Asymptotic Convergence of Some Metaheuristics Used for Multiobjective Optimization

Mario Villalobos-Arias[1,⋆], Carlos A. Coello Coello[2], and Onésimo Hernández-Lerma[1]

[1] CINVESTAV-IPN
Department of Mathematics
A. Postal 14-740
México, D.F. 07000, Mexico
{mava,ohernand}@math.cinvestav.mx
[2] CINVESTAV-IPN
Evolutionary Computation Group
Depto. de Ingeniería Eléctrica
Sección de Computación
Av. Instituto Politécnico Nacional No. 2508
Col. San Pedro Zacatenco
México, D. F. 07300, Mexico
ccoello@cs.cinvestav.mx

Abstract. This paper presents the asymptotic convergence analysis of Simulated Annealing, an Artificial Immune System and a General Evolutionary Algorithm for multiobjective optimization problems. In the case of a General Evolutionary Algorithm, we refer to any algorithm in which the transition probabilities use a uniform mutation rule. We prove that these algorithms converge if elitism is used.

1 Introduction

In nature, most problems have several objectives which we aim to optimize simultaneously. Such problems are called "multiobjective", and their solution requires a suitable definition of optimality (usually called "Pareto optimality"). Such problems normally have not one, but an infinite set of possible solutions, which represent possible trade-offs among the objectives (such solutions constitute the so-called "Pareto optimal set").

Diverse metaheuristics have been adopted to solve multiobjective optimization problems (MOP) [2]. In this paper, we study three of them: simulated annealing (SA) [10, 15], artificial immune systems (AIS) [14] and evolutionary algorithms (EA) [9, 6]. For these metaheuristics that use a uniform mutation rule (see end of Section 3.1) we show that the associated Markov chain converges geometrically to its stationary distribution, but not necessarily to the optimal solution set of the multiobjective optimization problem. Convergence to the optimal solution set is ensured if elitism (whose definition is provided in this paper) is used.

⋆ Permanent Address: Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica
mvillalo@cariari.ucr.ac.cr

A.H. Wright et al. (Eds.): FOGA 2005, LNCS 3469, pp. 95–111, 2005.
© Springer-Verlag Berlin Heidelberg 2005

Metaheuristics such as those indicated in this paper, have become a standard tool to solve both single-objective and multiobjective optimization problems. In the single-objective case, the convergence of a metaheuristic is reasonably well-understood [17]. However, when dealing with multiobjective optimization problems, there is not much work available in the literature, except for extremely particular cases (see for example [16]).

The remainder of this paper is organized as follows. Section 2 introduces the problem of our interest. The three specific algorithms studied in this paper are introduced in Section 3. In Section 4 we present some basic definitions related to Markov chain theory. Our main results (i.e., the corresponding proofs) are presented in Section 5. Section 6 provides our conclusions and some possible paths of future research.

2 The Multiobjective Optimization Problem

Let X be a set and $F : X \longrightarrow I\!R^d$ a given vector function with components $f_i : X \longrightarrow I\!R$ for each $i \in \{1, \ldots, d\}$. The multiobjective optimization problem (MOP) we are concerned with is to find $x^* \in X$ such that

$$F(x^*) = \min_{x \in X} F(x) = \min_{x \in X} [f_1(x), \ldots, f_d(x)], \tag{1}$$

where the minimum is understood in the sense of the standard Pareto order in which two vectors in $I\!R^d$ are compared as follows.

If $\boldsymbol{u} = (u_1, \ldots, u_d)$ and $\boldsymbol{v} = (v_1, \ldots, v_d)$ are vectors in $I\!R^d$, then

$$\boldsymbol{u} \preceq \boldsymbol{v} \iff u_i \leq v_i \; \forall \, i \in \{1, \ldots, d\}.$$

This relation is a partial order. We also write $\boldsymbol{u} \prec \boldsymbol{v} \iff \boldsymbol{u} \preceq \boldsymbol{v}$ and $\boldsymbol{u} \neq \boldsymbol{v}$.

Definition 1: A point $x^* \in X$ is called a *Pareto optimal solution* for the MOP (1) if there is no $x \in X$ such that $F(x) \prec F(x^*)$. The set

$$\mathcal{P}^* = \{x \in X \; : \; x \text{ is a Pareto optimal solution}\}$$

is called the *Pareto optimal set*, and its image under F, i.e.

$$F(\mathcal{P}^*) \; := \; \{F(x) \; : \; x \in \mathcal{P}^*\},$$

is called *Pareto front*.

In the remainder of the paper we will use the following well–known "scalarization" result.

Proposition 1: If $\boldsymbol{x}^* \in X$ is a solution of the weighted problem:

$$\min_{\boldsymbol{x} \in X} \sum_{s=1}^{d} w_s f_s(\boldsymbol{x}), \text{ where } w_s \geq 0 \; \forall s \in \{1, \ldots, n\} \text{ and } \sum_{s=1}^{d} w_s = 1,$$

then $\boldsymbol{x}^* \in \mathcal{P}^*$.

Proof. See, for instance, [13, p.78].

