Convergence Analysis of Self-adaptive Multi-objective Evolutionary Algorithm Based on Grids

Yuren Zhou\textsuperscript{a,1} and Jun He\textsuperscript{b,2}

\textsuperscript{a}School of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, China
\textsuperscript{b}School of Computer Science, University of Birmingham, Birmingham B15 2TT, United Kingdom.

Abstract

Evolutionary algorithms have been successfully applied to various multi-objective optimization problems. However, theoretical studies on multi-objective evolutionary algorithms, especially with self-adaption, are relatively scarce. This paper analyzes the convergence properties of a self-adaptive ($\mu+1$)-algorithm. The convergence of the algorithm is defined, and the general convergence conditions are studied. Under these conditions, it is proven that the proposed self-adaptive ($\mu+1$)-algorithm converges in probability or almost surely to the Pareto-optimal front.

Keywords: Analysis of Algorithms, Multi-Objective Optimization, Evolutionary Algorithms, Convergence.

1 Introduction

Multi-objective evolutionary algorithms (MOEAs) have been studied for more than ten years \cite{4,3}. It is generally recognized that Schaffer \cite{16} was the first researcher to use EAs to handle vector optimization problems. Today various MOEAs, e.g., NSGA \cite{17}, SPEA \cite{19}, PAES \cite{9} and NSGA-II \cite{6}, have

\textit{Email addresses:} zhouyuren@hotmail.com (Yuren Zhou), jun.he@ieee.org (Jun He).

\textsuperscript{1} Supported in part by the National Natural Science Foundation of China under Grant No.60673062 and Natural Science Foundation of Guangdong Province of China under Grant No.06025686.

\textsuperscript{2} Supported by the UK Engineering and Physical Research Council under Grant No. EP/C520696/1.

\textit{Preprint submitted to Elsevier Science} 24 April 2007
be proposed and applied in many practical fields. Like theoretical studies on single objective evolutionary algorithms [12,1], several rigorous analysis of MOEAs have been made in recent years [14,15,7,10], and it attracts more and more attentions. Rudolph and Agapie [15] analyzed and proved MOEAs’ convergence using Markov chain. Hanne [7] proposed a convergence theorem of function MOEAs with probability 1 under strict condition of “efficiency preserving”, a requirement that some current pragmatic MOEA do not meet. Laumanns [10] established a MOEA model which have both properties of converging to the Pareto-optimal front and maintaining a spread among obtained solutions, but he failed to rigorously define the convergence of MOEAs and prove that the MOEA model converges to the Pareto optimal sets. Recently Laumanns [11] presented the running time analysis of multi-objective EAs on pseudo-Boolean model problems.

This paper extends our previous discussion on the convergence from no-adaptive MOEAs [18] to self-adaptive ones. We will introduce a rigorous definition of the convergence of MOEAs and discuss general conditions that guarantee the convergence of MOEAs. Under these conditions, we show that the proposed MOEA converges almost surely to the Pareto optimal set. The remainder of this paper is arranged as follows: Section 2 describes a $(\mu+1)$ MOEA and introduces basic definitions and terms; Section 3 analyzes the proposed MOEA’s convergence; Section 4 concludes the paper.

2 Definitions and algorithm description

Consider a multi-objective optimization problem with $n$ decision variables and $m$ objectives:

\[
(MOP) \text{ maximize } f(x) = (f_1(x), \ldots, f_m(x)), \\
\text{subject to } g_i(x) \leq 0, \quad i = 1, \ldots, q,
\]

where $x = (x_1, \ldots, x_n) \in S_x \subset \mathbb{R}^n$, $y = f(x) = (y_1, \ldots, y_m) \in S_y \subset \mathbb{R}^m$. $x$ is called the decision (parameter) vector, $S_x$ is the decision space, $y$ is the objective vector and $S_y$ is the objective space, $g_i(x)$ is the constraint condition. For simplicity, this paper will not discuss constrained problems.

2.1 Dominance relation

Let’s introduce two basic definitions used in MOEAs: dominance relation and Pareto set.
Definition 1 Let \( f, g \in \mathbb{R}^m \) be two vectors, \( f \) is said to dominate \( g \) (written as \( f \succ g \)) if and only if

1. \( \forall i \in \{1, \ldots, m\} : f_i \geq g_i \);  
2. \( \exists j \in \{1, \ldots, m\} : f_j > g_j \).

If \( f \succ g \), it also means that \( g \) is dominated by \( f \), denoted as \( g \prec f \).

