View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by IntechOpen

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

4

Some Issues of ACO Algorithm Convergence

Lorenzo Carvelli[1] and Giovanni Sebastiani[1,2]
[1]*Mathematics department, Sapienza Università di Roma*
[2]*Istituto per le Applicazioni del Calcolo "Mauro Picone", CNR*
Italy

1. Introduction

The study of the convergence of ACO algorithms, or more in general of stochastic algorithms for solving combinatorial optimization problems is very important. In fact, it can provide information that can be useful in practice when applying such algorithms. This information can be of different kinds.
The most basic question of interest is about algorithm capability of solving the problem of interest. Given the stochastic nature of the kind of algorithms considered, this question can be properly formulated in terms of the "failure" probability, i.e. the probability that after the current iteration the algorithm has not yet found the solution. We are of course interested in ACO algorithms whose failure probability converges to zero. Another kind of convergence, that is stronger than simply having the failure probability to converge to zero, is when the whole ACO colony approaches, as time goes to infinity, a set of individuals all corresponding to one of the problem solutions. In addition to algorithm's effectiveness, other natural questions arise. In fact, the user is interested in the quantification of the time used by the algorithm to solve the problem. Since this time is random, the quantification will involve its expected value, its variance and ideally its distribution. Let us now go back to the failure probability. There are often situations where, by applying ACO algorithms, or more in general stochastic algorithms, to solve combinatorial optimization problems, the failure probability goes to zero too slowly. In those situations, one could ask the question if, instead of running the ACO algorithm for a certain time, it would be more convenient to stop it after another time T, smaller than the former, and to start it again from the beginning, and so on, until the original time is reached. In the case of a positive answer to this question, one could also study the problem of finding an optimal value for the time T.
In this chapter, we will illustrate some relevant known theoretical results on the former issues. Furthermore, we will provide some results on a our ingoing research on the last issue. Beside this, some numerical simulation results will also be introduced and discussed.

2. Algorithm's convergence

By definition, the failure probability is a non-increasing function of the number of iterations. The effectiveness of the algorithm can then be translated into the convergence to zero of the failure probability. Although this kind of convergence requirement is very basic, it is not always fullfilled. Therefore, theoretical studies on this kind of convergence are well motivated. Knowing in advance that the failure probability is decreasing to zero makes the user confident

that waiting long enough time, there are very good chances that the algorithm will solve the problem.
Let $X^i(t)$, $t = 1,2,\ldots$ be the stochastic process modeling the configuration assumed by the i-th of the A ants of the colony. Let X^* denote the optimal set for the function f to be maximized (or minimized).
We can distinguish two kinds of convergence:

convergence in value: when it holds

$$p_v(t) := P\left(\bigcap_{i=1}^{A}\bigcap_{k=1}^{t}\left\{X^i(k) \notin X^*\right\}\right) \to 0;$$

convergence in model: if for some $x^* \in X^*$ we have

$$p_m(x^*,t) := P\left(\bigcap_{i=1}^{A}\left\{X^i(t) = x^*\right\}\right) \to 1\,.$$

The convergence in value is important. This property tells us something about the way in which the algorithm is exploring the configuration space X. It is strongly connected to the strict positivity of the conditional probability to visit at the end of any iteration one point of X^* given that we are not currently in X^*. However, the convergence in model is stronger than the one in value. In the former, the probabilistic model itself evolves towards one that generates only optimal solutions. Not all algorithms converging in value are also converging in model. For example, this is the case for the algorithm that explores the configuration space in a uniform and independent way, known as Random Search (RS). In fact, we have $p_v(t) = \left(1 - \frac{|X^*|}{|X|}\right)^{At} \to 0$, while it holds $p_m(x^*,t) = \left(\frac{1}{|X|}\right)^{A}$.
In the following, we will only cope with the case where the ACO algorithm does not use the visibility matrix η_{ij}. In this case, the configuration of each ant of the colony is built by a path over the construction graph [Gutjahr (2000)]:

$$p(j|s^p) = \frac{\tau_{ij}^{\alpha}}{\sum_{j \in N(s^p)} \tau_{ij}^{\alpha}} \quad \forall\, j \in N(s^p)\,, \tag{1}$$

where (i,j) is the arc by which we continue the partial path s^p, τ_{ij} is its current pheromone value, and α is a positive parameter.
By simply imposing some constraints on the pheromone matrix, it is possible to have an ACO algorithm that converges in value. Indeed, for an ACO algorithm with $0 < \tau_{min} \leq \tau_{ij} \leq \tau_{max}$, given any $\epsilon > 0$, we have that, for t large enough it holds

$$p_v(t) \leq \epsilon\,,$$

that is, by definition,

$$\lim_{t\to\infty} p_v(t) = 0\,.$$

In fact, because of the bounds on the pheromone matrix, every choice made by the rule in Eq. (1) has a probability larger or equal to

$$p_{min} = \frac{\tau_{min}^{\alpha}}{(D_{max}-1)\tau_{max}^{\alpha} + \tau_{min}^{\alpha}} > 0$$

where D_{max}, is the maximum value of the degree of the nodes of the construction graph. It then follows that any configuration of the whole space X, including an optimal one x^*, can be visited at any iteration with a probability larger or equal than $\hat{p} = (p_{min})^{L_{max}} > 0$, where $L_{max} = max_{x \in X} L(x)$, and $L(x)$ is the length of the path by which we have built the configuration x. From this, it follows that

$$p_v(t) \leq (1 - \hat{p})^{At}, \tag{2}$$

i.e. there is convergence in value.
Different kinds of ACO algorithms are such that suitable bounds on the pheromone matrix values hold. Among them, the Max Min Ant System (MMAS), where lower and upper bounds on the pheromone matrix values are imposed explicitly when updating recursively them along iterations

$$\tau_{ij}(t+1) = \min\{\tau_{max}, \max\{(1-\rho)\tau_{ij}(t) + \rho 1\{(i,j) \in x_b(t)\}/L(x_b(t)), \tau_{min}\}\},$$

where $0 < \rho < 1$, and $1(\cdot)$ is the indicator function [Stützle & Hoos (1997)], [Stützle & Hoos (2000)]. We notice that the pheromone is reinforced only on arcs belonging to one configuration $x_b(t)$, e.g. the best one (w.r.t. the objective function f) that we have visited so far, i.e. $x_b(t) = \arg \max_{i=1,..,A,\, k=1,..,t} f(x^i(k))$. We could use another kind of update where no bounds on the pheromone are explicitely imposed:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho 1\{(i,j) \in x_b(t)\}/L(x_b(t)).$$

