Runtime Analysis of a Simple Ant Colony Optimization Algorithm

Frank Neumann
Inst. für Informatik und Prakt. Mathematik,
Christian-Albrechts-Univ. zu Kiel, 24098 Kiel, Germany
fne@informatik.uni-kiel.de

Carsten Witt*
FB Informatik, LS 2,
Univ. Dortmund, 44221 Dortmund, Germany
carsten.witt@cs.uni-dortmund.de

February 14, 2006

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 a finite amount of time. We present the first runtime analysis of a simple ACO algorithm that transfers many rigorous results with respect to the runtime of a simple evolutionary algorithm to our algorithm. In addition, we examine the choice of the evaporation factor, which is a crucial parameter in such an algorithm, in greater detail and analyze its effect with respect to the runtime.

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, Schäffer and Yannakakis (1990). The graph bisection problem has been subject to analysis of MA (Jerrum and Sorkin, 1998), 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 by Wegener (2005) 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 Colorni (1991), they have shown to be a powerful heuristic approach to solve combinatorial optimization problems (see Dorigo and Stützle, 2004, 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 (Dorigo and Blum, 2005), convergence results are so far the only results related to their runtimes. Dorigo and Blum (2005) 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 large values of ρ as the simplest evolutionary algorithm called (1+1) EA. This algorithm has been studied extensively with respect to its runtime on classes of pseudo-boolean functions (see, e. g. Droste, Jansen and Wegener, 2002) 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 (Giel and Wegener, 2003) and minimum spanning trees (Neumann and Wegener, 2004). It should be clear

2

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 (2005) 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 large ρ.

After having obtained these general results, we consider the effect of the evaporation factor ρ on the runtime of our ACO algorithm in detail. This analysis requires new techniques since it is the first one of its kind. We examine the simplest non-trivial pseudo-boolean function called ONEMAX and show that small values of ρ with high probability lead to an exponential optimization time even for this simple function. In addition, we examine for which choices of ρ the optimization time with high probability is still upper bounded by a small polynomial. To achieve these bounds, we consider the expected function value for the algorithm in the next step. It turns out that larger values of ρ change the pheromone values on the edges such that the expected value in the next step is determined by the function value of the best seen solution. Using results obtained by Hoeffding (1956), we show that an improvement will be achieved after an expected polynomial number of steps. In the case of small ρ, achieving an improvement does not increase the expected value in the next step that much. Here exponential lower bounds are obtained by showing that there is a large gap between the expected value and the best function value seen so far.

In Section 2, we introduce the simple ACO algorithm which we will consider. We investigate its relation to the (1+1) EA in Section 3 and transfer the results on this EA to our algorithm. In Section 4, we investigate the choice of the evaporation factor ρ for the function ONEMAX in greater detail and finish with some conclusions. In an appendix, it is shown how results by Hoeffding (1956) can be adapted to suit our needs.

2   The Algorithm

Gutjahr (2003) 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. 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 proportional 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 let
   $T := \sum_{(v,w)|w \in N_v} \tau_{v,w}$.

   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 respected 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 $f : \{0,1\}^n \to \mathbb{R}$ for $n \geq 3$. We investigate the construction graph $C_{\text{bool}} = (V, E)$ (see Figure 1) with $s = v_0$, which seems to be the most natural one in our setting. 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 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
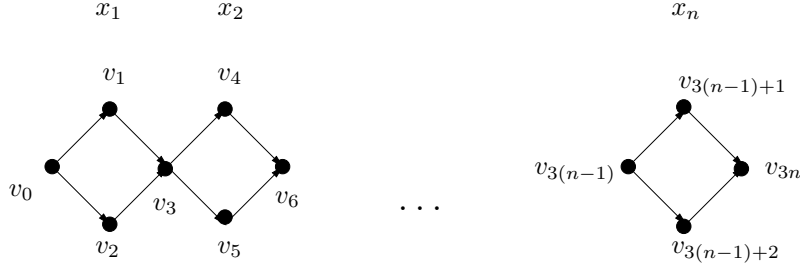
4