Definition 2 Let \( F \subset \mathbb{R}^m \) be a vector set. The set of vectors in \( F \) that are not dominated by any vector in \( F \) is called a Pareto set of \( F \), denoted as \( P(F) \), i.e.,

\[
P(F) := \{ g \in F | \forall f \in F : f \not\succ g \}.
\]

Definition 3 Let \( R_f \) be the range of function \( f \) in MOP (1), the Pareto set of \( R_f \) is called Pareto-optimal front. That is

\[
P(R_f) := \{ y \in R_f | \exists y' \in R_f : y' \succ y \}
\]

Definition 4 Let \( P(R_f) \) be the Pareto-optimal front of MOP(1), the image source of \( P(R_f) \) under mapping \( f \) is said to be Pareto-optimal set, denoted as \( P(f(x)) \), i.e.

\[
P(f(x)) := \{ x \in S_x | \exists x' \in S_x : f(x') \succ f(x) \}.
\]

The vector in \( P(f(x)) \) is called Pareto-optimal solution.

The concept of \( \epsilon \)-neighborhood, which is useful in discussing convergence of MOEAs, is defined as follows.

Definition 5 Let \( f = (f_1, f_2, \cdots, f_m) \in \mathbb{R}^m \) and \( \epsilon > 0 \), the \( \epsilon \)-neighborhood of \( f \) is defined as follows:

\[
N_\epsilon(f) := \{ y : y_i \in (f_i - \epsilon, f_i + \epsilon), i = 1, \ldots, m \}.
\]

Given a set \( F \subset \mathbb{R}^m \), the \( \epsilon \)-neighborhood of \( F \) is defined by:

\[
N_\epsilon(F) = \bigcup_{f \in F} N_\epsilon(f).
\]

2.2 Self-adaptive \((\mu + 1)\)-MOEA Based-on Grids

The grid-based MOEA is developed mainly by Knowles [9] and Laumanns [10]. Based on the \( \epsilon \)-dominance concept (or grid), Deb [5] proposed a steady-state MOEA that had a good compromise in terms of convergence near to Pareto-optimal front, diversity of solutions and computational time. The basic idea of
this MOEA is to divide the search (or objective) space into a number of hyper-boxes (i.e. grids) and to maintain the diversity by ensuring that a hyper-box can be occupied by only one solution.

In this paper we discuss a simple grids-based ($\mu + 1$) MOEA introduced in [9,10]. This ($\mu + 1$) MOEA is composed of algorithms 1-4 where the details of notations and description can be found in [10].

Algorithm 1: Iterative search algorithm

\begin{verbatim}
begin
  t := 0;
  A(0) := \emptyset;
  while ( terminate (A(t), t) = false ) do
    t := t + 1;
    f(t) := generate(A(t-1));
    A(t) := update(A(t-1), f(t));
  od
output: A(t).
end
\end{verbatim}

The counter $t$ denotes the generations. The set $A(t)$ is the population at generation $t$ with a variable population size $\mu$. The vector $f(t)$ is the $t$-generation objective vector. The new individual is yielded through “generate” function. “Update” function is used to choose the next generation population.

Algorithm 2: “generate” function

\begin{verbatim}
begin
  input : $A(t-1)$ = $\{a_1, \ldots, a_\mu\}$;
  $(x_1, \ldots, x_\mu) := (f^{-1}(a_1), \ldots, f^{-1}(a_\mu))$;
  $x_i$ := randomly choose from($x_1, \ldots, x_\mu$);
  $x'$ := mutate $x_i$;
  $f(t)$ := $f(x')$.
output: $f(t)$.
end
\end{verbatim}

Here the mutation is self-adaptive: $x' = x(t) + z$, where the joint probability density distribution of $z$ is $\phi(z, t)$, and $\phi(z, t) \in \mathbb{R}^n$ is changed as $t$.

Algorithm 3: “update” function
begin
input : $A, f$;
$D := \{ f' \in A \mid \text{box}(f) \succ \text{box}(f') \}$;
if $(D \neq \emptyset)$
then $A' := A \cup \{ f \} \setminus D$;
elsif $(\exists f' \in A : \text{box}(f') = \text{box}(f) \land f \succ f')$
then $A' := A \cup \{ f \} \setminus f'$;
elsif $\neg \exists f' \in A : \text{box}(f') = \text{box}(f) \lor \text{box}(f') \succ \text{box}(f))$
then $A' := A \cup \{ f \}$;
elsif $A' := A$;
fi
output: $A'$.
end

From the “update” function, it is seen once a solution enters the population $A^{(t)}$, the solution will stay in the population if no other solution can dominate it.