In this case, after t iterations, the pheromone values are bounded from above by

$$(1-\rho)^t \tau_0 + \rho \sum_{i=1}^{t} (1-\rho)^{t-i}/L_{min}.$$

The above quantity is bounded from above by $\tau_0 + 1/L_{min}$, where $L_{min} = min_{x \in X} L(x)$. In this case, we do not have in general any guarantee that also a lower bound holds. However, not having a lower bound sometimes can be a positive condition. In fact, as seen above, when we have both lower and upper bounds, convergence in value holds. When a lower bound holds, we cannot have convergence in model because at any iteration we have a lower bound for the conditional probability of reaching any configuration given any other. We will see now that if we impose the weaker condition that the lower bound of the pheromone matrix at time t, $\tau_{min}(t)$ (it always exists since there is a finite number of edges) goes to zero slowly enough, convergence in value still holds. This is stated by the following theorem [Dorigo & Stutzle (2004)].

Theorem 1. *Given an ACO algorithm with pheromone values having constant upper bound τ_{max} and lower bound $\tau_{min}(t)$*

$$\tau_{min}(t) = \Omega\left(\frac{1}{\ln(t+1)}\right),$$

then we have

$$p_v(t) \to 0.$$

In fact, similarly to Eq. (2), one can prove that

$$p_v(t) \leq \prod_{k=1}^{t} \left(1 - (p_{min}(k))^{L_{max}}\right)^A,$$

where

$$p_{min}(t) = \frac{\tau_{min}^{\alpha}(t)}{(D_{max}-1)\tau_{max}^{\alpha} + \tau_{min}^{\alpha}(t)} \geq \frac{\tau_{min}^{\alpha}(t)}{D_{max}\tau_{max}^{\alpha}}.$$

This implies that

$$p_v(t) \leq \prod_{k=1}^{t} \left(1 - \left(\frac{\tau_{min}^{\alpha}(k)}{D_{max}\tau_{max}^{\alpha}}\right)^{L_{max}}\right)^A = \prod_{k=1}^{t} \left(1 - K\,(\tau_{min}(k))^{\alpha L max}\right)^A. \quad (3)$$

Now, we prove that the infinite product on the final expression of Eq. (3) converges to zero. By the following lemma, it is sufficient to show that

$$\sum_{t=1}^{\infty} (\tau_{min}(t))^{\alpha L_{max}} \to +\infty.$$

Lemma 1. *Let* $\{a_n\}_{n \in IN}$ *be a sequence of real numbers converging to zero such that* $0 \leq a_n < 1$, $\forall n$. *Then, it follows that* $\sum_n a_n \to +\infty \Rightarrow \prod_n (1-a_n)^k \to 0$, $\forall k \geq 1$.

Since $\tau_{min}(t) = \Omega\left(\frac{1}{\ln(t+1)}\right)$, then the terms of the series $\sum_{t=1}^{\infty} (\tau_{min}(t))^{\alpha L_{max}}$ are asymptotically bounded from below by $\left(\frac{C}{\ln(t+1)}\right)^{\alpha L_{max}}$. Then, this series is diverging because $\sum_{t=1}^{\infty} \left(\frac{C}{\ln(t+1)}\right)^{\alpha L_{max}}$ is infinite. Finally, we have

$$\lim_{t \to \infty} p_v(t) = 0.$$

We will see now, that the variant of the MMAS algorithm such that τ_{max} is constant and $\tau_{min}(t) = \Omega\left(\frac{1}{\ln(t+1)}\right)$ converges in model. First, it is easy to see that, conditionally to the event $\{X_b(0) = x^* \in X^*\}$, the pheromone values of arcs that do not belong to x^* tends deterministically to zero [Dorigo & Stutzle (2004)]. In fact, $\forall t > 0$ we have $p(X_b(t) = x^* | X_b(0) = x^*) = 1$, and for any $(i,j) \notin x^*$, $\tau_{ij}(t) = \max\{\tau_{min}(t), (1-\rho)^t \tau_{ij}(0)\} \to 0$. Similarly, for any $(i,j) \in x^*$, $\tau_{ij}(t) \to 1/L(x^*)$.
We recall that $p_m(x^*,t)$ denotes the probability that each ant will build x^* at iteration t. A lower-bound for $p(x^*,t)$ is

$$\begin{aligned} p_m(x^*,t) &\geq p(X_b(\lfloor t/2 \rfloor) = x^*) \cdot p\left(\bigcap_{i=1}^{A} \{X^i(t) = x^*\} | X_b(\lfloor t/2 \rfloor) = x^*\right) \\ &= p(X_b(\lfloor t/2 \rfloor) = x^*) \cdot [p(X^1(t) = x^* | X_b(\lfloor t/2 \rfloor) = x^*)]^A. \end{aligned}$$