Now we introduce some notation that will be used later on. Let

$$\Sigma_{opt} := \{x \in X \ : \ \sum_{s=1}^{d} f_s(x) = \Sigma_m\},$$

where

$$\Sigma_m := \min_{x \in X} \sum_{s=1}^{d} f_s(x). \tag{2}$$

Then, by Proposition 1, the Pareto optimal set $\mathcal{P}^*$ contains Σ_{opt}, i.e.

$$\Sigma_{opt} \subset \mathcal{P}^*. \tag{3}$$

As we are concerned with computational aspects, in the remainder of the paper we will assume that the set X in (1) is *finite*. For an EA and the AIS, in which the elements are represented by strings of length l with 0 or 1 at each entry, we take $X = I\!B^l$, with $I\!B = \{0, 1\}$. For SA we only assume that X is finite.

3 Algorithms

3.1 Evolutionary Algorithms

Evolutionary algorithms are techniques that use a population which evolves over time (i.e., generations) applying some operations to the current population to obtain the next one. Some of these operations are

- mutation
- selection
- crossover
- reordering

The type of EAs we are interested in are modeled as Markov chains with transition probabilities that use uniform mutation and possibly other operations. This mutation is applied with a certain parameter or probability p_m, which is positive and less than $1/2$, i.e.

$$p_m \in (0, 1/2)\,. \tag{4}$$

Some examples of this type of EAs are the following:

- genetic algorithms (see [9]),
- evolution strategies (see [18]),
- evolutionary programming (see [8, 7]).

These types of algorithms can be modeled as a Markov chain $\{X_k : k \geq 0\}$ whose state space S is the set of all possible populations of n individuals, each one represented by a bit string of length l. Hence $S = (I\!B^l)^n = I\!B^{nl}$, where $I\!B = \{0, 1\}$ and so S is the set of all possible vectors of n entries, each of which is a string of length l with 0 or 1 at each entry.

Let $i \in S$ be a state, so that i can be represented as

$$i = (i_1, i_2, \ldots, i_n),$$

where each i_s is a string of length l of 0's and 1's.

The chain's transition probability is given by

$$P_{ij} = I\!P(X_{k+1} = j \mid X_k = i).$$

Thus the transition matrix is of the form

$$P = (P_{ij}) = LM, \tag{5}$$

where M is the transition matrix corresponding to the mutation operation and L represents the other operations.

Note that these matrices are stochastic, i.e. $L_{ij} \geq 0$, $M_{ij} \geq 0$ for all i, j, and for each $i \in S$

$$\sum_{j \in S} L_{ij} = 1 \text{ and } \sum_{j \in S} M_{ij} = 1. \tag{6}$$

The Mutation Probability

The mutation probability is very important in the convergence analysis of the EA. To calculate it from state i to state j we use that the individual i_s is transformed into the individual j_s applying *uniform mutation* (i.e. a flip mutation, with probability p_m, is applied to each entry of i_s) then each entry of i_s is transformed into the corresponding one of j_s with probability $1 - p_m$ or p_m depending on if the corresponding entries are equal or different, as in the following scheme.

$$\begin{array}{ccc} & 1\ 2\ \cdots\ n & \\ i & \boxed{i_1}\boxed{i_2}\boxed{\cdots}\boxed{i_n} & \\ \text{mutation} & \downarrow\ \downarrow\ \cdots\ \downarrow & \\ j & \boxed{j_1}\boxed{j_2}\boxed{\cdots}\boxed{j_n} & \end{array}$$

Thus, for each individual in the population the mutation probability can be calculated as

$$p_m^{H(i_s,j_s)}(1-p_m)^{l-H(i_s,j_s)} \ \ \forall s \in \{1,\ldots,n\},$$

where $H(i_s, j_s)$ is the Hamming distance between i_s and j_s. It follows that the mutation probability from i to j is:

$$M_{ij} = \prod_{s=1}^{n} p_m^{H(i_s,j_s)}(1-p_m)^{l-H(i_s,j_s)} \tag{7}$$

3.2 The Simulated Annealing Algorithm

Kirkpatrick et al. [10] and Černy [15] proposed an optimization algorithm based on some analogies with an annealing process in which a crystal is produced. This led to the development of an algorithm called "Simulated Annealing" which is a heuristic search technique that has been quite successful in combinatorial optimization problems (see [1] and [11] for details).

The simulated annealing algorithm generates a succession of possible solutions of the optimization problem. These possible solutions are the states of a Markov chain and the "energy" of a state is the evaluation of the possible solution that it represents.

The temperature is simulated with a sequence of positive control parameters c_k. A transition of the Markov chain occurs in two steps, given the value c_k of the control parameter. First, if the current state is i, a new state j is generated with a certain probability $G_{ij}(c_k)$, defined below. Then an "acceptance rule" $A_{ij}(c_k)$ is applied to j. Our main result hinges on a suitable selection of the acceptance rule, which we now discuss.

The generation probability. For each state i, let S_i be a subset of $S \setminus \{i\}$ called the neighborhood of i. We shall assume that the number of elements in S_i is the same, say Θ, for all $i \in S$, and also that the neighbor relation is symmetric, that is, $j \in S_i$ if and only if $i \in S_j$. Then, denoting by χ_{S_i} the indicator function of S_i (i.e. $\chi_{S_i}(j) := 1$ if $j \in S_i$ and 0 otherwise), we define the generation probability

$$G_{ij}(c_k) := \frac{\chi_{S_i}(j)}{\Theta} \quad \text{for all } i, j \in S. \tag{8}$$

The acceptance probability. This probability value is crucial for the behavior of the simulated annealing algorithm.

