Reasons of ACO's Success in TSP

Osvaldo Gómez and Benjamín Barán

Centro Nacional de Computación
Universidad Nacional de Asunción
Paraguay
{ogomez, bbaran}@cnc.una.py
http://www.cnc.una.py

Abstract. Ant Colony Optimization (ACO) is a metaheuristic inspired by the foraging behavior of ant colonies that has empirically shown its effectiveness in the resolution of hard combinatorial optimization problems like the Traveling Salesman Problem (TSP). Still, very little theory is available to explain the reasons underlying ACO's success. An ACO alternative called Omicron ACO (OA), first designed as an analytical tool, is presented. This OA is used to explain the reasons of elitist ACO's success in the TSP, given a globally convex structure of its solution space.

1 Introduction

Ant Colony Optimization (ACO) is a metaheuristic proposed by Dorigo et al. [3] that has been inspired by the foraging behavior of ant colonies. In the last years ACO has empirically shown its effectiveness in the resolution of several different NP-hard combinatorial optimization problems [3]; however, still little theory is available to explain the reasons underlying ACO's success. Birattari et al. [1] developed a formal framework of ant programming with the goal of gaining deeper understanding of ACO, while Meuleau and Dorigo [10] studied the relationship between ACO and the Stochastic Gradient Descent technique. Gutjahr [7] presented a convergence proof for a particular ACO algorithm called Graph-based Ant System (GBAS) that has an unknown empirical performance. He proved that GBAS converges, with a probability that could be made arbitrarily close to 1, to the optimal solution of a given problem instance. Later, Gutjahr demonstrated for a time-dependent modification of the GBAS that its current solutions converge to an optimal solution with a probability exactly equal to 1 [8]. Stützle and Dorigo presented a short convergence proof for a class of ACO algorithms called $\text{ACO}_{gb,\tau_{min}}$ [11], where gb indicates that the global best pheromone update is used, while τ_{min} indicates that a lower limit on the range of the feasible pheromone trail is forced. They proved that the probability of finding the optimal solution could be made arbitrarily close to 1 if the algorithm is run for a sufficiently large number of iterations. Stützle and Hoos [12] calculated a positive correlation between the quality of a solution and its distance to a global optimum for the TSP, studying search space characteristics. Hence, it seems reasonable to assume that the concentration of the search around the best solutions found so far is a key aspect that led to the improved performance shown by ACO algorithms. However, there is no clear understanding of the real reasons of ACO's success, as recognized by Dorigo and Stützle [4, 11]. They stated that although it has been experimentally shown to be highly effective, only limited knowledge is available to explain why ACO metaheuristic is so successful [4].

Considering that elitist versions of ACO outperform non-elitist ones [12], this paper concentrates only on elitists. In search of a new ACO analytical tool to study their success, a simple algorithm preserving certain characteristics of elitist ACO was developed for this work. This is how the Omicron ACO (OA) was conceived. This name comes from the main parameter used (Section 3.2), which is *Omicron* (O). The OA simplicity facilitates the study of the main characteristics of an ACO in the TSP context, as explained in the following Sections.

The TSP is summarized in Section 2, while the standard ACO approach and the OA are presented in Section 3. The behavior of the OA for the problems berlin52, extracted from TSPLIB[1], and for a small randomly chosen TSP are shown in Section 4. In Section 5, the core of this paper is presented, analyzing the reasons of ACO's success in the TSP. Finally, the conclusions and future work are given in Section 6.

2 Test Problem

In this paper the symmetric Traveling Salesman Problem (TSP) is used as a test problem to study the OA, given the recognized ACO success in solving it [3, 12].

The TSP can be represented by a complete graph $G = (N, A)$ with N being the set of nodes, also called cities, and A being the set of arcs fully connecting the nodes. Each arc (i, j) is assigned a value $d(i, j)$ which represents the distance between cities i and j. The TSP is stated as the problem of finding a shortest closed tour r^* visiting each of the $n = |N|$ nodes of G exactly once.

Suppose that r_x and r_y are TSP tours or solutions over the same set of n cities. For this work, $l(r_x)$ denotes the length of tour r_x. The distance $\delta(r_x, r_y)$ between r_x and r_y is defined as n minus the number of edges contained in both r_x and r_y.

3 Ant Colony Optimization

Ant Colony Optimization (ACO) is a metaheuristic inspired by the behavior of ant colonies [3]. In the last years, elitist ACO has received increased attention by the scientific community as can be seen by the growing number of publications and its different fields of application [12]. Even though there exist several ACO variants, what can be considered a standard approach is next presented [5].

