen later and show that the probability of obtaining the optimum in the phase is $2^{-\Omega(n^{\epsilon/3})}$. To this end, we study four properties that the whole phase simultaneously fulfills with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$ each. If any of the properties is not valid, we say that a failure occurs.

1. The best-so-far ONEMAX-value is in the interval $[n/2 - n^{1/2+\epsilon/3}, n/2 + n^{1/2+\epsilon/3}]$.
2. The 1-potential is $1/2 \pm O(1/\sqrt{n})$.
3. All pheromone values are $1/(4n) \pm o(1/n)$.
4. The total number of successful steps (i.e., steps that replace the best-so-far solution and subsequently update pheromone) is bounded by $O(n^{2\epsilon/3})$.

When we fix a property and prove that it holds with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$ at the next time step, we may work under the assumption that all four properties are fulfilled up to the current step. Since we consider only failure probabilities of $2^{-\Omega(n^{\epsilon/3})}$ and at most $2^{cn^{\epsilon/3}}$ time steps, the total failure probability is still bounded by $2^{-\Omega(n^{\epsilon/3})}$ for the whole phase if c is a small enough constant.

For the first property, Chernoff bounds show that the initial solution has a ONEMAX-value of at least $n/2 - n^{1/2+\epsilon/3}$ with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$. Since the value is non-decreasing in the run, this proves the lower bound of the interval. It has already been mentioned above that its upper bound follows from the second property by Chernoff bounds and that the failure probability is $2^{-\Omega(n^{\epsilon/3})}$.

We turn to the second property. In a successful step leading to ONEMAX-value $n/2 + i$ (where i may be negative), it holds that $n + 2i$ pheromone values on 1-edges are increased and $n - 2i$ are decreased. Assuming the third property, we apply Lemma 2. This yields that the 1-potential is changed by at most $4|i|(1 \pm o(1))\rho$ due to the considered successful step. Hence, assuming the first property, we bound $|i|$ by $n^{1/2+\epsilon/3}$, and using the fourth property, the total change of the 1-potential over all $O(n^{2\epsilon/3})$ successful steps is at most

$$O(n^{2\epsilon/3}) \cdot 4n^{1/2+\epsilon/3} \cdot (1 \pm o(1))\rho = O(n^{1/2+\epsilon}) \cdot O(n^{-1-\epsilon}) = O(1/\sqrt{n})$$

by our assumption on ρ . This proves the second property since the initial 1-potential is $1/2$.

The third property is easily proven assuming that the fourth property holds. Using the assumption on ρ , the total change of pheromone on any edge is bounded above by

$$O(n^{2\epsilon/3}) \cdot \rho = O(n^{2\epsilon/3}) \cdot O(n^{-1-\epsilon}) = o(1/n),$$

and the property follows since the initial pheromone values are $1/(4n)$.

We finally prove the fourth property. Let the best-so-far ONEMAX-value be k . We apply Lemma 3 with respect to the value of the next constructed solution; let μ be its expectation. By the second property, $\mu \geq n/2 - O(\sqrt{n})$. Using the first property, we obtain $k - \mu = O(n^{1/2+\epsilon/3})$. Moreover, we translate the pheromone value on the 1-edges of the i -th bit, $1 \leq i \leq n$, into the probability p_i of obtaining a 1 and use the third property. This means that $p_i = 1/2 \pm o(1)$ holds for all bits, hence $\sigma = \sqrt{\sum_{i=1}^n p_i(1-p_i)} = \Theta(n^{1/2})$. Thus, we have $\gamma_k = O(n^{\epsilon/3})$. The lemma now yields that with probability $\Omega(1)$, the next successful step increases k by at least $\Omega(n^{1/2-\epsilon/3})$. We apply Chernoff bounds on the number of successes that lead to such an increase and ignore the other successful steps. Therefore, if we choose an appropriately large constant c' then $c'n^{2\epsilon/3}$ successes with probability $1 - 2^{-\Omega(n^{2\epsilon/3})}$ increase the ONEMAX-value by at least $2n^{1/2+\epsilon/3}$. This would lead to a ONEMAX-value of more than $n/2 + n^{1/2+\epsilon/3}$, which does not happen according to the first property. Hence, with a probability of $1 - 2^{-\Omega(n^{\epsilon/3})}$, the number of successes is $O(n^{2\epsilon/3})$, which proves the fourth property. Since the sum of all failure probabilities is $2^{-\Omega(n^{\epsilon/3})}$, the proof is complete. \square