Algorithm 4: “box” function
begin
input : $f$;
for $(i = 1, \cdots, m)$ do
$\ b_i = \lfloor f_i / \varepsilon \rfloor$
end
$b := (b_1, \cdots, b_m)$.
output: $b$ (note: box index vector).
end

The “box” function is used to divide the objective space $S_y$ into hyper-boxes (grids) with size $\varepsilon$. It returns a vector to represent a box vertex. The box dominance relation is a kind of the vector dominance relation. Due to the “update” and “box” functions, up to one solution is in each box.

3 Convergence Analysis of $(\mu + 1)$ MOEA

Given a random variable sequence $X^{(t)}$, $(t = 1, 2, \cdots)$, there are definitions such as convergence almost surely and convergence in probability (See Chapter 4 in [2]).

Definition 6 Let $X, X^{(t)}, (t = 1, 2, \cdots)$ be random variables on a probability space, the random variable sequence $X^{(t)}$ is said
(1) to converge almost surely to random variable \( X \), if \( P(\lim_{t \to +\infty} X(t) = X) = 1 \);

(2) to converge in probability to \( X \), if \( \forall \epsilon > 0 : \lim_{t \to +\infty} P(\mid X(t) - X \mid \leq \epsilon) = 1 \);

The convergence almost surely implies the convergence in probability whereas the converse is wrong in general.

**Lemma 1** The following statements are equivalent:

(1) A random variable sequence \( \{X(t)\} \) converges almost surely to random variable \( X \);

(2) \( \forall \epsilon > 0, \lim_{\tau \to +\infty} P(\mid X(t) - X \mid < \epsilon, \text{ for all } t \geq \tau) = 1 \);

(3) \( \forall \epsilon > 0, \lim_{\tau \to +\infty} P(\mid X(t) - X \mid \geq \epsilon, \text{ for some } t \geq \tau) = 0 \);

(4) \( \forall \epsilon > 0, P(\mid X(t) - X \mid \geq \epsilon, \text{ appears infinite times}) = 0 \);

(5) \( \forall \epsilon > 0, P(\cap_{\tau=1}^{+\infty} \cup_{t=\tau}^{+\infty} \mid X(t) - X \mid \geq \epsilon) = 0 \);

**Proof:** See Theorem 4.1.1 in [2].

With the equivalent definitions in Lemma 1, we define the grids-based convergence as follows.

**Definition 7** The objective space \( S_y \) is divided into finite \( m \)-dimension hyper-boxes. Denote \( B_i \), \( i = 1, \cdots, s \) to be the hyper-boxes that contain the Pareto front of MOP(1), where \( s \) is the number of these hyper-boxes, and \( F_i \), \( i = 1, \cdots, s \) the Pareto-optimal front in \( B_i \). And \( F_i \)'s \( \epsilon \)-neighborhood is denoted as \( N_\epsilon(F_i) \) (refer to Definition 5). Let \( A(t) \), \( t = 1, 2, \cdots \) be the population sequence generated by the \((\mu + 1)\)-MOEA.

(1) The \((\mu + 1)\)-MOEA is said to converge almost surely to the Pareto front of MOP(1) if \( \forall \epsilon > 0, \forall i \in \{1, \cdots, s\} : \)

\[
\lim_{\tau \to +\infty} P(\bigcap_{t=\tau}^{+\infty} A(t) \cap N_\epsilon(F_i) \neq \emptyset \text{ for all } t \geq \tau) = 1.
\]  

(2) The \((\mu + 1)\)-MOEA is said to converge in probability to the Pareto front of MOP(1) if \( \forall \epsilon > 0, \forall i \in \{1, \cdots, s\} : \)

\[
\lim_{t \to +\infty} P(A(t) \cap N_\epsilon(F_i) \neq \emptyset) = 1.
\]

Note: (1) Since for any \( y \notin \cup_i N_\epsilon(F_i) \), there is some \( y' \in \cup_i N_\epsilon(F_i) \), where \( y' \succ y \), so according the “update” function, if both \( y \) and \( y' \) appear in the same population \( A(t) \), then \( y \) will be removed. From Definition 7(1),

\[
\lim_{\tau \to +\infty} P(A(t) \cap N_\epsilon(F_i) \neq \emptyset \text{ for all } t \geq \tau) = 1.
\]

it means
\[
\lim_{\tau \to +\infty} P(A(t) \setminus \bigcup_{i=1}^{s} N_{\epsilon}(F_i) = \emptyset \text{ for all } t \geq \tau) = 1. \tag{6}
\]