The last passage is justified because, conditionally to the event $\{X_b(\lfloor t/2 \rfloor) = x^*\}$ the pheromone evolution after time $\lfloor t/2 \rfloor$ will be deterministic and the A ants will evolve identically and independently from each other. Furthermore, it holds

$$\begin{aligned} p_m(x^*,t) &\geq p(X_b(\lfloor t/2 \rfloor) = x^*) \cdot \left[p(X^1(t - \lfloor t/2 \rfloor) = x^* | X_b(0) = x^*) \right]^A \\ &\geq p(X_b(\lfloor t/2 \rfloor) = x^*) \cdot \left(\frac{\min_{(i,j)\in x^*}(\tau_{ij}(t - \lfloor t/2 \rfloor))^\alpha}{\min_{(i,j)\in x^*}(\tau_{ij}(t - \lfloor t/2 \rfloor))^\alpha + \sum_{(i,h)\notin x^*}(\tau_{ih}(t - \lfloor t/2 \rfloor))^\alpha} \right)^{AL_{max}}, \end{aligned}$$

where $(i,j) \in x^*$. We notice that

$$\begin{aligned} & \lim_{t\to\infty} \frac{\min_{(i,j)\in x^*}(\tau_{ij}(t - \lfloor \frac{t}{2} \rfloor))^\alpha}{\min_{(i,j)\in x^*}(\tau_{ij}(t - \lfloor \frac{t}{2} \rfloor))^\alpha + \sum_{(i,h)\notin x^*}(\tau_{ih}(t - \lfloor \frac{t}{2} \rfloor))^\alpha} \\ = & \lim_{t\to\infty} \frac{\min_{(i,j)\in x^*}(\tau_{ij}(t))^\alpha}{\min_{(i,j)\in x^*}(\tau_{ij}(t))^\alpha + \sum_{(i,h)\notin x^*}(\tau_{ih}(t))^\alpha} \\ = & \frac{1}{1 + \sum_{(i,h)\notin x^*} \frac{\lim_{t\to\infty}(\tau_{ih}(t))^\alpha}{\min_{(i,j)\in x^*} \lim_{t\to\infty}(\tau_{ij}(t))^\alpha}} = 1, \end{aligned}$$

where we used the fact that $\lim_{t\to\infty}(\tau_{ij}(t))^\alpha > 0$ when $(i,j) \in x^*$. Since we also have that

$$\lim_{t\to\infty} p(X_b(\lfloor t/2 \rfloor) = x^*) = \lim_{t\to\infty} p(X_b(t) = x^*) = 1,$$

it finally follows that $\lim_{t\to\infty} p_m(x^*,t) = 1$, that is we have convergence in model.

3. Expected time to convergence

For combinatorial optimization problems, exaustive search can provide problem solution in a finite time, although this time grows tremendously with problem size. Therefore, it is natural to study the expected time needed for an algorithm to solve the problem as function of problem dimension. This problem can be studied for some classes of ACO algorithms, by using discrete time Markov chains on a countably infinite state space. We will describe now some general results which can be used to provide upper bounds for the expected time needed by a certain class of ACO algorithms to find an optimal solution [Gutjahr & Sebastiani (2008)]. To show their utility, we will also apply them to the MMAS algorithm [Gutjahr & Sebastiani (2008)]. The class of ACO algorithms involved are those such that the recursive rule for the pheromone only depends on the best-so-far-solution $x_b(t)$. This is the case for the MMAS algorithm. We also assume that, $x_b(t)$ is changed only when there is a strict increase of the objective function: $x_b(t+1) \neq x_b(t) \Leftrightarrow f(x_b(t+1)) > f(x_b(t))$.

Definition 1. *For any objective function $f(\cdot)$, we assume that the elements $f_1, ..., f_M$ of the finite set $\{f(x) | x \in X\}$ are ordered such that $f_1 < f_2 < ... < f_M$. We also define*

$$L_j = \{x \in S \,|\, f(x) = f_j\},$$

i.e. the level set *of index j, with $j = 1, ..., M$. The number M of level sets may vary from 1, in the case of a constant function, up to $|X|$ when there are no points in X with the same value of f.*

We notice, that if at time m we have that $x_b(m) \in L_j$ for some $j \in \{1, ..., M-1\}$, then $x_b(m')$ will be equal to $x_b(m)$ for $m' = m+1, ..., \bar{m}-1$ up to iteration $\bar{m}$ in which a solution with a

value of f higher than $x(m)$, so that $x_b(\bar{m}) \in L_k$ with $k > j$. Hence, at iteration m', the point $x_b(m')$ is identical to the one of the last time or it belongs to L_k with $k > j$. This means that transitions between $x_b(t) \in L_j$ and $x_b(t+1) \in L_k$ are allowed only if $k > j$.
Let us consider the homogeneous Markov process $X^t = (X_b(t), \tau(t))$, where τ denotes the pheromone matrix. Since the real numbers are approximated on a computer by a subset of the set $\mathcal{Q}$ of the rational numbers, it is not restrictive in practice to assume that this process takes values in the set $X \times \mathcal{Q}^{|\mathcal{A}|}$. This last set can be partitoned as

$$X \times \mathcal{Q}^{|A|} = \bigcup_{k=1}^{M} L_k \times \mathcal{Q}^{|A|} = \bigcup_{k=1}^{M} (L_k \times \mathcal{Q}^{|A|}) = \bigcup_{k=1}^{M} Y_k,$$

where $Y_k = L_k \times \mathcal{Q}^{|A|}$. Analogously as before, transitions between $Y(t) \in Y_j$ and $Y(t+1) \in Y_k$ are allowed only if $k > j$.
We can now introduce two lemmas [Gutjahr & Sebastiani (2008)].

