Convergence Properties of Some Multi-Objective Evolutionary Algorithms

Günter Rudolph
University of Dortmund
Department of Computer Science
44221 Dortmund / Germany
rudolph@LS11.cs.uni-dortmund.de

Alexandru Agapie
National Institute of Microtechnology
Computational Intelligence Lab
72225 Bucharest / Romania
agapie@imt.pub.ro

Abstract-
We present four abstract evolutionary algorithms for multi-objective optimization and theoretical results that characterize their convergence behavior. Thanks to these results it is easy to verify whether a particular instantiation of these abstract evolutionary algorithms offers the desired limit behavior or not. Several examples are given.

1 Introduction

Theoretical results on multi-objective evolutionary algorithms are hardly available. This work extends the results given in Rudolph (1998a) and van Veldhuizen (1999) for finitely large search spaces. Related work treating continuous search spaces may be found in Rudolph (1998b) and Hanne (1999).

The plan is as follows: It is assumed that the evolutionary algorithms are Markov processes which have to cope with partially ordered fitness values—this includes optimization under a single objective function as well as multiple objective functions. Therefore section 2 recalls some background material concerning partially ordered sets and finite Markov chains. Abstract versions of evolutionary algorithms and their convergence behavior are presented in section 3, whereas section 4 contains results how to verify the preconditions of the convergence results along with a variety of examples. Finally, we draw our conclusions in section 5.

2 Mathematical Prelude

2.1 Partially Ordered Sets

Let $\mathcal{F}$ be some set. A reflexive, antisymmetric, and transitive relation "$\preceq$" on $\mathcal{F}$ is termed a *partial order relation* whereas a *strict partial order relation* "$\prec$" must be antireflexive, asymmetric, and transitive. The latter relation may be obtained by the former one by setting $x \prec y := (x \preceq y) \wedge (x \neq y)$. After these preparations one is in the position to turn to the actual objects of interest.

Definition 1 *Let $\mathcal{F}$ be some set. If the partial order relation "$\preceq$" is valid on $\mathcal{F}$ then the pair $(\mathcal{F}, \preceq)$ is called a* partially ordered set *(or short:* poset*). If $x \prec y$ for some $x, y \in \mathcal{F}$ then x is said to* dominate *y. Distinct points $x, y \in \mathcal{F}$ are said to be* comparable *when either $x \prec y$ or $y \prec x$. Otherwise, x and y are* incomparable *which is denoted by $x \parallel y$. If each pair of distinct points of a poset $(\mathcal{F}, \preceq)$ is comparable then $(\mathcal{F}, \preceq)$ is called a* totally ordered set *or a* chain*. Dually, if each pair of distinct points of a poset $(\mathcal{X}\mathcal{F}, \preceq)$ are incomparable then $(\mathcal{F}, \preceq)$ is termed an* antichain. □

For example, $(\mathbb{R}^n, \preceq)$ with $n \geq 2$ is a partially ordered set when $x \preceq y$ means $x_i \leq y_i$ for all $i = 1, \ldots, n$. One obtains a strict partial order relation "$\prec$" from this partial order relation if it is additionally required that $x \neq y$. Notice that the poset $(\mathbb{R}^n, \preceq)$ is neither a chain nor an antichain. The situation changes for the poset $(\mathbb{R}, \preceq)$ with $x \preceq y$ if and only if $x \leq y$. Since each pair of distinct points in $\mathbb{R}$ is comparable the poset $(\mathbb{R}, \preceq)$ is totally ordered and therefore a chain. An example for an antichain is the set of "minimal elements" introduced next.

Definition 2 *An element $x^* \in \mathcal{F}$ is called a* minimal element *of the poset $(\mathcal{F}, \preceq)$ if there is no $x \in \mathcal{F}$ such that $x \prec x^*$. The set of all minimal elements, denoted $\mathcal{M}(\mathcal{F}, \preceq)$, is said to be* complete *if for each $x \in \mathcal{F}$ there is at least one $x^* \in \mathcal{M}(\mathcal{F}, \preceq)$ such that $x^* \preceq x$.* □

If the poset $(\mathcal{F}, \preceq)$ is finite then the completeness of $\mathcal{M}(\mathcal{F}, \preceq)$ is guaranteed—in contrast to infinitely large posets.