[1] Accessible at http://www.iwr.uni-heidelberg.de/iwr/comopt/software/TSPLIB95/

3.1 Standard Approach

ACO uses a pheromone matrix $\tau = \{\tau_{ij}\}$ for the construction of potential good solutions. It also exploits heuristic information using $\eta_{ij} = \frac{1}{d(i,j)}$. Parameters α and β define the relative influence between the heuristic information and the pheromone levels. While visiting city i, $\mathcal{N}_i$ represents the set of cities not yet visited and the probability of choosing a city j at city i is defined as

$$\mathcal{P}_{ij} = \begin{cases} \frac{\tau_{ij}^{\alpha} \cdot \eta_{ij}^{\beta}}{\sum_{\forall g \in \mathcal{N}_i} \tau_{ig}^{\alpha} \cdot \eta_{ig}^{\beta}} & \text{if } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

At every generation of the algorithm, each ant of a colony constructs a complete tour using (1). Pheromone evaporation is applied for all (i, j) according to $\tau_{ij} = (1 - \rho) \cdot \tau_{ij}$, where parameter $\rho \in (0, 1]$ determines the evaporation rate. Considering an elitist strategy, the best solution found so far r_{best} updates τ according to $\tau_{ij} = \tau_{ij} + \Delta\tau$, where $\Delta\tau = 1/l(r_{best})$ if $(i, j) \in r_{best}$ and $\Delta\tau = 0$ if $(i, j) \notin r_{best}$. For one of the best performing ACO algorithms, the $\mathcal{MAX}$-$\mathcal{MIN}$ Ant System ($\mathcal{MMAS}$) [12], minimum and maximum values are imposed to τ (τ_{min} and τ_{max}).

3.2 Omicron ACO

OA is inspired by $\mathcal{MMAS}$, an elitist ACO currently considered among the best performing algorithms for the TSP [12]. It is based on the hypothesis that it is convenient to search nearby good solutions [2, 12].

The main difference between $\mathcal{MMAS}$ and OA is the way the algorithms update the pheromone matrix. In OA, a constant pheromone matrix τ^0 with $\tau^0_{ij} = 1$, $\forall i, j$ is defined. OA maintains a population $P = \{P_x\}$ of m individuals or solutions, the best unique ones found so far. The best individual of P at any moment is called P^*, while the worst individual P_{worst}.

In OA the first population is chosen using τ^0. At every iteration a new individual P_{new} is generated, replacing $P_{worst} \in P$ if P_{new} is better than P_{worst} and different from any other $P_x \in P$. After K iterations, τ is recalculated. First, $\tau = \tau^0$; then, $\frac{O}{m}$ is added to each element τ_{ij} for each time an arc (i, j) appears in any of the m individuals present in P. The above process is repeated every K iterations until the end condition is reached (see pseudocode for details). Note that $1 \le \tau_{ij} \le (1 + O)$, where $\tau_{ij} = 1$ if arc (i, j) is not present in any P_x, while $\tau_{ij} = (1 + O)$ if arc (i, j) is in every P_x.

Similar population based ACO algorithms (P-ACO) [5, 6] were designed by Guntsch and Middendorf for dynamic combinatorial optimization problems. The main difference between the OA and the *Quality* Strategy of P-ACO [6] is that OA does not allow identical individuals in its population. Also, OA updates τ every K iterations, while P-ACO updates τ every iteration. Notice that any elitist ACO can be considered somehow as a population based ACO with a population that increases at each iteration and where older individuals have less influence on τ because of the evaporation.

Pseudocode of the main *Omicron ACO*

```
Input parameters: n, matrix D = {d_ij}, O, K, m, α, β
Output parameter: P  (m best found solutions)

P = Initialize population (τ^0)
REPEAT UNTIL end condition
   τ = Calculate pheromone matrix (P)
   REPEAT K TIMES
      Construct a solution P_new using equation (1)
      IF l(P_new) < l(P_worst) AND P_new ∉ P
         P = Update population (P_new , P)
```

Pseudocode of the function *Initialize population* (τ^0)

```
Initialize set P as empty
WHILE |P| < m
   Construct a solution P_x using equation (1)
   IF P_x ∉ P THEN include P_x in P
Sort P from worst to best considering l(P_x)
P_worst = P_0
```

Pseudocode of the function *Calculate pheromone matrix (P)*

```
τ = τ^0
REPEAT for every P_x of P
   REPEAT for every arc (i, j) of P_x
      τ_ij = τ_ij + O/m
```

Pseudocode of the function *Update population* (P_{new} , P)

```
P_0 = P_new
Sort P efficiently from worst to best considering l(P_x)
P_worst = P_0
```

4 Behavior of Omicron ACO

Definition 1. *Mean distance from a tour r to P.* $\delta(P,r) = \frac{1}{m}\sum_{i=1}^{m}\delta(P_i, r)$. If $r = r^*$ it gives a notion of how close a population is to the optimal solution r^*.