Lemma 2. *Let $X^0, X^1, \ldots$ be a homogeneous Markov process on a countably infinite state space Y, with partition $Y_1, \ldots, Y_M$ each of which is also countably infinite, such that, $\forall k \neq j$, we have*

$$p(X^{t+1} = y | X^t = x) > 0,\ \forall y \in Y_k,\ \forall x \in Y_j \ \Leftrightarrow\ k > j,$$

and let $E[T_{x \to Y_M}]$ be the expected time to reach the set Y_M, that corresponds to the optimal set, starting from $x \notin Y_M$, that is

$$E[T_{x \to Y_M}] := \sum_{t=1}^{\infty} t p(X^t \in Y_M, X^s \notin Y_M, 1 \leq s \leq t-1 | X^0 = x).$$

Then, for $x \in Y_j$ and $j = 1, \ldots, M-1$, it follows that

$$E[T_{x \to Y_M}] \leq E[T_{x \to \bar{Y}_j}] + \sum_{k=j+1}^{M-1} \sum_{l=1}^{\infty} E[T_{x_{l,k} \to Y_M}] p(x \to x_{l,k}),$$

where $\bigcup_{l=1}^{\infty} \{x_{l,k}\} = Y_k$,

$$E[T_{x \to \bar{Y}_j}] := \sum_{t=1}^{\infty} t p(X^t \notin Y_j, X^s \in Y_j, 1 \leq s \leq t-1 | X^0 = x),$$

$$p(x \to x_{l,k}) := \sum_{t=1}^{\infty} p(X^t = x_{l,k}, X^s \in Y_j, 1 \leq s \leq t-1 | X^0 = x).$$

Lemma 3. *Under the same hypothesis of the last lemma, it follows that the expected value of the time T_M to reach the set Y_M, that is*

$$E[T_M] := \sum_{t=1}^{\infty} t p(X^t \in Y_M, X^s \notin Y_M, 0 \leq s \leq t-1),$$

is bounded from above by

$$E[T_M] \leq \sum_{j=1}^{M-1} p(X^0 \in Y_j) \sum_{k=j}^{M-1} \sup_{y \in Y_k} \{E[T_{y \to \bar{Y}_k}]\} \leq \sum_{k=1}^{M-1} \sup_{y \in Y_k} \{E[T_{y \to \bar{Y}_k}]\}.$$

We will now apply the last lemma to the MMAS algorithm [Gutjahr & Sebastiani (2008)]. It fullfills the hypothesis of this lemma. In fact, because of the pheromone bounds, transitions between different $x_b(t)$ at any two consecutive times have a strictly positive probability iff the index of the set Y_j increases. To this aim, we first show a useful property of the algorithm MMAS. If $x_b(0) = y \in L_k$ and $\tau_{max} < 1$, then exists a deterministic time t^*, independent from the initial pheromone, such that, for any time $t > t^*$ where $x_b(t)$ is unchanged, the pheromone $\tau(t)$ is constant and equal to $\tau(t^*) = \tau^*(y)$, where $\tau^*_{ij}(y) = \tau_{max}$ for any arc (i,j) belongs to y, and $\tau^*_{ij}(y) = \tau_{min}$ otherwise. The time t^* is given by $t^* = \max\{t_1, t_2\}$, where

$$t_1 = \left\lceil \frac{\log \tau_{min} - \log \tau_{max}}{\log(1-\rho)} \right\rceil,$$

and

$$t_2 = \left\lceil \frac{\log(1-\tau_{max}) - \log(1-\tau_{min})}{\log(1-\rho)} \right\rceil.$$

Based on this property, it is possible to prove the following lemma.

Lemma 4. *For the algorithms MMAS and for $\tau_{max} < 1$ we have that*

$$\sup_{y \in Y_k} \{E[T_{y \to \bar{Y}_k}]\} \leq t^* + 1/\bar{p}_k,$$

where $\bar{p}_k$ it is the smallest probability to get out from Y_k after the pheromone became constant

$$\bar{p}_k := \inf_{y \in L_k} p(X^1 \notin Y_k | X^0 = (y, \tau^*(y))),$$

where $1 \leq k < M$.

Finally, using the last two lemmas, we have that the expected value of T_M to reach the set Y_M is bounded from above

$$E[T_M] \leq \sum_{k=1}^{M-1} (t^*_k + 1/\bar{p}_k) = \sum_{k=1}^{M-1} t^*_k + \sum_{k=1}^{M-1} 1/\bar{p}_k.$$

The last lemma has been used to provide upper bounds of T_M for the MMAS when used to optimized some simple pseudo-boolean functions [Gutjahr & Sebastiani (2008)]. These bounds provided information useful to choose among different alternatives for the parameter ρ as function of the length of the boolean strings.
In another work, a theoretical study was performed on the time of convergence of two variants of the MMAS algorithms when maximizing some pseudo-Boolean functions [Neumann et al. (2009)]. The study is asymptotical with respect to the number of binary variables. The two algorithms differ to each other only because of the updating of the best-so-far solution, depending whether the function value is either increased or not decreased. Moreover, some lower bounds for the expected time to convergence are provided.

4. Restarting the algorithm

Let X_t be the stochastic process describing a given algorithm to find an optimal point of a function f (w.l.g. we deal with the case of maximization). We consider now a new algorithm

which consists of the former one restarted from beginning every T iterations. Let $\tilde{X}_t$ be the process that describes the new algorithm. Before of providing some results on the expected time to convergence, we introduce two examples. In particular, in the first example, due to the sub-exponential decay to zero of the failure probability $p_v(t)$, the algorithm with restart is successful. Instead, in the second example the failure probability goes to zero at exponential rate or even faster. Then, the restart is not convenient.

Example 1. If $p_v(t) = \frac{c}{t^\alpha}$, then, for any t sufficiently long, there exists $T < t$ such that the failure probability $p_v(t)$ of the algorithm after t iterations is larger than $p_v(T)^{\left\lfloor \frac{t}{T} \right\rfloor}$, that is the failure probability of the algorithm restarted $\left\lfloor \frac{t}{T} \right\rfloor$ times with restart time equal to T. To this aim, it will be sufficient to prove that

$$p_v(t) > p_v(T)^{\left(\frac{t}{T}-1\right)}.$$

To show this, we compute the derivate of $p_v(T)^{\frac{1}{T}}$:

$$\begin{aligned}\frac{d}{dT}\left(p_v(T)^{\frac{1}{T}}\right) &= \left(\frac{c}{T^\alpha}\right)^{\frac{1}{T}}\left[-\frac{1}{T^2}\ln\left(\frac{c}{T^\alpha}\right)+\frac{1}{T}\left(-\alpha\frac{cT^\alpha}{T^{\alpha+1}c}\right)\right] \\ &= -\left(\frac{c}{T^\alpha}\right)^{\frac{1}{T}}\frac{1}{T^2}\left(\ln\left(\frac{c}{T^\alpha}\right)+\alpha\right).\end{aligned}$$

