

A Simple Ant Colony Optimizer for Stochastic Shortest Path Problems

Dirk Sudholt · Christian Thyssen

Received: date / Accepted: date

Abstract Ant Colony Optimization (ACO) is a popular optimization paradigm inspired by the capabilities of natural ant colonies of finding shortest paths between their nest and a food source. This has led to many successful applications for various combinatorial problems. The reason for the success of ACO, however, is not well understood and there is a need for a rigorous theoretical foundation.

We analyze the running time of a simple ant colony optimizer for stochastic shortest path problems where edge weights are subject to noise that reflects delays and uncertainty. In particular, we consider various noise models, ranging from general, arbitrary noise with possible dependencies to more specific models such as independent gamma-distributed noise. The question is whether the ants can find or approximate shortest paths in the presence of noise. We characterize instances where the ants can discover the real shortest paths efficiently. For general instances we prove upper bounds for the time until the algorithm finds reasonable approximations. Contrariwise, for independent gamma-distributed noise we present a graph where the ant system needs exponential time to find a good approximation. The behavior of the ant system changes dramatically when the noise is perfectly correlated as then the ants find shortest paths efficiently. Our results shed light on trade-offs between the noise strength, approximation guarantees, and expected running times.

Keywords Ant colony optimization · combinatorial optimization · running time analysis · shortest path problems · stochastic optimization

A preliminary version of this work appeared in Horoba and Sudholt (2010).

Dirk Sudholt

CERCIA, School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK

E-mail: d.sudholt@cs.bham.ac.uk

Christian Thyssen, né Horoba

Fakultät für Informatik, LS 2, Technische Universität Dortmund, 44221 Dortmund, Germany

E-mail: christian.thyssen@ls2.cs.tu-dortmund.de

1 Introduction

Ant Colony Optimization (ACO) is a modern optimization paradigm inspired by the foraging behavior of ant colonies. When an ant searches the environment for food, it deposits pheromones on the ground while moving around. Ants can smell pheromones and they tend to be attracted by pheromone trails. If an ant finds a food source, it tends to walk back on its own trail, investing it with more pheromones. Other ants are then attracted towards this trail. If the path is short, the pheromone trail is reinforced quickly. This way, if several ants find different paths from the nest to a food source, it is likely that the whole colony converges to the shortest path. This remarkable behavior is an example of *swarm intelligence* where simple agents are capable of solving complex problems without centralized control.

The collective intelligence of ant colonies has inspired the ACO paradigm (Dorigo, Maniezzo, and Coloni, 1991). The basic idea is that artificial ants search for good candidate solutions, guided by artificial pheromones. Artificial ants thereby perform a random walk through a graph and the next edge to be chosen is selected according to pheromones that represent the attractiveness of an edge. This leads to a construction procedure for problems that can be represented as finding good paths in graphs; examples are shortest paths or the TSP (Dorigo and Gambardella, 1997). Furthermore, the path formation of ants can be used to construct solutions for various combinatorial problems via so-called construction graphs. In such a graph the choice which edges are taken by an artificial ant are mapped to decisions about components of a candidate solution. This makes ACO a general paradigm for the design of metaheuristics. These algorithms have been applied to many problems from various domains, such as the Quadratic Assignment Problem, network routing problems, and scheduling problems. For details and further applications see the book by Dorigo and Stützle (2004).

Metaheuristics such as ACO algorithms are popular for practitioners as they are easy to implement and they usually produce good solutions in short time. They often produce better solutions than problem-specific approximation algorithms with proven performance guarantees and they are applicable in settings where there is not enough knowledge, time, or expertise to design a problem-specific algorithm. For some problems like the sequential ordering problem or open shop scheduling ACO algorithms are regarded as state-of-the-art (Dorigo and Stützle, 2004).

Despite many successful applications and empirical studies on benchmarks, the success behind ACO is not well understood. There is little insight into how, when, and why the collective intelligence of an artificial ant colony efficiently finds good solutions for a particular problem. In particular, studies making rigorous formal statements about ACO are very rare. Many design questions for ACO systems remain unanswered and finding appropriate parameter settings is often done by trial and error. Leading researchers have therefore called out to theoretical computer science to provide analyses that give insight into the dynamic behavior and the performance of ACO (Dorigo and Blum, 2005).

A number of researchers have followed this call and established a rapidly growing theory of ACO algorithms. The motivation is to assess the performance of ACO on interesting and well-understood problems, using techniques from the analysis of randomized algorithms. The goal is to shed light on the working principles of ACO, identify situations where ACO algorithms perform well and where they do not, and to give hints on how to design better ACO algorithms.

The theory of ACO started with investigations of ACO on simple example functions (Doerr, Neumann, Sudholt, and Witt, 2011; Gutjahr and Sebastiani, 2008; Neumann and Witt, 2009; Neumann, Sudholt, and Witt, 2009). The methods developed in these works have then enabled researchers to perform analyses for more complex settings. This includes hybridizations with local search (Neumann, Sudholt, and Witt, 2008), broader function classes (Kötzing, Neumann, Sudholt, and Wagner, 2011), problems from combinatorial optimization, and different pheromone update schemes that lead to systems which are harder to analyze (Neumann, Sudholt, and Witt, 2010; Sudholt, 2011). Concerning combinatorial optimization, Neumann and Witt (2008) considered ACO for finding minimum spanning trees for two different construction graphs. Attiratanasunthron and Fakcharoenphol (2008) as well as the present authors (Horoba and Sudholt, 2009) considered the classical problem of finding shortest paths in graphs. Zhou (2009) as well as Kötzing, Neumann, Röglin, and Witt (2010b) analyzed ACO for the TSP and Kötzing, Lehre, Oliveto, and Neumann (2010a) considered ACO for the minimum cut problem.

In this work, we take a further step and consider combinatorial problems in the presence of uncertainty. In particular, we focus on the performance of ACO for stochastic shortest path problems. Shortest path problems closely reflect the biological inspiration for ACO and they represent a fundamental problem in computer science and many other areas. Algorithmic research on these problems is still an active field (Abraham, Fiat, Goldberg, and Werneck, 2010; Bast, Funke, Sanders, and Schultes, 2007; Chan, 2007; Orlin, Madduri, Subramani, and Williamson, 2010). Shortest paths have also been investigated in the context of other metaheuristics as described in the following.

1.1 Related Work

Shortest path problems have been investigated for various evolutionary algorithms. First studies focused on evolutionary algorithms that only use mutation, without crossover (Baswana, Biswas, Doerr, Friedrich, Kurur, and Neumann, 2009; Doerr, Happ, and Klein, 2007; Scharnow, Tinnefeld, and Wegener, 2004). Remarkably, it turned out that the all-pairs shortest path problem is a rare example where using crossover significantly speeds up optimization (Doerr and Johannsen, 2010; Doerr and Theile, 2009; Doerr, Happ, and Klein, 2008). Also results for NP-hard multi-objective shortest path problems have been obtained. Horoba (2010) proved that a simple multi-objective evolutionary algorithm using only mutation and a mechanism to maintain diversity in the population represents a fully polynomial randomized approximation scheme (FPRAS). Neumann and Theile (2010) extended this result towards the use of crossover, showing that crossover leads to improved running times. In addition, their algorithm works with a smaller population size, compared to Horoba (2010).

The first investigation of ACO for shortest path problems has been made by Attiratanasunthron and Fakcharoenphol (2008). They considered the single-destination shortest path problem (SDSP) on directed acyclic graphs where one is looking for shortest paths from all vertices to a single destination vertex. In their algorithm n -ANT, on each vertex an artificial ant heads out in search for the destination. For each vertex v the best path found so far is recorded and the pheromones on the edges leaving v are updated according to the best-so-far path. The update also involves a parameter $0 \leq \rho \leq 1$ called *evaporation rate* that determines the strength of a pheromone update. Note that all ants update disjoint sets of edges. The collaborative effort of all ants leads

to the construction of shortest paths. Ants whose shortest paths contain only few edges tend to find their shortest paths first. By marking the respective edges at their start vertices with pheromones they pave the way for other ants whose shortest paths contain more edges. This implies that shortest paths propagate through the graph, similar to the algorithmic idea of the Bellman-Ford algorithm (Cormen, Leiserson, Rivest, and Stein, 2001, Section 24.1).

Attiratanasunthron and Fakcharoenphol proved a bound of $O(m\Delta\ell \log(\Delta\ell)/\rho)$ for the expected optimization time (i. e., the expected number of iterations) of n -ANT for the SDSP that holds for every directed acyclic graph with m edges, maximum outdegree Δ and maximum number of edges ℓ on any shortest path. These results were extended and improved in Horoba and Sudholt (2009). The authors considered a modified variant of n -ANT called $\text{MMAS}_{\text{SDSP}}$ that—unlike n -ANT—can deal with arbitrary directed graphs containing cycles. We gave an improved running time bound for $\text{MMAS}_{\text{SDSP}}$ on the SDSP of $O(\Delta\ell^2 + \ell \log(\Delta\ell)/\rho)$ that holds under some mild conditions on the graph. The same bound on the number of iterations also holds for the all-pairs shortest path problem (APSP) and an algorithm $\text{MMAS}_{\text{APSP}}$ using distinct ants and distinct pheromones for each destination. We also showed that a simple interaction mechanism between ants heading for different destinations yields a significant speed-up. The resulting bound of $O(n^3 \log^3 n)$ constructed solutions was—at that time—lower than the worst-case expected optimization time of evolutionary algorithms using only mutation (Doerr et al, 2007) as well as evolutionary algorithms with crossover (Doerr et al, 2008). For the latter an upper bound of $O(n^{3.25} \log^{0.75} n)$ holds and examples are given where this is tight (Doerr and Theile, 2009). Only after our previous work was published, Doerr, Johannsen, Kötzing, Neumann, and Theile (2010) presented an evolutionary algorithm with a modified parent selection that has a better upper bound of $O(n^3 \log n)$ constructed solutions.

In an independent line of research, Borkar and Das (2009) presented convergence proofs and empirical results for an ACO system that contains an additional learning component. In terms of shortest paths they only considered specific classes of graphs containing layers of nodes such that the edge set contains exactly all pairs of nodes in subsequent layers. This is referred to as multi-stage shortest path problem. Kolavali and Bhatnagar (2009) extended this work towards four variants of the basic ACO algorithm from Borkar and Das (2009).

1.2 Our Contribution

In this work we extend our previous work (Horoba and Sudholt, 2009) on the SDSP towards a stochastic variant of the SDSP on directed acyclic graphs. The motivation is to investigate the robustness of ACO in stochastic settings and to see under which conditions the ants are still able to find shortest paths efficiently. Several different variants of stochastic shortest path problems have been investigated in the literature on optimization (see, e. g., Miller-Hooks and Mahmassani (2000); Nikolova, Brand, and Karger (2006)). One variant is to find a path with the least expected time (LET) (Bertsekas and Tsitsiklis, 1991; Papadimitriou and Yannakakis, 1991). Another problem is to maximize the probability of arriving at the destination within a given time bound (Boyan and Mitzenmacher, 2001; Fan, Kalaba, and Moore, 2005).

We consider a setting where noise is added to edges. The noise is non-negative and the task is to find or approximate the real shortest paths, i. e., the shortest paths

without noise. The main question is in which settings the ants are still able to locate shortest paths efficiently, while being subjected to noise. The reader might think of noise reflecting errors of measurement that occur when trying to evaluate the quality of a candidate solution. In the special case where the expected noisy path length and the real path length differ by a fixed factor, for all paths, the task of finding the shortest real path is equivalent to finding the path with the least expected time. This then yields an instance of the LET problem. The described property holds for some of the investigated settings, hence our analyses address special cases of the LET problem. To our knowledge this is the first running time analysis of a randomized search heuristic on a stochastic combinatorial problem.

We describe our results and give an outline of this paper (for an overview of our theorems see Table 1). In Section 2 we formally introduce our setting, the problem, and the ant system. Section 3 presents general upper bounds for the time until a reasonable approximation is found. This includes graphs with gaps between the shortest-path lengths and the lengths of non-optimal paths that allow the ants to efficiently compute the real shortest paths. For arbitrary weights an upper bound for general and possibly dependent noise is presented. This result is refined for general independent noise. Section 4 deals with the gamma distribution, which is a common choice for modeling noise. In Section 5 we consider independent gamma-distributed noise and prove an exponential lower bound for the time until a good approximation is found on a constructed graph. Section 6 is dedicated to the case that the noise values on different edges are strongly correlated. We show that the negative result for the graph considered in Section 5 breaks down. This demonstrates that correlations can make a difference between polynomial and exponential times for finding a good approximation. We also prove a general upper bound for finding shortest paths under gamma-distributed noise that holds for graphs with gaps between the shortest-path lengths and those of non-optimal paths. In particular, there is a trade-off between the size of the gap and the upper bound on the expected running time. We conclude in Section 7.

2 Problem and Algorithm

Consider a weighted directed acyclic graph $G = (V, E, w)$ with $V = \{1, \dots, n\}$, $E = \{e_1, \dots, e_m\} \subseteq V \times V$, and $w: E \rightarrow \mathbb{R}_0^+$. We are interested in finding a shortest path from each source $v \in V$ to a single destination. Throughout this work, ℓ denotes the maximum number of edges on any shortest path to the destination. Similarly, L denotes the maximum number of edges on any path to the destination.

We consider a sequence $p = (v_0, \dots, v_s)$ of vertices $v_i \in V$ with $(v_{i-1}, v_i) \in E$ for each $1 \leq i \leq s$ as a *path* from v_0 to v_s with s edges. We also consider the corresponding sequence $p = ((v_0, v_1), \dots, (v_{s-1}, v_s))$ of edges $(v_{i-1}, v_i) \in E$ as a *path*. In the remainder of the paper we utilize both representations for convenience. We define the *length* $w(p)$ of a path $p = (e_1, \dots, e_s)$ as $w(p) := \sum_{i=1}^s w(e_i)$ if it ends with the destination and $w(p) := \infty$ otherwise. By $\deg(v)$ we denote the outdegree of v and by Δ we denote the maximum outdegree of any vertex in the graph.

We investigate a stochastic version of the described shortest path problem. The term “stochastic” means that whenever we evaluate the length of a path p at time t we do not get the *real length* $w(p)$ but a *noisy length* $\tilde{w}(p, t)$. The resulting random variables $\tilde{w}(p, 1), \tilde{w}(p, 2), \dots$ for a fixed path are i. i. d.

Thm	Graph Noise	Optimization Goal Time Bound
1	instances with gaps ¹ $\eta(e) \geq 0$	shortest paths $O((L \log n)/\tau_{\min} \cdot \alpha/(\alpha - 1) + L \log(\tau_{\max}/\tau_{\min})/\rho)$ in expect.
2	all instances $\eta(e) \geq 0$	multiplicative error $(1 + \alpha \cdot \eta_{\max})^L$, $\alpha > 1$ $O((L \log n)/\tau_{\min} \cdot \alpha/(\alpha - 1) + L \log(\tau_{\max}/\tau_{\min})/\rho)$ in expect.
3	all instances $\eta(e) \geq 0$ ind.	additive error $(L + 1)^2/2 \cdot \tilde{w}_{\max}$ $O((L \log n)/\tau_{\min} + L \log(\tau_{\max}/\tau_{\min})/\rho)$ in expect.
4	example gr. $G_{n,k\theta,\varepsilon}$ $\eta(e) \sim \Gamma(k, \theta)$ ind.	multiplicative error $(1 + r)$, $r > 0$ arbitrary $\Omega(n/\tau_{\min} + \sqrt{n} \log(\tau_{\max}/\tau_{\min})/\rho)$ w. h. p.
5	example gr. $G_{n,k\theta,\varepsilon}$ $\eta(e) \sim \Gamma(k, \theta)$ ind.	multiplicative error $(1 + \varepsilon)$, $\varepsilon \leq 1$ with $\varepsilon/(k\theta) = 1/e - \Omega(1)$ $\Omega(e^{c\sqrt{n}})$ w. h. p., $c > 0$ constant
6	example gr. $G_{n,k\theta,\varepsilon}$ $\eta \sim \Gamma(k, \theta)$	shortest paths $O(n/\tau_{\min} + (k\theta/\varepsilon)^k \cdot e^{\varepsilon/\theta}/\tau_{\min} + n \log(\tau_{\max}/\tau_{\min})/\rho)$ in exp.
7	instances with gaps ² $\eta \sim \Gamma(k, \theta)$	shortest paths $O(((z + 1)^k L \log n)/\tau_{\min} + L \log(\tau_{\max}/\tau_{\min})/\rho)$ in expect.

¹ length of every non-optimal path from v at least $(1 + \alpha \cdot \mathbb{E}(\eta(\text{opt}_v))) \cdot \text{opt}_v$, $\alpha > 1$

² length of every non-optimal path from v at least $(1 + k\theta/z) \cdot \text{opt}_v$, $z \in \mathbb{N}$

Table 1 Overview of the theorems. $\eta(e) = \eta(e, p, t)$ and $\eta = \eta(p, t)$ are shorthands for noise added to an edge e , after being multiplied with the edge's weight $w(e)$ (see Section 2). Independence ("ind.") refers to independence across edges. L denotes the maximum number of edges on any path to the destination, opt_v is the length of a shortest path from v , $\eta_{\max} := \max_{1 \leq i \leq m} \mathbb{E}(\eta(e_i))$, and $\tilde{w}_{\max} := \max_{e \in E} \mathbb{E}(\eta(e)) \cdot w(e)$. The constraints on the parameters of the algorithm τ_{\min} , τ_{\max} and ρ (see Section 2) and the parameters of the gamma-distributed noise k and θ are given in the corresponding theorem.

Each random variable $\tilde{w}(p, t)$ is determined as follows. Assume a family $(\eta(e, p, t))_{e \in E}$ of nonnegative random variables $\eta(e, p, t) \geq 0$. The noisy length $\tilde{w}(e, t)$ of an edge $e \in p$ is then computed as $(1 + \eta(e, p, t)) \cdot w(e)$. The noisy length $\tilde{w}(p, t)$ of a path $p = (e_1, \dots, e_s)$ is then defined as

$$\sum_{i=1}^s (1 + \eta(e_i, p, t)) \cdot w(e_i).$$

Note that $\tilde{w}(e, t) \geq w(e)$ holds for each edge e and $\tilde{w}(p, t) \geq w(p)$ holds for each path p . Also note that the strength of the noise depends on the weight of the corresponding edge.

We already mentioned that noisy path length evaluations are independent across time, i. e., $\eta(e, p, t)$ and $\eta(e, p, t')$ are i. i. d. random variables for all $t \neq t'$. Similarly, we assume that $\eta(e, p, t)$ and $\eta(e, p', t)$ are i. i. d. for all $p' \neq p$. Two ants, constructing paths p and p' , respectively, may experience different noise on the same edge, in the same iteration. (Many analyses also apply to noise models where all ants in one iterations experience the same noise: $\eta(e, p, t) = \eta(e, p', t)$ for all $p \neq p'$.) When speaking of *independent noise*, in the following we refer to independence across edges, i. e., $\eta(e, p, t)$ and $\eta(e', p, t)$ are independent. While independence across time and across paths/ants is always assumed, our results cover noise models with non-independent noise across edges as well as models where $\eta(e, p, t)$ and $\eta(e', p, t)$ are i. i. d. for all edges $e \neq e'$.

We consider the ant system MMAS_{SDSP} introduced in Horoba and Sudholt (2009). To ease the presentation, we describe a simplification of the system for directed acyclic graphs as the results in this work will be limited to acyclic graphs. If the underlying graph is acyclic, the ant system can be formulated as follows (see Algorithms 1 and 2).

Algorithm 1 Path Construction from u to v

```

1: Initialize  $i \leftarrow 0$ ,  $p_0 \leftarrow u$ ,  $V_1 \leftarrow \{p \in V \mid (p_0, p) \in E\}$ 
2: while  $p_i \neq v$  and  $V_{i+1} \neq \emptyset$  do
3:    $i \leftarrow i + 1$ 
4:   Choose  $p_i \in V_i$  with prob.  $\tau((p_{i-1}, p_i)) / \sum_{p \in V_i} \tau((p_{i-1}, p))$ 
5:    $V_{i+1} \leftarrow \{p \in V \mid (p_i, p) \in E\}$ 
6: end while
7: return  $(p_0, \dots, p_i)$ 

```

Algorithm 2 MMAS_{SDSP}

```

1: Initialize pheromones  $\tau$  and best-so-far paths  $p_1^*, \dots, p_n^*$ 
2: for  $t = 1, 2, \dots$  do in parallel
3:   for  $u = 1$  to  $n$  do in parallel
4:     Construct a simple path  $p_u$  from  $u$  to the destination w. r. t.  $\tau$ 
5:     Sample  $\tilde{w}(p_u, t)$  from the distribution underlying the noisy path length of  $p_u$ 
6:     if  $\tilde{w}(p_u, t) \leq \tilde{w}(p_u^*, t_u^*)$  then  $p_u^* \leftarrow p_u$ ,  $t_u^* \leftarrow t$  end if
7:     Update pheromones  $\tau$  on all edges  $(u, \cdot)$  w. r. t.  $p_u^*$ 
8:   end for
9: end for

```

In every iteration, from each vertex, an ant starts and tries to construct a path to the destination. It does so by performing a random walk through the graph guided by pheromones. Pheromones are positive real values associated with the edges of a graph. They are denoted by a function $\tau: E \rightarrow \mathbb{R}^+$. The next edge is always chosen with a probability that is proportional to the amount of pheromone on the edge. If the ant gets stuck because there are no more outgoing edges, the length of the resulting path is ∞ by definition of w . This can only happen if there are vertices from which the destination is not reachable. The path construction is detailed in Algorithm 1.

Once an ant with start vertex u has created a path, its best-so-far path p_u^* is updated with the newly constructed path p_u if the latter is not worse than the former. Note that this decision is made according to the noisy path lengths $\tilde{w}(p_u^*, t_u^*)$ and $\tilde{w}(p_u, t)$, t_u^* being the iteration in which p_u^* was stored as best-so-far path. This means that we store the noisy length of p_u^* at the time t_u^* and use this value for the comparison, instead of re-evaluating the noisy path length of p_u^* at time t . In the following, we always use short-hands $\tilde{w}(p_u^*) := \tilde{w}(p_u^*, t_u^*)$ and $\tilde{w}(p_u) := \tilde{w}(p_u, t)$, t referring to the current iteration. We use similar abbreviations for $\eta(e, p, t)$ where appropriate.

Finally, the ant updates pheromones on the edges leaving its start vertex according to the best-so-far path p_u^* and the following formula. Pheromones on an edge e are denoted by $\tau(e)$. Initially we set $\tau(e) = 1/\deg(v)$ for all $e = (v, \cdot)$ as well as $p_v^* = ()$, $t_v^* = 0$ and $\tilde{w}(p_v^*, 0) = \infty$ for all $v \in V$. Using equal pheromones for all edges (v, \cdot) implies that in the first iteration all ants make uniform random decisions at all vertices. The evaporation factor $0 \leq \rho \leq 1$ as well as τ_{\min} and τ_{\max} are parameters of the algorithm. The pheromone update formula is given by

$$\tau(e) \leftarrow \begin{cases} \min \{(1 - \rho) \cdot \tau(e) + \rho, \tau_{\max}\} & \text{if } e = (u, v) \in p_u^*, \\ \max \{(1 - \rho) \cdot \tau(e), \tau_{\min}\} & \text{if } e = (u, v) \notin p_u^*. \end{cases}$$

The so-called pheromone borders τ_{\min} and τ_{\max} ensure that the pheromone for each edge is always bounded away from 0, so that there is always a chance of reverting a decision once made. As in Horoba and Sudholt (2009) we fix $\tau_{\max} := 1 - \tau_{\min}$ and only vary τ_{\min} , subject to the constraints $0 < \tau_{\min} \leq 1/\Delta$. The latter inequality is required

to make the initialization work properly. Note that the algorithm can be parallelized easily as the path constructions are independent and the pheromone update concerns disjoint sets of edges. The complete algorithm is described in Algorithm 2.

In the theory of randomized search heuristics often the number of iterations is taken as performance measure. As we have n ants constructing solutions in one iteration, it makes sense to also consider the number of constructed solutions. This holds in particular when comparing our results to other search heuristics. For more detailed runtime measures as common in the theory of algorithms the number of operations needed for path constructions and pheromone updates has to be taken into account. The probabilities for choosing specific edges only need to be computed once in each iteration. The effort for each vertex is proportional to its outdegree, hence the total effort for computing the probabilities is $O(|E|)$. One path construction can be implemented such that it runs in time $\tilde{O}(L)$ (\tilde{O} hiding logarithmic factors). Hence, one iteration of the ant system can be implemented using $\tilde{O}(|E| + n \cdot L) = \tilde{O}(n^2)$ elementary operations.

In the following, we bound the number of iterations until the subsequent goal is met. Let opt_v denote the length $w(p)$ of a shortest path p from $v \in V$ to n . Let P_v denote the set containing all paths from $v \in V$ to n and $P_v(\alpha)$ the set containing each path p from $v \in V$ to n with $w(p) \leq \alpha \cdot \text{opt}_v$, where $\alpha \in \mathbb{R}$ with $\alpha \geq 1$. Call a vertex v α -approximated if and only if $w(p_v^*) \leq \alpha \cdot \text{opt}_v$. Call a vertex v *permanently* α -approximated if and only if $\tilde{w}(p_v^*) \leq \alpha \cdot \text{opt}_v$. The difference between these notions can be explained as follows. If v is α -approximated, the length of the best-so-far path might be larger than an α -approximation as we store the noisy length of the best-so-far path, $\tilde{w}(p_v^*)$. The ant may then later switch to another path with smaller noisy length, but larger real length, so that v loses its α -approximation property. On the other hand, if a vertex is permanently α -approximated, then it will always remain permanently α -approximated.

We say that $\text{MMAS}_{\text{SDSP}}$ has found an α -approximation if all $v \in V$ are α -approximated at the same time. Note that we require the best-so-far paths to reflect good approximations. It may happen that ants create paths with better real length, but if their noisy length is worse than the current best-so-far path length, this better path will be lost immediately. Therefore, it makes sense to base the goal of the optimization on best-so-far paths as this is much more permanent information. Also we should expect that a well-performing ACO system is able to store information about good solutions in the pheromones.

Note that with a 1-approximation all shortest paths have been found. However, in general it does not make sense to hope for a permanent 1-approximation as the probability of experiencing a noise value of 0 might be 0.

3 General Upper Bounds

We begin with a very general upper bound for the optimization time holding for all noise distributions. Our arguments are based on the following statements that are implicitly proven in Horoba and Sudholt (2009). For the sake of completeness, we give a proof.

Lemma 1 *The ant system $\text{MMAS}_{\text{SDSP}}$ features the following properties on directed acyclic graphs with destination n .*

1. The probability of an ant at $u \neq n$ with $\deg(u) > 1$ following the edge $e = (u, v)$ is at most $\tau(e)$ and at least $\tau(e)/2 \geq \tau_{\min}/2$.
2. If the pheromone on an edge has not been increased for at least $T^* := \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations, the pheromone on the edge is τ_{\min} .
3. Consider a set \mathcal{S}_v of paths from any vertex $v \in V$ to the destination such that every edge leaving \mathcal{S}_v has pheromone τ_{\min} . If every path in \mathcal{S}_v has at most $\ell - 1$ edges and $\tau_{\min} \leq 1/(\Delta\ell)$, then the probability of an ant following a path in \mathcal{S}_v from v to the destination is at least $1/e$.

Proof The first statement follows from Lemma 1 in Horoba and Sudholt (2009). It says that probabilities and pheromones are closely related as they differ only by a factor of at most 2. In many cases, like at initialization, the pheromones on edges (v, \cdot) for some vertex v sum up to 1. If v has degree larger than 2, this sum can exceed 1 if pheromones on several edges are set to τ_{\min} by the max-term in the pheromone update formula. Taking the maximum can be thought of as pheromones being artificially raised after being multiplied with $1 - \rho$. This can increase the sum of pheromones beyond 1. However, a simple calculation shows that if $\tau_{\min} \leq 1/\Delta$ then the sum of pheromones is always bounded by 2. This directly implies the claimed lower bound $\tau(e)/2$; the upper bound $\tau(e)$ follows as the sum of pheromones is always at least 1.

The second statement holds since any initial pheromone $\tau(e)$ attains the value τ_{\min} after not having been increased for at least $\ln(\tau_{\max}/\tau_{\min})/\rho$ iterations. The reason is that the pheromone is always multiplied by $1 - \rho$ (unless the lower border τ_{\min} is hit) and $(1 - \rho)^{\ln(\tau_{\max}/\tau_{\min})/\rho} \cdot \tau(e) \leq e^{-\ln(\tau_{\max}/\tau_{\min})} \cdot \tau(e) \leq \tau_{\min}$.

The third statement holds since at every vertex v the probability of choosing a “wrong” edge (i. e. an edge leaving \mathcal{S}_v) is at most $\deg(v) \cdot \tau_{\min}$. This follows from the assumption on the pheromones and the first statement. The probability of not choosing any wrong edge on a path of at most $\ell - 1$ edges is then at least $(1 - \deg(v) \cdot \tau_{\min})^{\ell-1} \geq (1 - 1/\ell)^{\ell-1} \geq 1/e$ since $\tau_{\min} \leq 1/(\deg(v)\ell)$ by assumption. \square

Intuitively, our ant system should be able to identify the real shortest paths in case there is a certain gap between weights of all shortest paths and all non-optimal paths. The following theorem gives insight into how large this gap has to be. The variable α describes a trade-off between the strength of the preconditions and the expected running time.

Theorem 1 *Consider a graph $G = (V, E, w)$ with weight $w(e) > 0$ and noise $\eta(e) = \eta(e, p, t) \geq 0$ for each $e \in E$. Choose $\tau_{\min} \leq 1/(\Delta L)$. Let opt_v denote the real length of a shortest path from v and $\eta(\text{opt}_v)$ denote the random noise on all edges of this path. If for each vertex $v \in V$ and some $\alpha > 1$ it holds that every non-optimal path has length at least $(1 + \alpha \cdot \mathbb{E}(\eta(\text{opt}_v))) \cdot \text{opt}_v$ then in expectation $\text{MMAS}_{\text{SDSP}}$ finds a 1-approximation after $O((L \log n)/\tau_{\min} \cdot \alpha/(\alpha - 1) + L \log(\tau_{\max}/\tau_{\min})/\rho)$ iterations.*

For $\tau_{\min} = 1/(\Delta L)$ and $\alpha = 1 + \Omega(1)$, the running time bound simplifies to $O(\Delta L^2 \log n + (L \log n)/\rho)$. If additionally $\rho = \Omega(1/(\Delta L))$, we get $O(\Delta L^2 \log n)$.

Proof Adapting notation from Attiratanasunthron and Fakcharoenphol (2008), we call a vertex v *optimized* if the ant starting at v has found a shortest path and $\tilde{w}(p_v^*) < (1 + \alpha \cdot \mathbb{E}(\eta(\text{opt}_v))) \cdot \text{opt}_v$. Along with our condition on path lengths, we have that then the ant at v will never accept a non-optimal path as its best-so-far path. Hence, it will always reinforce some shortest path with pheromones. We call vertex v *processed* if it is optimized and all pheromones on edges (v, \cdot) that are not part of shortest paths

from v have decreased to τ_{\min} . Due to Lemma 1 this happens at most $\ln(\tau_{\max}/\tau_{\min})/\rho$ iterations after v has become optimized.

We define a partition of V according to the maximum number of edges on any path to the destination. Let

$$V_i := \{v \in V \mid \exists p \in P_v : |p| \geq i \text{ and } \forall p \in P_v : |p| \leq i\}$$

with $0 \leq i \leq L$ where $|p|$ denotes the number of edges of p . (Recall that P_v denotes the set containing all paths from v to the destination, see Section 2.) Then V_0, \dots, V_L form a partition of V . Note that for each $(u, v) \in E$ there are indices i and j with $u \in V_i, v \in V_j$, and $i > j$.

Consider a vertex $u \in V_i$. Assume that for each index $0 \leq j < i$ each vertex $v \in V_j$ is processed. We estimate the expected time until u becomes processed. For each edge $(v, \cdot), v \in V_j, 0 \leq j < i$, that is not part of a shortest path from v we have $\tau((v, \cdot)) = \tau_{\min}$. Denote by a_u the ant starting at vertex u . The probability of ant a_u choosing the first edge $(u, v), v \in V_j, 0 \leq j < i$, of a path from $P_u(1)$ (i.e. a shortest path) is lower bounded by $\tau_{\min}/2$ due to the first statement of Lemma 1. Invoking Lemma 1 with $\mathcal{S}_v = P_v(1)$, the set of all shortest paths, the probability of ant a_u continuing from v on some path from $P_v(1)$ to the destination is at least $1/e$. Together, the probability of ant a_u finding some shortest path is at least $\tau_{\min}/(2e)$.

In addition to finding such a path $p = (e_1, \dots, e_k)$, the noisy length evaluation of p must not be too poor. The vertex u becomes optimized if the noise sampled on opt_u is less than $\alpha \cdot E(\eta(\text{opt}_u))$. By Markov's inequality this happens with probability at least $1 - 1/\alpha$. Hence, the probability of optimizing vertex $u \in V_i$ is lower bounded by $\tau_{\min}/(2e) \cdot (1 - 1/\alpha)$. The expected waiting time until this happens is at most $O(1/\tau_{\min} \cdot \alpha/(\alpha - 1))$.

By standard arguments, the expected time until the last vertex in V_i has become optimized is bounded by $O((\log n)/\tau_{\min} \cdot \alpha/(\alpha - 1))$. After an additional waiting time of $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations all pheromones for edges leaving V_i have been adapted appropriately and we continue our considerations with the set V_{i+1} . Summing up the expected times for all L sets yields the claimed upper bound. \square

Note that the statement of Theorem 1 can be improved by replacing L with $\ell \leq L$ and adapting the partition V_0, \dots, V_L towards edges on shortest paths. We refrain from such a modification to be consistent with upcoming results.

The condition on the length of non-optimal paths is stronger for vertices that are “far away” from the destination. Imagine a multigraph where some vertex v has two edges e_1, e_2 of different weight such that e_1 is part of a shortest path and both edges lead to the same vertex w . In order for each non-optimal path to have length at least $(1 + \alpha \cdot E(\eta(\text{opt}_v))) \cdot \text{opt}_v$, it must be that e_2 has a large weight. This effect becomes more pronounced, the larger the real length of the shortest path from w is. This stronger requirement makes sense because the ants must still be able to distinguish short paths from long paths efficiently.

In the case of arbitrary weights the situation becomes more complicated. If second-best paths are only slightly longer than shortest paths, it may be difficult for artificial ants to distinguish between these paths in the presence of noise. In this case we cannot always rely on the optimality of sub-paths as done in Theorem 1. The next theorem provides a trade-off between the desired approximation ratio and the required optimization time according to the variable α .

Theorem 2 Consider a graph $G = (V, E, w)$ with weight $w(e) > 0$ and noise $\eta(e) = \eta(e, p, t) \geq 0$ for each $e \in E$. Choose $\tau_{\min} \leq 1/(\Delta L)$ and $\alpha > 1$ and let $\chi = (1 + \alpha \cdot \eta_{\max})^L$, where $\eta_{\max} := \max_{1 \leq i \leq m} \mathbb{E}(\eta(e_i))$. Then in expectation $\text{MMAS}_{\text{SDSP}}$ finds a χ -approximation in $O((L \log n)/\tau_{\min} \cdot \alpha/(\alpha-1) + L \log(\tau_{\max}/\tau_{\min})/\rho)$ iterations.

The running time bound can be simplified in the same way as described after Theorem 1. Note that the approximation ratio $\chi \leq e^{\alpha \cdot \eta_{\max} \cdot L}$ converges to 1 if $\eta_{\max} = o(1/(\alpha \cdot L))$. However, the approximation ratio quickly deteriorates with larger noise.

Proof Recall the partition V_0, \dots, V_L from Theorem 1. Consider a vertex $u \in V_i$. Assume that for each index $0 \leq j < i$ each vertex $v \in V_j$ has been permanently $\chi^{j/L}$ -approximated for at least $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations. We estimate the expected time until u becomes permanently $\chi^{i/L}$ -approximated. For each edge (v, \cdot) , $v \in V_j$, $0 \leq j < i$, that is not extendable to a path from $P_v(\chi^{j/L})$ we have $\tau((v, \cdot)) = \tau_{\min}$ by the second statement of Lemma 1. Denote by a_u the ant starting at vertex u . The probability of ant a_u choosing the first edge (u, v) , $v \in V_j$, $0 \leq j < i$, of a path from $P_u(1)$ (i. e. a shortest path) is lower bounded by $\tau_{\min}/2$ due to the first statement of Lemma 1. Invoking Lemma 1 with $\mathcal{S}_v = P_v(\chi^{j/L})$, the probability of ant a_u continuing from v on some path from $P_v(\chi^{j/L})$ to the destination is at least $1/e$. Together, the probability of ant a_u finding some path p from $P_u(\chi^{j/L})$ is at least $\tau_{\min}/(2e)$.

In addition to finding such a path $p = (e_1, \dots, e_k)$, the noisy length evaluation of p must not be too poor. The vertex u becomes permanently $\chi^{i/L}$ -approximated if $\tilde{w}(p) \leq \chi^{i/L} \cdot \text{opt}_u$. We have

$$\begin{aligned} \chi^{i/L} \cdot \text{opt}_u &= \chi^{i/L} \cdot (w(e_1) + \text{opt}_v) \\ &\geq \chi^{i/L} \cdot \left(w(e_1) + \frac{\sum_{i=2}^k w(e_i)}{\chi^{j/L}} \right) \\ &= \left(\chi^{i/L} - \chi^{(i-j)/L} \right) \cdot w(e_1) + \chi^{(i-j)/L} \cdot \sum_{i=1}^k w(e_i) \\ &\geq \chi^{1/L} \cdot \sum_{i=1}^k w(e_i). \end{aligned}$$

Hence,

$$\begin{aligned} &\text{Prob}\left(\tilde{w}(p) \geq \chi^{i/L} \cdot \text{opt}_u\right) \\ &\leq \text{Prob}\left(\sum_{i=1}^k (1 + \eta(e_i)) \cdot w(e_i) \geq \chi^{1/L} \cdot \sum_{i=1}^k w(e_i)\right) \\ &= \text{Prob}\left(\sum_{i=1}^k \eta(e_i) \cdot w(e_i) \geq (\chi^{1/L} - 1) \cdot \sum_{i=1}^k w(e_i)\right). \end{aligned}$$

By Markov's inequality, this probability is at most

$$\frac{\mathbb{E}\left(\sum_{i=1}^k \eta(e_i) \cdot w(e_i)\right)}{(\chi^{1/L} - 1) \cdot \sum_{i=1}^k w(e_i)} = \frac{\sum_{i=1}^k \mathbb{E}(\eta(e_i)) \cdot w(e_i)}{\alpha \cdot \eta_{\max} \cdot \sum_{i=1}^k w(e_i)} \leq \frac{1}{\alpha} < 1.$$

Therefore we have $\tilde{w}(p) < \chi^{i/L} \cdot w(p)$ with probability at least $1 - 1/\alpha > 0$. Hence, the probability of vertex $u \in V_i$ becoming permanently $\chi^{i/L}$ -approximated is lower bounded by $\tau_{\min}/(2e) \cdot (1 - 1/\alpha)$. The expected waiting time until this happens is at most $O(1/\tau_{\min} \cdot \alpha/(\alpha - 1))$.

By standard arguments, the expected time until the last vertex in V_i has become permanently $\chi^{i/L}$ -approximated is bounded by $O((\log n)/\tau_{\min} \cdot \alpha/(\alpha - 1))$. After an additional waiting time of T^* iterations all pheromones for edges leaving V_i have been adapted appropriately and we continue our considerations with the set V_{i+1} . Summing up the expected times for all L sets yields the claimed upper bound. \square

In the following we assume that the random variables $\eta(e)$, $e \in E$, are independent. Each time a new path is constructed, new random variables $\eta(e)$ are used to derive its noisy weight; all $\eta(e)$ -values are independent for each edge, each ant and each iteration. This means that in one iteration different ants may experience different noise, even if they follow the same edges. Recall, however, that the length of a best-so-far path is not re-evaluated whenever it is compared against a new path.

In contrast to Theorem 2 we formulate the result in terms of additive errors instead of multiplicative approximation ratios. While the approximation guarantee in Theorem 2 depends exponentially on L , the additive error in the upcoming Theorem 3 only depends quadratically on L . We use the following lemma, which is an immediate implication of Theorem 1 in Feige (2006).

Lemma 2 (Feige (2006)) *Let X_1, \dots, X_n be arbitrary nonnegative independent random variables, with expectations μ_1, \dots, μ_n , respectively. Let $X = \sum_{i=1}^n X_i$, and let μ denote the expectation of X (hence, $\mu = \sum_{i=1}^n \mu_i$). Let $\mu_{\max} := \max\{\mu_1, \dots, \mu_n\}$. Then for every $\alpha > 0$,*

$$\text{Prob}(X \leq \mu + \alpha\mu_{\max}) \geq \min\{\alpha/(1 + \alpha), 1/13\}.$$

Theorem 3 *Consider a graph $G = (V, E, w)$ with weight $w(e) > 0$ where the noise variables $\eta(e) = \eta(e, p, t) \geq 0$ for each $e \in E$ are independent. Choose $\tau_{\min} \leq 1/(\Delta L)$ and let $\tilde{w}_{\max} := \max_{e \in E} E(\eta(e)) \cdot w(e)$. $\text{MMAS}_{\text{SDSP}}$ finds an approximation within an additive error of $(L + 1)^2/2 \cdot \tilde{w}_{\max}$ within $O((L \log n)/\tau_{\min} + L \log(\tau_{\max}/\tau_{\min})/\rho)$ expected iterations.*

Proof The basic proof idea is the same as in Theorem 2. Recall the partition V_0, \dots, V_L from Theorem 1 and consider a vertex $u \in V_i$. Assume that for each index $0 \leq j < i$ each vertex $v \in V_j$ has been permanently approximated within an additive error of $(j + 1)^2/2 \cdot \tilde{w}_{\max}$ for at least $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations. We estimate the expected time until u becomes permanently approximated within an additive error of $(i + 1)^2/2 \cdot \tilde{w}_{\max}$. For $i = L$ this implies the claim.

Let $P_v^+(\varepsilon) := P_v(1 + \varepsilon/\text{opt}_v)$ for all $v \in V$ and $\varepsilon \geq 0$. For each edge (v, \cdot) , $v \in V_j$, $0 \leq j < i$, that is not extendable to a path from $P_v^+((j + 1)^2/2 \cdot \tilde{w}_{\max})$ we have $\tau((v, \cdot)) = \tau_{\min}$ by the second statement of Lemma 1. Denote by a_u the ant starting at vertex u . The probability of ant a_u choosing the first edge (u, v) , $v \in V_j$, $0 \leq j < i$, of a path from $P_u^+(0)$ (i.e. a shortest path) is lower bounded by $\tau_{\min}/2$ due to the first statement of Lemma 1. Invoking Lemma 1 with $\mathcal{S}_v = P_v^+((j + 1)^2/2 \cdot \tilde{w}_{\max})$, the probability of ant a_u continuing from v on some path from $P_v^+((j + 1)^2/2 \cdot \tilde{w}_{\max})$ to the destination is at least $1/e$. Together, the probability of ant a_u finding some path $p = (e_1, \dots, e_k)$ from $P_u^+((i + 1)^2/2 \cdot \tilde{w}_{\max})$ is at least $\tau_{\min}/(2e)$.

In addition to finding such a path p , the noisy length evaluation of p must not be too poor. Let u be the first and v be the second vertex on the path. Assuming $v \in P_v^+((j+1)^2/2 \cdot \tilde{w}_{\max})$ implies that

$$\sum_{i=2}^k w(e_i) \leq \text{opt}_v + \frac{(j+1)^2}{2} \cdot \tilde{w}_{\max}. \quad (1)$$

The vertex u becomes permanently approximated within an additive error of $(i+1)^2/2 \cdot \tilde{w}_{\max}$ if $\tilde{w}(p) \leq \text{opt}_u + (i+1)^2/2 \cdot \tilde{w}_{\max}$. Using $(i+1)^2/2 - i^2/2 = i + 1/2$ and (1), we have

$$\begin{aligned} \text{opt}_u + \frac{(i+1)^2}{2} \cdot \tilde{w}_{\max} &= w(e_1) + \text{opt}_v + \frac{(i+1)^2}{2} \cdot \tilde{w}_{\max} \\ &\geq w(e_1) + \text{opt}_v + \frac{(j+1)^2}{2} \cdot \tilde{w}_{\max} + \left(i + \frac{1}{2}\right) \tilde{w}_{\max} \\ &\geq \sum_{r=1}^k w(e_r) + \left(i + \frac{1}{2}\right) \cdot \tilde{w}_{\max}. \end{aligned}$$

Hence,

$$\begin{aligned} &\text{Prob}\left(\tilde{w}(p) \leq \text{opt}_u + \frac{(i+1)^2}{2} \cdot \tilde{w}_{\max}\right) \\ &\geq \text{Prob}\left(\sum_{r=1}^k (1 + \eta(e_r)) \cdot w(e_r) \leq \sum_{r=1}^k w(e_r) + \left(i + \frac{1}{2}\right) \cdot \tilde{w}_{\max}\right) \\ &\geq \text{Prob}\left(\sum_{r=1}^k \eta(e_r) \cdot w(e_r) \leq \left(k + \frac{1}{2}\right) \cdot \tilde{w}_{\max}\right) \\ &\geq \text{Prob}\left(\sum_{r=1}^k \eta(e_r) \cdot w(e_r) \leq \sum_{r=1}^k \mathbb{E}(\eta(e_r)) \cdot w(e_r) + \frac{1}{2} \cdot \tilde{w}_{\max}\right) \end{aligned}$$

as $\mathbb{E}(\eta(e_r)) \cdot w(e_r) \leq \tilde{w}_{\max}$ for every edge e_r . Invoking Feige's inequality (Lemma 2) with $\alpha = 1/2$, we have that

$$\text{Prob}\left(\sum_{r=1}^k \eta(e_r) \cdot w(e_r) \leq \sum_{r=1}^k \mathbb{E}(\eta(e_r)) \cdot w(e_r) + \frac{1}{2} \cdot \tilde{w}_{\max}\right) \geq \frac{1}{13}.$$

We conclude that $\tilde{w}(p) \leq \text{opt}_u + (i+1)^2/2 \cdot \tilde{w}_{\max}$ holds with probability at least $1/13$. The probability of vertex $u \in V_i$ becoming permanently approximated within an additive error of $(i+1)^2/2 \cdot \tilde{w}_{\max}$ is at least $\tau_{\min}/(26e)$. The expected waiting time until this happens is at most $O(1/\tau_{\min})$.

By standard arguments, the expected time until the last vertex in V_i has become permanently approximated within an additive error of $(i+1)^2/2 \cdot \tilde{w}_{\max}$ is bounded by $O((\log n)/\tau_{\min})$. After an additional waiting time of $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations all pheromones for edges leaving V_i have been adapted appropriately and we continue our considerations with the set V_{i+1} . Summing up the expected times for all L sets yields the claimed upper bound. \square

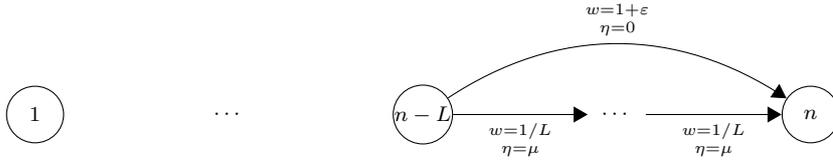


Fig. 1 Example graph with parameters n, μ, L , deterministic noise, and destination n .

In the proofs of Theorems 2 and 3 we argue that a bound on the noisy path length implies the same bound for the real path length. This way of reasoning may seem wasteful as in many noise models the real path length will be shorter than the noisy path length. However, there are examples where ants may switch from a short real path with noise to a longer path without noise and both paths have almost the same noisy path length.

There is a family of graphs parameterized by n, L , and an arbitrary value $\mu \geq 0$ where we cannot hope for the ant system to achieve an approximation ratio $1 + \varepsilon$ with $0 \leq \varepsilon < \mu$. Consider the graph $G = (V, E)$ with $V = \{1, \dots, n\}$ and $E = \{(i, i+1) \mid n-L \leq i < n\} \cup \{(n-L, n)\}$. We have $w(e) = 1/L$ and $\text{Prob}(\eta(e) = \mu/L) = 1$ for $e = (i, i+1), n-L \leq i < n$, as well as $w(e) = 1 + \varepsilon$ and $\text{Prob}(\eta(e) = 0) = 1$ for $e = (n-L, n)$. See Figure 1 for an illustration. Consider $p' = (n-L, \dots, n)$ and $p'' = (n-L, n)$. Hence, $w(p') = 1 < 1 + \varepsilon = w(p'')$ and $\text{Prob}(\tilde{w}(p') = 1 + \mu > 1 + \varepsilon = \tilde{w}(p'')) = 1$. When ant a_{n-L} follows p'' for the first time it replaces the current best-so-far path p_{n-L}^* and the approximation ratio remains $1 + \varepsilon$ forever.

Since deterministic noise is independent, we can use Theorem 3. In our example we have $\tilde{w}_{\max} = \mu/L$, hence the ant system finds an approximation within an additive error of $\frac{(L+1)^2}{2L} \cdot \mu$ in polynomial expected time. This corresponds to a multiplicative error of $1 + \frac{(L+1)^2}{2L} \cdot \mu$, in comparison to the lower bound $1 + \varepsilon$. Our upper bound is asymptotically tight if L is a constant. In general, as ε can be made arbitrarily close to μ the additive terms in the two bounds are basically off by a factor of around $L/2$. It is an open problem to determine whether this factor is really necessary or whether it is just an artifact of our proof.

This example also shows that the considered setting is too general since deterministic noise can transform any problem instance into any other instance with larger or equal weights. This illustrates the major obstacle on the way to stronger and more detailed results. For arbitrary noise, a guarantee on the noisy best-so-far path length may not contain much reliable information—there can be an arbitrarily large discrepancy between noisy and real path lengths. Better results on what approximation ratio is achievable within a certain expected time require further restrictions on the noise model or the considered instances.

In the following, we will consider gamma-distributed noise $\eta(e, p, t) \sim \Gamma(k, \theta)$ for all edges e . In this case for every path p it holds $E(\tilde{w}(p)) = (1 + k\theta) \cdot w(p)$, hence we are looking for least expected time (LET) paths.

4 The Gamma Distribution

Preparing for the upcoming analyses, we discuss the gamma distribution many of our results are based on. The gamma distribution has been introduced for modeling

stochastic travel times by Fan et al (2005). The motivation is due to the following observation. In collision-free traffic cars that arrive at a particular landmark are due to a Poisson process. As the gamma distribution models the time between events in a Poisson process, travel times follow a gamma distribution (Nikolova et al, 2006).

Note that we do not directly use a gamma-distributed random variable X as edge weight. Instead, we use $(1 + X) \cdot w(e)$ as noisy length for the edge e . The addition of 1 can be seen as incorporating a minimum travel time. We multiply by the real weight of the edge to make delays proportional to the length of an edge. Besides being motivated by physical models, the gamma distribution also has nice structural properties that make it well suited for a theoretical analysis.

The gamma distribution is parameterized by a shape parameter $k \in \mathbb{R}^+$ and a scale parameter $\theta \in \mathbb{R}^+$. A gamma-distributed random variable X is denoted by $X \sim \Gamma(k, \theta)$. Its probability density function is given by

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} & \text{if } x > 0 \end{cases}$$

where Γ denotes the gamma function $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$. For $k \in \mathbb{N}$, the probability density function simplifies to

$$f_X(x) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k (k-1)!}$$

for $x > 0$. In the special case $k = 1$, the gamma distribution equals the exponential distribution with mean θ . In general, with $k \in \mathbb{N}$ the gamma distribution reflects the sum of k such exponentially distributed random variables. This distribution is also known as Erlang distribution. The expectation of X is $E(X) = k\theta$; the variance is $k\theta^2$.

The cumulative distribution function for $k \in \mathbb{N}$ and $x \geq 0$ is

$$F_X(x) = \text{Prob}(X \leq x) = 1 - \sum_{i=0}^{k-1} \frac{(x/\theta)^i}{i!} e^{-x/\theta}.$$

The gamma distribution exhibits the following properties.

1. The sum of m independent gamma distributed random variables with the same scaling parameter is again gamma distributed: if for $1 \leq i \leq m$ we have independent variables $X_i \sim \Gamma(k_i, \theta)$ then

$$\sum_{i=1}^m X_i \sim \Gamma\left(\sum_{i=1}^m k_i, \theta\right).$$

2. Scaling a gamma-distributed random variable results in another gamma-distributed random variable: if $X \sim \Gamma(k, \theta)$ and $\alpha > 0$ then $\alpha X \sim \Gamma(k, \alpha\theta)$.

The ‘‘invariance’’ with respect to summation makes sense as we sum over sums of i. i. d. exponentially distributed variables. This property will prove useful for estimating the total noise on a path of edges with equal weights. The ‘‘invariance’’ with respect to scaling implies that the absolute added noise on an edge e , $w(e) \cdot \Gamma(k, \theta)$, is again a gamma-distributed random variable according to $\Gamma(k, w(e) \cdot \theta)$.

Considering the gamma distribution in noise models also makes sense because of the following. Assume we would try to reduce the (unweighted) noise $\Gamma(k, \theta)$ by sampling an

edge s times, for some integer s , and taking the average noise. The resulting distribution of the added noise is then again gamma-distributed according to $\Gamma(ks, \theta/s)$. This does not change the expected noise as $k\theta = ks \cdot \theta/s$, but it leads to a reduction of variance from $k\theta^2$ to $k\theta^2/s$. Note that these arguments also hold if the noisy length of an edge e is given by a function $w(e) + \Gamma(k, \theta)$ for a deterministic weight $w(e)$. One important conclusion is that results for arbitrary gamma-distributed noise also apply when this resampling strategy is employed. Conditions on the expected noise $k\theta$ apply for both settings in an equal manner.

The following lemma provides estimates for the cumulative distribution function of a gamma-distributed random variable and will prove useful later on. We present a self-contained proof as we did not find appropriate references in the literature. The proof technique used is well known.

Lemma 3 *Consider a gamma-distributed random variable X with $X \sim \Gamma(k, \theta)$ where $k \in \mathbb{N}$ and $\theta \in \mathbb{R}^+$. Then for every $x \in \mathbb{R}^+$*

$$\frac{(x/\theta)^k}{k!} \cdot e^{-x/\theta} \leq \text{Prob}(X \leq x) \leq \frac{(x/\theta)^k}{k!}.$$

Proof Consider the n -th order Taylor approximation to f at a

$$T_n(x) = \sum_{i=0}^n \frac{f^{(i)}(a)}{i!} \cdot (x-a)^i$$

where $f^{(i)}$ denote the i -th derivative of f . In general, we have $f(x) = T_n(x) + R_n(x)$ where

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x-a)^{n+1}$$

denotes the Lagrange form of the remainder term for some $\xi \in [-x/\theta, 0]$. In particular, putting $f := e^x$, $a := -x/\theta$, and $n := k-1$ we have

$$e^0 = \left[\sum_{i=0}^{k-1} \frac{e^{-x/\theta}}{i!} \cdot (0 - (-x/\theta))^i \right] + \left[\frac{e^\xi}{k!} \cdot (0 - (-x/\theta)^k) \right].$$

Hence, we conclude

$$\text{Prob}(X \leq x) = 1 - \sum_{i=0}^{k-1} \frac{(x/\theta)^i}{i!} \cdot e^{-x/\theta} = 1 - \left(e^0 - \frac{e^\xi}{k!} \cdot (x/\theta)^k \right)$$

for some $\xi \in [-x/\theta, 0]$. We derive

$$\frac{e^{-x/\theta}}{k!} \cdot (x/\theta)^k \leq \text{Prob}(X \leq x) \leq \frac{e^0}{k!} \cdot (x/\theta)^k.$$

□

We also prove the following tail bound. Again, the proof is self-contained. A similar tail bound can be found in Dubhashi and Panconesi (2009, Chapter 1).

Lemma 4 *Consider a gamma-distributed random variable X with $X \sim \Gamma(k, \theta)$ where $k \in \mathbb{N}$ and $\theta \in \mathbb{R}^+$. Then for every $x \in \mathbb{R}^+$ with $x \geq k\theta = \mathbb{E}(X)$*

$$\text{Prob}(X > x) \leq e^{-x/\theta} (x/\theta)^{k-1} \frac{k}{(k-1)!}.$$

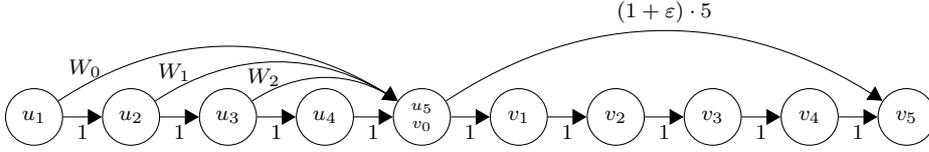


Fig. 2 Example graph $G_{n,k,\theta,\varepsilon}$ from Definition 1 for $n = 10$ with $W_i := 2n(2 + 2k\theta n)^{i+1}$.

Proof Using $x/\theta \geq k \geq 1$,

$$\begin{aligned} \sum_{i=0}^{k-1} \frac{(x/\theta)^i}{i!} &= \sum_{i=0}^{k-1} \frac{(x/\theta)^{k-1}}{i!(x/\theta)^{k-i-1}} \\ &\leq (x/\theta)^{k-1} \sum_{i=0}^{k-1} \frac{1}{i!k^{k-i-1}} \\ &\leq (x/\theta)^{k-1} \sum_{i=0}^{k-1} \frac{1}{(k-1)!} = (x/\theta)^{k-1} \frac{k}{(k-1)!}. \end{aligned}$$

Plugging this into the cumulative distribution function yields the claimed bound. \square

5 A Lower Bound for Independent Noise

In this section we will establish a lower bound on the random time until a good approximation is found. The result holds for a worst-case graph that will be described in the following (for an example, see Figure 2). First of all, the graph contains the following subgraph, referred to as *right part*. The right part has a unique source and a unique sink. There are two paths to traverse the right part: a chain of $n/2$ vertices and a single edge that directly connects the source with the sink. The edge weight for the long edge is chosen such that the edge is by a factor of $(1 + \varepsilon)$ more expensive than the whole chain.

The main observation now is as follows. When an ant traverses the chain, the noisy length of the chain will be concentrated around the expectation as the length of the chain is determined by the sum of many i. i. d. variables. Contrarily, the variance for the noisy length of the long edge will be quite large. This means that, although in expectation taking the chain leads to a shorter noisy path, chances are high that the ant experiences a small noise when taking the long edge. If the value of ε is not too large, this will result in the ant storing the long edge as best-so-far path and reinforcing it with pheromones. As long as this best-so-far path is maintained, the approximation ratio is no better than $1 + \varepsilon$.

To establish our lower bound, we will prove that after some time, with high probability the best-so-far path of the ant starting at the source of the right part will be the long edge. Furthermore we will show that the stored noisy length of the best-so-far path will be relatively small. This implies that the ant will not accept the chain as best-so-far path with high probability. This will establish the claimed lower bound on the approximation ratio.

However, the construction of the worst-case instance is not yet complete. As mentioned, the ant starting in the source of the right part needs some time in order to

traverse the long edge and sample it with a noise value that is significantly smaller than its expectation. In other words, we need some time to trick the ant into believing that the long edge leads to a shorter path. Therefore, we extend our graph by a subgraph where the ants will need a specific minimum amount of time in order to find a good approximation. This subgraph is prepended to the right part, so that it does not change the behavior of ants starting at the source of the right part. We give a formal definition for our graph. Figure 2 gives an illustration for $n = 10$.

Definition 1 Let $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\theta \in \mathbb{R}^+$, and $\varepsilon \in \mathbb{R}^+$. W.l.o.g. we assume that $n/2$ is an integer. Let $G_{n,k\theta,\varepsilon} = (V, E, w)$ with the vertices $V = \{u_1, \dots, u_{n/2} = v_0, v_1, \dots, v_{n/2}\}$ and the following edges. We have (u_i, u_{i+1}) , $1 \leq i \leq n/2 - 1$ and (v_i, v_{i+1}) , $0 \leq i \leq n/2 - 1$, with weight 1. Furthermore we have $(u_i, u_{n/2})$, $1 \leq i \leq n/2 - 2$, with weight $W_i := 2n(2 + 2k\theta n)^{i+1}$ and $(v_0, v_{n/2})$ with weight $(1 + \varepsilon) \cdot n/2$. The destination is given by the vertex $v_{n/2}$.

With regard to Figure 2, we refer to edges with weight W_i or $(1 + \varepsilon) \cdot n/2$ as upper edges and to the remaining edges as lower edges (or the lower path in the latter case). We also speak of a left part containing all vertices u_i and a right part containing all vertices v_i ($u_{n/2} = v_0$ belongs to both parts).

The first step for proving a lower bound is to show that, given enough time, the ant starting at v_0 will eventually favor the upper path. The condition $\varepsilon/(k\theta) = 1/e - \Omega(1)$ can be read as a trade-off between the expected noise strength $k\theta$ and the gap between the edge weights on the upper and lower paths, respectively. The larger the expected noise $k\theta$, the larger the gap can be.

Lemma 5 Consider the graph $G_{n,k\theta,\varepsilon} = (V, E, w)$ with independent noise $\eta(e) = \eta(e, p, t) \sim \Gamma(k, \theta)$ for each $e \in E$. Let $\varepsilon = O(1)$, $k = o(\log n)$, $k \in \mathbb{N}$, and $\varepsilon/(k\theta) = 1/e - \Omega(1)$. Choose the parameters $1/\text{poly}(n) \leq \tau_{\min} \leq 1/2$ and $\rho \geq 1/\text{poly}(n)$. Then with probability $1 - \exp(-\Omega(\sqrt{n}))$ after $\kappa n/\tau_{\min}$ iterations, we have a situation where for the ant starting from v_0 the following holds:

1. the ant's best-so-far path is the upper path and
2. the probability of changing the best-so-far path within a specific iteration towards the lower path is $\exp(-\Omega(n))$

where $\kappa > 0$ is an arbitrary constant.

Proof Consider the ant a_{v_0} and the edge $(v_0, v_{n/2})$. Since $\tau((v_0, v_{n/2})) \geq \tau_{\min}$, the probability of the ant following this edge is lower bounded by $\tau_{\min}/2$ due to Lemma 1. Consider the first $\kappa n/\tau_{\min}$ iterations and assume w.l.o.g. that this number is an integer. We count the number of times the ant follows the path. Define $X = \sum_{i=1}^{\kappa n/\tau_{\min}} X_i$ where X_i are independent Bernoulli-distributed random variables with $\text{Prob}(X_i = 1) = \tau_{\min}/2$ and $\text{Prob}(X_i = 0) = 1 - \tau_{\min}/2$. It is obvious that the real number of times the ant chooses the upper edge stochastically dominates X . We have $\mathbb{E}(X) = \kappa n/\tau_{\min} \cdot \tau_{\min}/2 = \kappa n/2$. Then the probability of following this path less than $\kappa n/4$ times is at most

$$\begin{aligned} \text{Prob}(X < \kappa n/4) &\leq \text{Prob}(X \leq (1 - 1/2) \cdot \mathbb{E}(X)) \\ &\leq e^{-(1/2)^2 \cdot \mathbb{E}(X)/2} = e^{-\kappa n/16} \end{aligned}$$

due to a well-known Chernoff bound. Hence, with probability at least $1 - \exp(-\Omega(n))$ the ant chooses the upper edge at least $\kappa n/4$ times during the first $\kappa n/\tau_{\min}$ iterations of the phase.

We now argue that there is a threshold b such that with high probability the ant finds a path of noisy length at most b when following the upper path. Contrarily, experiencing a noisy length at most b when following the lower path has probability $\exp(-\Omega(n))$. This proves both statements.

The noise on the upper path is distributed according to $U \sim \Gamma(k, (1 + \varepsilon) \cdot n/2 \cdot \theta)$ and the noise on the lower path is distributed according to $L \sim \Gamma(n/2 \cdot k, \theta)$. Hence, the noisy length of the upper path is $(1 + \varepsilon) \cdot n/2 + U_i$ and the noisy length of the lower path is $n/2 + L_i$ where U_i and L_i denote the i -th instantiation of U and L , respectively.

The precondition $\varepsilon/(k\theta) \leq 1/e - \Omega(1)$ implies the existence of a constant $c, 0 < c < 1$, such that $\varepsilon \leq k\theta \cdot \frac{2c-1}{e}$. We choose the threshold

$$b := \frac{n}{2} + \frac{n}{2} \cdot k\theta \cdot \frac{c}{e}.$$

First of all, consider $\text{Prob}((1 + \varepsilon) \cdot n/2 + \min\{U_i\} \leq b)$ with $1 \leq i \leq \kappa n/4$. In order for the above event to happen, we must have one instantiation U_i where $U_i \leq b - (1 + \varepsilon) \cdot n/2$ or, equivalently,

$$U_i \leq \frac{n}{2} \cdot \left(k\theta \cdot \frac{c}{e} - \varepsilon \right).$$

We estimate the probability of this event. Due to Lemma 3 and the inequality $n! \leq n^n$,

$$\begin{aligned} \text{Prob}\left(U_i \leq \frac{n}{2} \cdot \left(k\theta \cdot \frac{c}{e} - \varepsilon \right)\right) &\geq \frac{1}{k!} \cdot \left(\frac{n/2 \cdot (k\theta \cdot c/e - \varepsilon)}{(1 + \varepsilon)n/2 \cdot \theta} \right)^k \cdot e^{-\frac{n/2 \cdot (k\theta \cdot c/e - \varepsilon)}{(1 + \varepsilon)n/2 \cdot \theta}} \\ &= \frac{1}{k!} \cdot \left(\frac{k\theta \cdot c/e - \varepsilon}{(1 + \varepsilon)\theta} \right)^k \cdot e^{-\frac{k\theta \cdot c/e - \varepsilon}{(1 + \varepsilon)\theta}} \end{aligned}$$

and using $k\theta c/e - \varepsilon \geq k\theta(1 - c)/e$ we bound this as

$$\geq \frac{1}{k!} \cdot \left(\frac{k(1 - c)}{(1 + \varepsilon) \cdot e} \right)^k \cdot e^{-\frac{k\theta \cdot c/e - \varepsilon}{(1 + \varepsilon)\theta}}.$$

The e -term is bounded from below by $e^{-k/e}$ since

$$\frac{k\theta c/e - \varepsilon}{(1 + \varepsilon)\theta} \leq \frac{k\theta c/e}{\theta} \leq \frac{k}{e}.$$

Using $k! \cdot e^{k/e} \leq 2k^k$ for all $k \in \mathbb{N}$, we conclude

$$\text{Prob}\left(U_i \leq \frac{n}{2} \cdot \left(k\theta \cdot \frac{c}{e} - \varepsilon \right)\right) \geq \frac{1}{2} \cdot \left(\frac{1 - c}{(1 + \varepsilon) \cdot e} \right)^k \geq \frac{1}{2\sqrt{n}}$$

where the last inequality holds due to $k = o(\log n)$, if n is not too small.

Hence,

$$\begin{aligned} &\text{Prob}((1 + \varepsilon) \cdot (n/2) + \min\{U_i\} \leq b) \\ &= 1 - (1 - \text{Prob}(U_1 \leq b - (1 + \varepsilon) \cdot (n/2)))^{\kappa \cdot (n/4)} \\ &\geq 1 - \left(1 - \frac{1}{2\sqrt{n}}\right)^{\kappa \cdot (n/4)} \\ &= 1 - \exp(-\Omega(\sqrt{n})). \end{aligned}$$

The probability that an ant on the chain finds a path with costs greater than the threshold b is given by $\text{Prob}(n/2 + L_i > b)$. By Lemma 3 and $n! \geq (n/e)^n$,

$$\begin{aligned} \text{Prob}(n/2 + L_i \leq b) &= \text{Prob}(L_i \leq b - n/2) \\ &\leq \frac{1}{(n/2 \cdot k)!} \cdot \left(\frac{b - n/2}{\theta}\right)^{n/2 \cdot k} \\ &= \frac{(n/2 \cdot kc/e)^{n/2 \cdot k}}{(n/2 \cdot k)!} \\ &\leq \left(\frac{n/2 \cdot kc}{n/2 \cdot k}\right)^{n/2 \cdot k} \\ &= c^{n/2 \cdot k} = e^{-\Omega(n)}. \end{aligned}$$

This completes the proof. \square

The following lemma establishes a lower bound on the time until shortest paths are found for the left part of the graph. This will also give a lower bound for the time until a good approximation is reached for the left part and, as a consequence, also for the whole graph.

Lemma 6 *Consider the graph $G_{n,k,\theta,\varepsilon} = (V, E, w)$ with independent noise $\eta(e) = \eta(e, p, t) \sim \Gamma(k, \theta)$ for each $e \in E$. Let $\varepsilon \leq 1$ and $k \in \mathbb{N}$. Choose $1/\text{poly}(n) \leq \tau_{\min} \leq 1/2$ and $1/\text{poly}(n) \leq \rho \leq 1/2$. With probability $1 - \exp(-\Omega(\sqrt{n}/\log n))$ $\text{MMAS}_{\text{SDSP}}$ does not find a 2-approximation within $n/(6\tau_{\min}) + \sqrt{n} \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations.*

Proof In Horoba and Sudholt (2009) the expected optimization time of the ant system $\text{MMAS}_{\text{SDSP}}$ on a graph being very similar to the left part of G in a setting without noise is lower bounded by $\Omega(n/\tau_{\min} + n/(\rho \log(1/\rho)))$. We partly base our analysis on this proof and focus on the left part of G . A common trait for the graph in Horoba and Sudholt (2009) without noise and the left part of our graph with noise is that the shortest path through the subgraph is clearly to follow the chain of lower edges to vertex $u_{n/2}$. In addition, when starting from vertex u_i the second best choice is to take the edge $(u_i, u_{n/2})$ as the weights for the upper edges increase from left to right. As a consequence, many ants are tricked into placing pheromones on their first upper edges, which makes it hard for other ants to follow the chain of lower edges.

We first prove that $(u_i, u_{n/2})$ is the second best choice in the noisy setting, with overwhelming probability. This enables us to argue with basically the same ordering of paths as in the deterministic setting of our prior work. Then we prove that the ant system needs at least the claimed amount of time to subsequently find good approximate paths from right to left. The basic idea is that at the beginning pheromones are laid such that the probability of choosing an edge of the chain is $1/2$. For vertices whose ants have chosen the direct edge to $u_{n/2}$, this probability steadily drops over time, until it reaches the lower pheromone border τ_{\min} . In this situation, after a first phase of $\sqrt{n} \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations, there is still a linear number of vertices for which no good approximation has been found. The remaining time until these vertices become well approximated is at least $n/(6\tau_{\min})$ with the claimed probability.

On second-best paths: First observe that, by Lemma 4, if $\eta(e) \sim \Gamma(k, \theta)$ we have $\text{Prob}(\eta(e) \geq k\theta n) \leq e^{-kn} (kn)^{k-1} k / (k-1)! \leq e^{-\Omega(n)}$. Assuming that this event does not happen for any edge during the considered time period of $n/(6\tau_{\min}) + \sqrt{n}T^*$ with

$T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations only introduces an exponentially small error. Thus, with overwhelming probability we can assume for every path p

$$w(p) \leq \tilde{w}(p) < (1 + k\theta n)w(p) = \beta w(p)$$

where $\beta := 1 + k\theta n$. For $1 \leq i \leq n/2 - 2$ and $0 \leq j \leq n/2 - i - 2$ let $p_{i,j}$ denote the path that starts with u_i and follows j lower edges before taking the upper edge to $u_{n/2}$:

$$p_{i,j} := (u_i, u_{i+1}), \dots, (u_{i+j-1}, u_{i+j}), (u_{i+j}, u_{n/2}).$$

Recall that $(u_j, u_{n/2})$ has weight $W_j = 2n(2 + 2k\theta n)^{j+1} = 2n(2\beta)^{j+1}$. Also note that every path from $u_{n/2}$ to the destination has real length in between $n/2$ and $(1 + \varepsilon)n/2 \leq n$ as $\varepsilon \leq 1$. For every i and every j' with $j < j' \leq n/2 - i - 2$ and for every two paths p^*, p^{**} from $u_{n/2}$ to the destination we have (with \circ denoting the concatenation of paths)

$$\begin{aligned} \tilde{w}(p_{i,j} \circ p^*) &< \beta w(p_{i,j} \circ p^*) \\ &\leq \beta(j + 2n(2\beta)^{j+1} + n) \leq 2n(2\beta)^{j+2} \\ &\leq w(p_{i,j'} \circ p^{**}) \leq \tilde{w}(p_{i,j'} \circ p^{**}). \end{aligned}$$

Hence, an ant will always prefer $p_{i,j}$ over $p_{i,j'}$; in particular, $p_{i,0}$ is the second best path. Note that the second best path for each vertex in the left part has approximation ratio at least 2. Hence, with the claimed probability the ants on all vertices in the left part need to find shortest paths to $u_{n/2}$ in order to arrive at a 2-approximation.

Having established this relation, we can re-use some arguments from Theorem 4 in Horoba and Sudholt (2009). However, as the mentioned result only makes a statement about an expected time, more effort is needed to turn this into a high-probability result.

If $\tau_{\min} \geq 1/\sqrt{n}$ the following simple argument proves the claim. The ant at the leftmost vertex u_1 has to make a correct decision between two edges for $n/2 - 2$ times in order to find a 2-approximation. Even when the pheromones are best possible, τ_{\min} is still so large such that the probability of always choosing the right edge is at most $(1 - \tau_{\min}/2)^{n/2-2} = e^{-\Omega(\sqrt{n})}$. This proves the claim. In the remainder of the proof we hence assume $\tau_{\min} < 1/\sqrt{n}$.

As described before, we first investigate a first phase of $\sqrt{n} \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations, followed by a second phase of $n/(6\tau_{\min})$ steps.

Analysis of Phase 1: Call a vertex *wrong* if the best-so-far path starts with the upper edge. By Chernoff bounds, with probability $1 - \exp(-\Omega(n))$ we initially have at least $4/9 \cdot n/2$ wrong vertices. Also observe that for a wrong vertex the probability of taking the first edge of a shortest path is initialized with $1/2$ and it decreases over time towards τ_{\min} as long as the vertex remains wrong (cf. the upper bound on the probability from the first statement of Lemma 1). After T^* iterations, the border τ_{\min} is reached.

Call a vertex *optimized* if the corresponding ant has found a shortest path to $u_{n/2}$. Consider a vertex v with at least $8 \log(1/\rho) + \kappa\sqrt{n}/\log n$ wrong successors on its shortest path for some small constant $\kappa > 0$. In order to optimize v , the ant starting at v has to make the correct decisions for all wrong successors. Lemma 4 in Sudholt and Thyssen (2011) states that the probability of optimizing v within $1/\rho - 1 \geq 1/(2\rho)$ iterations is $\exp(-\Omega(\sqrt{n}/\log n))$. The intuitive reason is that even if successors of v become

optimized, the pheromones need some time to adapt (recall that the probability of choosing a correct edge is at most $1/2$). Note that $\text{poly}(n) \cdot \exp(-\Omega(\sqrt{n}/\log n))$ is still of order $\exp(-\Omega(\sqrt{n}/\log n))$ as we can decrease the hidden constant in the exponent to account for arbitrary polynomials $\text{poly}(n)$.

Taking the union bound for all vertices v , we have that during $1/(2\rho)$ iterations with probability $1 - \exp(-\Omega(\sqrt{n}/\log n))$ only $8 \log(1/\rho) + \kappa\sqrt{n}/\log n$ wrong vertices are corrected.

After $2\sqrt{n}T^*$ phases, each consisting of $1/(2\rho)$ iterations, with probability $1 - \exp(-\Omega(\sqrt{n}/\log n))$ at most $2\sqrt{n}T^*(8 \log(1/\rho) + \kappa\sqrt{n}/\log n) \leq 1/18 \cdot n/2$ wrong vertices have been corrected, where the inequality holds if κ is chosen small enough and n is large enough.

Analysis of Phase 2: Assume the described “typical” events in Phase 1 have happened. Then we are in a situation where there are still $4/9 \cdot n/2 - 1/18 \cdot n/2 = 7/18 \cdot n/2$ wrong vertices left. For these vertices the probability of taking the lower edge has decreased to τ_{\min} as pheromone has been decreased for more than T^* iterations. Now, if v has i wrong successors on its shortest path, the probability of optimizing v in the next iteration is at most τ_{\min}^i by the first statement of Lemma 1.

The following argument is borrowed from Theorem 17 in Droste, Jansen, and Wegener (2002). Imagine a 0-1-string of unbounded length where each bit is set to 1 independently with probability τ_{\min} . Then the random number of ones before the first zero follows the same geometric distribution as the number of optimized vertices (in fact, the former stochastically dominates the latter as the probabilities are not exactly τ_{\min}^i but smaller). The probability of optimizing $7/18 \cdot n/2$ wrong vertices in $n/(6\tau_{\min})$ iterations is thus bounded by the probability of having at least $7/18 \cdot n/2 = 7/36 \cdot n$ ones among the first $n/(6\tau_{\min}) + 7/36 \cdot n$ bits of the 0-1-string. Recall $\tau_{\min} \leq 1/\sqrt{n}$, hence the expected number of ones is $n/6 + O(n \cdot \tau_{\min}) = n/6 + O(\sqrt{n})$. This is by a constant factor smaller than $7/36 \cdot n$, if n is large enough. By a straightforward application of a Chernoff bound, the mentioned probability is $\exp(-\Omega(n))$.

Adding up all error probabilities, with probability $1 - \exp(-\Omega(\sqrt{n}/\log n))$ not all vertices have been optimized after $n/(6\tau_{\min}) + \sqrt{n}T^*$ iterations. \square

Note that we can easily achieve an arbitrarily bad approximation ratio in the specified time period by increasing all weights W_i by some arbitrarily large factor. In fact, we have proven the following theorem stating that we cannot hope to achieve any specific approximation ratio within less than $n/(6\tau_{\min}) + \sqrt{n} \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations.

Theorem 4 *For every $r \in \mathbb{R}^+$ there is a graph with n vertices and independent noise $\eta(e) = \eta(e, p, t) \sim \Gamma(k, \theta)$ for each edge e , $k \in \mathbb{N}$, such that with overwhelming probability MMAS_{SDSP} with $1/\text{poly}(n) \leq \rho \leq 1/2$ and $1/\text{poly}(n) \leq \tau_{\min} \leq 1/2$ does not find a $(1+r)$ -approximation within the first $n/(6\tau_{\min}) + \sqrt{n} \ln(\tau_{\max}/\tau_{\min})/\rho$ iterations.*

The results for MMAS_{SDSP} in Horoba and Sudholt (2009) state that the algorithm can indeed find all shortest paths when given a little more time: an upper bound of $O(n/\tau_{\min} + n \log(\tau_{\max}/\tau_{\min})/\rho)$ holds without noise. For stochastic shortest paths, the situation is much different. Putting Lemmas 5 and 6 on the different parts of G together, we obtain that the waiting time to obtain any approximation ratio better than $1 + \varepsilon$ is exponential.

Theorem 5 *Consider the graph $G_{n,k\theta,\varepsilon} = (V, E, w)$ with independent noise $\eta(e) = \eta(e, p, t) \sim \Gamma(k, \theta)$ for each $e \in E$. Let $\varepsilon \leq 1$, $k = o(\log n)$, $k \in \mathbb{N}$, and $\varepsilon/(k\theta) =$*

$1/e - \Omega(1)$. Choose the parameters $1/\text{poly}(n) \leq \tau_{\min} \leq 1/2$ and $1/\text{poly}(n) \leq \rho \leq 1/2$. Then with probability $1 - \exp(-\Omega(\sqrt{n}/\log n))$ $\text{MMAS}_{\text{SDSP}}$ does not achieve an approximation ratio better than $(1 + \varepsilon)$ within the first $e^{c\sqrt{n}}$ iterations where $c > 0$ is a small enough constant.

In particular, we get the following.

Corollary 1 *Under the conditions given in Theorem 5, with overwhelming probability $\text{MMAS}_{\text{SDSP}}$ does not find a $(1 + k\theta/3)$ -approximation within $e^{c\sqrt{n}}$ iterations.*

Proof (of Theorem 5) Let $\delta < \varepsilon$. According to Lemma 6 the algorithm does not find a $(1 + \delta)$ -approximation within the first $n/(6\tau_{\min}) + \sqrt{n}T^*$ iterations where $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$. By the first statement of Lemma 5, after this time period the ant starting at the source of the right part has stored the upper path as best-so-far path. Furthermore, the probability that this best-so-far path is changed in one iteration is at most $\exp(-\Omega(n))$ by the second statement of Lemma 5. Recall that the upper path has weight $(1 + \varepsilon) \cdot (n/2)$ in $G_{n,k\theta,\varepsilon}$ while the lower path has weight $n/2$. Also recall that a $(1 + \delta)$ -approximation requires all best-so-far paths to be $(1 + \delta)$ -approximations of the shortest paths. Hence, by the union bound the probability that the ant a_{v_0} achieves a $(1 + \delta)$ -approximation within e^{cn} iterations is still $\exp(-\Omega(n))$ if c is small enough. \square

The result from Theorem 5 is due to fact that the ant at v_0 cannot store a $(1 + \delta)$ -approximation as best-so-far path. It can easily construct a shortest real path, but it does not realize it as being a shortest path. Our negative results can easily be extended towards more relaxed notions of approximations that accept an approximate path if it is only found temporarily. Replace v_0 by a chain of $\Omega(n)$ vertices such that all edges leading to $u_{n/2} = v_0$ lead to the start of the chain instead and all edges leaving v_0 leave from the end of the chain. All edges on the chain receive a tiny weight. In this setting, all ants of the chain are basically in the same situation as the single ant on v_0 in the original graph. In order to achieve a $(1 + \delta)$ -approximation, all ants on the chain must choose to follow their shortest paths via $v_0, v_1, \dots, v_{n/2}$. The probability that in one iteration all ants decide to do so is exponentially small.

6 Perfectly Correlated Noise

In many stochastic optimization settings noise is not independent, but correlated. In this section we look at a setting that is opposed to independent noise: we assume the same noise for all edges, i. e., for each ant there is a single value $\eta = \eta(p, t)$ for each path p (i. e., for each ant) and each time t such that the noisy length of an edge e is given by $\tilde{w}(e) := (1 + \eta)w(e)$. The reader might think of ants reflecting traveling agents and each agent traveling at an individual speed. Formally, we may still think of gamma-distributed η -values for all edges: $\eta(e_1, p, t), \dots, \eta(e_m, p, t) \sim \Gamma(k, \theta)$ for $k \in \mathbb{N}$ and $\theta \in \mathbb{R}^+$, but they are all equal: $\eta(e_1, p, t) = \dots = \eta(e_m, p, t) = \eta(p, t) = \eta$. Note that this corresponds to a perfect correlation among the delays. The noisy length $\tilde{w}(p)$ of a path $p = (e_1, \dots, e_s)$ then equals

$$\tilde{w}(p) = \sum_{i=1}^s (1 + \eta(e_i)) \cdot w(e_i) = (1 + \eta) \cdot w(p).$$

When reconsidering the graph $G_{n,k\theta,\varepsilon}$ from Definition 1, we expect strongly correlated noise to be helpful as the noise values for the chain of lower edges in the right part are likely to show similar deviations from their expectation. This enables the ants to sample the lower path with small total noise. In fact, with perfectly correlated noise we prove that the ants indeed can observe the shorter paths and find shortest paths efficiently. The following theorem states an upper bound that depends on the noise parameters and the value of ε that determines the gap between the weights on the upper and lower paths in G , respectively.

Theorem 6 *Consider the graph $G_{n,k\theta,\varepsilon} = (V, E, w)$ with the same noise $\eta = \eta(p, t) \sim \Gamma(k, \theta)$ for each $e \in E$. Choose $\tau_{\min} \leq 1/(2n)$, then in expectation $\text{MMAS}_{\text{SDSP}}$ finds a 1-approximation after at most $O(n/\tau_{\min} + (k\theta/\varepsilon)^k \cdot e^{\varepsilon/\theta}/\tau_{\min} + n \log(\tau_{\max}/\tau_{\min})/\rho)$ iterations.*

Proof We first consider the right part and estimate the expected time until the ant at v_0 has sampled a shortest path with such a low noise that the noisy length of its best-so-far path is less than the real length of the sub-optimal path. By Lemma 1 the ant follows the optimal path with probability at least $\tau_{\min}/(2e)$. The noisy length $\tilde{w}(p)$ of the shortest path is $(1+X)n/2$ with $X \sim \Gamma(k, \theta)$. The real length of the upper path is $(1+\varepsilon)n/2$. By Lemma 3 we have

$$\begin{aligned} \text{Prob}((1+X)n/2 \leq (1+\varepsilon)n/2) &= \text{Prob}(X \leq \varepsilon) \\ &\geq \frac{(\varepsilon/\theta)^k}{k!} e^{-\varepsilon/\theta} \geq \left(\frac{\varepsilon}{k\theta}\right)^k e^{-\varepsilon/\theta}. \end{aligned}$$

The expected waiting time for following the shortest path and sampling it with low noise is $O(1/\tau_{\min} \cdot (k\theta/\varepsilon)^k \cdot e^{\varepsilon/\theta})$. After $T^* = \ln(\tau_{\max}/\tau_{\min})/\rho$ more steps we have $\tau((v_0, v_1)) = \tau_{\min}$ and this property is maintained forever. In particular, we have found a 1-approximation for all vertices in the right part.

A similar argument holds for the left part. We show that the ants in the left part find shortest paths from right to left. Consider the ant starting from $u_{n/2-i}$ with $1 \leq i \leq n/2$. Assume that all ants at $u_{n/2-i+1}, \dots, u_{n/2}$ have already found their shortest paths and sampled it with such a low noise that they will stick to their shortest paths forever. Also assume that the pheromones on their incorrect edges have decreased to τ_{\min} . Under these assumptions the ant at $u_{n/2-i}$ follows the optimal path with probability at least $\tau_{\min}/(2e)$. The noisy length of the shortest path from $u_{n/2-i}$ is $(1+X)(i+n/2) \leq (1+X)n$ with $X \sim \Gamma(k, \theta)$ and the noisy length $\tilde{w}(p)$ of a path p starting with an upper edge is larger than $2n(1+k\theta n) + n/2$. Using Lemma 4,

$$\begin{aligned} &\text{Prob}((1+X)n \leq 2n(1+k\theta n) + n/2) \\ &= \text{Prob}(X \leq 3/2 + 2k\theta n) \\ &\geq \text{Prob}(X \leq k\theta n) \geq 1 - \exp(-\Omega(n)). \end{aligned}$$

This yields an expected number of $O(1/\tau_{\min})$ steps until the desired property also holds for the best-so-far path from $u_{n/2-i}$. After T^* more steps we have $\tau((u_{n/2-i}, u_{n/2})) = \tau_{\min}$. Summing up all expected waiting times for the right part and for all vertices $u_{n/2-1}, \dots, u_1$ in the left part yields the claimed time bound. \square

In particular, we conclude the following special case.

Corollary 2 For $k = O(1)$ and $\varepsilon = \Omega(\theta/n)$ Theorem 6 yields a polynomial upper bound, while Theorem 5 yields an exponential lower bound for the expected running time of $\text{MMAS}_{\text{SDSP}}$.

This example demonstrates that correlated noise really helps the ants to locate shortest paths on the considered graph.

Finally, we give a general upper bound for perfectly correlated noise. We consider graphs with gaps between the weights of optimal and non-optimal paths, respectively. In contrast to the general upper bound from Theorem 1 we have a smoother trade-off between the gap size and the expected optimization time. This is expressed by the parameter z in the following statement.

Theorem 7 Consider a graph $G = (V, E, w)$ with weight $w(e) > 0$ and with the same noise $\eta = \eta(p, t) \sim \Gamma(k, \theta)$ for each $e \in E$. Choose $\tau_{\min} \leq 1/(\Delta L)$ and let opt_v denote the real length of a shortest path from v . If for each vertex $v \in V$ and some $z \in \mathbb{N}$ it holds that every non-optimal path has length at least $(1 + k\theta/z) \cdot \text{opt}_v$ then in expectation $\text{MMAS}_{\text{SDSP}}$ finds a 1-approximation after $O(((z+1)^k L \log n)/\tau_{\min} + L \log(\tau_{\max}/\tau_{\min})/\rho)$ iterations.

Proof The proof is very similar to the proof of Theorem 1. We call a vertex v optimized if $\tilde{w}(p_v^*) < (1 + \theta/z) \cdot \text{opt}_v$ and processed if all pheromones on edges that are not part of shortest paths have decreased to τ_{\min} . Call an iteration *good* if the global noise is less than $k\theta/z$. Define $\eta_1, \dots, \eta_k \sim \Gamma(1, \theta)$ and note that η has the same distribution as $\sum_{i=1}^k \eta_i$. In particular, we have $\eta < k\theta/z$ if $\eta_i < \theta/z$ for all i . The probability for a good iteration is at least

$$\text{Prob}\left(\eta < \frac{k\theta}{z}\right) \geq \prod_{i=1}^k \text{Prob}\left(\eta_i < \frac{\theta}{z}\right) \geq (1 - e^{-1/z})^k \geq \left(1 - \frac{1}{1+1/z}\right)^k = \frac{1}{(z+1)^k}.$$

By Chernoff bounds, the probability that in $(z+1)^k \cdot (c \log n)/\tau_{\min}$ iterations we have at least $2 \ln n \cdot 2e/\tau_{\min}$ good iterations is at least $1 - 1/n$, if $c > 0$ is a constant chosen large enough. In the following we assume that this happens.

Recall the partition V_0, \dots, V_L of V from Theorem 1. Assume that all vertices in V_0, \dots, V_{i-1} are processed and fix a vertex $u \in V_i$. The probability of finding a shortest path from u is at least $\tau_{\min}/(2e)$. The probability that this happens at least once in $2 \ln n \cdot 2e/\tau_{\min}$ good iterations is

$$1 - \left(1 - \frac{\tau_{\min}}{2e}\right)^{2 \ln n \cdot 2e/\tau_{\min}} \geq 1 - e^{-2 \ln n} = 1 - \frac{1}{n^2}.$$

So, with a failure probability of at most $1/n + |V_i| \cdot 1/n^2 \leq 2/n$ all vertices in V_i become optimized within $(z+1)^k \cdot (c \log n)/\tau_{\min}$ iterations. Repeating these arguments in case of a failure establishes a bound of $O((z+1)^k \cdot (\log n)/\tau_{\min})$ on the expected time until all vertices in V_i are optimized. Summing up all these times and adding terms $\ln(\tau_{\min}/\tau_{\min})/\rho$ for optimized vertices to become processed yields the claimed bound. \square

7 Conclusions

We have presented a first analysis of ACO for a stochastic combinatorial problem: finding shortest paths with noisy edge weights. Different settings of (nonnegative) noisy edge weights have been examined: general noise, independent noise, independent gamma-distributed noise, and perfectly correlated noise. We have characterized on which instances ACO can still find shortest paths in case there is a gap between the shortest-path length and the lengths of non-optimal paths. For general weights we have given general upper bounds on the approximation ratio that can be obtained efficiently.

For gamma-distributed noise we have constructed a setting where noise and the different variances for upper and lower paths trick the ants into following the path with the larger variance, but larger real length. The ants can be seen to become risk-seeking in this scenario. The expected time until a good approximation is found then becomes exponential. Another insight is that this effect vanishes when considering perfectly correlated noise. Many results have established trade-offs between the expected running times on the one hand and the expected noise strength or approximation guarantees on the other hand.

Future work could deal with other variants of stochastic shortest path problems. Also other noise functions could be investigated, including functions with negative noise or noise with moderate correlation. It is also not clear whether and in which settings it is possible to prove better general upper bounds that avoid the term $(L + 1)/2$ in the approximation ratio of Theorem 3 or the level-wise blow-up of the approximation ratio in Theorem 2, respectively. Finally, an open question is whether ACO shows similar degrees of robustness for other stochastic problems.

Acknowledgments

The authors thank Samitha Samaranyake and Sébastien Blandin from UC Berkeley for references and discussions on stochastic shortest path problems and Thomas Sauerwald for pointing us to Feige (2006). Dirk Sudholt was partly supported by EPSRC grant EP/D052785/1 and a postdoctoral fellowship from the German Academic Exchange Service while visiting the International Computer Science Institute in Berkeley, CA, USA.

References

- Abraham I, Fiat A, Goldberg AV, Werneck RFF (2010) Highway dimension, shortest paths, and provably efficient algorithms. In: Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10), SIAM Press, pp 782–793
- Attiratanasunthron N, Fakcharoenphol J (2008) A running time analysis of an ant colony optimization algorithm for shortest paths in directed acyclic graphs. *Information Processing Letters* 105(3):88–92
- Bast H, Funke S, Sanders P, Schultes D (2007) Fast routing in road networks with transit nodes. *Science* 316(5824):566
- Baswana S, Biswas S, Doerr B, Friedrich T, Kurur PP, Neumann F (2009) Computing single source shortest paths using single-objective fitness functions. In: Proceedings

-
- of the International Workshop on Foundations of Genetic Algorithms (FOGA '09), ACM Press, pp 59–66
- Bertsekas DP, Tsitsiklis JN (1991) An analysis of stochastic shortest path problems. *Mathematics of Operations Research* 16(3):580–595
- Borkar V, Das D (2009) A novel ACO algorithm for optimization via reinforcement and initial bias. *Swarm Intelligence* 3(1):3–34
- Boyan JA, Mitzenmacher M (2001) Improved results for route planning in stochastic transportation. In: *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '01)*, SIAM Press, pp 895–902
- Chan TM (2007) More algorithms for all-pairs shortest paths in weighted graphs. In: *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC '07)*, ACM Press, pp 590–598
- Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) *Introduction to Algorithms*, 2nd edn. MIT Press
- Doerr B, Johannsen D (2010) Edge-based representation beats vertex-based representation in shortest path problems. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '10)*, ACM Press, pp 759–766
- Doerr B, Theile M (2009) Improved analysis methods for crossover-based algorithms. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '09)*, ACM Press, pp 247–254
- Doerr B, Happ E, Klein C (2007) A tight analysis of the (1+1)-EA for the single source shortest path problem. In: *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '07)*, IEEE Press, pp 1890–1895
- Doerr B, Happ E, Klein C (2008) Crossover can provably be useful in evolutionary computation. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '08)*, ACM Press, pp 539–546
- Doerr B, Johannsen D, Kötzing T, Neumann F, Theile M (2010) More effective crossover operators for the all-pairs shortest path problem. In: *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN '10)*, Springer-Verlag, pp 184–193
- Doerr B, Neumann F, Sudholt D, Witt C (2011) Runtime analysis of the 1-ANT ant colony optimizer. *Theoretical Computer Science* 412(17):1629–1644
- Dorigo M, Blum C (2005) Ant colony optimization theory: A survey. *Theoretical Computer Science* 344(2–3):243–278
- Dorigo M, Gambardella LM (1997) Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation* 1(1):53–66
- Dorigo M, Stützle T (2004) *Ant Colony Optimization*, 1st edn. MIT Press
- Dorigo M, Maniezzo V, Colomi A (1991) The ant system: An autocatalytic optimizing process. Tech. Rep. 91-016 Revised, Politecnico di Milano
- Droste S, Jansen T, Wegener I (2002) On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science* 276(1–2):51–81
- Dubhashi D, Panconesi A (2009) *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press
- Fan YY, Kalaba RE, Moore JE (2005) Arriving on time. *Journal of Optimization Theory and Applications* 127(3):497–513
- Feige U (2006) On sums of independent random variables with unbounded variance and estimating the average degree in a graph. *SIAM Journal on Computing* 35(4):964–984

- Gutjahr WJ, Sebastiani G (2008) Runtime analysis of ant colony optimization with best-so-far reinforcement. *Methodology and Computing in Applied Probability* 10(3):409–433
- Horoba C (2010) Exploring the runtime of an evolutionary algorithm for the multi-objective shortest path problem. *Evolutionary Computation* 18(3):357–381
- Horoba C, Sudholt D (2009) Running time analysis of ACO systems for shortest path problems. In: *Proceedings of the International Workshop on Engineering Stochastic Local Search Algorithms (SLS '09)*, Springer-Verlag, pp 76–91
- Horoba C, Sudholt D (2010) Ant colony optimization for stochastic shortest path problems. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '10)*, Springer-Verlag, pp 1465–1472
- Kolavali SR, Bhatnagar S (2009) Ant colony optimization algorithms for shortest path problems. In: Altman E, Chaintreau A (eds) *Network Control and Optimization*, Springer-Verlag, pp 37–44
- Kötzing T, Lehre PK, Oliveto PS, Neumann F (2010a) Ant colony optimization and the minimum cut problem. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '10)*, ACM Press, pp 1393–1400
- Kötzing T, Neumann F, Röglin H, Witt C (2010b) Theoretical properties of two ACO approaches for the traveling salesman problem. In: *Proceedings of the International Conference on Ant Colony Optimization and Swarm Intelligence (ANTS '10)*, Springer-Verlag, pp 324–335
- Kötzing T, Neumann F, Sudholt D, Wagner M (2011) Simple Max-Min ant systems and the optimization of linear pseudo-Boolean functions. In: *Proceedings of the International Workshop on Foundations of Genetic Algorithms (FOGA '11)*, ACM Press, to appear
- Miller-Hooks ED, Mahmassani HS (2000) Least expected time paths in stochastic, time-varying transportation networks. *Transportation Science* 34(2):198–215
- Neumann F, Theile M (2010) How crossover speeds up evolutionary algorithms for the multi-criteria all-pairs-shortest-path problem. In: *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN '10)*, Springer-Verlag, pp 667–676
- Neumann F, Witt C (2008) Ant colony optimization and the minimum spanning tree problem. In: *Proceedings of the International Conference on Learning and Intelligent Optimization (LION '07)*, Springer-Verlag, pp 153–166
- Neumann F, Witt C (2009) Runtime analysis of a simple ant colony optimization algorithm. *Algorithmica* 54(2):243–255
- Neumann F, Sudholt D, Witt C (2008) Rigorous analyses for the combination of ant colony optimization and local search. In: *Proceedings of the International Conference on Ant Colony Optimization and Swarm Intelligence (ANTS '08)*, Springer-Verlag, pp 132–143
- Neumann F, Sudholt D, Witt C (2009) Analysis of different MMAS ACO algorithms on unimodal functions and plateaus. *Swarm Intelligence* 3(1):35–68
- Neumann F, Sudholt D, Witt C (2010) A few ants are enough: ACO with iteration-best update. In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '10)*, ACM Press, pp 63–70
- Nikolova E, Brand M, Karger DR (2006) Optimal route planning under uncertainty. In: *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS '06)*, AAAI Press, pp 131–141

-
- Orlin JB, Madduri K, Subramani K, Williamson M (2010) A faster algorithm for the single source shortest path problem with few distinct positive lengths. *Journal of Discrete Algorithms* 8(2):189–198
- Papadimitriou CH, Yannakakis M (1991) Shortest paths without a map. *Theoretical Computer Science* 84(1):127–150
- Scharnow J, Tinnefeld K, Wegener I (2004) The analysis of evolutionary algorithms on sorting and shortest paths problems. *Journal of Mathematical Modelling and Algorithms* 3(4):349–366
- Sudholt D (2011) Using Markov-chain mixing time estimates for the analysis of ant colony optimization. In: *Proceedings of the International Workshop on Foundations of Genetic Algorithms (FOGA '11)*, ACM Press, to appear
- Sudholt D, Thyssen C (2011) Running time analysis of ant colony optimization for shortest path problems. *Journal of Discrete Algorithms* <http://dx.doi.org/10.1016/j.jda.2011.06.002>, to appear
- Zhou Y (2009) Runtime analysis of an ant colony optimization algorithm for TSP instances. *IEEE Transactions on Evolutionary Computation* 13(5):1083–1092