(2) Similarly, Definition 7(2) means
\[
\lim_{t \to +\infty} P(A(t) \setminus \bigcup_{i=1}^{s} N_{\epsilon}(F_i) = \emptyset) = 1. \tag{7}
\]

Definition 7 requires that in each \( N_{\epsilon}(F_i) \), there exist exactly one individual (solution). Other early similar definition can be found in [13].

Similar to Lemma 1, we have the following equivalents.

**Lemma 2** The following statements are equivalent: \( \forall \epsilon > 0, \forall i \in \{1, \cdots, s\} \):

1. \( \lim_{\tau \to +\infty} P(A(t) \cap N_{\epsilon}(F_i) = \emptyset \text{ for some } t \geq \tau) = 0; \)
2. \( P(A(t) \cap N_{\epsilon}(F_i) = \emptyset \text{ appears infinite times}) = 0; \)
3. \( P(\cap_{\tau=1}^{+\infty} \cup_{t=\tau}^{+\infty} A(t) \cap N_{\epsilon}(F_i) = \emptyset) = 0. \)

**Lemma 3** (Borel-Cantelli) Let \( X(t), (t = 1, 2, \cdots) \) be a random sequence, then
\[
\sum_{t=1}^{+\infty} P(X(t)) < +\infty \Rightarrow P(\cap_{\tau=1}^{+\infty} \cup_{t=\tau}^{+\infty} X(t)) = 0.
\]

**Proof:** See Theorem 4.2.1 in [2].

The first convergence theorem is an extension of Theorems 3 in [8].

**Theorem 1** Let \( A(t), (t = 1, 2, \cdots) \) be the population sequence generated by the \((\mu + 1)\)-MOEA, \( N_{\epsilon}(F_i), (i = 1, \cdots, s) \) the \( \epsilon \)-neighborhood as specified in Definition 7. For \( t \geq 1, i = 1, \cdots, s \), let
\[
\alpha_{i}^{(t)} := P(A(t) \cap N_{\epsilon}(F_i) = \emptyset \mid A(t) \cap N_{\epsilon}(F_i) \neq \emptyset);
\]
\[
\beta_{i}^{(t)} := P(A(t) \cap N_{\epsilon}(F_i) = \emptyset \mid A(t) \cap N_{\epsilon}(F_i) = \emptyset);
\]
\[
\gamma_{i}^{(t)} := \beta_{i}^{(1)} \times \cdots \beta_{i}^{(t)}. \]

Then \( \forall \epsilon > 0, \forall i \in \{1, \cdots, s\}, \)

1. if \( \lim_{t \to +\infty} \gamma_{i}^{(t)} = 0, (i = 1, \cdots, s) \), then the \((\mu + 1)\)-MOEA converges in probability to the Pareto-optimal front;
2. if \( \sum_{t=1}^{+\infty} \gamma_{i}^{(t)} < +\infty \), then the \((\mu + 1)\)-MOEA converges almost surely to the Pareto-optimal front.

**Proof:** For any \( i \in \{1, 2, \cdots s\} \), assume \( A(t) \cap N_{\epsilon}(F_i) \neq \emptyset \), and \( x \in A(t) \cap N_{\epsilon}(F_i) \), then only one of the following events can happen: (1) \( x \) appears in the next
generation \(x \in A^{(t+1)}\); (2) an individual \(x' \in N_\epsilon(F_i)\) is generated and dominate \(x\), then it replace \(x\) in the new generation. In both cases, \(A^{(t+1)} \cap N_\epsilon(F_i) \neq \emptyset\), and we have \(\alpha_i^{(t)} = 0\).

Denote \(P_i^{(t)} := P(A^{(t)} \cap N_\epsilon(F_i) = \emptyset)\).