Definition 2. *Mean distance of P.* $\delta(P) = \frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\delta(P_i, P_j)$. It gives an idea of the convergence degree of a population.

Boese studied in [2] the space of solutions of the TSP att532 of 532 cities. 2,500 runs of different local search heuristic were made. For each heuristic, h different tours were stored in a set $H = \{H_i\}$. Each stored tour H_i has a length $l(H_i)$, a distance to r^* denoted as $\delta(H_i, r^*)$ and a mean distance to the other solutions of H called $\delta(H, H_i)$. Boese calculated a positive correlation between all these 3 variables. Given that the set $\{l(H_i)\}$ has a positive correlation with the set $\{\delta(H,H_i)\}$, Boese suggested a globally convex structure of the TSP solution space. In other words, the more central the position of a solution H_i within the set of solutions H is, the smaller its mean distance to the other solutions; therefore, the smaller is its expected length $l(H_i)$, i.e. the better the solution is. Global convexity is not convexity in the strict sense [9]. Boese suggested the analogy with a big valley structure, in which viewed from afar may appear to have a single minimum, but which up close has many local minima [2, Fig. 1]. Boese found similar results for two random geometric instances with 100 and 500 cities. At the same time, the authors of the present work are studying TSPLIB problems with identical conclusions. Also Stützle and Hoos calculated a positive correlation between the quality of a solution and its distance to a global optimum for the problems rat783 and fl1577 [12]. All these experimental results support the conjecture of a globally convex structure of the TSP's search space.

Based on the studies on local search heuristics mentioned above, the present work uses the globally convex structure of the TSP solution space concept as the main idea to explain the reasons of ACO's success.

It is also interesting to observe the length of P^*, $l(P^*)$; the mean length of a population, $l(P) = \frac{1}{m}\sum_{i=1}^{m} l(P_i)$ and the number of individuals $\zeta(P)$ that entered a population. Their mean values for several runs of the OA are denoted as $l(P^*)_M$, $l(P)_M$ and $\zeta(P)_M$ respectively. Accordingly, $\delta(P,r^*)_M$ and $\delta(P)_M$ represent the mean value of $\delta(P,r^*)$ and $\delta(P)$.

To maintain the number of possible tours to a manageable value, a random TSP called omi1 was designed with 8 cities of coordinates (58,12), (2,73), (14,71), (29,8), (54,50), (0,7), (2,91) and (44,53). Fig. 1 shows the evolution of the mean variables above defined as a function of the number of iterations in 10 runs of the OA. The left side of Fig. 1 presents the graphics for the TSP berlin52 (using the parameters $O = 600$, $m = 25$, $K = 1{,}000$, $\alpha = 1$ and $\beta = 2$), while the right side presents the graphics for the TSP omi1 (using the parameters $O = 30$, $m = 8$, $K = 10$, $\alpha = 1$ and $\beta = 2$).

The typical behaviors for both problems are similar to the mean behaviors shown in Fig. 1 respectively. The correlation values between $\delta(P,r^*)_M$, $\delta(P)_M$, $l(P^*)_M$, $l(P)_M$ and $\zeta(P)_M$ for the problem berlin52 are summarized in Table 1.

Table 1. Correlation of the OA behavior variables studied for the problem berlin52

	$\delta(P,r^*)_M$	$l(P^*)_M$	$l(P)_M$	$\zeta(P)_M$
$\delta(P)_M$	0.990	0.957	0.977	-0.972
$\delta(P,r^*)_M$		0.928	0.957	-0.995
$l(P^*)_M$			0.996	-0.900
$l(P)_M$				-0.934