The idea of this acceptance rule is that any new state that improves the actual state will be accepted with probability 1 and the others are accepted with certain probability that tends to zero as time goes to infinity.

When dealing with MOPs there are different options to define the acceptance rule. For instance, Serafini [20] proposes to use the L_∞–Tchebycheff norm given by

$$A'_{ij}(c) = \min\left\{1, \exp\left(\max_{s\in\{1,\ldots,d\}} \frac{\lambda_s(f_s(i) - f_s(j))}{c}\right)\right\},$$

where the λ_s are given positive parameters.

On the other hand, Ulungu and coworkers [21, 22, 24, 23] use

$$\begin{aligned} A''_{ij}(c) &:= \min\left\{1, \exp\left(\sum_{s=1}^{d} \frac{\lambda_s(f_s(i) - f_s(j))}{c}\right)\right\} \\ &= \exp\left\{-\left(\sum_{s=1}^{d} \frac{\lambda_s(f_s(j) - f_s(i))}{c}\right)^+\right\}. \end{aligned} \tag{9}$$

where as usual, a^+ denotes the positive part of a number $a \in \mathbb{R}$, namely

$$a^+ := \begin{cases} a & \text{if } a > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Here, we will use the acceptance probability presented in [20]:

$$A_{ij}(c) := \prod_{s=1}^{d} \min\left\{1, \exp\left(\frac{f_s(i) - f_s(j)}{c}\right)\right\},$$

which can be expressed in the simpler form

$$A_{ij}(c) = \exp\left(-\frac{\sum_{s=1}^{d}(f_s(j) - f_s(i))^+}{c}\right). \tag{10}$$

For the last two acceptance rules, we have shown somewhere else that the SA for MOP converges (see [25]).

The transition probability. Having the generation and the acceptance probabilities, we can now define the *transition probability* from i to j as

$$P_{ij}(c_k) := \begin{cases} G_{ij}(c_k)A_{ij}(c_k) & \text{if } i \neq j, \\ 1 - \sum_{l \in S, l \neq i} P_{il}(c_k) & \text{if } i = j, \end{cases} \tag{11}$$

where A_{ij} is as in (10) (or as in (9)).

3.3 Artificial Immune System

The Artificial Immune System (AIS) algorithm is a technique that, as its name indicates, simulates in a computer certain aspects of an immune system. When an antigen enters our immune system, it is immediately detected and generates a response from the immune system. As a consequence, antibodies are generated by the immune system. Antibodies are molecules that play the main role in the immune response. They are capable of adhering to the antigens in order to neutralize and mark them for elimination by other cells of the immune system. Successful antibodies are cloned and hypermutated. This is called the *clonal selection principle* and has been the basis for developing the algorithm on which we base the work reported in this paper [4].

For our mathematical model, we will consider the AIS (based on clonal selection theory [4]) for multiobjective optimization proposed in [3]. From here on, we will refer to this approach using the same name adopted by the authors of this algorithm: "Multiobjective Immune System Algorithm" (MISA for short). Next, we will focus our discussion only on the aspects that are most relevant for its mathematical modelling. For a detailed discussion on this algorithm, readers should refer to [3].

In MISA the antigens are simulated with a population of strings of 0's and 1's. The population is divided in two parts, a primary set and a secondary set; the primary set contains the "best" individuals (or elements) of the population. The transition of one population to another is made by means of two mutation rules and a reordering operation. First, the elements of the primary set are copied several times, then in each of these copies a fixed number of bits are mutated, at random. Regarding the secondary set, a uniform mutation with parameter p_m is applied. This parameter is positive and less than $1/2$, i.e. $p_m \in (0, 1/2)$.

After that, the elements are reordered, moving the "best" individuals to the primary set. MISA can be modeled with a Markov chain $\{X_k : k \geq 0\}$, with state space $S = \mathbb{B}^{nl}$, where $\mathbb{B} = \{0, 1\}$. In this case a individual can be represented as:

$$i = (i^1, i^2) = (i_1, i_2, \ldots, i_{n_1}; i_{n_1+1}, \ldots, i_n),$$

i^1 represents the primary set and i^2 the secondary.

3.4 Using Elitism

In our case, when dealing with MOPs, we say that we are using *elitism* in an algorithm if we use an extra set, called the *elite* set, in which we put the "best" elements (nondominated elements of the state in our case) found. This elite set usually does not participate in the evolution (although, there are multi-objective evolutionary algorithms that use the elite set in the selection process, such as the Strength Pareto Evolutionary Algorithm [27]), since it is used only to store the nondominated elements.

After each transition we apply an *elitism operation* that accepts a new state if there is an element in the population that improves some element in the elite set (i.e., if there is an element in the population that dominates, in the Pareto sense, some element in the elite set).

If we are using elitism, the representation of the states changes to the following form:

$$\hat{i} = (i^e; i) = (i_1^e, \cdots, i_r^e; i_1, \cdots, i_n),$$

where $i_1^e, \cdots, i_r^e$ are the members of the elite set of the state, r is the number of elements in the elite set and we assume that the cardinality of $\mathcal{P}^*$ is greater than or equal to r. In addition we assume that $r \leq n$.

Note that in general $i_1^e, \cdots, i_r^e$ are not necessarily the "best" elements of the state $\hat{i}$, but after applying the elitism operation in i^e they are the "best" elements of the state.