This derivate vanishes when $\ln\left(\frac{c}{T^\alpha}\right)+\alpha = 0$ and hence for $T = ec^{\frac{1}{\alpha}}$. Of course we will choose a restart time equal to $\bar{T} = \left\lceil ec^{\frac{1}{\alpha}} \right\rceil = \beta ec^{\frac{1}{\alpha}}$, where $\beta \geq 1$. After having calculated $p_v(\bar{T}) = \frac{1}{(\beta e)^\alpha}$, we consider $p_v(\bar{T})^{t/\bar{T}-1}$:

$$p_v(T)^{t/\bar{T}-1} = \left(\frac{1}{(\beta e)^\alpha}\right)^{t/\beta e\sqrt[\alpha]{c}-1} < \frac{c}{t^\alpha} = p_v(t),$$

where the last inequality is true for t sufficiently great. Therefore, in this case there is an advantage to consider the process with restart.

Example 2. An example for which we have instead $p_v(T)^{\left\lfloor \frac{t}{T} \right\rfloor} \geq p_v(t)\ \forall T \leq t$ is when $p_v(t) = c^{t^\alpha}$ with $c < 1$ and $\alpha \geq 1$. In this case, it holds

$$p_v(T)^{\left\lfloor \frac{t}{T} \right\rfloor} \geq p_v(T)^{\frac{t}{T}} = c^{T^{\alpha-1}t} \geq c^{t^\alpha} = p_v(t) \quad \forall\, T \leq t.$$

We will now study the restart algorithm in terms of the expected value of the first hitting time T_C of the process $\tilde{X}_t$ into the optimal set X^* of the function f. If each state of X^* is "absorbent" (i.e. $P(X_t \in X^* \mid X_s \in X^*) = 1$ for any s,t such that $s < t$), then it is easy to see that the expected value of T_C is

$$E[T_C] = \sum_{k=1}^{\infty} P(X_k \notin X^*).$$

In fact, for any non-negative random variable X it's possible to write

$$E[X] = \int_0^\infty P(X > t)\, dt.$$

In our case, the random variable T_C is discrete and the integral in the above equation will be replaced by a series whose generic term is $P(T_C > t)$. We remark that the event $T_C > t$ is equal to the event $X_t \notin X^*$. In fact if the first hitting time will be greater than t, then $X_t \notin X^*$. Viceversa, since the points in X^* are absorbent, if $X_t \notin X^*$, then X_s for $s < t$ does not belong to X^* and therefore the first hitting time will be greater than t.
Let T_R first hitting time of $\tilde{X}_t$ into X^*, which can written as

$$E[T_R] = \sum_{k=1}^{\infty} P(X_T \notin X^*)^{\left\lfloor \frac{k-1}{T} \right\rfloor} P(X_{k-\left\lfloor \frac{k-1}{T} \right\rfloor T} \notin X^*).$$

In general, we notice that a necessary condition for $E[T_C]$ to be finite is that $P(X_t \in X^*)$ is infinitesimal. Instead, the expected value of T_R is always finite because

$$E[T_R] \leq \sum_{k=1}^{\infty} P(X_T \notin X^*)^{\frac{k-1}{T}-1},$$

and the series is convergent. Below, there is an example of a case where $P(X_t \in X^*)$ is not going to zero as t goes to infinity. Let us focus now on the algorithm known as (1+1)EA with single flip proposed for maximizing pseudo-boolean functions [Gutjahr & Sebastiani (2008)]. This algorithm, consists at each iteration first of the flip of a single component, randomly chosen. Then, the proposed flip is accepted if it corresponds to a non-decrease of f. For this algorithm, the probability $P(X_k \notin X^*)$ has a positive lower-bound in the case when there are more than one local maximum of the objective function f. With local maximum we mean the existence of a point $\bar{x}$ and a set $R(\bar{x})$ containing it such that

$$f(\bar{x}) \geq f(x) \qquad \forall x \in R(\bar{x}), \qquad\qquad f(\bar{x}) > f(x) \qquad \forall x \in \partial R(\bar{x}),$$

where $\partial R(\bar{x})$ is the border of the set $R(\bar{x})$. This algorithm can be modeled through a Markov chain such that $P(X_{t+1} = y \mid X_t = x) > 0$ if and only if $f(y) \geq f(x)$. We consider a function f with at least two local maxima and only one global maximum. If the initial probability to be in the interior $\bar{R}$ of one of the regions not corresponding to the global maximum is positive then

$$P(X_k \notin X^*) \geq P(X_k \notin X^* \mid X_0 \in \bar{R})\, P(X_0 \in \bar{R}) > 0.$$

In the last inequality we have exploited the fact that $P(X_k \notin X^* \mid X_0 \in \bar{R}) = 1$. In fact, the considered algorithm only allows transitions towards points at Hamming distance equal to 1 and with values of the objective not decreased. Therefore, if the initial point belongs to the region $\bar{R}$, it will never go outside it. Since the generic term of the series that gives $E[T_C]$ is not infinitesimal, the series diverges and $E[T_C] = \infty$.
A sufficient condition to have $E[T_R] < E[T_C]$ is obviuosly

$$P(X_T \notin X^*)^{\left\lfloor \frac{k-1}{T} \right\rfloor} P(X_{k-\left\lfloor \frac{k-1}{T} \right\rfloor T} \notin X^*) < P(X_k \notin X^*). \tag{4}$$

for any $t > T$. We notice that, the above condition becomes an identity for $t \leq T$. In the following, we provide different kinds of sufficient conditions for Eq. (4), that therefore also imply $E[T_R] < E[T_C]$.