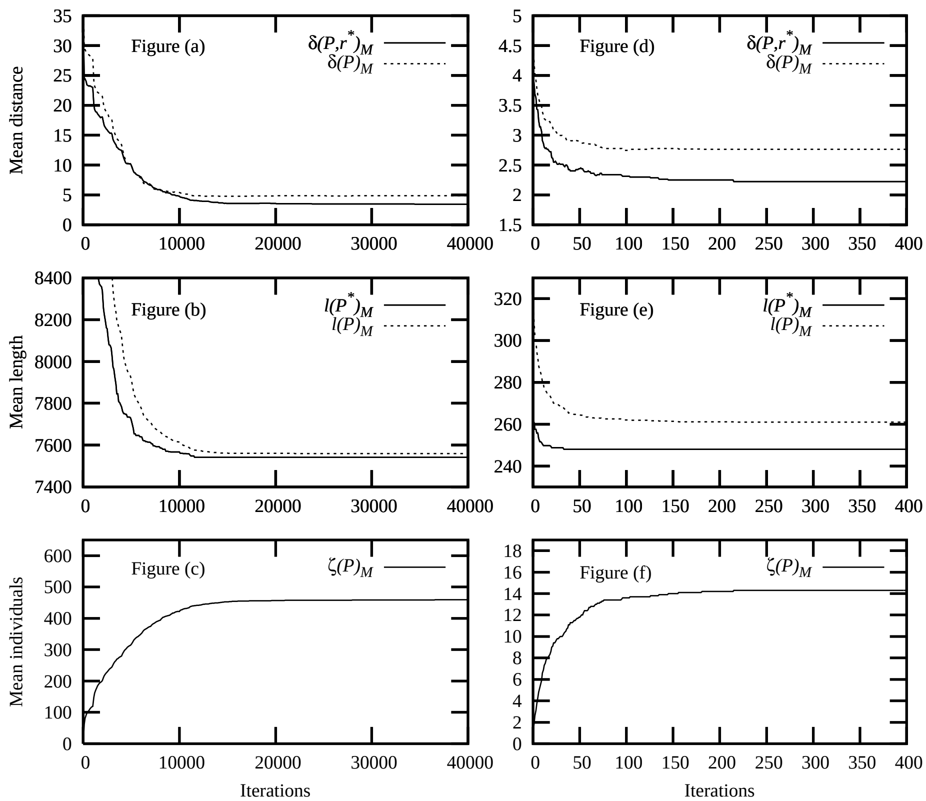

Fig. 1. Evolution as a function of iterations of $\delta(P, r^*)_M$, $\delta(P)_M$, $l(P^*)_M$, $l(P)_M$ and $\zeta(P)_M$ for 10 runs. Left, for the problem berlin52. Right, for the problem omi1

It can be observed in Fig. 1 that $\delta(P, r^*)_M$, $\delta(P)_M$, $l(P^*)_M$ and $l(P)_M$ decrease in the initial phase, while $\zeta(P)_M$ increases. In other words, new individuals with shorter length enter P at the beginning of a run; these individuals get closer to each other and at the same time they get closer to r^*. In the final phase the variables remain almost constant. It can be said that almost no individuals enter P and that $\delta(P, r^*)_M$ results smaller than $\delta(P)_M$, which means that the individuals finish closer to r^* than to the other individuals of P. These results motivate the analysis of the reasons of ACO's success in the next section.

5 Reasons of ACO's Success in TSP

The following exhaustive study is presented using the problem omi1 with 8 cities considering the space restrictions of this publication. The same exhaustive study was made using other randomly chosen problems with 7 and 9 cities and the results were very similar, making unnecessary any repetition in this paper.

5.1 Geometry of the Problem

Definition 3. $S = \{r_x\}$, *i.e. the whole discrete search space of a TSP.* Ω *will denote a subspace of* S, *i.e.* $\Omega \subseteq S$.

Definition 4. $\Omega_P = \{r_x | \ \delta(P, r_x) < \delta(P)\}$, *i.e. the set of tours* r_x *with a mean distance to population* P *shorter than the mean distance of* P. Ω_P is a central zone of P, as illustrated in Section 5.3, Fig. 5 (b).

Definition 5. $\Omega(e)$. Ω *conformed by the* e *best solutions of* S*;* e.g. $\Omega(100)$ denotes the set of the 100 shortest tours.

Inspired by [2], Fig. 2 presents $l(r_x)$ as a function of $\delta(r_x, r^*)$ for the whole space S of the test problem omi1. As in previous works [2, 12], a positive correlation can be observed. For this omi1 problem a correlation of 0.7 was calculated.