Let $\hat{P}$ be the transition matrix associated with the new states. If all the elements in the elite set of a state are Pareto optimal, then any state that contains an element in the elite set that is not a Pareto optimal will not be accepted, i.e.

$$\text{if } \{i_1^e, \cdots, i_r^e\} \subset \mathcal{P}^* \text{ and } \{j_1^e, \cdots, j_r^e\} \not\subset \mathcal{P}^* \text{ then } \hat{P}_{ij} = 0. \tag{12}$$

4 Markov Chain Theory

We provide here some standard definitions and results.

We first introduce the definition of convergence of an algorithm, which uses the following notation: if $V = (v_1, v_2, \ldots, v_n)$ is a vector, then $\{V\}$ denotes the set of entries of V, i.e.

$$\{V\} = \{v_1, v_2, \ldots, v_n\}.$$

Definition 2: Let $\{X_k : k \geq 0\}$ be the Markov chain associated to an algorithm. We say that the algorithm converges with probability 1 if

$$I\!P(\{X_k\} \subset \mathcal{P}^*) \to 1 \text{ as } k \to \infty.$$

In the case in which we are using elitism we replace X_k by X_k^e, the elite set of the state (i.e. if $X_k = i$ then $X_k^e = i^e$).

The next result gives an upper bound on the rate of convergence of P^k as $k \to \infty$. We will use it to show the existence of the stationary distribution in Theorem 2.

Lemma 1: Let N be the cardinality of S, and let P_{ij}^k be the entry ij of P^k. Suppose that there exists an integer $\nu > 0$ and a set J of $N_1 \geq 1$ values of j such that

$$\min_{\substack{1 \leq i \leq N \\ j \in J}} P_{ij}^{\nu} = \delta > 0.$$

Then there are numbers $\pi_1, \pi_2, \ldots, \pi_{N_1}$ such that

$$\lim_{k \to \infty} P_{ij}^k = \pi_j \; \forall i = 1, \ldots, N, \; \forall \, j \in J, \text{ with } \; \pi_j \geq \delta > 0,$$

and $\pi_1, \pi_2, \ldots, \pi_{N_1}$ form a set of stationary probabilities. Moreover

$$|P_{ij}^k - \pi_j| \leq (1 - N_1 \delta)^{\frac{k}{\nu} - 1} \quad \forall \, i = 1, \ldots, N, \; \forall \, j \in J, \; \forall \, k = 1, 2, \ldots.$$

Proof. See, for example, [5, p. 173].

We will need some properties of the limiting distribution, which we present next. Recall that a probability distribution $\boldsymbol{q}$ is called the *limiting distribution* of a Markov chain with transition probability P if

$$q_i = \lim_{k \to \infty} I\!P(X_k = i | X_0 = j) \text{ for all } i, j \in S.$$

If such a limiting distribution $\boldsymbol{q}$ exists and $a_i(k) = I\!P(X_k = i)$, for $i \in S$, denotes the distribution of X_k, then

$$\lim_{k \, \longrightarrow \, \infty} a_i(k) = q_i \text{ for all } i \in S.$$

Moreover, $\boldsymbol{q}$ is an *invariant* (or *stationary*) distribution of the Markov chain, which means that

$$\boldsymbol{q} = \boldsymbol{q} \, P; \tag{13}$$

that is, $\boldsymbol{q}$ is a left eigenvector of P with eigenvalue 1. A converse to this result (which is true for *finite* Markov chains) is given in Lemma 2 below.

Observe that (13) trivially holds if $\boldsymbol{q}$ is a probability distribution satisfying

$$q_i P_{ij} = q_j P_{ji} \quad \forall i, j \in S. \tag{14}$$

This equation is called the *detailed balance equation*, and (13) is called the *global balance equation*.

Lemma 2:[12, p.19] Let P be the transition matrix of a finite, irreducible and aperiodic Markov chain. Then the chain has a unique stationary distribution $\boldsymbol{q}$ (that is $\boldsymbol{q}$ is the unique distribution that satisfies (13)) and, in addition, $\boldsymbol{q}$ is the chain's limiting distribution.

Definition 3: Let X be as in problem (1). We say that X is *complete* if for each $x \in X \setminus \mathcal{P}^*$ there exists $x^* \in \mathcal{P}^*$ such that $F(x^*) \preceq F(x)$.

For instance, if X is finite then X is complete.

Let $i, j \in S$ be two arbitrary states, we say that i *leads* to j, and write $i \to j$, if there exists an integer $k \geq 1$ such that $P_{ij}^k > 0$. If i does not lead to j, then we write $i \not\to j$.

We call a state i *inessential* if there exists a state j such that $i \to j$ but $j \not\to i$. Otherwise the state i is called *essential*.

We denote the set of essential states by E and the set of inessential states by I. Clearly,

$$S = E \cup I.$$

We say that P is in *canonical form* if it can be written as

$$P = \begin{pmatrix} P_1 & 0 \\ R & Q \end{pmatrix}.$$

Observe that P can be put in this form by reordering the states, that is, the essential states at the beginning and the inessential states at the end. In this case, P_1 is the matrix associated with the transitions between essential states, R with transitions from inessential to essential states, and Q with transitions between inessential states.

Note that P^k has a Q^k in the position of Q in P, i.e.

$$P^k = \begin{pmatrix} P_1^k & 0 \\ R_k & Q^k \end{pmatrix},$$

where R_k is a matrix that depends of $P_1,\ Q$ and R.