Proposition 1. *Let $\{X_t\}_{t \in IN}$ be a process such that it exists $T > 0$ and for $t > T$ it holds*
$P(X_{t+1} \notin X^* \mid X_t \notin X^*) > P(X_T \notin X^*)^{\frac{1}{T}}$,
where each point of X^* is an absorbent state for the process X_t,

then

$$E[T_R] < E[T_C].$$

Proof.

$$\begin{aligned} P(X_t \notin X^*) &= P(X_t \notin X^* \mid X_{t-1} \notin X^*)\, P(X_{t-1} \notin X^*) \\ &> P(X_T \notin X^*)^{\frac{1}{T}}\, P(X_{t-1} \notin X^*) \\ &> \left(P(X_T \notin X^*)^{\frac{1}{T}}\right)^{\left\lfloor \frac{t-1}{T} \right\rfloor T} P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*). \end{aligned}$$

Hence, we have

$$P(X_t \notin X^*) > P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*) P(X_T \notin X^*)^{\left\lfloor \frac{t-1}{T} \right\rfloor},$$

that is (4). □

Proposition 2. *Let $\{X_t\}_{t \in IN}$ be a process such that it exists $T > 0$ for which it holds*

$$\frac{P(X_{t+1} \notin X^*)}{P(X_t \notin X^*)} > \frac{P(X_{t+1-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*)}{P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*)} \qquad \forall t > T, t \neq mT \quad \forall m \in I\!N \tag{5}$$

$$\frac{P(X_{mT+1} \notin X^*)}{P(X_{mT} \notin X^*)} > P(X_1 \notin X^*), \tag{6}$$

where each point of X^* is an absorbent state for the process X_t,
then

$$E[T_R] < E[T_C].$$

Proof. To obtain the thesis, it is sufficient to show that

$$P(X_T \notin X^*)^{\left\lfloor \frac{t-1}{T} \right\rfloor} P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*) < P(X_t \notin X^*) \qquad \forall t > T. \tag{7}$$

We proceed by induction on t distinguishing two cases: the first one where $\left\lfloor \frac{t}{T} \right\rfloor = \left\lfloor \frac{t-1}{T} \right\rfloor$ and the second where $\left\lfloor \frac{t}{T} \right\rfloor = \left\lfloor \frac{t-1}{T} \right\rfloor + 1$. The inequality (7) it's true for $t = T + 1$ because it is a particular case of inequality (6). We assume the inequality (7) to be true for t and we show that it is true also for $t + 1$. We analyze first the case where $\left\lfloor \frac{t}{T} \right\rfloor = \left\lfloor \frac{t-1}{T} \right\rfloor$. We have

$$\begin{aligned} P(X_T \notin X^*)^{\left\lfloor \frac{t}{T} \right\rfloor} P(X_{t+1-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*) &= P(X_T \notin X^*)^{\left\lfloor \frac{t-1}{T} \right\rfloor} \frac{P(X_{t+1-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*)}{P(X_{t-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*)} P(X_{t-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*) \\ &< P(X_t \notin X^*) \frac{P(X_{t+1-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*)}{P(X_{t-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*)} \leq P(X_{t+1} \notin X^*). \end{aligned}$$

In the case where $\left\lfloor \frac{t}{T} \right\rfloor = \left\lfloor \frac{t-1}{T} \right\rfloor + 1$ we have

$$\begin{aligned} P(X_T \notin X^*)^{\left\lfloor \frac{t}{T} \right\rfloor} P(X_{t+1-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*) &= P(X_T \notin X^*) P(X_T \notin X^*)^{\left\lfloor \frac{t-1}{T} \right\rfloor} P(X_{t+1-\left\lfloor \frac{t}{T} \right\rfloor T} \notin X^*) \\ &= P(X_T \notin X^*) P(X_T \notin X^*)^{m-1} P(X_1 \notin X^*) \\ &< P(X_{mT} \notin X^*) P(X_1 \notin X^*) \\ &< P(X_{mT+1} \notin X^*) = P(X_{t+1} \notin X^*). \end{aligned}$$

For the last inequalities we have used that in the considered case $t = mT$ with $m \in N$. □

We remark that, since X^* is absorbent for the chain $\{X_t\}_{t\in IN}$ then $\{X_{t+1} \notin X^*\} \subset \{X_t \notin X^*\}$ so that $\{X_{t+1} \notin X^*\} \cap \{X_t \notin X^*\} = \{X_{t+1} \notin X^*\}$. Hence, the left handside of Eq. (5) is equal to $P(X_{t+1} \notin X^* \mid X_t \notin X^*)$, and similarly for the right handside and for Eq. (6).

Proposition 3. *Let $\{X_t\}_{t\in IN}$ be a process such that it exists $T > 0$ for which it holds*

$$\forall m \qquad P(X_1 \notin X^*)P(X_T \notin X^*)^m < P(X_{(m+1)T} \notin X^*),$$

where each point of X^* is an absorbent state for the process X_t,
then

$$P(X_T \notin X^*)^{\left[\frac{t-1}{T}\right]} P(X_{t-\left[\frac{t-1}{T}\right]T} \notin X^*) < P(X_t \notin X^*).$$

Proof. Let us consider the function $\forall t \to m(t) \mid t \in \{m(t)T+1, \ldots, (m(t)+1)T\}$. Then, we have

$$\begin{aligned} P(X_T \notin X^*)^{\left\lfloor \frac{t-1}{T} \right\rfloor} P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*) &= P(X_T \notin X^*)^{m(t)} P(X_{t-\left\lfloor \frac{t-1}{T} \right\rfloor T} \notin X^*) \\ &\leq P(X_T \notin X^*)^{m(t)} P(X_1 \notin X^*) < P(X_{(m(t)+1)T} \notin X^*) \\ &\leq P(X_t \notin X^*). \end{aligned}$$

The last inequality follows from the non-increasing property of $P(X_t \notin X^*)$. □

We will illustrate now a simulation study where restarting an ACO algorithm is successful. We want to maximize the following pseudo-Boolean function

$$f(x) = \left| \sum_{i=1}^{N} x_i - \frac{N-1}{2} \right|, \tag{8}$$

with respect to all binary strings of length N. In Fig. 1, the function considered is plotted as function of the number of 1s in the case of $N = 20$.