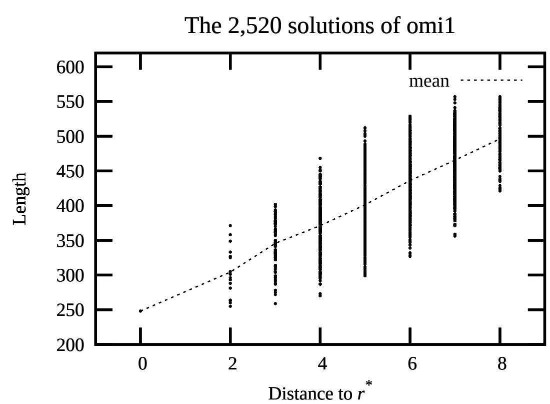

Fig. 2. $l(r_x)$ versus $\delta(r_x, r^*)$ $\forall r_x \in S$

Fig. 3 shows $l(r_x)$ as a function of $\delta(\Omega(e), r_x)$ $\forall r_x \in S$. For $e = 2,520$ ($\Omega(2,520) = S$), Fig. 3 (a) clearly shows that the correlation of the variables is 0 since the mean distance from any solution to all the others is the same. Fig. 3 (b) shows the same graph for $e = 2,519$, i.e. eliminating the worst solution from S. For this case the correlation increases to 0.521. Finally, Fig. 3 (c) draws the graph for $e = 1,260$ (best half solutions) and the correlation between the variables is 0.997. These results are consistent with the suggestion of a globally convex structure of the TSP solution space, since the smaller the distance of a solution r_x to a set of good solutions (and thus, more central its position in $\Omega(e) \subset S$), the smaller its expected tour length is.

Definition 6. $Q(e) = \{Q(e)_x\}$ *is defined as a set of randomly chosen elements of* $\Omega(e)$ *with cardinality* $|Q(e)|$.

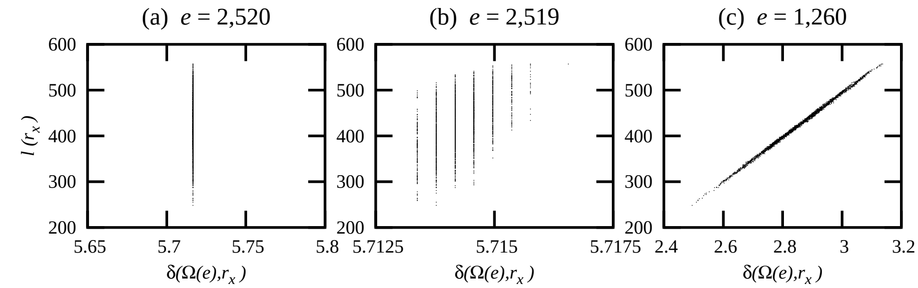

Fig. 3. $l(r_x)$ versus $\delta(\Omega(e), r_x)$ $\forall r_x \in S$ for different values of e

Given the interesting geometrical characteristic of $\Omega(e)$, a good question is if this globally convex property is maintained for $Q(e)$. To understand the importance of this question, it should be noticed that a population P of an OA may be considered as $Q(e)$. Fig. 4 shows $l(r_x)$ as a function of $\delta(Q(e), r_x)$ $\forall r_x \in S$. Randomly selected $Q(e)$ with $|Q(e)| = 25$ for different values of e are presented in figures 4 (a) to (d). The figure shows individuals of $Q(e)$, denoted as $Q(e)_x$ and the elements of Ω_P for $P = Q(e)$, denoted as $\Omega_{Q(e)}$. As can be seen in Fig. 4 (b) to (d) the best solutions are in $\Omega_{Q(e)}$; therefore, it seems convenient to explore $\Omega_{Q(e)}$.

To interpret Fig. 4 better, Table 2 presents the correlation ϱ between $l(r_x)$ and $\delta(Q(e), r_x)$ $\forall r_x \in S$ for the four different $Q(e)$ of Fig. 4. To compare the experimental results of Fig. 4 with average values for 1,000 randomly chosen $Q(e)$, Table 2 also presents the calculated average ϱ_M for the same parameters e, showing that Fig. 4 represents pretty well an average case.