Now we present some results that will be essential in the proof of Theorem 3.

Lemma 3: Let P be a stochastic matrix, and let Q be the submatrix of P associated with transitions between inessential states. Then, as $k \to \infty$,

$$Q^k \to 0 \text{ elementwise geometrically fast.}$$

Proof. See, for instance, [19, p.120]. ■

As a consequence of Lemma 3 we have the following.

Corollary 1: For any initial distribution,

$$I\!P(X_k \in I) \to 0 \ \text{ as } \ k \to \infty.$$

Proof. For any initial distribution vector p_0, let $p_0(I)$ be the subvector that corresponds to the inessential states. Then, by Lemma 3,

$$I\!P(X_k \in I) = p_0(I)' Q^k \mathbf{1} \to 0 \text{ as } k \to \infty.$$ ■

5 Main Results

In this section we present some recent results on the convergence of the algorithms introduced in Section 3, for multiobjective optimization problems (MOPs).

5.1 Convergence of Simulated Annealing

Following the ideas of Laarhoven, Aarts and Korst in [1, 11] we developed a convergence proof of SA for MOPs, which is presented in the following Theorem.

Theorem 1: Let $P(c)$ be the transition matrix associated with the SA algorithm defined by (8), (10), (11) and, moreover, suppose that $G(c)$ is irreducible. Then:

(a) The Markov chain has a stationary distribution $\boldsymbol{q}(c)$ whose components are given by

$$q_i(c) = \frac{1}{N_0(c)} \exp\left(-\frac{\sum_{s=1}^{d} f_s(i)}{c}\right), \tag{15}$$

where

$$N_0(c) = \sum_{j \in S} \exp\left(-\frac{\sum_{s=1}^{d} f_s(j)}{c}\right) \tag{16}$$

(b) For each $i \in S$

$$q_i^* := \lim_{c \searrow 0} q_i(c) = \frac{1}{|\Sigma_{opt}|} \chi_{\Sigma_{opt}}(i),$$

where $|\Sigma_{opt}|$ denotes the number of elements in Σ_{opt}.

(c) The SA algorithm converges with probability 1.

These results remain valid if (10) is replaced with (9).

Proof of Theorem 1.

(a) Since G is irreducible, using Lemma 2 it can be seen that the Markov chain is irreducible and aperiodic (see [1, p.39]). Hence, by Lemma 2 there exists a unique stationary distribution. We now use (8) and (11) to see that (14) holds for all $i \neq j$. First note that

$$\begin{aligned} q_i(c)P_{ij}(c) &= q_i(c)G_{ij}(c)A_{ij}(c) \\ &= \begin{cases} \frac{1}{\Theta} q_i(c)A_{ij}(c) & \text{if } j \in S_i \\ 0 & \text{if } j \notin S_i. \end{cases} \end{aligned}$$

Similarly,

$$\begin{aligned} q_j(c)P_{ji}(c) &= q_j(c)G_{ji}(c)A_{ji}(c) \\ &= \begin{cases} \frac{1}{\Theta} q_j(c)A_{ji}(c) & \text{if } i \in S_j \\ 0 & \text{if } i \notin S_j. \end{cases} \end{aligned}$$

Thus, since $i \in S_j$ if and only if $j \in S_i$, to obtain (14) we only have to prove that

$$q_i(c)A_{ij}(c) = q_j A_{ji}(c).$$

But this follows from (10), (15) and using that for any real numbers $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$, we have

$$\sum_{k=1}^{n}(a_k - b_k) + \left(\sum_{k=1}^{n}(b_k - a_k)\right)^+ = \left(\sum_{k=1}^{n}(a_k - b_k)\right)^+,$$

$$\sum_{k=1}^{n}(a_k - b_k) + \sum_{k=1}^{n}(b_k - a_k)^+ = \sum_{k=1}^{n}(a_k - b_k)^+.$$

because

$$\begin{aligned}
&q_i(c)A_{ij}(c) \;= \\
&= \frac{1}{N_0(c)} \exp\left(-\frac{\sum_{s=1}^{n} f_s(i)}{c}\right) \exp\left(-\frac{\sum_{s=1}^{n}(f_s(j) - f_s(i))^+}{c}\right) \\
&= \frac{1}{N_0(c)} \exp\left(-\frac{\sum_{s=1}^{n} f_s(j)}{c}\right) \\
&\quad \exp\left(-\frac{\sum_{s=1}^{n}(f_s(i) - f_s(j)) + \sum_{s=1}^{n}(f_s(j) - f_s(i))^+}{c}\right) \\
&= \frac{1}{N_0(c)} \exp\left(-\frac{\sum_{s=1}^{n} f_s(j)}{c}\right) \exp\left(-\frac{\sum_{s=1}^{n}(f_s(i) - f_s(j))^+}{c}\right) \\
&= q_j(c)A_{ji}(c).
\end{aligned}$$

This shows that (14) holds, which in turn yields part (a) in Theorem 1. (Note that this proof, with obvious changes, remains valid if the acceptance probability is given by (9) rather than (10)).