This function has two local maxima but only one of them is a global maximum. The ACO algorithm considered is the MMAS. The presence of the pheromone bounds τ_{min} and $\tau_{\max}$ ensures convergence in value of the MMAS algorithm. However, if the algorithm visits a configuration with few 1s it takes a very long time in average to move towards the global maximum. Therefore, we expect that in this case the restart will be successfull.
By using as construction graph, the chain graph [Gutjahr (2000)], and by setting the initial values of the pheromone matrix equal to 0.5, the pheromone matrix coincides with the matrix of the probability of transitions [Gutjahr & Sebastiani (2008)]. The initial string was chosen uniformly in the state space $\{0,1\}^N$. The algorithms were implemented in Matlab. The values of the MMAS parameters were $\rho = 0.01$, $\tau_{min} = 0.3$, $\tau_{max} = 0.7$. We used one thousands runs of the algorithm each with 20000 iterations. Based on these simulations, we estimated the failure probability.
As suggested by Example 1, to find a good value of the restart time, we have computed the power $1/t$ of the estimated failure probability and then we have minimized it. In Fig. 2, the function $\hat{p}_v(t)^{\frac{1}{t}}$ is plotted. A minimum of this function is clearly visible at iteration 2900 ca.

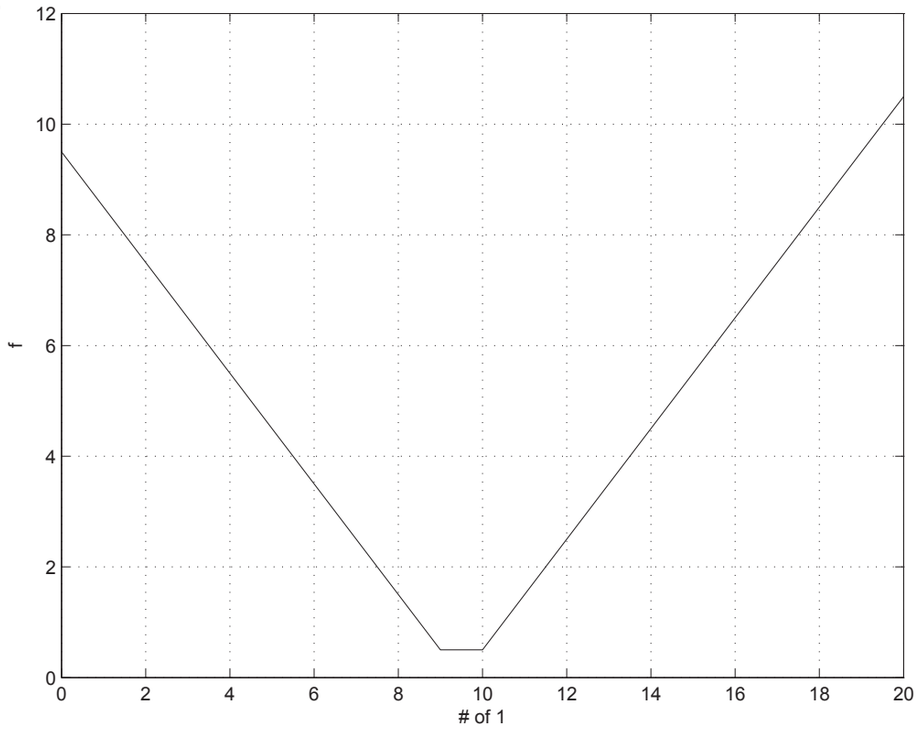


Fig. 1. The plot of the considered pseudo-Boolean function versus the number of 1s of the binary string

Finally, in Fig. 3 we show the estimated failure probability $\hat{p}_v(t)$ for the MMAS algorithm with chain graph to maximize the pseudo-binary function of Fig. 1 with $N = 20$ (continuous line). On the same figure, the failure probability of the restart algorithm with $T = 2900$ ca. is plotted (dashed line). As seen in the last figure, there is a clear advantage to use the restart MMAS algorithm when compared to the standard MMAS.