Table 2. Correlation between $l(r_x)$ and $\delta(Q(e), r_x)$ $\forall r_x \in S$ for different values of e

	$e = 2,520$	$e = 1,890$	$e = 1,260$	$e = 630$
Correlation ϱ (for Fig. 4)	-0.201	0.387	0.641	0.862
Experimental mean correlation ϱ_M	0.001	0.425	0.683	0.836

As seen in Table 2 there is no meaningful correlation ϱ_M in $Q(2,520)$, as happened with $\Omega_{2,520} = S$. When decreasing the value of e, ϱ_M increases as happened with the correlation calculated for $\Omega(e)$. Thus, with good probability, $Q(e)$ is also globally convex (this probability increases with $|Q(e)|$). Considering the global convexity property, it can be stated that given a population $P = Q(e)$ of good solutions, it is a good idea to search in a central zone of P and specially in Ω_P, which contains the solutions with the shortest distance to P, given the positive correlation between the quality of a solution and its distance to P.

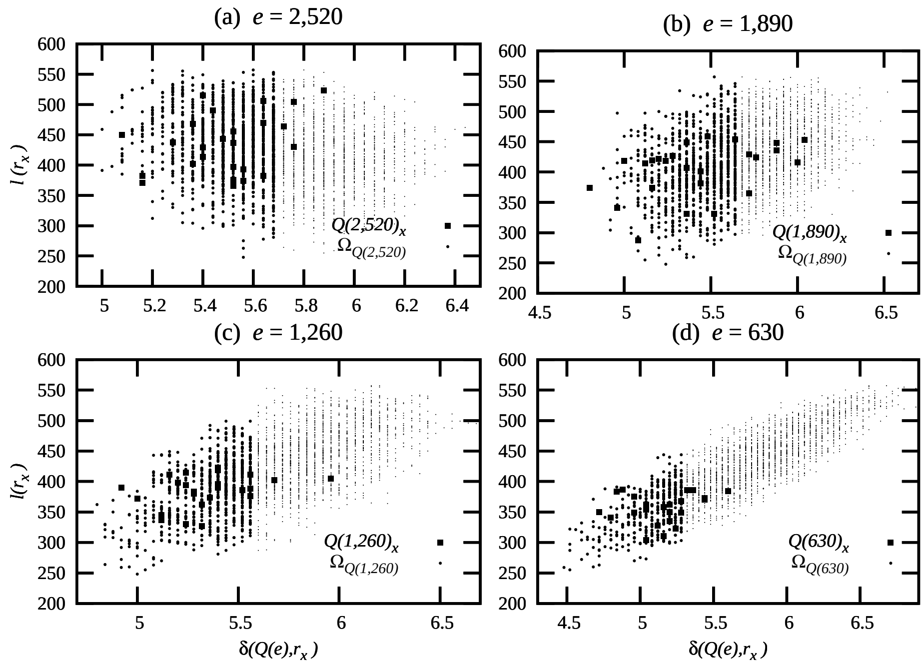

Fig. 4. $l(r_x)$ versus $\delta(Q(e), r_x)$ $\forall r_x \in S$ for different values of e

5.2 OA in a Globally Convex Geometry

OA concentrates an important proportion of its search of new solutions in Ω_P. This can be understood because in the construction of a new solution P_{new}, a larger probability is assigned to the arcs of each individual of P. This can be seen as a search made close to each individual of P. As a consequence, P_{new} is expected to have several arcs of $P_x \in P$, which means that the expected $\delta(P, P_{new})$ should be smaller than $\delta(P)$, i.e. P_{new} is expected to be in Ω_P.

Experimental results ratify this theory. 1,000 populations $Q(e)$ were taken randomly and 1,000 new solutions $Q(e)_{new}$ were generated with each population, using equation (1) with $O = 600$, $|Q(e)| = 25$, $\alpha = 1$, $\beta = 2$ for different values of e. Table 3 shows the proportion p of $Q(e)_{new}$ which lies inside $\Omega_{Q(e)}$, the mean cardinality of $\Omega_{Q(e)}$ (denoted w) and the relation $\frac{p}{w}$ (that may be understood as the proportion of $\Omega_{Q(e)}$ explored in average when generating each $Q(e)_{new}$).

At the very beginning of an OA computation, e is very large and there is a good probability of generating a solution in Ω_P (see p in Table 3 for $e = 2,520$). After progressing in its calculation, e decreases and so does $|\Omega_P|$; therefore, it becomes harder to find a new solution in Ω_P as shown in Table 3 (see how w and p decreases with e). Even though p decreases, it should be noticed that $\frac{p}{w}$, which is the proportion of Ω_P explored with each new individual, increases, i.e. OA searches more efficiently as computation continues.