Figure 1: Construction graph for pseudo-boolean optimization

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*. 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 $(v_{3(i-1)+2}, v_{3i})$ are called *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., Dorigo and Blum, 2005) 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 $\left[\frac{1}{2n^2}, \frac{n-1}{2n^2}\right]$ 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 $v = v_{3i+1}$, $1 \leq i \leq n$. 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. 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 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 $\mathrm{Prob}(x_i = 1) = \mathrm{Prob}(x_i = 0) = 1/2$ 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 up to now best solution 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 $n \geq 3$ and $\rho \geq (n-2)/(3n-2)$, 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

6

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. $\square$

4 1-ANT on OneMax

In the following, we inspect the choice of ρ in greater 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 explicity demanded by Dorigo and Blum (2005). Note that due to results on the (1+1) EA by Droste, Jansen and Wegener (2002), 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 resp. 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* iff 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 resp. τ_2 be their current pheromone values in the 1-ANT. Let τ_1' resp. 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

$$\tau_1' - \tau_2' \leq \rho(\tau_2 - \tau_1) + (\tau_1 - \tau_2) \quad \Rightarrow \quad (\tau_1' - \tau_1) - (\tau_2' - \tau_2) \leq \rho|\tau_1 - \tau_2|.$$

Taking the absolute value of $(\tau_1' - \tau_1) - (\tau_2' - \tau_2)$, the claim follows. $\square$

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 of the normalization factor $1 - \rho + 2n\rho$ diverges for $\rho = \omega(1/n)$ and is $1 - \rho - o(1)$ for $\rho = o(1/n)$. We will make precise that the behavior of the 1-ANT on OneMax changes drastically when ρ is asymptotically smaller resp. 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 proability 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.

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

Proof: Since the X_i are bounded and σ diverges, Lindeberg's generalization of the Central Limit Theorem (Feller, 1971, Chapter 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 $\tilde{p}_k := \mathrm{Prob}(X \geq k + \sigma/\gamma_k) = \Omega(1)$. 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 $p_k := \mathrm{Prob}(X \geq k)$, $d_k := (k - \mu)/\sigma$, 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

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

(see Feller, 1968, Chapter 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 are $\Omega(1)$. Finally, the e-term is $\Omega(1)$ since $(\tilde{d}_k)^2 - (d_k)^2 \leq (d_k + 1/d_k)^2 - (d_k)^2 \leq 2 + 1/(d_k)^2 \leq 3$. $\square$

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 so-called 1-*potential*, defined as the sum of pheromone values on 1-edges. 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 above 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 as long as the ONEMAX-value of the so far best solution is bounded above by $n/2 + n^{1/2+\epsilon/3}$.

Starting with initialization, we consider a phase of length $s := \lfloor 2^{cn^{\epsilon/3}} \rfloor$ for some constant c to be chosen later and show that the success probability in the phase is $2^{-\Omega(n^{\epsilon/3})}$. A main task is to bound the number of successful steps of the phase, i.e., of steps where the new solution is accepted and a pheromone update occurs. In a success with ONEMAX-value $n/2 + i$, $n + 2i$ pheromone values on 1-edges are increased and $n - 2i$ are decreased. Suppose all pheromone values are $1/(4n) \pm o(1/n)$ in the phase. Then Lemma 2 yields that the 1-potential is changed by at most $4i(1 \pm o(1))\rho$ due to the considered success. Hence, if the best solution always had ONEMAX-value at most $n/2 + n^{1/2+\epsilon/3}$, the total change of the 1-potential due to at most $O(n^{2\epsilon/3})$ successes would be at most

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

by our assumption on ρ. This would prove the theorem since initially, the 1-potential is $1/2$.

Under the assumption on the pheromone values, we want to show that with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$, at most $c'n^{2\epsilon/3}$ successes occur in the phase, where c' is an appropriate constant. We already know that then the probability of a success with value at least $n/2 + n^{1/2+\epsilon/3}$ is $2^{-\Omega(n^{\epsilon/3})}$ in each step of the phase. If c is chosen small enough, this probability is $2^{-\Omega(n^{\epsilon/3})}$ for the whole phase. Moreover, the initial value is at least $n/2 - n^{1/2+\epsilon/3}$ with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$.