4.2 Polynomial Upper Bounds

In the following, we consider for which values of ρ the optimization time of the 1-ANT on ONEMAX with high probability is still polynomial. We will show that the function value of the last accepted solution determines the expected value of the next solution almost exactly if $\rho = \Omega(n^{-1+\epsilon})$, $\epsilon > 0$ an arbitrary constant. To determine the expected time to reach an improvement, we give a lower bound on the probability of overshooting the expected value by at least a small amount. Again the asymptotics should be understood with respect to n , which in this case is used to define intervals for success probabilities, and which at the same time bounds the number of variables from above.

Lemma 5 *Let $X_1, \dots, X_t \in \{0, 1\}$, $t \leq n$, be independent Poisson trials with success probabilities $p_i \in [1/n, 1 - 1/n]$, $1 \leq i \leq t$. Let $X := X_1 + \dots + X_t$ and $\mu := E(X) = p_1 + \dots + p_t$. If $\mu \leq t - 1$ then $\text{Prob}(X \geq \mu + 1/2) = \Omega(1/n)$.*

Proof We look for a choice of the p_i that minimizes $\text{Prob}(X \geq \mu + 1/2)$. The work by Gleser [8] establishes Schur-convexity of this probability w.r.t. the p_i , hence a minimum is obtained if as many p_i as possible take their extremal values. It follows that at most one p_i is distinct from $1/n$ and $1 - 1/n$. We obtain three sets of probabilities, up to two of which can be empty: Let t_ℓ be the number of p_i that are $1/n$, t_h be the number of those that are $1 - 1/n$ and $t_a \in \{0, 1\}$ be the number of those that take a third value a , $1/n < a < 1 - 1/n$. Let X_ℓ , X_h and X_a be the sums of the random variables belonging to the corresponding sets, then $X = X_\ell + X_h + X_a$, and let $\mu_\ell = t_\ell/n$, $\mu_h = t_h(1 - 1/n)$ and $\mu_a = t_a a$ be the corresponding expectations.

It always holds that $X_h = t_h \geq \mu_h$ with probability $(1 - 1/n)^{t_h} = \Omega(1)$. We distinguish several cases according to the expectations defined by the other two sets. Note that $\mu_\ell, \mu_a \leq 1$.

Case 1a: $\mu_\ell = 0$ and $\mu_a = 0$. Hence, $t_a = t_\ell = 0$ and $t_h = t$. Therefore, $X_h = t \geq \mu + 1$ holds with probability $(1 - 1/n)^t = \Omega(1)$.

Case 1b: $\mu_\ell = 0$ and $\mu_a > 0$. Hence, $t_\ell = 0$ and $t_a = 1$. Therefore, we have $X_a = 1$ with probability $a = \Omega(1/n)$. Altogether, $X_h + X_a \geq t_h + t_a + t_\ell = t \geq \mu + 1$ holds with probability $\Omega(1/n)$.

Case 2a: $0 < \mu_\ell < 1/4$ and $\mu_a < 1/4$. Then we have $X_\ell \geq 1 \geq \mu_\ell + 3/4$ with probability $\Omega(1/n)$ and altogether, $X_h + X_\ell \geq t_h + \mu_\ell + 3/4 \geq \mu_h + \mu_\ell + 3/4 + (\mu_a - 1/4) = \mu + 1/2$ with probability $\Omega(1/n)$.

Case 2b: $0 < \mu_\ell < 1/4$ and $\mu_a \geq 1/4$. Then $X_\ell \geq 1 \geq \mu_\ell + 3/4$ with probability $\Omega(1/n)$ and $X_a = 1 \geq \mu_a$ with probability $a = \Omega(1)$, altogether $X_h + X_a + X_\ell \geq \mu_h + \mu_a + \mu_\ell + 3/4$ with probability $\Omega(1)$.

Case 3: $\mu_\ell \geq 1/4$. Then $t_\ell \geq n/4$ and $\text{Prob}(X_\ell \geq 3) \geq \binom{n/4}{3}(1/n)^3(1 - 1/n)^{n/4-3} = \Omega(1)$. Altogether, $X_h + X_\ell \geq t_h + 3 \geq \mu_h + \mu_\ell + 2 \geq \mu_h + \mu_a + \mu_\ell + 1 = \mu + 1$ with probability $\Omega(1)$. □

The result of Lemma 5 is asymptotically tight as shows the following instance: let $t = n/2$, $p_1 = \dots = p_{t-1} = 1 - 1/n$ and $p_t = 1/n$. Since $\lceil \mu + 1/2 \rceil = t$, it holds that $\text{Prob}(X \geq \mu + 1/2) \leq 1/n$.

To prove the following theorem, we need another tail inequality on the sum of independent trials. It is in a Chernoff-Hoeffding style, however, requires knowledge of a variance.

Lemma 6 *Let X_1, \dots, X_n be independent random variables such that $X_i - E(X_i) \leq b$ for $1 \leq i \leq n$. Then for any $t > 0$*

$$\text{Prob}(X_1 + \dots + X_n \geq t) \leq e^{-\frac{t^2}{2\text{Var}(X) + 2bt/3}}.$$

See, e.g., [14] for further references.

Theorem 7 *Choosing $\rho = \Omega(n^{-1+\epsilon})$, $\epsilon > 0$ a constant, the optimization time of the 1-ANT on ONEMAX is $O(n^2)$ with probability $1 - 2^{-\Omega(n^{\epsilon/2})}$.*

Proof We assume $\rho \leq 1/2$ since the result follows from Theorem 1 otherwise. In contrast to previous definitions, an edge is called saturated if its pheromone value is $\frac{n-1}{2n^2}$ and called unsaturated otherwise. Let x^* be a newly accepted solution and denote by \mathcal{S} the set of saturated 1-edges and by \mathcal{U} the set of unsaturated 1-edges after the pheromone update. Let $k = \text{ONEMAX}(x^*)$ and decompose k according to $k = k_s + k_u$, where k_s denotes the number of ones in x^* whose corresponding 1-edges belong to \mathcal{S} and k_u to the number of ones in x^* whose 1-edges belong to \mathcal{U} . The probability that the edges of \mathcal{S} contribute at least k_s to the next (not necessarily accepted) solution x is at least $(1 - 1/n)^{k_s} = \Omega(1)$.

Consider the 1-potential (i.e., the sum of pheromone values) P^* of all edges of \mathcal{U} before x^* updates the pheromone values. Let $\mu^* = P^*n$ be the expected ONEMAX-value w.r.t. these edges before the update. Depending on P^* and k_u , we compute

$P(\rho)$, their 1-potential *after* the update:

$$P(\rho) = \frac{(1 - \rho)P^* + 2k_u\rho}{(1 - \rho) + 2n\rho}.$$

We denote by $\mu = P(\rho) \cdot n$ the expected ONEMAX-value w.r.t. the edges of \mathcal{U} *after* the update has occurred. Under certain assumptions, we will prove that with probability $1 - 2^{-\Omega(n^\epsilon)}$, $\mu + 1/2 > k_u$. Since k_u is an integer, Lemma 5 shows that the probability of producing in the next solution x at least $\lceil \mu + 1/2 \rceil \geq k_u + 1$ ones by the \mathcal{U} -edges is at least $\Omega(1/n)$. Consider the difference

$$\mu - k_u \geq \frac{(1 - \rho)P^* + 2k_u\rho}{(1 - \rho) + 2n\rho} \cdot n - k_u = \frac{(\mu^* - k_u)(1 - \rho)}{(1 - \rho) + 2n\rho}.$$

We exploit that $\rho \leq 1/2$, implying $1 - \rho \geq 0$. Hence, if $\mu^* - k_u \geq 0$ then $\mu \geq k_u - k_u - 1/2$ anyway. Assuming $\mu^* - k_u < 0$, we can lower bound the last fraction by $(\mu^* - k_u)/(2n\rho)$. Hence, if we can prove that $k_u - \mu^* < n\rho$, we obtain $\mu > k_u - 1/2$ as desired. We will bound the probability of a large deviation $k_u - \mu^*$ keeping track of the variance of the random number of ones on the \mathcal{U} -edges. Let v^* be the variance *before* the pheromone values have been updated with respect to x^* and denote by v the variance *after* the update. If $v^* \leq (n\rho)^{3/2}$, then Lemma 6 yields

$$\text{Prob}(k_u - \mu^* \geq n\rho) \leq e^{-\frac{(n\rho)^2}{2v^*(1+n\rho/(3v^*))}} = 2^{-\Omega(\sqrt{n\rho})} = 2^{-\Omega(n^{\epsilon/2})}.$$

However, we cannot show that $v^* \leq (n\rho)^{3/2}$ is likely for all points of time. Therefore, we will prove $v \geq v^*/(4n\rho)$ for any time step. This will show that v is large enough to compensate a large $k_u - \mu^*$ in the following step (constructing x).

Suppose $v^* > (n\rho)^{3/2}$. Then $v^* \geq \sqrt{v^*n\rho}$, and by Lemma 6,

$$\text{Prob}(k_u - \mu^* \geq \sqrt{v^*n\rho}) \leq e^{-\frac{(\sqrt{v^*n\rho})^2}{2v^*+2\sqrt{v^*n\rho/3}}} \leq e^{-\frac{v^*n\rho}{2v^*+2v^*/3}} = 2^{-\Omega(n^\epsilon)}.$$

Hence, with probability $1 - 2^{-\Omega(n^\epsilon)}$, $(k_u - \mu^*)/(2n\rho) \leq \sqrt{v^*/(2n\rho)}$, implying $\mu \geq k_u - \sqrt{v^*/(2n\rho)}$. Due to the assumptions $v^* > (n\rho)^{3/2}$, $v \geq v^*/(4n\rho)$ and $n\rho = \Omega(n^\epsilon)$, it follows that $v \rightarrow \infty$. Hence, we can apply Lindeberg’s generalization of the Central Limit Theorem for the number of ones on \mathcal{U} . The probability of producing at least $k_u + 1$ ones on these edges is bounded below by the probability of producing at least $1 + \mu + \sqrt{v^*/(2n\rho)}$ ones on these edges. By the Central Limit Theorem, this has probability $\Omega(1)$ since $\sqrt{v} \geq \sqrt{v^*/(2n\rho)}$.

We still have to show that $v \geq v^*/(4n\rho)$. It is sufficient to show a statement on the success probability for each edge (u, v) of the construction graph. Consider the expression $\tau'_{(u,v)} \geq \frac{(1-\rho)\tau_{(u,v)}}{1-\rho+2n\rho}$. The last fraction is at least $\frac{\tau_{(u,v)}}{4n\rho}$ since $\rho \leq 1/2$.