(b) Note that for each $a \leq 0$

$$\lim_{x \searrow 0} e^{\frac{a}{x}} = \begin{cases} 1 & \text{if } a = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Now, by (2), (15) and (16)

$$\begin{aligned}
q_i(c) &= \frac{\exp\left(-\frac{\sum_{s=1}^{n} f_s(i)}{c}\right)}{\sum_{j \in S} \exp\left(-\frac{\sum_{s=1}^{n} f_s(j)}{c}\right)} \\
&= \frac{\exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(i)}{c}\right)}{\sum_{j \in S} \exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(j)}{c}\right)} \\
&= \frac{\exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(i)}{c}\right)}{\sum_{j \in S} \exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(j)}{c}\right)} \left(\chi_{\Sigma_{opt}}(i) + \chi_{S - \Sigma_{opt}}(i)\right) \\
&= \frac{1}{\sum_{j \in S} \exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(j)}{c}\right)} \chi_{\Sigma_{opt}}(i) \\
&\quad + \frac{\exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(i)}{c}\right)}{\sum_{j \in S} \exp\left(\frac{\Sigma_m - \sum_{s=1}^{n} f_s(j)}{c}\right)} \chi_{S - \Sigma_{opt}}(i).
\end{aligned}$$

Now let $c \searrow 0$. Then, by (17), the second term of the latter sum goes to 0, whereas the denominator of the first term goes to $|\Sigma_{opt}|$. Hence

$$\lim_{c \searrow 0} q_i(c) = \frac{1}{|\Sigma_{opt}|} \chi_{\Sigma_{opt}}(i) \; + \; 0 \; = \; q_i^*,$$

which completes the proof of part (b).

(c) By (b) and Lemma 2

$$\lim_{c \searrow 0} \lim_{k \to \infty} I\!P\{X_k = i\} = \lim_{c \searrow 0} q_i(c) = q_i^*,$$

and so by (3)

$$\lim_{c \searrow 0} \lim_{k \to \infty} I\!P\{X_k \in \mathcal{P}^*\} \geq \lim_{c \searrow 0} \lim_{k \to \infty} I\!P\{X_k \in \Sigma_{opt}\} = 1. \tag{18}$$

Thus

$$\lim_{c \searrow 0} \lim_{k \to \infty} I\!P\{X_k \in \mathcal{P}^*\} = 1,$$

and (c) follows. ■

5.2 Convergence of Evolutionary Algorithms

In this subsection we present convergence results for the EA for solving MOPs, in which we show that the use of elitism is necessary to guarantee the convergence of this kind of algorithms.

The first result is related to the existence of a stationary distribution for the Markov chain of the EA.

Theorem 2: Let P be the transition matrix of an EA. Then P has a stationary distribution π such that

$$|P_{ij}^k - \pi_j| \leq \left(1 - 2^{nl} p_m^{nl}\right)^{k-1} \quad \forall i, j \in S \;\; \forall k = 1, 2, \ldots. \tag{19}$$

Moreover, π has all entries positive.

Theorem 2 states that P^k converges geometrically to π. Nevertheless, in spite of this result, the convergence of the EA to the Pareto optimal set cannot be guaranteed. In fact, from Theorem 2 and using the fact that π has all entries positive, we immediately deduce the following.

Corollary 2: The EA does not converge.

To ensure convergence of the EA we need to use elitism.

Theorem 3: The EA using elitism converges.

The next lemma will be used in the proof of Theorem 2.

Lemma 4: Let P be the transition matrix of the EA. Then

$$\min_{i,j \in S} P_{ij} = p_m^{nl} > 0 \ \ \forall i, j \in S, \tag{20}$$

and therefore P is primitive.

Proof. By (4) we have

$$p_m < \frac{1}{2} < 1 - p_m.$$

Thus, from (7),

$$\begin{aligned} M_{ij} &= \prod_{s=1}^{n} p_m^{H(i_s,j_s)} (1-p_m)^{l-H(i_s,j_s)} \\ &> \prod_{s=1}^{n} p_m^{H(i_s,j_s)} p_m^{l-H(i_s,j_s)} \ = \ \prod_{s=1}^{n} p_m^{l} \\ &= p_m^{nl} \end{aligned}$$

On the other hand, by (5) and (6)

$$\begin{aligned} P_{ij} &= \sum_{s \in S} R_{is} M_{sj} \\ &\geq p_m^{nl} \sum_{s \in S} R_{is} \\ &= p_m^{nl} \ > \ 0, \end{aligned}$$

To verify (20), observe that P_{ij} attains the minimum in (20) if i has 0 in all entries and j has 1 in all entries.Thus the desired conclusion follows. ■

Proofs

Proof of Theorem 2. Because (20) holds for all $j \in S$ we have that $J = S$, $N_1 = N = 2^{nl}$ and $\nu = 1$. Thus, by Lemma 1, P has a stationary distribution π with all entries positive and we get (19). ■

Despite the fact that Theorem 3 is an extension of a result originally presented by Rudolph [17], our proof is more general. Additionally, we do not have to make any assumptions regarding the existence of a single optimal point (i.e., our proof is simpler), due to the use of essential and inessential states.

Proof of Theorem 3. By Corollary 1, it suffices to show that the states that contain elements in the elite set that are not Pareto optimal are inessential states. To this end, first note that $X = \mathbb{B}^l$ is complete, because it is finite.

Now suppose that there is a state $\hat{i} = (i^e; i)$ in which the elite set contains elements $i^e_{s_1}, \ldots, i^e_{s_k}$ that are not Pareto optimal. Then, as X is complete, there are elements, say $j^e_{s_1}, \ldots, j^e_{s_k} \in \mathcal{P}^*$, that dominate $i^e_{s_1}, \ldots, i^e_{s_k}$, respectively.