Let $f : \mathcal{X} \to \mathcal{F}$ be a mapping from some set $\mathcal{X}$ to the poset $(\mathcal{F}, \preceq)$. For some $A \subseteq \mathcal{X}$ the set

$$\mathcal{M}_f(A, \preceq) = \{a \in A : f(a) \in \mathcal{M}(f(A), \preceq)\}$$

contains those elements from A whose images are minimal elements in the image space $f(A) = \{f(a) : a \in A\}$.

2.2 Finite Markov Chains

If S is a finite set and $\{X_t : t \in \mathbb{N}_0\}$ an S-valued random sequence with the property

$$\mathsf{P}\{X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0\} = \mathsf{P}\{X_{t+1} = j \mid X_t = i\} = p_{ij}$$

for all $t \geq 0$ and for all pairs $(i, j) \in S \times S$ then the sequence $\{X_t : t \in \mathbb{N}_0\}$ is called a homogeneous finite Markov chain with state space S. Since S is finite the transition probabilities can be gathered in the transition matrix $P = (p_{ij})_{i,j \in S}$. The row vector $\pi(t)$ with $\pi_i(t) = \mathsf{P}\{X_t = i\}$ denotes the distribution of the Markov chain at step $t \geq 0$. Since

$$\pi(t) = \pi(t-1)\,P = \pi(0)\,P^t$$

for all $t \geq 1$, a homogeneous finite Markov chain is completely specified by its initial distribution $\pi(0)$ and its transition matrix P. The kth step transition probabilities are

$$p_{ij}(k) = \mathsf{P}\{\, X_k = j \mid X_0 = i \} = e_i \, P^k \, e_j' \,,$$

where e_i is the ith unit vector, such that

$$\pi_j(t) = \sum_{i \in S} \pi_i(0) \cdot p_{ij}(t) \,.$$

A matrix $A : n \times m$ is termed *nonnegative* if $a_{ij} \geq 0$ and *positive* if $a_{ij} > 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. A nonnegative matrix is called *stochastic* if each row sum equals one. Thus, transition matrices are stochastic. A stochastic square matrix P is *irreducible* if

$$\forall i, j \in S : \exists k \in \mathbb{N} : p_{ij}(k) > 0$$

and it is *primitive* or *regular* if

$$\exists k \in \mathbb{N} : \forall i, j \in S : p_{ij}(k) > 0 \,.$$

Therefore, every positive matrix P is regular and every regular matrix P is irreducible.

Lemma 1 (Iosifescu 1980, p. 89)
A homogeneous Markov chain with finite state space and irreducible transition matrix visits every state infinitely often with probability one regardless of the initial distribution. □

In order to clarify the notion of "stochastic convergence to the set of minimal elements" we need measures on the distances between finite point sets. The first measure used here is characterized as follows:

Remark 1 *If A and B are subsets of a finite ground set X then $d(A, B) = |A \cup B| - |A \cap B|$ is a metric on the power set of X.*
Proof: Let $X = \{X_1, X_2, \ldots, X_n\}$ and $a \in \{0, 1\}^n$ with $a_i = 1_A(X_i)$ be the incidence vector of set $A \subseteq X$. Since

$$\begin{aligned} |A \cap B| &= \sum_{i=1}^{n} a_i \, b_i \quad \text{and} \\ |A \cup B| &= \sum_{i=1}^{n} (a_i + b_i - a_i \, b_i) \end{aligned}$$

one obtains

$$\begin{aligned} d(A, B) &= \sum_{i=1}^{n} (a_i - 2\, a_i \, b_i + b_i) \\ &= \sum_{i=1}^{n} [(1 - b_i)\, a_i + (1 - a_i)\, b_i] \\ &= \sum_{i=1}^{n} |a_i - b_i| \;=\; \|a - b\|_1 \,. \end{aligned}$$

Thus, $d(A, B)$ is equivalent to the Hamming distance between the associated incidence vectors and therefore a metric on X. □

The second measure uses the quantity $\delta_B(A) = |A| - |A \cap B|$ counting the number of elements that are in set A but not in set B.