Table 3. Mean values p, w y $\frac{p}{w}$ for different values e

	$e = 2,520$	$e = 1,890$	$e = 1,260$	$e = 630$
p	0.758	0.718	0.636	0.516
w	1,282.06	1,098.42	761.83	372.06
$\frac{p}{w}$	5.91e-4	6.53e-4	8.34e-4	13.86e-4

5.3 Two-Dimension Interpretation of OA Exploration Space

For didactic reasons, an analogy between the n-dimensional TSP search space S and a two-dimension interpretation is presented. First, Fig. 5 (a) shows that the search nearby 3 points implies a concentration of the exploration in their central area. The intersection of the areas close to each point is their central area, i.e. the search nearby every $P_x \in P$, done by OA, implies an exploration in the central zone of P, a recommended search region according to Section 5.1.

Experimentally, the area composed by the points where its geometrical distance to a randomly chosen population of 25 points is smaller than the mean distance of the population is shown in Fig. 5 (b). This is the two-dimension interpretation of Ω_P and it is a central area in the population. As a consequence, OA's ability to search mainly in the central zone of P, means that it searches with a good probability in Ω_P.

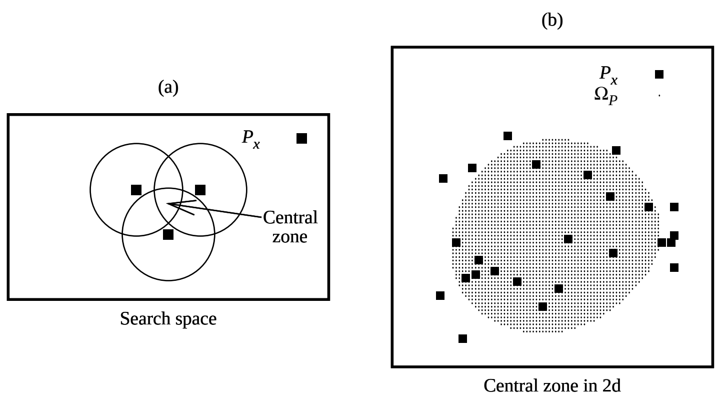

Fig. 5. (a) Simplified view of the search zone nearby all solutions of P (Ω_P) (b) Geometrical central zone of a population of 25 randomly chosen points

5.4 Reasons Underlying ACO's Success

In Fig. 6 the n-dimensional TSP search space is simplified to two dimensions for a geometrical view of the OA behavior. To understand the typical behavior of OA after the initial phase, a population $P1 = \{P1_x\} = Q(e)$ for $Q(e) \subset S$

of good solutions uniformly distributed is assumed in Fig. 6. As seen before, OA concentrates the search of new solutions in Ω_{P1} and replaces the worst solution of $P1$ ($P1_{worst}$) by a new solution P_{new} of smaller length $l(P_{new})$. A new population $P2$ is created including P_{new}. This is shown in Fig. 6 with a dotted line arrow. As a consequence, it is expected that $\delta(P2, r^*) < \delta(P1, r^*)$ because there is a positive correlation between $l(r_x)$ and $\delta(r_x, r^*)$. Similarly, $\delta(P, P_{new}) < \delta(P, P_{worst})$ because there is a positive correlation between $l(r_x)$ and $\delta(P, r_x)$, therefore $\delta(P2) < \delta(P1)$, i.e. it is expected that the subspace where the search of potential solutions is concentrated decreases, as experimentally verified in Section 5.2. Another easily measured property is that $l(P2) < l(P1)$.

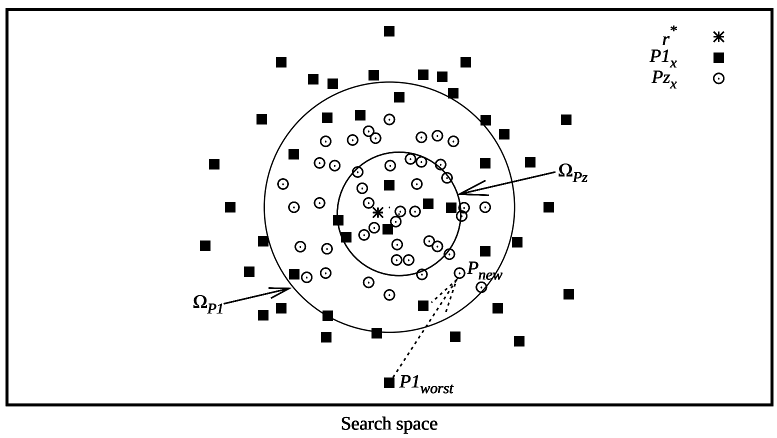

Fig. 6. Simplified view of OA behavior

OA performs this procedure repeatedly to decrease the search zone where promising solutions are located. Considering population $Pz = \{Pz_x\}$ for $z >> 2$, Fig. 6 shows how Ω_{Pz} has decreased considerably as a consequence of the globally convex structure of the TSP search space. At this point it should be clear that the main reason of OA's success is its ability to search in a central zone of P, where usually better solutions lie. This analysis is consistent with the empirical behavior of the OA observed in Section 4 for the problems berlin52 and omi1.