Take $\hat{j} = (j^e; j)$ such that all Pareto optimal points of i^e are in j^e and replace the other elements of i^e with the corresponding $j^e_{s_1}, \ldots, j^e_{s_k}$. Thus, all the elements in j^e are Pareto optimal.

Now let

$$j = (j^e_1, \ldots, j^e_r, \underbrace{i^e_{s_1}, \ldots, i^e_{s_1}}_{n-r \text{ copies}}).$$

By Lemma 4 we have $i \to j$. Hence, with positive probability we can pass from (i^e, i) to (i^e, j), and then we apply the elitism operation to pass from (i^e, j) to (j^e, j). This implies that $\hat{i} \to \hat{j}$. On the other hand, using (12), $\hat{j} \not\to \hat{i}$ and therefore $\hat{i}$ is an inessential state.

Finally, from Corollary 1 we have

$$I\!P(\{X^e_k\} \subset \mathcal{P}^*) = I\!P(X_k \in E) = 1 - I\!P(X_k \in I) \to 1 - 0 = 1$$

as $k \to \infty$.

This completes the proof of Theorem 3. ■

5.3 Convergence of an Artificial Immune System Algorithm

A previous proof for a version of MISA was presented in [26], in which some constraints were imposed on the way in which one could go from one state to another. Here, we present a proof of a more general version of MISA. The idea is the same for the EA, and is presented in the next lemma.

Lemma 5: If any state in MISA has in its elite set an element that is not a Pareto optimal, this state is an inessential state.

Proof. Note that $X = I\!B^l$ is complete, because it is finite.

Let $\hat{i} = (i^e; i^1, i^2)$ be a state in which the elite set contains elements that are not Pareto optimal.

1. From i^1, a set of clones is generated. Next, a fixed number of (randomly chosen) string positions of these clones are mutated. Then we change the initial positions in all the strings of the clones (there exists a positive probability of doing this). The set obtained from this previous process is called $ClonesM(i^1)$.
2. Since a uniform mutation is applied to i^2, we change whatever is necessary in all the elements within this set, so that we can obtain the worst element of $ClonesM(i^1)$. As before, there exists a positive probability of doing this, so that none of these elements enters the primary set.
3. Then, all the elements are rearranged and we select the nondominated elements and they are placed in j^1. Now, let j^2 contain a number of individuals of the remainder of the elements available, until completing N (N is the population size).
4. When we apply elitism we obtain the set j^e.
5. To the clones of j^1, we mutate the same initial string positions. Then $ClonesM(j^1) \subseteq ClonesM(i^1)$. Therefore, the best elements of $ClonesM(j^1)$ will be in j^1 again. When we apply elitism to the elements of j^1, we do not modify the set j^e.

6. Let $j^e_{s_1}, \ldots, j^e_{s_k}$ be the elements of j^e that are not Pareto optimal. As X is complete, there exist elements $i^\star_{s_1}, \ldots, j^\star_{s_k} \in \mathcal{P}^*$ that dominate $j^e_{s_1}, \ldots, j^e_{s_k}$, respectively.
7. Now, since we apply uniform mutation to j^2, we can obtain from $j^2_1, \ldots, j^2_k$ to $i^\star_{s_1}, \ldots, j^\star_{s_k}$ respectively, and the other elements of j^2 are left as they were before.
8. Like $ClonesM(j^1)$ and $\{j_{k+1}, \ldots, j_{n_2}\}$ had already been modified j^e when applying elitism, we will not modify again j^e. Thus, the only part of j^e that is modified will be $i^\star_{s_1}, \ldots, j^\star_{s_k}$ and they will replace the nondominated elements of j^e.
9. Finally, let $i^\dagger$ be the resulting state of this process. Using the previous process, we can go from $\hat{i}$ to $i^\dagger$ $(\hat{i} \to i^\dagger)$, but as in $i^{\dagger e}$ there are only Pareto optimal solutions, from (12) $P_{i^\dagger \hat{i}} = 0$ $(i.e\ i^\dagger \not\to \hat{i})$. This proves that $\hat{i}$ is an inessential state. ■

From Lemma 5, the convergence of MISA is easily obtained as follows.

Theorem 4: The MISA algorithm using elitism converges.

Proof. From Lemma 5 and Corollary 1 we have

$$I\!P(\{X^e_k\} \subset \mathcal{P}^*) = I\!P(X_k \in E) = 1 - I\!P(X_k \in I) \to 1 - 0 = 1$$

as $k \to \infty$. This completes the proof. ■

6 Conclusions and Future Work

We have presented the convergence proofs of three meta-heuristics that have been used for solving MOPs: simulated annealing, an artificial immune system (based on clonal selection theory), and a general evolutionary algorithm.

It is worth noting that in the case of the general EA, our convergence proof extends previous proofs of convergence presented for genetic algorithms used for single-objective optimization (e.g., [17]). Actually, our proof is valid for a more general class of evolutionary algorithms that use uniform mutation.

Regarding the artificial immune system, the proof included here, together with some of our previous work [26], constitute the only attempts currently known to prove convergence of such metaheuristic.

Finally, regarding simulated annealing, our proof relies on previous work by Laarhoven, Aarts and Korst [1, 11], but it constitutes (to the best of our knowledge), the first proof of convergence of simulated annealing in multiobjective optimization problems.