Definition 3 *Let A_t be the population of some evolutionary algorithm at iteration $t \geq 0$ and $F_t = f(A_t)$ its associated image set. The evolutionary algorithm is said to* converge with probability 1 to the entire set of minimal elements *if*

$$d(F_t, \mathcal{F}^*) \to 0 \quad \textit{with probability 1 as } t \to \infty$$

whereas it is said to converge with probability 1 to the set of minimal elements *if*

$$\delta_{\mathcal{F}^*}(F_t) \to 0 \quad \textit{with probability 1 as } t \to \infty$$

Here, $\mathcal{F}^*$ denotes the set of minimal elements. □

It is clear that $d(F_t, \mathcal{F}^*) \to 0$ implies $\delta_{\mathcal{F}^*}(F_t) \to 0$. But in the first case the populations size will eventually grow at least to the size of the set $\mathcal{F}^*$. If $(\mathcal{F}, \preceq)$ is a chain (e.g., if there is only a single objective) then $|\mathcal{F}^*| = 1$. But if $(\mathcal{F}, \preceq)$ is not a chain (e.g., if there are multiple objectives) then $|\mathcal{F}^*|$ may be almost as large as the search space. As a consequence, the population size should have a manageable maximum size. Unfortunately, this requirements makes the selection procedure more complicated. The next section illustrates this fact.

3 Base Algorithms & Their Analysis

Let $\mathcal{X}$ be the finite search space and $f : \mathcal{X} \to \mathcal{F} = \{f(x) : x \in \mathcal{X}\}$ the fitness function where $(\mathcal{F}, \preceq)$ is a partially ordered set. The target of the evolutionary search is the detection of some or all members of the set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq)$. Each of the following subsections presents an "evolutionary base algorithm" and its convergence property subject to certain conditions. We use the term "base algorithm" because they may be instantiated with many variation and selection operators without affecting the convergence properties in the negative.

3.1 Base Algorithm VV

This base algorithm matches the preconditions of a result in van Veldhuizen (1999) in the broadest sense.

```
B(0) is drawn at random from X^n
A(0) = M_f(B(0), ⪯)
t = 0
repeat
  B(t+1) = generate(B(t))
  A(t+1) = M_f(A(t) ∪ B(t+1), ⪯)
  t ← t + 1
until stopping criterion fulfilled
```

The proof of the result given in van Veldhuizen (1999) is based on a result presented in Bäck (1996), p. 129, who reproduced an unpublished result of Hartl (1990) incompletely:

Additional conditions imposed on the selection procedure are missing. Moreover, Hartl's proof tacitly presupposes a totally ordered fitness set so that some special features of partially (not totally) ordered sets are inevitably not taken into account. For these reasons we have provided a new proof here.

Proposition 1
If the sequence $(B_t)_{t\geq 0}$ is a homogeneous finite Markov chain with irreducible transition matrix then $d(f(A_t), \mathcal{F}^*) \to 0$ with probability one and in mean as $t \to \infty$.
Proof: By construction of the algorithm it is guaranteed that the image set $f(A_t)$ of A_t is an antichain and therefore the set of minimal elements of the poset $(f(A_t), \preceq)$ for all $t \geq 0$. As soon as an element of $\mathcal{F}^* = \mathcal{M}(\mathcal{F}, \preceq)$ has entered $f(A_t)$ it will stay there forever. It remains to show that all elements of $\mathcal{F}^*$ will be contained in $f(A_\tau)$ for some random time τ with $\mathsf{P}\{\tau < \infty\} = 1$. Let $B^*(t) = \mathcal{M}_f(B(t), \preceq)$ and notice that $\mathcal{M}_f(A(t) \cup B(t+1), \preceq) = \mathcal{M}_f(A(t) \cup B^*(t+1), \preceq)$. Let $a \in A(t_0)$ with $f(a) \notin \mathcal{F}^*$. Since $(\mathcal{F}, \preceq)$ is complete it is guaranteed that there exists an elements $x \in \mathcal{X}$ such that $f(x) \prec f(a)$, i.e., a non-optimal element will be discarded by a better (and finally optimal) one provided that such an element will occur in the sequence $(B(t))_{t\geq 0}$ at some iteration t with $t > t_0$. Since the Markov chain is irreducible Lemma 1 ensures that every element of $\mathcal{X}^n$ will be visited infinitely often. This implies that the waiting time of the first occurrence as well as between two consecutive occurrences of x is finite with probability one. Therefore non-optimal elements will be eliminated after a finite number of iterations with probability one. Moreover, each element $b \in B^*(t+1)$ that is incomparable to all elements in $A(t)$ will enter $A(t+1)$. If such a b is optimal it will be member of each set $A(\cdot)$ after iteration t. If it is not optimal then it will be replaced in finite time by an optimal element (see above). The appearance of such incomparable elements b is ensured by the irreducibility of the Markov chain $(B_t)_{t\geq 0}$. Summing up: All optimal elements will enter the set $A(\cdot)$ in finite time with probability one and as soon as this has happened all non-optimal elements have been discarded. Since optimal elements cannot get lost one gets $d(f(A_t), \mathcal{F}^*) \to 0$ with probability one and, due to the boundedness of $d(\cdot, \mathcal{F}^*)$, also in mean. □