The \mathcal{S} -edges contribute with probability $\Omega(1)$ at least k_s to the next solution, and (if no failure of probability $2^{-\Omega(n^{\epsilon/2})}$ occurs) with probability $\Omega(1/n)$, the \mathcal{U} -edges contribute at least $k_u + 1$. At most $n - 1$ improvements suffice, and, by Chernoff bounds, cn^2 steps contain at least $n - 1$ improvements with probability $1 - 2^{-\Omega(n)}$ for an appropriate constant c . Since $\rho \leq 1/2$, $\epsilon \leq 1$ must hold. Hence, the sum of all failure probabilities in $O(n^2)$ steps is $2^{-\Omega(n^{\epsilon/2})}$. □

5 Conclusions

For the first time, bounds on the runtime of a simple ACO algorithm have been obtained. Choosing a large evaporation factor, it behaves like the $(1 + 1)$ EA and all results on this algorithm transfer directly to our ACO algorithm. In addition, we have inspected the effect of the evaporation factor in detail for the function ONEMAX and figured out the border between a polynomial and an exponential optimization time almost completely. Thereby, we have developed new techniques for the analysis of randomized search heuristics.

References

1. Dorigo, M., Blum, C.: Ant colony optimization theory: A survey. *Theor. Comput. Sci.* **344**, 243–278 (2005)
2. Dorigo, M., Stützle, T.: *Ant Colony Optimization*. MIT, Cambridge (2004)
3. Dorigo, M., Maniezzo, V., Colormi, A.: The ant system: An autocatalytic optimizing process. *Tech. Rep. 91-016 Revised*, Politecnico di Milano, Italy (1991)
4. Droste, S., Jansen, T., Wegener, I.: On the analysis of the $(1 + 1)$ evolutionary algorithm. *Theor. Comput. Sci.* **276**, 51–81 (2002)
5. Feller, W.: *An Introduction to Probability Theory and Its Applications*, 3rd edn., vol. 1. Wiley, New York (1968)
6. Feller, W.: *An Introduction to Probability Theory and Its Applications*, 2nd edn., vol. 2. Wiley, New York (1971)
7. Giel, O., Wegener, I.: Evolutionary algorithms and the maximum matching problem. In: *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science. Lecture Notes in Computer Science*, vol. 2607, pp. 415–426. Springer, Berlin (2003)
8. Gleser, L.J.: On the distribution of the number of successes in independent trials. *Ann. Probab.* **3**(1), 182–188 (1975)
9. Gutjahr, W.J.: A generalized convergence result for the graph-based ant system metaheuristic. *Probab. Eng. Inf. Sci.* **17**, 545–569 (2003)
10. Gutjahr, W.J.: On the finite-time dynamics of ant colony optimization. *Methodol. Comput. Appl. Probab.* **8**, 105–133 (2006)
11. Jerrum, M., Sorkin, G.B.: The Metropolis algorithm for graph bisection. *Discrete Appl. Math.* **82**(1–3), 155–175 (1998)
12. Neumann, F., Wegener, I.: Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '04). Lecture Notes in Computer Science*, vol. 3102, pp. 713–724. Springer, Berlin (2004)
13. Papadimitriou, C.H., Schäffer, A.A., Yannakakis, M.: On the complexity of local search. In: *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC '90)*, pp. 438–445. ACM Press, Cambridge (1990)
14. Scheideler, C.: *Probabilistic methods for coordination problems*. HNI-Verlagsschriftenreihe 78. Habilitation Thesis, University of Paderborn. Available at <http://www14.in.tum.de/personen/scheideler/index.html.en> (2000)
15. Wegener, I.: Simulated annealing beats Metropolis in combinatorial optimization. In: *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP '05). Lecture Notes in Computer Science*, vol. 3580, pp. 589–601. Springer, Berlin (2005)
16. Witt, C.: Worst-case and average-case approximations by simple randomized search heuristics. In: *Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science (STACS '05). Lecture Notes in Computer Science*, vol. 3404, pp. 44–56. Springer, Berlin (2005)