As part of our future work, we intend to extend these results to a more general case in which not even uniform mutation is required. We also plan to analyze other types of heuristics used for multiobjective optimization, and to try to determine bounds of convergence for such algorithms.

Acknowledgments

The first author acknowledges support from the Universidad de Costa Rica through a scholarship to pursue graduate studies at the Department of Mathematics of

CINVESTAV-IPN. The second author acknowledges support from NSF-CONACyT project No. 42435-Y. The last author acknowledges partial support from CONACyT grant 37355-E.

References

1. E.H. Aarts and J.H. Korst. *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley, Chichester, UK, 1989.
2. Carlos A. Coello Coello, David A. Van Veldhuizen, and Gary B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, 2002.
3. Nareli Cruz Cortés and Carlos A. Coello Coello. Multiobjective Optimization Using Ideas from the Clonal Selection Principle. In Erick Cantú-Paz et al., editor, *Genetic and Evolutionary Computation—GECCO 2003. Proceedings, Part I*, pages 158–170. Springer. Lecture Notes in Computer Science Vol. 2723, July 2003.
4. Leandro Nunes de Castro and F. J. Von Zuben. Learning and Optimization Using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251, 2002.
5. J. L. Doob. *Stochastic Processes*. Wiley, New York, 1953.
6. David B. Fogel. *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. The Institute of Electrical and Electronic Engineers, New York, 1995.
7. Lawrence J. Fogel. *Artificial Intelligence through Simulated Evolution*. Wiley, New York, 1966.
8. Lawrence J. Fogel. *Artificial Intelligence through Simulated Evolution. Forty Years of Evolutionary Programming*. Wiley, New York, 199.
9. David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 1989.
10. S. Kirkpatrick, C.D. Gellatt, and M.P. Vecchi. Optimization by Simulated Annealing. *Science*, 220:671–680, 1983.
11. P. Laarhoven and E.H. Aarts. *Simulated Annealing: Theory and Applications*. D. Reidel, Boston, MA, 1987.
12. G.F. Lawler. *Introduction to Stochastic Processes*. Chapman & Hall/CRC, Boca Raton, FLA, 1995.
13. Kaisa M. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, 1998.
14. Leandro Nunes de Castro and Jonathan Timmis. *An Introduction to Artificial Immune Systems: A New Computational Intelligence Paradigm*. Springer-Verlag, 2002.
15. V. Černy. A Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, 1985.
16. G. Rudolph and A. Agapie. Convergence Properties of Some Multi-objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, NJ, July 2000. IEEE Press.
17. Günther Rudolph. Convergence Analysis of Canonical Genetic Algorithms. *IEEE Transactions on Neural Networks*, 5:96–101, 1994.
18. Hans-Paul Schwefel. *Evolution and Optimum Seeking*. Wiley, New York, 1995.
19. E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer-Verlag, New York, second edition, 1981.
20. Paolo Serafini. Simulated Annealing for Multiple Objective Optimization Problems. In G.H. Tzeng, H.F. Wang, U.P. Wen, and P.L. Yu, editors, *Proceedings of the Tenth International Conference on Multiple Criteria Decision Making: Expand and Enrich the Domains of Thinking and Application*, volume 1, pages 283–292, Berlin, 1994. Springer-Verlag.

21. E.L. Ulungu. *Optimisation combinatoire multicritere: Determination de l'ensemble des solutions efficaces et methodes interactives*. PhD thesis, Faculté des Sciences, Université de Mons-Hainaut, Mons, Belgium, 1993.
22. E.L. Ulungu, J. Teghem, and Ph. Fortemps. Heuristics for Multi-Objective Combinatorial Optimization by Simulated Annealing. In J. Gu, G. Chen, Q. Wei, and S. Wang, editors, *Multiple Criteria Decision Making: Theory and Applications. Proceedings of the 6th National Conference on Multiple Criteria Decision Making*, pages 228–238, Windsor, UK, 1995. Sci-Tech.
23. E.L. Ulungu, J. Teghem, Ph. Fortemps, and D. Tuyttens. MOSA Method: A Tool for Solving Multiobjective Combinatorial Optimization Problems. *Journal of Multi-Criteria Decision Analysis*, 8(4):221–236, 1999.
24. E.L. Ulungu, J. Teghem, and Ch. Ost. Efficiency of Interactive Multi-Objective Simulated Annealing Through a Case Study. *Journal of the Operational Research Society*, 49:1044–1050, 1998.
25. M. Villalobos-Arias, C. A. Coello Coello, and O. Hernández-Lerma. Asymptotic Convergence of a Simulated Annealing Algorithm for Multiobjective Optimization Problems. Technical Report EVOCINV-02-2004, Evolutionary Computation Group at CINVESTAV, Sección de Computación, Departamento de Ingeniería Eléctrica, CINVESTAV-IPN, México, D.F., March 2004. available at:
`http://delta.cs.cinvestav.mx/~ccoello/2004.html`.
26. M. Villalobos-Arias, C. A. Coello Coello, and O. Hernández-Lerma. Convergence Analysis of a Multiobjective Artificial Immune System Algorithm. In *Artificial Immune Systems: Third International Conference (ICARIS 2004). Proceedings*, pages 226–235, Catania, Sicily, Italy, September 2004. Springer. Lecture Notes in Computer Science Vol. 3239.
27. Eckart Zitzler and Lothar Thiele. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, November 1999.