This base algorithm has a disadvantage: The size of the sets A_t will finally grow to the size of the set of minimal elements. Since this size may be huge this base algorithm offers only limited usefulness in practice.

3.2 Algorithm AR1

We now describe a variation of base algorithm VV. In order to constrain the size of the sets A_t the selection procedure must be altered considerably. Let $n = |B_t|$ and $m \geq n$ where m denotes the maximum size of the sets A_t.

```
B(0) is drawn at random from X^n
A(0) = M_f(B(0), ⪯)
t = 1
repeat
  B(t) = generate(B(t − 1))
  B*(t) = M_f(B(t), ⪯)
  C(t) = ∅
  foreach b ∈ B*(t) do
    D_b = {a ∈ A(t) : f(b) ≺ f(a)}
    if D_b ≠ ∅ then A(t) ← (A(t) \ D_b) ∪ {b}
    if ∀a ∈ A(t) : f(a) || f(b) then C(t) ← C(t) ∪ {b}
  endfor
  k = min{m − |A(t)|, |C(t)|}
  A(t + 1) = A(t) ∪ draw(k, C(t))
  t ← t + 1
until stopping criterion fulfilled
```

Here, the function *draw*(k, C) returns a set of at most k distinct elements from set C drawn by an arbitrary method.

Proposition 2
If the sequence $(B_t)_{t\geq 0}$ is a homogeneous finite Markov chain with irreducible transition matrix then $\delta_{\mathcal{F}^*}(f(A_t)) \to 0$ and $|A_t| \to \min\{m, |\mathcal{F}^*|\}$ with probability one and in mean as $t \to \infty$.
Proof: By construction of the algorithm it is guaranteed that the image set $f(A_t)$ of A_t is an antichain and therefore the set of minimal elements of the poset $(f(A_t), \preceq)$ for all $t \geq 0$. An element $a \in A_t$ is deleted if and only if there is an element in B_t (resp. B^*_t) whose image dominates $f(a)$. Therefore an element of $\mathcal{F}^* = \mathcal{M}(\mathcal{F}, \preceq)$ will be a member of the sequence $f(A_t)_{t\geq\tau}$ as soon as it has entered $f(A_\tau)$. If such an element b enters A_t then at least one member of A_t is discarded. Elements in C_t are incomparable to all members of A_t. Since the size of C_t may be as large as n it is necessary to include only that many elements of C_t in A_t such that $|A_t|$ does not exceed m. This is realized by the operation $A_{t+1} = A_t \cup \text{draw}(k, C_t)$. It remains to show that non-optimal elements in the sequence $f(A_t)$ will be replaced by optimal elements in finite time. This can be verified by the same argumentation as in the proof of proposition 1: Since $(B_t)_{t\geq 0}$ is an irreducible Markov chain optimal elements will be generated infinitely often with probability 1. These elements can enter the set $A(\cdot)$ directly, if they dominate elements therein, or via the sets $C(\cdot)$. Consequently, $\delta_{\mathcal{F}^*}(f(A_t)) \to 0$ with probability 1 and in mean. □

Remark 2 *If the image set $f(\mathcal{X})$ is totally ordered then the base algorithms VV and AR1 are identical.* □

3.3 Algorithm PR

The base algorithms considered so far were using the sets $A(\cdot)$ as an archive and not as the sets of parents. The next base algorithm originating from Peschel and Riedel (1977) makes $A(\cdot)$ to the set of parents.

$B(0)$ is drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 0$
repeat
  $B(t+1) = \text{generate}(A(t))$
  $A(t+1) = \mathcal{M}_f(A(t) \cup B(t+1), \preceq)$
  $t \leftarrow t + 1$
until stopping criterion fulfilled

Proposition 3
Let G be the homogeneous stochastic matrix describing the transition behavior from $A(t)$ to $B(t+1)$. If matrix G is positive then $d(f(A_t), \mathcal{F}^*) \to 0$ with probability one and in mean as $t \to \infty$.
Proof: See Rudolph (1998a), p. 351. □

Two points deserve special mention: First, the size of the sets $A(\cdot)$ will grow to $|\mathcal{F}^*|$ limiting the practical use of this algorithm in general (especially if $|\mathcal{F}^*|$ is large). Second, if matrix G is an irreducible (or primitive) but *non-positive* transition matrix then convergence cannot be guaranteed in general. To see this consider the following example: Let $\mathcal{X} = \{0,1\}^3$ whose partially ordered image set $f(\mathcal{X})$ obeys the relations

$$\begin{tikzcd}
f(000) & f(110) \arrow[l, "\prec"'] & f(010) \arrow[l, "\prec"'] & f(100) \arrow[l, "\prec"'] \arrow[dl, "\prec"'] \\
f(001) & f(111) \arrow[l, "\prec"'] & f(101) \arrow[l, "\prec"'] & f(011) \arrow[l, "\prec"']
\end{tikzcd}$$

Thus, $f(000)$ and $f(001)$ are minimal and therefore incomparable elements whereas incomparable and non-minimal elements are, for example, $f(111)$ and $f(010)$.

Now suppose that $A(t) = \{110, 111\}$ for some $t \geq 0$ and that the variation operation, represented by the function *generate(·)*, inverses exactly one bit uniformly at random in the population. In this case transition matrix G is irreducible but not positive. Similarly, if either exactly one bit is inverted at random in the individual or the individual remains unaltered then matrix G is primitive but not positive. Since $B(t+1)$ can contain only elements from $\{010, 011, 100, 101, 110, 111\}$ and neither of them dominates 110 or 111, one obtains $A(t+1) = A(t) = \{110, 111\}$ in the next step. Thus, the set of minimal elements is not reachable from population $\{110, 111\}$ which in turn precludes convergence to $\mathcal{F}^*$.

3.4 Algorithm AR2

Now we describe a variant of base algorithm PR; actually, we simply plug in the selection method already used in algorithm AR1. Again, let $n = |B_t|$ and $m \geq n$ where m denotes the maximum size of the sets A_t.

$B(0)$ is drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 1$
repeat
  $B(t) = \text{generate}(A(t-1))$
  $B^*(t) = \mathcal{M}_f(B(t), \preceq)$
  $C(t) = \emptyset$
  foreach $b \in B^*(t)$ **do**
    $D_b = \{a \in A(t) : f(b) \prec f(a)\}$
    if $D_b \neq \emptyset$ **then** $A(t) \leftarrow (A(t) \setminus D_b) \cup \{b\}$
    if $\forall a \in A(t) : f(a) \parallel f(b)$ **then** $C(t) \leftarrow C(t) \cup \{b\}$
  endfor
  $k = \min\{m - |A(t)|, |C(t)|\}$
  $A(t+1) = A(t) \cup \text{draw}(k, C(t))$
  $t \leftarrow t + 1$
until stopping criterion fulfilled

Again, the function *draw*(k, C) returns a set of at most k distinct elements from set C drawn by an arbitrary method.

Before we state and prove the limit properties of this base algorithm it is useful to collect a number of facts. The first two follow immediately from the construction of the algorithm.

Fact 1
If an optimal element has entered $A(\cdot)$ it stays there forever.

Fact 2
If $b \in B^*(\cdot)$ and $f(b)$ dominates elements of $f(A(\cdot))$ then b moves to $A(\cdot)$ and the dominated elements leave $A(\cdot)$.

The next fact captures a limitation of this base algorithm.

Fact 3
If $b \in B^*(\cdot)$ is optimal then it moves either to $A(\cdot)$ or $C(\cdot)$.

Thus, in case of $A(\cdot) = m$ an optimal element cannot be included even if $A(\cdot)$ contains non-optimal elements. This happens if the non-optimal elements are incomparable to all other members of $A(\cdot)$ as well as the optimal element just found.

The last fact follows from the completeness of the poset $(f(\mathcal{X}), \preceq)$.

Fact 4
If there is a non-optimal element in $A(\cdot)$ there exists a dominating element.

After these preparations we can state our result:

Proposition 4
Let G be the homogeneous stochastic matrix describing the transition behavior from $A(t)$ to $B(t+1)$. If matrix G is positive then $\delta_{\mathcal{F}^*}(f(A_t)) \to 0$ and $|A_t| \to \min\{m, |\mathcal{F}^*|\}$ with probability one and in mean as $t \to \infty$.
Proof: Let $m \leq |\mathcal{F}^*|$ and suppose that not all members of $A(\cdot)$ are optimal. Fact 4 ensures that there exists at least one element in $\mathcal{X}$ whose image dominates an image of an element in $A(\cdot)$. Since G is positive there exists a positive minimum

probability that an arbitrary element of $\mathcal{X}$ (and even $\mathcal{X}^n$) is created by the operation *generate($\cdot$)* in one step. Owing to the Borel-Cantelli Lemma (see e.g. Feller 1970, p. 201) it is guaranteed that this arbitrary (and therefore every) element will be generated infinitely often and that the waiting time for the first occurrence as well as for the second, third, and so forth will be finite with probability 1. Consequently, a dominating element is generated in finite time with probability one and, according to Fact 2, it will enter the set $A(\cdot)$. By Fact 1, this element will stay in $A(\cdot)$ forever iff this element is optimal. If it is not optimal, then it will be replaced by an optimal one after finite time by a repetition of the arguments given so far. Summing up: The size of $A(\cdot)$ increases up to m by including optimal or non-optimal incomparable elements. Non-optimal elements in $A(\cdot)$ will be replaced by optimal elements in finite time.

Let $|\mathcal{F}^*| < m$. As long as $|A(\cdot)| \leq |\mathcal{F}^*|$ the dynamics of the algorithm is as in the previous case. The only interesting difference surfaces if $|\mathcal{F}^*| < |A(\cdot)| \leq m$. If these inequalities are valid then the number of optimal elements in $A(\cdot)$ is necessarily less than $|\mathcal{F}^*|$. Thus, $A(\cdot)$ contains several non-optimal incomparable elements. But owing to Fact 4 and the positive transition matrix G it is ensured that optimal elements dominating the non-optimal ones will be generated. It is clear that the optimal elements replace the non-optimal ones. As soon as all optimal elements have been generated (which happens in finite time), set $A(\cdot)$ has size $|\mathcal{F}^*|$ with $f(A(\cdot)) = \mathcal{F}^*$ and no other element can enter anymore. Since optimal elements cannot get lost by Fact 1 we have established convergence in the sense of the proposition. □

Remark 3 *If the image set $f(\mathcal{X})$ is totally ordered then the base algorithms PR and AR2 are identical. They reduce to an evolutionary algorithm with a $(1 + n)$-selection scheme.* □

4 Instantiations

The proofs of the previous section have shown that we only need to check whether the transition matrices are irreducible (in case of base algorithms VV and AR-1) or positive (cases PR and AR-2) in order to get convergence results. Since the transition matrix, as it appears in the previous section, is usually a product of several other transition matrices (describing e.g. mutation, crossover, pre-selection et cetera) it is useful to find certain characteristics of stochastic matrices that imply positiveness or irreducebility of the product of such matrices. Here, we need two additional definitions: A stochastic matrix is termed *diagonal-positive* if every diagonal entry is nonzero, whereas it is called *column-allowable* if each column contains at least one positive entry. Thus, every diagonal-positive matrix is column-allowable.

Lemma 2
Let I, D, C, P, A be stochastic matrices where I is irreducible, D is diagonal-positive, C column-allowable, P positive, and A arbitrary. Then the products
(a) AP and PC are positive,
(b) ID and DI are irreducible.

Proof:
(a) Rudolph (1994), p. 97. (b) Agapie (1998a), p. 188. □

Remark 4
The conjecture that IC or CI are irreducible is wrong as can be seen from the following counter-example: Let

$$I = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Notice that both stochastic matrices are irreducible as well as column-allowable. Since C is the inverse of I we obtain $IC = CI = U$ where U is the unit matrix which is of course not irreducible. □

Hereinafter the following assumptions have been made: First, the search space $\mathcal{X}$ is a product space $\mathcal{X} = \mathcal{A}^\ell$ where $\mathcal{A}$ is a finite set. Second, the final transition matrix is a product of three transition matrices describing the stochastic effects of crossover, mutation, and "pre-selection." We use the term "pre-selection" to denote any form of favoring certain individuals prior to the elitism-like selection procedure in the main loop of the base algorithms, i.e., we only consider evolutionary operators that are subsumed in the function *generate($\cdot$)*.

Since neither crossover nor pre-selection possess irreducible (or primitive or positive) transition matrices—unusual constructions excluded—the mutation operator must establish such a property. We now generalize the special case with $\mathcal{A} = \{0, 1\}$ already considered in section 3.3.

(a) *Single spot mutation*
Let $x \in \mathcal{X} = \mathcal{A}^\ell$ be some individual. Choose an index k between 1 and ℓ at random. The entry x_k is now mutated according to some probability distribution on $\mathcal{A}$. If every element of $\mathcal{A}$ is accessible with positive probability, then the transition matrix for mutation is primitive (and not positive). If every element of $\mathcal{A}$ except the value of x_k may occur with positive probability then the transition matrix for mutation is irreducible (and not primitive).

(b) *Multiple spot mutation*
Instead of choosing a single index at random, the mutation operation is now applied at each index $k = 1, \ldots, \ell$. If x_k may assume every element of $\mathcal{A}$ with positive probability, then the transition matrix for mutation is positive. If every element of $\mathcal{A}$ except the value of x_k may occur with positive probability then the transition matrix for mutation is irreducible (and not primitive).

Consequently, in order to establish the desired convergence of base algorithms PR or AR-2 we choose the first version of multiple spot mutation (positive transition matrix), an arbitrary crossover operator, and a column-allowable pre-selection operator (see Lemma 2 and Propositions 3 resp. 4).

It remains to characterize column-allowable pre-selection operators. If pre-selection is omitted, i.e., pre-selection is the identity operation, then the transition matrix is the unit matrix and therefore column-allowable. Alternatively, one may proceed as follows: The population is partitioned into antichains $A_1, A_2, \ldots$ such that members of A_i dominate members of A_j with $j > i$. Then individuals of antichain A_i are given rank i and one may use traditional selection operators based on totally ordered fitness/ranks (like proportional or tournament selection) in order to get a pre-selection of individuals. Since proportional and tournament selection may choose every individual with positive probability, there is a positive probability that pre-selection leaves the population unaltered. As a consequence, the transition matrix is diagonal-positive and therefore column-allowable. This remains true if the individuals within each antichain are additionally ranked by the number of individuals they are dominating. See the surveys of Fonseca and Fleming (1995) or Horn (1997) for a variety of similar approaches for introducing a total order on the individuals.

Finally, we have to look for crossover operations with diagonal-positive transition matrices. Under the assumption that the individuals participating in the crossover operation are drawn with replacement we obtain immediately a diagonal-positive transition matrix because there is a positive probability that a ρ-ary crossover operator draws the same individual ρ times such that the preliminary offspring is identical to its parents (e.g., single or multi-point crossover, gene pool recombination and others).

5 Conclusions

Many versions of multi-criteria evolutionary algorithms fit into the theoretical framework developed here. The conditions for convergence are "user-friendly" in the sense that it suffices in many cases to verify properties of single operators in lieu of properties of the transition function of the entire evolutionary algorithm. Needless to say, there are still many versions of evolutionary algorithms that we did not examine here (and which possibly do not match our basic assumptions). All kinds of fitness sharing mechanisms are just one example.

Even though base algorithms VV and PR (the same with RA1 and RA2) look very much alike, the stochastic models behind them are quite different: the "working population" of VV was $(B_t)_{t\geq 0}$ – a homogeneous Markov chain with transition matrix $\bar{G}$, whereas for PR the sequence of the "working population" is $\{B_0, A_0, B_1, A_1, \ldots\}$ which is no longer Markovian, i.e., the irreducibility of transition matrix G from B_t to A_t is no longer a convergence condition because G itself characterizes only "half" of the Markov chain.

Rather, the entire process is a *random system with complete connections* (Iosifescu and Grigorescu 1990) which generalizes the notion of a Markov chain and may be an appropriate model for more sophisticated evolutionary algorithms (Agapie 1998b).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the *Collaborative Research Center "Computational Intelligence" (SFB 531)*. The second author gratefully acknowledges financial support of the DFG while being guest researcher of the SFB at the University of Dortmund during Oct.-Dec. 1999.

Bibliography

A. Agapie (1998a). Genetic algorithms: Minimal conditions for convergence. In J. K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers (Eds.), *Artificial Evolution: Third European Conference; selected papers / AE '97*, pp. 183–193. Berlin: Springer.

A. Agapie (1998b). Modelling genetic algorithms: From markov chains to dependance with complete connections. In T. Bäck, A. E. Eiben, M. Schoenauer, and H.-P. Schwefel (Eds.), *Parallel Problem Solving From Nature — PPSN V*, pp. 3–12. Berlin: Springer.

T. Bäck (1996). *Evolutionary Algorithms in Theory and Practice*. New York: Oxford University Press.

W. Feller (1970). *An Introduction to Probability Theory and Its Applications Vol. 1* (3rd (revised) ed.). Singapore: Wiley.

C. M. Fonseca and P. J. Fleming (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation 3*(1), 1–16.

T. Hanne (1999). On the convergence of multiobjective evolutionary algorithms. *European Journal of Operational Research 117*(3), 553–564.

R. F. Hartl (1990). A global convergence proof for a class of genetic algorithms. Technical report, Institute of Econometrics, Operations Research and Systems Theory, Vienna University of Technology.

J. Horn (1997). Multicriterion decision making. In T. Bäck, D. B. Fogel, and Z. Michalewicz (Eds.), *Handbook of Evolutionary Computation*, pp. F1.9:1–15. New York and Bristol (UK): IOP Publishing and Oxford University Press.

M. Iosifescu (1980). *Finite Markov Processes and Their Applications*. Chichester: Wiley.

M. Iosifescu and S. Grigorescu (1990). *Dependence with complete connections and its applications*. Cambridge: Cambridge University Press.

M. Peschel and C. Riedel (1977). Use of vector optimization in multiobjective decision making. In D. E. Bell, R. L. Keeney, and H. Raiffa (Eds.), *Conflicting Objectives in Decisions*, pp. 97–121. Chichester: Wiley.

G. Rudolph (1994). Convergence properties of canonical genetic algorithms. *IEEE Transactions on Neural Networks 5*(1), 96–101.

G. Rudolph (1998a). Evolutionary search for minimal elements in partially ordered finite sets. In V. W. Porto, N. Saravanan, D. Waagen, and A. E. Eiben (Eds.), *Evolutionary Programming VII, Proceedings of the 7th Annual Conference on Evolutionary Programming*, pp. 345–353. Berlin: Springer.

G. Rudolph (1998b). On a multi–objective evolutionary algorithm and its convergence to the Pareto set. In *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation*, pp. 511–516. Piscataway (NJ): IEEE Press.

D. A. van Veldhuizen (1999). *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. Ph. D. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.