Given that any P-ACO maintains a population of individuals P, as the presented OA, similar analysis applies to explain its success. In other words, the search is mainly oriented towards the central zone of P, where good solutions are usually found. Finally, as already mentioned in Section 3.2, any elitist ACO may be considered as a P-ACO and therefore the same explanation applies.

6 Conclusions and Future Work

OA concentrates the search in a central zone Ω_P of its population P. In globally convex problems, good solutions are usually found in this region; therefore, OA concentrates its search in a promising subspace. Every time a good solution is found, it enters the population reducing the promising search zone iteratively. Thus, this work explains the main reasons of OA's success, and any elitist ACO in general (e.g. $\mathcal{MMAS}$).

OA does not use positive feedback. Hence, elitist ACO does not necessarily share the same reasons of success with real ants, even though ACO was *inspired* by real ants behavior. This suggests not to limit the study of useful ACO properties to real ants behavior.

This work was limited to the analysis of elitist ACO in globally convex structures. Based on the presented framework, a future work will study local convergence, as well as other evolutionary algorithms and different problems. The authors are also studying all TSPLIB problems with known optimal solutions to experimentally confirm the globally convex property of the TSP search space.

References

[1] M. Birattari, G. D. Caro, and M. Dorigo. For a Formal Foundation of the Ant Programming Approach to Combinatorial Optimization. Part 1: The problem, the representation, and the general solution strategy. Technical Report TR-H-301, ATR-Human Information Processing Labs, Kyoto, Japan, 2000.

[2] K. D. Boese. Cost Versus Distance in the Traveling Salesman Problem. Technical Report 950018, Univ. of California, Los Angeles, Computer Science, May 19, 1995.

[3] M. Dorigo and G. Di Caro. The Ant Colony Optimization Meta-Heuristic. In D. Corne, M. Dorigo, and F. Glover, editors, *New Ideas in Optimization*, pages 11–32. McGraw-Hill, London, 1999.

[4] M. Dorigo and T. Stützle. An Experimental Study of the Simple Ant Colony Optimization Algorithm. In *2001 WSES International Conference on Evolutionary Computation (EC'01)*. WSES-Press International, 2001.

[5] M. Guntsch and M. Middendorf. A Population Based Approach for ACO. In S. Cagnoni, J. Gottlieb, E. Hart, M. Middendorf, and G. Raidl, editors, *Applications of Evolutionary Computing, Proceedings of EvoWorkshops2002: EvoCOP, EvoIASP, EvoSTim*, volume 2279, pages 71–80, Kinsale, Ireland, 3-4 2002. Springer-Verlag.

[6] M. Guntsch and M. Middendorf. Applying Population Based ACO to Dynamic Optimization Problems. In *Ant Algorithms, Proceedings of Third International Workshop ANTS 2002*, volume 2463 of *LNCS*, pages 111–122, 2002.

[7] W. J. Gutjahr. A graph-based Ant System and its convergence. *Future Generation Computer Systems*, 16(8):873–888, June 2000.

[8] W. J. Gutjahr. ACO Algorithms with Guaranteed Convergence to the Optimal Solution. *Information Processing Letters*, 82(3):145–153, May 2002.

[9] T. C. Hu, V. Klee, and D. Larman. Optimization of globally convex functions. *SIAM Journal on Control and Optimization*, 27(5):1026–1047, Sept. 1989.

[10] N. Meuleau and M. Dorigo. Ant colony optimization and stochastic gradient descent. *Artificial Life*, 8(2):103–121, 2002.

[11] T. Stützle and M. Dorigo. A Short Convergence Proof for a Class of Ant Colony Optimization Algorithms. *IEEE Transactions on Evolutionary Computation*, 6:358–365, Aug. 2002.

[12] T. Stützle and H. H. Hoos. $\mathcal{MAX}$-$\mathcal{MIN}$ Ant System. *Future Generation Computer Systems*, 16(8):889–914, June 2000.