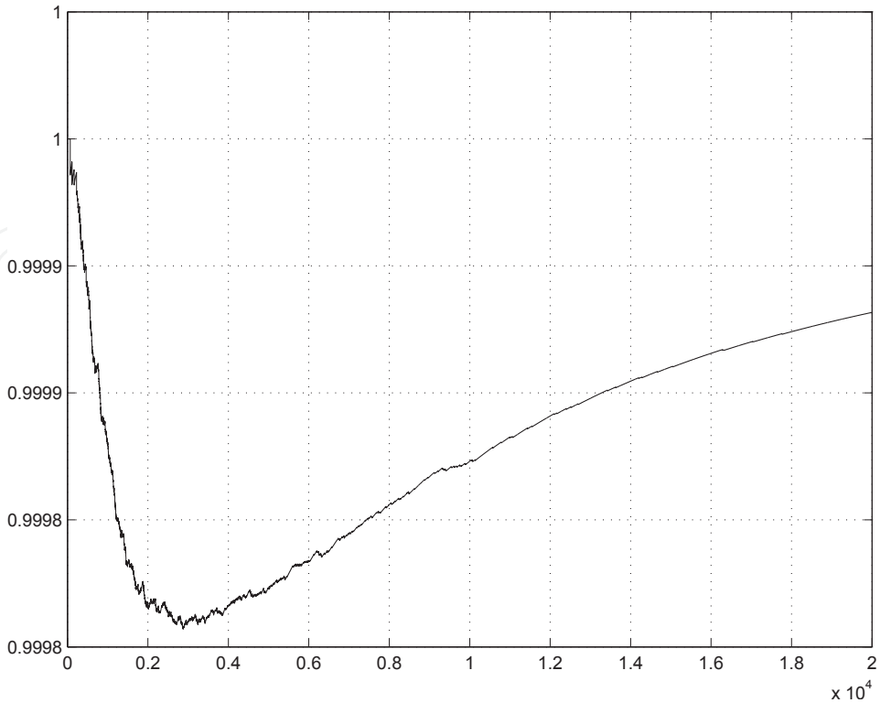


Fig. 2. The estimated failure probability raised to the power $1/t$

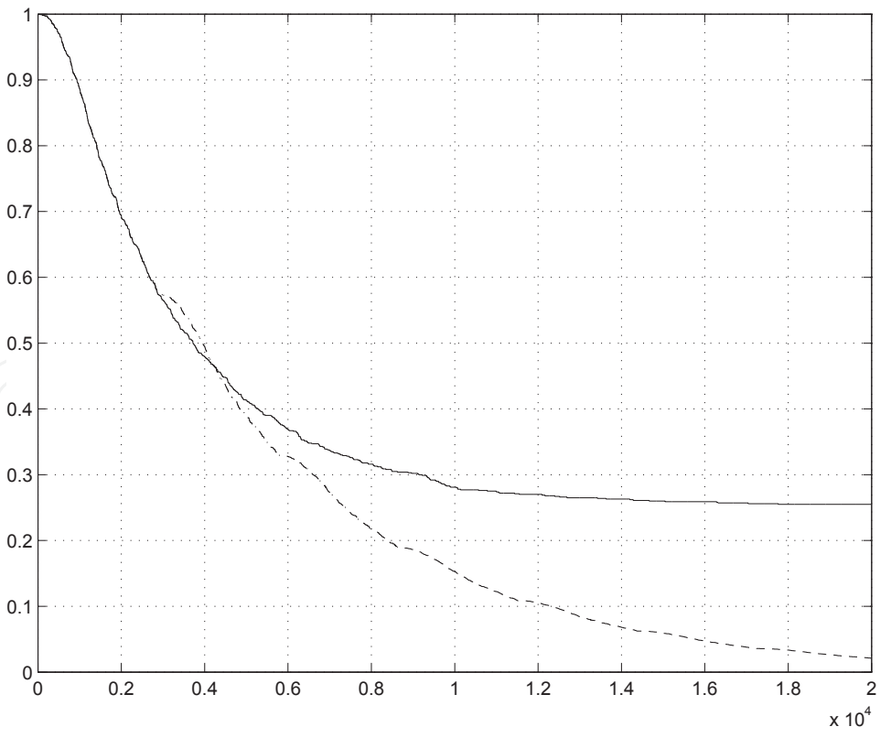


Fig. 3. The estimated failure probability for the standard MMAS (continuous line) and the restarted MMAS (dashed line)

5. References

Dorigo M., and Stutzle T. *Ant Colony Optimization* MIT Press, 2004.

Gutjahr W.J., "A graph–based ant system and its convergence", *Fut. Gener. Comput. Sys.* Vol. 16, pp. 873-888, 2000.

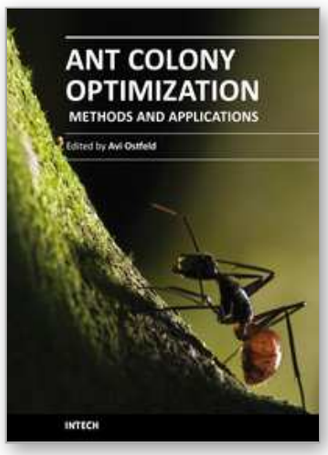
Gutjhar W.J., Sebastiani G. "Runtime Analysis of Ant Colony Optimization with Best-So-Far Reinforcement" *Meth. Comput. Appl. Prob.* Vol. 10, pp. 409-433, 2008.

Neumann F., Sudholt D., and Witt C. "Analysis of different MMAS ACO algorithms on unimodal functions and plateaus" *Swarm Intelligence* Vol. 3, pp. 35-68, 2009.

Stützle T., and Hoos H.H. "The MAX-MIN Ant System and local search for the travelling salesman problem" in Proc. ICEC '97 Int. Conf. on Evolutionary Computation (Baeck T., Michalewicz Z., and Yao X., eds) pp. 309-314 1997.

Stützle T., and Hoos H.H. "MAX-MIN Ant System" *Fut. Gener. Comput. Sys.* Vol. 16 pp. 889-914 2000.

IntechOpen



Ant Colony Optimization - Methods and Applications

Edited by Avi Ostfeld

ISBN 978-953-307-157-2
Hard cover, 342 pages
Publisher InTech
Published online 04, February, 2011
Published in print edition February, 2011

Ants communicate information by leaving pheromone tracks. A moving ant leaves, in varying quantities, some pheromone on the ground to mark its way. While an isolated ant moves essentially at random, an ant encountering a previously laid trail is able to detect it and decide with high probability to follow it, thus reinforcing the track with its own pheromone. The collective behavior that emerges is thus a positive feedback: where the more the ants following a track, the more attractive that track becomes for being followed; thus the probability with which an ant chooses a path increases with the number of ants that previously chose the same path. This elementary ant's behavior inspired the development of ant colony optimization by Marco Dorigo in 1992, constructing a meta-heuristic stochastic combinatorial computational methodology belonging to a family of related meta-heuristic methods such as simulated annealing, Tabu search and genetic algorithms. This book covers in twenty chapters state of the art methods and applications of utilizing ant colony optimization algorithms. New methods and theory such as multi colony ant algorithm based upon a new pheromone arithmetic crossover and a repulsive operator, new findings on ant colony convergence, and a diversity of engineering and science applications from transportation, water resources, electrical and computer science disciplines are presented.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Lorenzo Carvelli and Giovanni Sebastiani (2011). Some Issues of ACO Algorithm Convergence, Ant Colony Optimization - Methods and Applications, Avi Ostfeld (Ed.), ISBN: 978-953-307-157-2, InTech, Available from: http://www.intechopen.com/books/ant-colony-optimization-methods-and-applications/some-issues-of-aco-algorithm-convergence

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen