Evolutionary Search for Minimal Elements in Partially Ordered Finite Sets

Günter Rudolph

Universität Dortmund, Fachbereich Informatik, D–44221 Dortmund / Germany

Abstract. The task of finding minimal elements of a partially ordered set is a generalization of the task of finding the global minimum of a real–valued function or of finding pareto–optimal points of a multicriteria optimization problem. It is shown that evolutionary algorithms are able to converge to the set of minimal elements in finite time with probability one, provided that the search space is finite, the time–invariant variation operator is associated with a positive transition probability function and that the selection operator obeys the so–called 'elite preservation strategy.'

1 Introduction

Traditionally, evolutionary algorithms (EAs) were used to find or approximate the global minimum of a real–valued objective function $f : \mathcal{S} \to \mathbb{R}$ defined on some non–empty set $\mathcal{S}$. The development of EAs that can cope with more than a single objective function began in the mid–1980s but for a long time this field of application did not receive the resonance that it should have deserved. The situation changed during the last five years: As can be learned from recent surveys [1–3] there are now numerous suggestions of how to design multi–objective EAs. But these activities were not accompanied by the development of a theoretical foundation. A similar time lag between practice and theory can be observed in case of single–objective EAs. It is therefore the goal of this paper to seed a starting point regarding a theory of multi–objective EAs that may initiate further research on that subject.

The main difference between single– and multi–objective optimization rests on the fact that two elements are not guaranteed to be comparable in the latter case. To understand the problem to full extent it is important to keep in mind that the values $f_1(x), \ldots, f_m(x)$ of the $m \geq 2$ objective functions represent *incommensurable* quantities that cannot be minimized simultaneously: While f_1 may measure costs, f_2 may measure the level of pollution, f_3 the pressure of some boiler, and so forth. As a consequence, the notion of the "optimality" of some solution needs a more general formulation as in the single–criterion case. It seems reasonable to regard those elements as being optimal which cannot be improved with respect to one criterion without getting a worse value in another criterion. Elements with this property are said to be *Pareto–optimal* in this context.

From a more general point of view single– as well as multi–objective optimization can be seen as special cases of the task to find *minimal elements* of *partially ordered sets*. The meaning of these terms will be made rigorous in Section 2. After these preparations it is shown in Section 3 that EAs with finite search space and partially ordered fitness values do stochastically converge to minimal elements provided that the selection mechanism employs some kind of elitism and that the time–invariant variation operator's support is identical to the search space. This result includes earlier established convergence results regarding single–objective EAs with finite search space (e.g. [4–7]) as special cases. Moreover, it also includes a *new* convergence result for multi–objective EAs. Section 4 is devoted to transcribe the general result into the terminology of these special cases. Finally, some directions towards an extension of the presented theory are discussed in Section 5.

2 Partially Ordered Sets

A prerequisite to introduce 'partially ordered sets' is the notion of the 'relation.' The definitions presented in this section are extracted from [8] and [9].

Definition 1. *Let $\mathcal{X}$ be some set. The subset $\mathcal{R} \subseteq \mathcal{X} \times \mathcal{X}$ is called a* binary relation *in $\mathcal{X}$. Let $x, y \in \mathcal{X}$. If $(x, y) \in \mathcal{R}$, also denoted $x\mathcal{R}y$, then x is said to be in relation $\mathcal{R}$ to y. A relation $\mathcal{R}$ in $\mathcal{X}$ is said to be*

(a) reflexive *if $x\mathcal{R}x$ is true for all $x \in \mathcal{X}$,*
(b) antireflexive *if $x\mathcal{R}y \Rightarrow x \neq y$ is true for all $x, y \in \mathcal{X}$,*
(c) symmetric *if $x\mathcal{R}y \Rightarrow y\mathcal{R}x$ is true for all $x, y \in \mathcal{X}$,*
(d) antisymmetric *if $x\mathcal{R}y \wedge y\mathcal{R}x \Rightarrow x = y$ is true for all $x, y \in \mathcal{X}$,*
(e) asymmetric *if $x\mathcal{R}y \Rightarrow \overline{y\mathcal{R}x}$ is true for all $x, y \in \mathcal{X}$,*
(f) transitive *if $x\mathcal{R}y \wedge y\mathcal{R}z \Rightarrow x\mathcal{R}z$ is true for all $x, y, z \in \mathcal{X}$.* □

Some relations that possess several of the properties above simultaneously bear their own names. For example, if $\mathcal{R}$ is a reflexive, symmetric and transitive relation then $\mathcal{R}$ is called an *equivalence relation*. In this case it is common to use the symbol "$\sim$" in lieu of $\mathcal{R}$. A reflexive, antisymmetric and transitive relation "$\preceq$" is termed a *partial order relation* whereas a *strict partial order relation* "$\prec$" must be antireflexive, asymmetric and transitive. The result below shows how to obtain a strict partial order relation from a partial order relation.

Lemma 1. *Let $\mathcal{X}$ be some set and let "$\preceq$" denote a reflexive, antisymmetric and transitive relation on $\mathcal{X}$. Then the relation "$\prec$" defined by*

$$x \prec y \Leftrightarrow (x \preceq y) \wedge (x \neq y) \tag{1}$$

is antireflexive, asymmetric and transitive on $\mathcal{X}$.

Proof.
(a) $x \prec y \Rightarrow x \neq y$ (antireflexive):
Follows directly from the definition of relation $\prec$.

(b) $x \prec y \Rightarrow \overline{y \prec x}$ (asymmetric):
Let $x \prec y$ be valid and assume $y \prec x$. By definition of $\prec$ it follows that $x \preceq y \wedge y \preceq x$ is true. Since $\preceq$ is antisymmetric we obtain $x = y$ which contradicts the validity of $x \prec y$.

(c) $x \prec y \wedge y \prec z \Rightarrow x \prec z$ (transitive):
Since $\preceq$ is transitive one immediately obtains $x \preceq y \wedge y \preceq z \Rightarrow x \preceq z$. It remains to prove $x \neq z$. Notice that antireflexivity of $\prec$ implies $x \neq y$ and $y \neq z$. Assume $x = z$. It follows that $x \preceq y \wedge y \preceq x$ is true. Since $\preceq$ is antisymmetric we obtain $x = y$ which contradicts $x \neq y$. □

As a convention, every appearance of "$\prec$" in context with a poset tacitly assumes its definition via eqn. (1) and the partial order relation of the poset. After these preparations one is in the position to turn to the actual objects of interest.

Definition 2. *Let $\mathcal{X}$ be some set. If the reflexive, antisymmetric and transitive relation "$\preceq$" is valid on $\mathcal{X}$ then the pair $(\mathcal{X}, \preceq)$ is called a* partially ordered set *(or short:* poset*). Distinct points $x, y \in \mathcal{X}$ 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 $(\mathcal{X}, \preceq)$ is comparable then $(\mathcal{X}, \preceq)$ is called a* totally ordered set *or a* chain. *Dually, if each pair of distinct points of a poset $(\mathcal{X}, \preceq)$ are incomparable then $(\mathcal{X}, \preceq)$ 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 $i = 1, \ldots, n$. According to Lemma 1 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 3. *An element $x^* \in \mathcal{X}$ is called a* minimal element *of the poset $(\mathcal{X}, \preceq)$ if there is no $x \in \mathcal{X}$ such that $x \prec x^*$. The set of all minimal elements, denoted $\mathcal{M}(\mathcal{X}, \preceq)$, is said to be* complete *if for each $x \in \mathcal{X}$ there is at least one $x^* \in \mathcal{M}(\mathcal{X}, \preceq)$ such that $x^* \preceq x$.* □

Minimal elements are the targets of the evolutionary search studied in the next section. Since the analysis presented in the next section assumes the completeness of $\mathcal{M}(\mathcal{X}, \preceq)$ it is useful to know under which circumstances this assumption is fulfilled.

Lemma 2. *([10], p. 91) If $(\mathcal{X}, \preceq)$ is a poset with $0 < |\mathcal{X}| < \infty$ then the set of minimal elements $\mathcal{M}(\mathcal{X}, \preceq)$ is complete.* □

This result shows that the set of minimal elements may be incomplete only if the poset is infinitely large. But since many evolutionary algorithms operate in finite search spaces and in favor of an easy presentation the more general case will not be considered here. Nevertheless, the next result does not require the finiteness of the poset.

Lemma 3. *Let $\mathcal{M}(\mathcal{X}, \preceq) \neq \emptyset$ be the set of minimal elements of some partially ordered set $(\mathcal{X}, \preceq)$ and $\mathcal{G}(x) = \{y \in \mathcal{X} : y \preceq x\}$ with $x \in \mathcal{X}$.*

(a) $x \in \mathcal{M}(\mathcal{X}, \preceq) \iff \mathcal{G}(x) \setminus \{x\} = \emptyset$.

(b) If $\mathcal{M}(\mathcal{X}, \preceq)$ is complete and $x \notin \mathcal{M}(\mathcal{X}, \preceq)$ then $(\mathcal{G}(x) \setminus \{x\}) \cap \mathcal{M}(\mathcal{X}, \preceq) \neq \emptyset$.

Proof.
(a) By definition, $x \in \mathcal{M}(\mathcal{X}, \preceq) \Leftrightarrow (\nexists\, y \in \mathcal{X} : y \prec x) \Leftrightarrow (\nexists\, y \in \mathcal{X} : y \preceq x \wedge y \neq x)$ are valid equivalences. The rightmost expression is equivalently rewriteable as $\nexists\, y \in \mathcal{G}(x) : y \neq x$ which in turn is equivalent to $\mathcal{G}(x) \setminus \{x\} = \emptyset$.

(b) Let $x \notin \mathcal{M}(\mathcal{X}, \preceq)$. Owing to part (a) we know that $\mathcal{G}(x) \setminus \{x\}$ is not empty. It remains to show that at least one element of $\mathcal{M}(\mathcal{X}, \preceq)$ is also in $\mathcal{G}(x) \setminus \{x\}$. Since $\mathcal{M}(\mathcal{X}, \preceq)$ is complete there must exist an $x^* \in \mathcal{M}(\mathcal{X}, \preceq)$ with $x^* \neq x$ such that $x^* \preceq x$. By definition of $\mathcal{G}(\cdot)$ it follows that necessarily $x^* \in \mathcal{G}(x) \setminus \{x\}$. $\square$

3 Evolutionary Search in Partially Ordered Sets

Hereinafter it is assumed that the set $\mathcal{S} \neq \emptyset$ is finite and that $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\}$ is a mapping where $(\mathcal{F}, \preceq)$ is a poset. Trivially, since $\mathcal{S}$ is finite so is $\mathcal{F} = f(\mathcal{S})$. Owing to Lemma 2 it is guaranteed that the set of minimal elements is complete. This property plays a key role in the subsequent analysis. At first only a simple individual–based EA is considered before the result is extended to the population–based case. This individual–based EA may be seen as a generalized version of the $(1 + 1)$–EA. More specifically, the algorithm runs as follows:

(1) Generate an individual $x_0 \in \mathcal{S}$ at random and set $k = 0$.
(2) Apply some variation operator to obtain an offspring $y_k \in \mathcal{S}$ from x_k.
(3) If $f(y_k) \preceq f(x_k)$ then set $x_{k+1} = y_k$ otherwise $x_{k+1} = x_k$.
(4) Increase k and goto (2) unless some termination condition is fulfilled.

Needless to say that the purpose of this EA is to generate a sequence $(x_k : k \geq 0)$ such that the sequence $(f(x_k) : k \geq 0)$ enters the set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq)$ in a finite number of steps and then stays there forever. Indeed, this EA can accomplish this task if it is assumed that the variation operator is characterized by the property that its associated transition probability function (called *variation kernel* hereinafter) fulfills the inequality $P_v(x, y) \geq \delta > 0$ for all $x, y \in \mathcal{S}$. Before proving this claim one needs the following result.

Lemma 4. *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\}$ be a mapping where $(\mathcal{F}, \preceq)$ is a partially ordered set. Then the following statements are valid:*

(a) $\mathcal{T}(f(x)) = \mathcal{G}(f(x)) \cap \mathcal{M}(\mathcal{F}, \preceq) \neq \emptyset$ for all $x \in \mathcal{S}$.

(b) $\mathcal{I}(f(x)) = \{y \in \mathcal{S} : f(y) \in \mathcal{T}(f(x))\} \neq \emptyset$ for all $x \in \mathcal{S}$.

Proof.
Since $\mathcal{M}(\mathcal{F}, \preceq)$ is complete Lemma 3(b) ensures that the set $\mathcal{T}(f(x))$ is not empty for each $x \in \mathcal{S}$. As a consequence, the inverse image set $\mathcal{I}(f(x))$ of $\mathcal{T}(f(x))$ must be non–empty as well. $\square$

Since the variation kernel is strictly bounded from zero, the probability of generating a specific offspring $y \in \mathcal{S}$ with $f(y) \in \mathcal{M}(\mathcal{F}, \preceq)$ by a variation of an arbitrary parent $x \in \mathcal{S}$ is at least $\delta > 0$. Thus, every minimal element can be visited within *one* step from every $x \in \mathcal{S}$. Assume that this event has happened, i.e., there was a transition from

$x \in \mathcal{S}$ to $y \in \mathcal{S}$ with $f(y) \in \mathcal{M}(\mathcal{F}, \preceq)$. There is, however, no guarantee of acceptance for it can be seen from the third step of the EA that an offspring y is accepted if and only if $f(y) \in \mathcal{G}(f(x))$. Thus, it may happen that the offspring $y \in \mathcal{S}$ is rejected although $f(y)$ is a minimal element. But Lemma 4(a) ensures the existence of minimal elements that are also contained in $\mathcal{G}(f(x))$ for every specific $x \in \mathcal{S}$. These minimal elements are collected in the set $\mathcal{T}(f(x))$ for every $x \in \mathcal{S}$ whereas the set $\mathcal{I}(f(x)) \neq \emptyset$ contains all offspring which will be accepted and whose image is minimal in $\mathcal{F}$. Therefore, the probability that the EA generates and accepts an offspring whose image is minimal in $\mathcal{F}$ within a single iteration can be bounded via

$$P(x, \mathcal{I}(f(x))) \geq \delta \cdot |\mathcal{I}((f(x))| \geq \delta > 0 \tag{2}$$

for each $x \in \mathcal{S}$. Next it is investigated what happens as soon as the sequence $(f(x_k) : k \geq 0)$ has entered the set of minimal elements for the first time. Let $f(x_{k_0}) \in \mathcal{M}(\mathcal{F}, \preceq)$ for some $k_0 \geq 0$. Owing to Lemma 3(a) it is guaranteed that $f(x_k) \in \mathcal{M}(\mathcal{F}, \preceq)$ for all $k \geq k_0$, because the set of acceptable elements is $\mathcal{G}(f(x_{k_0})) = \{f(x_{k_0})\}$. This in turn implies $f(x_k) = f(x_{k_0})$ for all $k \geq k_0$. As a consequence, the probability to leave the set of minimal elements once it was entered is zero. Taking into account this result and the result summarized in inequality (2) it follows from a theorem in [11] that

$$\mathsf{P}\{\, f(x_k) \in \mathcal{M}(\mathcal{F}, \preceq) \,\} \geq 1 - (1 - \delta)^k \tag{3}$$

for $k \geq 0$. To proceed—and in anticipation of potential generalizations of these results to arbitrary metrizable spaces—the set $\mathcal{F}$ is equipped with a metric. Notice that every non–empty set $\mathcal{F}$ may be endowed with the *discrete metric*

$$d(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

with $a, b \in \mathcal{F}$. Using this metric the distance $d(a, \mathcal{A}) = \min\{d(a, b) : b \in \mathcal{A}\}$ from a point $a \in \mathcal{F}$ to a subset $\mathcal{A} \subseteq \mathcal{F}$ is well–defined. With the definition of the nonnegative random variable $D_k = d(f(x_k), \mathcal{M}(\mathcal{F}, \preceq))$ inequality (3) may be equivalently expressed by

$$\mathsf{P}\{\, D_k \leq \varepsilon \,\} \geq 1 - (1 - \delta)^k \tag{4}$$

for every $\varepsilon > 0$. Since the r.h.s of inequality (3) converges to one as $k \to \infty$ it has been shown that the sequence $(D_k : k \geq 0)$ converges in probability to zero. Since the rate of approach in inequality (4) is geometrically fast it follows that $(D_k : k \geq 0)$ converges even completely to zero (implying convergence with probability 1). Moreover, convergence in mean is ensured by the boundedness of D_k and convergence in probability. Thus, it was proven:

Theorem 1. *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\}$ be a mapping where $(\mathcal{F}, \preceq)$ is a partially ordered set that is additionally endowed with the discrete metric $d(\cdot, \cdot)$. Let $(x_k : k \geq 0)$ with $x_k \in \mathcal{S}$ be the random sequence generated by the generalized $(1+1)$–EA whose time–invariant variation kernel has the property $P_v(x, y) \geq \delta > 0$ for all $x, y \in \mathcal{S}$. Then the sequence $d(f(x_k), \mathcal{M}(\mathcal{F}, \preceq))$ of random distances between $f(x_k)$ and the set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq))$ converges completely and in mean to zero as $k \to \infty$.* □

The generalization of this theorem to the population–based case is almost trivial. Let the population of the EA consist of an n–tuple of individuals where $n < \infty$. The input of the variation operator is now the entire population whereas its output is an N–tuple of offsprings with $0 < N < \infty$. Suppose that the associated variation kernel has the property $P_v(p, q) \geq \delta > 0$ for all $p \in \mathcal{S}^n$ and $q \in \mathcal{S}^N$. Then it is guaranteed that the image of at least one offspring enters the set of minimal elements with minimum probability $\delta > 0$ in one iteration, regardless of the actual population of parents. The mechanism to compile the new population of parents may be arbitrary provided that the procedure obeys the *elite preservation strategy:*

Let $(x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)})$ be the population of parents and $\mathcal{O}_k = \{y_k^{(1)}, y_k^{(2)}, \ldots, y_k^{(N)}\}$ be the collection of offspring at generation $k \geq 0$. Without loss of generality let parent $x_k^{(1)}$ be the *elitist parent*. If $\mathcal{B}_k = \{y \in \mathcal{O}_k : f(y) \in \mathcal{G}(f(x_k^{(1)})) \setminus \{f(x_k^{(1)})\}\}$ is empty then set $x_{k+1}^{(1)} = x_k^{(1)}$. Otherwise, choose an arbitrary $y_k^* \in \mathcal{B}_k$ whose image is minimal in the poset $(f(\mathcal{B}_k), \preceq)$ and set $x_{k+1}^{(1)} = y_k^*$.

Evidently, the sequence $(x_k^{(1)} : k \geq 0)$ behaves like the generalized $(1 + 1)$–EA considered previously and without much effort it was proven:

Theorem 2. *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\}$ be a mapping where $(\mathcal{F}, \preceq)$ is a partially ordered set that is additionally endowed with the discrete metric $d(\cdot, \cdot)$. An evolutionary algorithm with $n < \infty$ parents and $N < \infty$ offspring in $\mathcal{S}$ and fitness function $f(\cdot)$ is guaranteed to generate at least one sequence $(x_k : k \geq 0)$ of parents such that the sequence $d(f(x_k), \mathcal{M}(\mathcal{F}, \preceq))$ converges completely and in mean to zero, provided that the time–invariant variation kernel is positive on $\mathcal{S}^n \times \mathcal{S}^N$ and that the selection operator obeys the elite preservation strategy.* □

An early example of a probabilistic algorithm resembling an evolutionary algorithm that obeys the elite preservation strategy was proposed by Peschel & Riedel in 1977 [12]. Originally, their method was designed for multi–objective optimization over $\mathcal{S} = \mathbb{R}^\ell$ but it can be easily formulated in the framework presented here. For this purpose let $\mathcal{P}_k$ be the collection of n_k parents and $\mathcal{O}_k$ be the collection of N individuals at generation $k \geq 0$. Then the algorithm runs as follows:

1. Generate a collection $\mathcal{P}_0$ of $n_0 = N$ parents at random and set $\mathcal{O}_0 = \mathcal{P}_0$.
2. Select those offspring whose images are minimal in the poset $(f(\mathcal{O}_0), \preceq)$. This yields n_1 distinct parents after deletion of potential duplicates. Set $k = 1$.
3. Generate a collection $\mathcal{O}_k$ of N offspring from the collection $\mathcal{P}_k$ of n_k parents by variation.
4. Select those individuals from $\mathcal{P}_k \cup \mathcal{O}_k$ whose images are minimal in the poset $(f(\mathcal{P}_k \cup \mathcal{O}_k), \preceq)$. This yields n_{k+1} distinct parents after deletion of potential duplicates.
5. Increase k and goto (3) unless some termination criterion is fulfilled.

Notice that the number of offspring $N < \infty$ was fixed whereas the number of parents n_k may vary for each generation $k \geq 0$. Actually, the number of parents is a random variable. But since $\mathcal{S}$ is finite and the population consists of distinct individuals one immediately obtains $n_k \leq |\mathcal{S}| < \infty$ and hence $\mathsf{P}\{\max\{n_k : k \geq 0\} < \infty\} = 1$. If the variation kernel is positive on $\mathcal{S}^n \times \mathcal{S}^N$ for arbitrary $n \in \{1, 2, \ldots, |\mathcal{S}|\}$ then the argumentation in the proof of Theorem 2 remains valid. The elite preservation strategy is fulfilled since a parent $x \in \mathcal{P}_k$ is not contained in $\mathcal{P}_{k+1}$ if and only if there exists a $y \in \mathcal{P}_{k+1}$ with $f(y) \preceq f(x)$. Therefore Theorem 2 ensures that the generalized Peschel/Riedel–method generates at least one sequence that converges to the set of minimal elements. But an even stronger result than Theorem 2 can be proven for this method.

Theorem 3. *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\}$ be a mapping where $(\mathcal{F}, \preceq)$ is a partially ordered set. The set–valued sequence $(f(\mathcal{P}_k) : k \geq 0)$ generated by the generalized Peschel/Riedel–method with positive time–invariant variation kernel converges completely and in mean to the set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq)$ as $k \to \infty$, i.e., $|f(\mathcal{P}_k) \cap \mathcal{M}(\mathcal{F}, \preceq)| \to |\mathcal{M}(\mathcal{F}, \preceq)|$ completely and in mean as $k \to \infty$.*

Proof.

Let $m = |\mathcal{M}(\mathcal{F}, \preceq)|$ and $M_k = |f(\mathcal{P}_k) \cap \mathcal{M}(\mathcal{F}, \preceq)|$ for $k \geq 0$. Since the variation kernel is positive the transition probability of the stochastic process $(M_k : k \geq 0)$ can be bounded via $p_{i,i+1} = \mathsf{P}\{M_{k+1} = i + 1 \,|\, M_k = i\} \geq \delta\,(m - i)$ for $0 \leq i < m$ and some $\delta > 0$. Notice that $p_{ij} = 0$ if $0 \leq j < i \leq m$ by construction of the algorithm, i.e., M_k increases monotonically. Although there are also transitions with $p_{i,i+j} > 0$ for $j > 1$, these short cuts will be ignored. Thus, it is assumed that M_k attains all values from 0 to m consecutively. It is clear that this modified process requires more time to reach the value m than the original one. Larger values than m cannot be attained because the completeness of $\mathcal{M}(\mathcal{F}, \preceq)$ guarantees that $|f(\mathcal{P}_k) \cap \mathcal{M}(\mathcal{F}, \preceq)| \leq |\mathcal{M}(\mathcal{F}, \preceq)|$ for every $k \geq 0$. As a consequence, $p_{ii} = 1$ if and only if $i = m$. By setting $p_{i,i+1} = \delta\,(m - i)$ for $0 \leq i < m$ and $p_{ii} = 1 - p_{i,i+1}$ for $i < m$ one obtains a finite absorbing Markov chain that requires more time to reach its only absorbing state m than the original stochastic process. The eigenvalues of the associated transition matrix are $\{1, \delta\, m, \delta\,(m-1), \ldots, 2\,\delta, \delta\}$. Owing to a result in [13], p. 391, one obtains $\mathsf{P}\{M_k < m\} = O(\lambda_*^k)$ where $\lambda_* = \delta\, m$ is the second largest eigenvalue. This proves complete convergence of M_k to m; convergence in mean follows from the boundedness of M_k. □

4 Transcription of Result into Specialized Terminology

4.1 Real–valued Function Minimization

Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\} \subset \mathrm{I\!R}$ be a real–valued mapping. In this case the poset $(\mathcal{F}, \leq)$ with the usual order relation $\leq$ in $\mathrm{I\!R}$ is totally ordered. As a consequence, the set of minimal elements in $\mathcal{F}$ consists of a single element f^*, which is called the *global minimum* of the objective function $f(\cdot)$. The *lower level set* $\mathcal{G}(f(x)) = \{f(y) : f(y) \leq f(x)\}$ contains the unique global minimum f^* for every $x \in \mathcal{S}$.

In this case the elite preservation strategy is simplified as follows: Without loss of generality let $x_k^{(1)}$ be the elitist parent at generation $k \geq 0$. If $\mathcal{B}_k$ is empty then set $x_{k+1}^{(1)} = x_k^{(1)}$, i.e., there was no offspring with an objective function value less than $f(x_k^{(1)})$. If $\mathcal{B}_k$ is not empty then choose an arbitrary offspring $y_k^* \in \mathcal{B}_k$ among those possessing the least objective function value, and set $x_{k+1}^{(1)} = y_k^*$.

Moreover, since the set $\mathcal{F}$ is finite stochastic convergence to zero in the discrete metric $d(f(x_k), f^*)$ is equivalent to $|f(x_k) - f^*| \to 0$ as well as to $f(x_k) \to f^*$. Now one is in the position to transcribe the main result of the previous section into the terminology of real–valued function minimization.

Corollary 1 (to Theorem 2). *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\} \subset \mathbb{R}$ be the objective function. An evolutionary algorithm with $n < \infty$ parents and $N < \infty$ offspring in $\mathcal{S}$ and fitness function $f(\cdot)$ is guaranteed to generate at least one sequence $(x_k : k \geq 0)$ of parents such that the sequence $(f(x_k) : k \geq 0)$ converges completely and in mean to the global minimum f^*, provided that the time–invariant variation kernel is positive on $\mathcal{S}^n \times \mathcal{S}^N$ and that the selection operator obeys the elite preservation strategy.* □

4.2 Multicritera Optimization

Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\} \subset \mathbb{R}^m$ with $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))'$ a vector–valued function, where each of the $m \geq 2$ real–valued objective functions $f_1, \ldots, f_m$ are to be minimized simultaneously. The partial order relation "$\preceq$" on $\mathcal{F}$ is given by

$$f(x) \preceq f(y) \iff f_i(x) \leq f_i(y) \text{ for all } i = 1, \ldots, m.$$

In this particular context the set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq)$ is called the *Pareto set* while its elements are termed *pareto–optimal* or *efficient points*.

Corollary 2 (to Theorem 2). *Let $\mathcal{S} \neq \emptyset$ be a finite set and $f : \mathcal{S} \to \mathcal{F} = \{f(x) : x \in \mathcal{S}\} \subset \mathbb{R}^m$ where $f(\cdot)$ is a vector of $m \geq 2$ real–valued objective functions. An evolutionary algorithm with $n < \infty$ parents and $N < \infty$ offspring in $\mathcal{S}$ and objective function vector $f(\cdot)$ is guaranteed to generate at least one sequence $(x_k : k \geq 0)$ of parents such that the sequence of distances (in discrete metric) between $f(x_k)$ and the Pareto set converges completely and in mean to zero, provided that the time–invariant variation kernel is positive on $\mathcal{S}^n \times \mathcal{S}^N$ and that the selection operator obeys the elite preservation strategy.* □

5 Potential Directions of Further Research

There are several directions in which the main results may be extended. First, the preconditions regarding the variation kernel and the requirement of the elitist preservation property can be weakened certainly. Second, a generalization to infinitely large search and image sets should be possible. Metrizable topological spaces ought to be sufficiently general, although some complications regarding measure theory can be expected. Third, if the binary relation on $\mathcal{F}$ is reflexive and antisymmetric but not necessarily transitive, then one has to cope with quasi–ordered sets in lieu of partially ordered sets. Last but certainly not least, the finite time behavior of the evolutionary algorithms is of significant practical importance. This includes the expected first entry times to the set of minimal elements as well as the expected achievable quality of the solution under some specific stopping rules.

Acknowledgments

This work is a result of the *Collaborative Research Center "Computational Intelligence" (SFB 531)* supported by the German Research Foundation (DFG).

References

1. C. M. Fonseca and P. J. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.
2. H. Tamaki, H. Kita, and S. Kobayashi. Multi–objective optimization by genetic algorithms: a review. In *Proceedings of the 3rd IEEE International Conference on Evolutionary Computation*, pages 517–522. IEEE Press, Piscataway (NJ), 1996.
3. J. Horn. Multicriterion decision making. In T. Bäck, D. B. Fogel, and Z. Michalewicz, editors, *Handbook of Evolutionary Computation*, pages F1.9:1–15. IOP Publishing and Oxford University Press, New York and Bristol (UK), 1997.
4. A. E. Eiben, E. H. L. Aarts, and K. M. Van Hee. Global convergence of genetic algorithms: A markov chain analysis. In H.-P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*, pages 4–12. Springer, Berlin and Heidelberg, 1991.
5. D. B. Fogel. Asymptotic convergence properties of genetic algorithms and evolutionary programming: Analysis and experiments. *Cybernetics and Systems*, 25(3):389–407, 1994.
6. G. Rudolph. Convergence properties of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.
7. J. Suzuki. A markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4):655–659, 1995.
8. J. Ester. *Systemanalyse und mehrkriterielle Entscheidung*. VEB Verlag Technik, Berlin, 1987.
9. W. T. Trotter. Partially ordered sets. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics, Vol. 1*, pages 433–480. Elsevier Science, Amsterdam, 1995.
10. W. Klingenberg and P. Klein. *Lineare Algebra und analytische Geometrie, Band 1*. Bibliographisches Institut, Mannheim, 1971.
11. G. Rudolph. Convergence of evolutionary algorithms in general search spaces. In *Proceedings of the Third IEEE Conference on Evolutionary Computation*, pages 50–54. IEEE Press, 1996.
12. M. Peschel and C. Riedel. Use of vector optimization in multiobjective decision making. In D. E. Bell, R. L. Keeney, and H. Raiffa, editors, *Conflicting Objectives in Decisions*, pages 97–121. Wiley, Chichester, 1977.
13. J. S. Rosenthal. Convergence rates for markov chains. *SIAM Review*, 37(3):387–405, 1995.