Let the so far best value be k. We apply Lemma 3 with respect to the expected ONEMAX-value μ of the next constructed solution. Note that $k - \mu = O(n^{1/2+\epsilon/3})$ holds at each time step we consider. Moreover, $p_i = 1/2 \pm o(1)$ is assumed to hold for all bits, implying $\sigma = \Theta(n^{1/2})$. Hence, with probability $\Omega(1)$ the next success leads to a value at least $k + \Omega(n^{1/2-\epsilon/3})$. Using Chernoff bounds, with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$, $c'n^{2\epsilon/3}$ successes increase the ONEMAX-value by at least $c''n^{1/2+\epsilon/3}$, where c'' is an appropriate constant.

We still have to show the statement on the pheromone values. This is not too difficult for our choice of ρ if the number of successes is bounded by $O(n^{2\epsilon/3})$. Then the total change of pheromone on any fixed edge is bounded above by

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

with probability $1 - 2^{-\Omega(n^{\epsilon/3})}$. Since the number of edges is bounded by $4n$, this holds also for all edges together. Since the sum of all failure probabilities is $2^{-\Omega(n^{\epsilon/3})}$, this completes the proof. $\qquad \square$

9

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.

Lemma 5 *Let $X_1, \ldots, X_n \in \{0, 1\}$ be independent Poisson trials with success probabilities $p_i \in [1/n, 1 - 1/n]$, $1 \le i \le n$. Let $X := X_1 + \cdots + X_n$ and $\mu := E(X) = p_1 + \cdots + p_n$. Then $\mathrm{Prob}(X \ge \mu + 1/2) = \Omega(1/n)$.*

Proof: It follows from the work by Hoeffding (1956) that $\mathrm{Prob}(X \ge \mu + 1/2)$ is minimized if the p_i take on at most 3 different values, only one of which is distinct from $1/n$ and $1 - 1/n$. (See Lemma 9 in Appendix A.)

Let n_ℓ be the number of p_i that are $1/n$, n_h be the number that are $1 - 1/n$ and n_a be the number that take a different value a, $1/n < a < 1 - 1/n$. Let the random variables belonging to each of the three sets be called ℓ-variables, h-variables and a-variables, respectively. Let X_ℓ, X_h and X_a be the sums of the variables from these sets, i.e., $X = X_\ell + X_h + X_a$, and let $\mu_\ell = n_\ell/n$, $\mu_h = n_h(1 - 1/n)$ and $\mu_a = n_a a$ be the corresponding expectations. In the following arguments we also cover the case that up to two sets are empty.

It always holds that $X_h = n_h \ge \mu_h$ with probability $(1 - 1/n)^{n_h} = \Omega(1)$. We distinguish several cases according to the variables from the other two sets. If $n_\ell = n_a = 0$ then $n_h = n$ and $X = n = \mu + 1$ with probability $\Omega(1)$. Now we concentrate on the most complicated case that $n_\ell \ne 0 \ne n_a$, implying $\mu_\ell \ne 0 \ne \mu_a$. If $0 < \mu_\ell \le 1/4$ and $0 < \mu_a \le 1/4$, we exploit that $X_\ell \ge 1$ with probability at least $1/n$. Hence $X_h + X_\ell \ge n_h + 1 \ge \mu_h + (\mu_\ell + 3/4) \ge \mu + 1/2$ with probability $\Omega(1/n)$.

Now let $\mu_a > 1/4$ and $0 < \mu_\ell \le 1/4$. We distinguish four cases depending on n_a, μ_a and $\sigma_a = \sqrt{n_a a(1 - a)}$. In all cases, we exploit that $X_\ell \ge 1 \ge \mu_\ell + 3/4$ with probability $\Omega(1/n)$.

Case 1: $n_a = O(1)$. Since $\mu_a \ge 1/4$, $a = \Omega(1)$. Hence, we have $X_a = n_a \ge \mu_a$ with probability $\Omega(1)$, implying $X = X_h + X_a + X_\ell \ge \mu_h + \mu_a + \mu_\ell + 3/4 = \mu + 3/4$ with probability $\Omega(1/n)$.

Case 2: $n_a = \omega(1)$ and $\mu_a = O(1)$. Hence, X_a can be approximated by means of the Poisson distribution with parameter μ_a, implying $X_a \ge \mu_a$ with probability at least $(1 - o(1)) \cdot e^{-\mu_a} (\mu_a)^{\lceil \mu_a \rceil} / (\lceil \mu_a \rceil)! = \Omega(1)$. Hence, we conclude as in Case 1 that $X \ge \mu + 3/4$ with probability $\Omega(1/n)$.

Case 3: $\mu_a = \omega(1)$ (implying $n_a = \omega(1)$) and $\sigma_a = \omega(1)$. Using the Central Limit Theorem (Feller, 1968), we approximate X_a by a Normal distribution, implying $X_a \ge \mu_a$ with probability $\Omega(1)$. We go on as in Case 1.

Case 4: $\mu_a = \omega(1)$ and $\sigma_a = O(1)$. Since $\sigma_a^2 = \mu_a(1-a) = O(1)$ implies $a \geq 1/2 - o(1)$, we obtain $1 - a = O(1/\mu_a) = O(1/n_a)$. Hence, $X_a = n_a \geq \mu_a$ with probability at least $(1 - O(1/n_a))^{n_a} = \Omega(1)$. We go on as in Case 1.

The case that $0 < \mu_a \leq 1/4$ and $\mu_\ell > 1/4$ can be handled by an analogous case distinction according to n_ℓ and μ_ℓ. Here some cases are even impossible.

We still have to study the situation that $\mu_a > 1/4$ and $\mu_\ell > 1/4$. Then we study to which of the four cases the a-variables and ℓ-variables belong. If both lead to one of the Cases 1 or 4, we even obtain $X = n \geq \mu + 1$ (since $p_i \leq 1 - 1/n$) with probability $\Omega(1)$. Otherwise, at least one of the two sets of variables lead to Case 2 or 3. W.l.o.g., let this be the a-variables. However, both the approximation by the Poisson and the Normal distribution show that even $X_a \geq \mu_a + 1/2$ still holds with probability $\Omega(1)$ then.

Finally, we have to consider the case that $n_\ell = 0 \neq n_a$ or $n_a = 0 \neq n_\ell$. It suffices to study the case $n_a \neq 0$. Considering the above four cases and the extra case $\mu_a \leq 1/4$, the lemma follows by the same arguments as above. $\square$

Theorem 6 *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 potential P of all edges of $\mathcal{U}$ before x^* updates the pheromone values. Let $\mu = Pn$ be the expected OneMax-value w.r.t. these edges before the update. Depending on P and k_u, we compute $P^*(\rho)$, the new 1-potential on these edges:

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

We denote by $\mu^* = P^*(\rho) \cdot n$ the expected OneMax-value w.r.t. to edges of $\mathcal{U}$ after the update has occured. 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 mentioned 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

11

the (negative) last fraction by $(\mu - k_{\mathrm{u}})/(2n\rho)$. Hence, if we can prove that $k_{\mathrm{u}} - \mu < n\rho$, we obtain $\mu^* > k_{\mathrm{u}} - 1/2$ as desired. We will bound the probability of a large deviation $k_{\mathrm{u}} - \mu$ keeping track of the variance of the random ONEMAX-value of x^*. 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 $v \le (n\rho)^{3/2}$, then a Chernoff-Hoeffding-type bound

(Theorem 3.44 in Scheideler, 2000) yields

$$\mathrm{Prob}(k_{\mathrm{u}} - \mu \ge n\rho) \; \le \; e^{-\frac{(n\rho)^2}{2v(1+n\rho/(3v))}} \; = \; 2^{-\Omega(\sqrt{n\rho})} \; = \; 2^{-\Omega(n^{\epsilon/2})}.$$

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

Suppose $v > (n\rho)^{3/2}$. Then $v \ge \sqrt{vn\rho}$, and the above bound yields

$$\mathrm{Prob}(k_{\mathrm{u}} - \mu \ge \sqrt{vn\rho}) \; \le \; e^{-\frac{(\sqrt{vn\rho})^2}{2v+2\sqrt{vn\rho}/3}} \; \le \; e^{-\frac{vn\rho}{2v+2v/3}} \; = \; 2^{-\Omega(n^\epsilon)}.$$

Hence, with probability $1 - 2^{-\Omega(n^\epsilon)}$, $(k_{\mathrm{u}} - \mu)/(2n\rho) \le \sqrt{v/(2n\rho)}$, implying $\mu^* \ge k_{\mathrm{u}} - \sqrt{v/(2n\rho)}$. Due to the assumptions $v > (n\rho)^{3/2}$, $v^* \ge v/(4n\rho)$ and $n\rho = \Omega(n^\epsilon)$, it follows that $v^* = \omega(1)$. Hence, we can apply Lindeberg's generalization of the Central Limit Theorem for the value of x. The probability of producing at least $k_{\mathrm{u}} + 1$ ones on the edges of $\mathcal{U}$ 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 $\sqrt{v^*} \ge \sqrt{v/(2n\rho)}$.

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

The edges of $\mathcal{S}$ contribute with probability $\Omega(1)$ at least k_{s}, and (if no failure of probability $2^{-\Omega(n^{\epsilon/2})}$ occurs) with probability $\Omega(1/n)$, the value of the bits corresponding to edges of $\mathcal{U}$ is at least $k_{\mathrm{u}} + 1$. At most $n - 1$ improvements are needed, 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 $\rho \le 1/2$, $\epsilon \le 1$ must hold. Hence, the sum of all failure probabilities for $O(n^2)$ steps is $2^{-\Omega(n^{\epsilon/2})}$. $\qquad\Box$

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 greater detail for the function ONEMAX and figured out the border between a polynomial and an exponential optimization time. Thereby, we have developed new techniques for the analysis of randomized search heuristics.

References

Dorigo, M. and Blum, C. (2005). Ant colony optimization theory: A survey. *Theoretical Computer Science*, **344**, 243–278.

Dorigo, M., Maniezzo, V., and Colorni, A. (1991). The ant system: An autocatalytic optimizing process. Tech. Rep. 91-016 Revised, Politecnico di Milano, Italy.

Dorigo, M. and Stützle, T. (2004). *Ant Colony Optimization*. MIT Press.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, vol. 1. Wiley, 3rd ed.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, 2nd ed.

Giel, O. and Wegener, I. (2003). Evolutionary algorithms and the maximum matching problem. In *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science*, vol. 2607 of *Lecture Notes in Computer Science*, 415–426. Springer.

Gutjahr, W. J. (2003). A generalized convergence result for the graph-based ant system metaheuristic. *Probability in the Engineering and Informational Sciences*, **17**, 545–569.

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Annals of Mathematical Statistics*, **27**, 713–721.

Jerrum, M. and Sorkin, G. B. (1998). The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, **82**(1–3), 155–175.

Neumann, F. and Wegener, I. (2004). Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 04)*, vol. 3102 of *Lecture Notes in Computer Science*, 713–724. Springer.

Papadimitriou, C. H., Schäffer, A. A., and Yannakakis, M. (1990). On the complexity of local search. In *Proceedings of the 22nd annual ACM symposium on Theory of Computing (STOC '90)*, 438–445. ACM Press.

Scheideler, C. (2000). *Probabilistic Methods for Coordination Problems*. HNI-Verlagsschriftenreihe 78, University of Paderborn. Habilitation Thesis, available at `http://www14.in.tum.de/personen/scheideler/index.html.en`.

Wegener, I. (2005). Simulated annealing beats metropolis in combinatorial optimization. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP '05)*, vol. 3580 of *Lecture Notes in Computer Science*, 589–601.

Witt, C. (2005). 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)*, vol. 3404 of *Lecture Notes in Computer Science*, 44–56. Springer.

A   Appendix

Hoeffding Lemma

In this section, we repeat Hoeffding's technique, leading to Lemma 9. Note that the following statements constitute only minor modifications of the first pages in Hoeffding (1956).

The expected value of a function $g(S)$ is

$$f(p) = E(g(S)) = \sum_{k=0}^{n} g(k) A_{nk}(p), \tag{1}$$

where $p = (p_1, \ldots, p_n)$ and the probability A_{nk} of $S = k$ is given by

$$A_{nk}(p) = \sum_{\substack{(i_1,\ldots,i_n) \in \{0,1\}^n, \\ i_1 + \cdots + i_n = k}} \prod_{j=1}^{n} p_j^{i_j} (1 - p_j)^{1-i_j}, \quad k = 0, 1 \ldots, n.$$

The function $f(p)$ is symmetric in the components of p and linear in each component. Any function with these two properties can be represented in form (1). We consider the problem of finding the maximum and the minimum of $f(p)$ in the section D of the hyperplane

$$p_1 + p_2 + \cdots + p_n = np \quad (1/n < p < 1 - 1/n).$$

We denote by $p^{i_1, i_2, \ldots, i_m}$ the point in the $(n - m)$-dimensional space, which is obtained from p by omiting the coordinates $p_{i_1}, p_{i_2}, \ldots, p_{i_m}$.

Since $f(p)$ is symmetric, and linear in each component, we can write

$$f(p) = f_{n-1,0}(p^j) + p_j f_{n-1,1}(p^j), \quad j = 1, 2, \ldots, n, \tag{2}$$

where the functions $f_{n-1,0}$ and $f_{n-1,1}$ are independent of the index j and symmetric and linear in the components of p^j.

We define the functions $f_{n-k,i}$ by $f_{n,0}(p) = f(p)$ and

$$f_{n-k,i}(p^{1,2,\ldots,k}) = f_{n-k-1,i}(p^{1,2,\ldots,k+1}) + p_{k+1} f_{n-k-1,i+1}(p^{1,2,\ldots,k}),$$
$$i = 0, 1, \ldots, k, \quad k = 0, 1, \ldots n - 1. \tag{3}$$

We obtain

$$f(p) = \sum_{i=1}^{m} C_{mi}(p_1, p_2, \ldots, p_m) f_{n-m,i}(p^{1,\ldots,m}), \quad m = 1, 2, \ldots, n, \tag{4}$$

where $C_{m0}, C_{m1}, \ldots, C_{m,m}$ are the symmetric sums

$$C_{m0}(p_1, p_2, \ldots, p_m) = 1 \tag{5}$$

and

$$C_{mi}(p_1, p_2, \ldots, p_m)$$
$$= (p_1 p_2 \cdots p_i) + (p_1 p_2 \cdots p_{i-1} p_{i+1}) + \cdots + (p_{m-i+1} p_{m-i+2} \cdots p_m)$$

for $i > 0$.

Theorem 7 *Let $a = (a_1, a_2, \ldots, a_n)$ be a point in D at which $f(p)$ attains its maximum. Then for every two distinct indices i, j, we have*

$$f_{n-2,2}(a^{ij}) \leq 0 \quad \text{if } a_i \neq a_j, \tag{6}$$

$$f_{n-2,2}(a^{ij}) = 0 \quad \text{if } a_i \neq a_j, 1/n < a_i, a_j < 1 - 1/n, \tag{7}$$

$$f_{n-2,2}(a^{ij}) \geq 0 \quad \text{if } 1/n < a_i = a_j < 1 - 1/n. \tag{8}$$

Proof: Let a' denote the point which is obtained from a if a_i and a_j are replaced by $a_i + x$ and $a_j - x$. The point a' is in D for all x in the interval I defined by $1/n \leq a_i + x \leq 1 - 1/n$, $1/n \leq a_j - x \leq 1 - 1/n$. We have

$$f(a') = f_{n-2,0}(a^{ij}) + (a_i + a_j)f_{n-2,1}(a^{ij}) + (a_i + x)(a_j - x)f_{n-2,2}(a^{ij}).$$

Hence,

$$f(a') - f(a) = x(a_j - a_i - x)f_{n-2,2}(a^{ij}). \tag{9}$$

Since $f(a)$ is a maximum, the right side must be negative or zero for all x in I. We may assume $a_i \leq a_j$. If $a_i \neq a_j$, we can choose x positive and sufficiently small such that x is in I and (6) holds. If $1/n < a_i < 1 - 1/n$ and $1/n < a_j < 1 - 1/n$ then the point $x = -a_i + 1/n$ is in the interior of I and (8) must hold. Moreover, if $a_i \neq a_j$, together with (6), we obtain (7). If the maximum is not attained at a' when x is in I and is different and sufficiently close to zero, the inequalities (6) and (8) must be strict. $\qquad\square$

In general, the maximum or minimum of $f(p)$ can be attained at more than one point in D. The following theorem gives some information about the set of points at which an extremum is attained.

Theorem 8 *Let a be a point in D at which $f(p)$ attains its maximum or its minimum. Suppose that a has at least two unequal coordinates which are distinct from $1/n$ and $1 - 1/n$. Then $f(p)$ attains its maximum (or minimum) at any point in D which has the same number of $1/n$ coordinates and the same number of $1 - 1/n$ coordinates as a has.*

Proof: Let $m = n - r - s$ be the number of coordinates of $a = (a_1, \ldots, a_n)$ which are distinct from $1/n$ and $1 - 1/n$. We may take $a_1, \ldots, a_m$ to be these coordinates and assume $a_1 \neq a_2$. We first show

$$f_{n-k,i}(a_{k+1}, \ldots, a_n) = 0, \quad i = 2, \ldots, k. \tag{10}$$

We prove this equation by induction on k. Due to Theorem 7 this holds for $k = 2$. Let

$$b_k = (b_1, \ldots, b_k, a_{k+1}, \ldots, a_n), \tag{11}$$

where

$$b_1 + \cdots + b_k = a_1 + \cdots + a_k, \quad 1/n \leq b_i \leq 1 - 1/n, \quad i = 1, \ldots, k. \tag{12}$$

16

The point b_k is in D. By (4) and the induction hypothesis,

$$f(b_k) = f_{n-k,0}(a_{k+1}, \ldots, a_n) + (a_1 + \cdots + a_k) f_{n-k,1}(a_{k+1}, \ldots, a_n) = f(a). \quad (13)$$

Thus, the maximum is attained at every point b_k which satisfies (11) and (12). In particular (12) can be satisfied with $b_1 \neq b_2$, $b_1 \neq a_{k+1}$, $b_2 \neq a_{k+1}$, $1/n < b_i < 1 - 1/n$, $i = 1, \ldots, k$ (since $1/n < a_j < 1 - 1/n$ for $j = 1, \ldots, m$). Under these assumptions, we can apply the induction hypothesis (10) with a replaced by the point b_k, whose first $k+1$ coordinates can by suitably rearranged. Hence,

$$f_{n-k,i}(b_1, a_{k+2}, \ldots, a_n) = 0, \quad f_{n-k,i}(b_2, a_{k+2}, \ldots, a_n) = 0, \quad i = 2, \ldots k.$$

Applying (3) to the left sides of these equations, we obtain

$$f_{n-k-1,i}(a_{k+2}, \ldots, a_n) + b_h f_{n-k-1,i+1}(a_{k+2}, \ldots, a_n) = 0,$$
$$i = 2, \ldots, k, \quad h = 1, 2. \quad (14)$$

Since $b_1 \neq b_2$, we find that (10) is satisfied with k replaced by $k+1$. Thus (10) holds for $k = 2, \ldots, m$. Equation (13) holds for every b_m that satisfies (11) and (12) and f is symmetric which completes the proof. $\square$

Corollary 1 *The maximum and minimum of $f(p)$ in D are attained at points whose coordinates take on at most three different values, only one of which is distinct from $1/n$ and $1 - 1/n$.*

Lemma 9 *Let $X_1, \ldots, X_n \in \{0, 1\}^n$ be independent Poisson trails with success probabilities $p_i \in [1/n, 1 - 1/n]$, $1 \leq i \leq n$. Let $X = X_1 + \cdots + X_n$ and $\mu = p_1 + \cdots + p_n$. Then $Prob(X \geq \mu + 1/2)$ is minimized if the p_i take on at most three different values, only one which is distinct from $1/n$ and $1 - 1/n$.*

Proof: Consider (1) and set $g(k) = 1$ if $k \geq \mu + 1/2$ and $g(k) = 0$ otherwise. Hence $f(p)$ computes in this case the probability of obtaining a value at least $\mu + 1/2$ and we can apply Corollary 1. $\square$