According to Bayesian formula, we obtain

\[
P_i^{(t+1)} = P(A^{(t+1)} \cap N_\epsilon(F_i) = \emptyset) \\
= P(A^{(t+1)} \cap N_\epsilon(F_i) = \emptyset | A^{(t)} \cap N_\epsilon(F_i) \neq \emptyset) P(A^{(t)} \cap N_\epsilon(F_i) \neq \emptyset) \\
+ P(A^{(t+1)} \cap N_\epsilon(F_i) = \emptyset | A^{(t)} \cap N_\epsilon(F_i) = \emptyset) P(A^{(t)} \cap N_\epsilon(F_i) = \emptyset) \\
= \alpha_i^{(t)} P(A^{(t)} \cap N_\epsilon(F_i) \neq \emptyset) + \beta_i^{(t)} P_i^{(t)} \\
= \beta_i^{(t)} \cdots \beta_i^{(1)} P_i^{(1)} \\
= \gamma_i^{(t)} P_i^{(1)}
\]

(1) If Condition (1) holds, then we get \(\lim_{t \to +\infty} P_i^{(t+1)} = 0\).

Hence, according to Definition 6, the MOEA converges in probability to Pareto-optimal front.

(2) If Condition (2) holds, then from \(P_i^{(t+1)} = \gamma_i^{(t)} P_i^{(1)}\), and Lemma 3, we know that the MOEA converges almost surely to Pareto-optimal front.

The second theorem is an application of the above theorem for the \((\mu + 1)\) MOEA with self-adaptive mutations.

**Theorem 2** Assume that the decision space \(S_x\) is a compact set in \(\mathbb{R}^n\), the objective function \(f(x)\) is continue on \(S_x\). The mutation is \(x' = x + z\), where the joint probability density function \(\phi(z, t)\) of \(z\) satisfies \(\phi(z, t) \geq \sigma(t) \varphi(z)\) for all \(z \in S_x\).

(1) if \(\forall \epsilon > 0, \exists t_0 > 0 \text{ and } 0 < \theta < 1 \text{ with } \lim_{t \to +\infty} (t + 1)^{\theta/2} = +\infty: \sigma(t) \geq (t + 1)^{(\theta-1)}\),

(2) If \(\varphi(z) > 0\) is continuous on \(S_x\).

Then the \((\mu + 1)\)-MOEA converges almost surely (and also in probability) to the Pareto-optimal front.

**Proof:** \(\forall \epsilon > 0, \text{ and } i \in \{1, 2, \cdots, s\}\), according to Theorem 1(2), it is sufficient to prove \(\sum_{i=1}^{s} \gamma_i^{(t)} < +\infty\).

Let \(y^*\) be a point on Pareto front in \(B_s\), where \(y^* = f(x^*), x^* = (x^*_1, \cdots, x^*_n) \in S_x\).
Because $f(x)$ is continuous on $S_x$, there exists a positive $\delta > 0$ such that for any $x$ satisfying $\| x - x^* \|_\infty \leq \delta$, it holds $\| f(x) - f(x^*) \|_\infty \leq \epsilon$, therefore $f(x) \in N_\epsilon(F_i)$.

Let $D(x^*, \delta) := \{ x \in S_x \ | \ \| x - x^* \| \leq \delta \}$.

For $x(t) = (x_1, \cdots, x_n) \in S_x$, we have $\forall t \geq t_0$,

$$P(x(t) + z \in D(x^*, \delta)) = \int_{z \in D(x^* - x(t), \delta)} \phi(z, t) dz \geq (t + 1)^{(\theta - 1)} \int_{z \in D(x^* - x(t), \delta)} \varphi(z) dz.$$ 

Let $\omega(x, x^*, \delta) := \int_{z \in D(x^* - x, \delta)} \varphi(z) dz$. From Condition (2), and $S_x$ being a compact set,

$$\omega_{\text{min}} = \min_{x, x^*} \omega(x, x^*, \delta)$$

exists and is greater than 0.

From Condition (1), there exists some $t_1 \geq t_0$, $\forall t \geq t_1$,

$$P(x(t) + z \in D(x^*, \delta)) \geq (1 + t)^{(\theta - 1)} \omega_{\text{min}} \geq (t + 1)^{\theta/2 - 1}.$$ 

$$P(A^{(t+1)} \cap N_\epsilon(F_i) \neq \emptyset \mid A^{(t)} \cap N_\epsilon(F_i) = \emptyset) \geq (t + 1)^{\theta/2 - 1}.$$ 

$$\beta_i^{(t)} = P(A^{(t+1)} \cap N_\epsilon(F_i) = \emptyset \mid A^{(t)} \cap N_\epsilon(F_i) = \emptyset) \leq t^2 (t + 1)^{-2}.$$ 

$$\gamma_i^{(t)} \leq (t + 1)^{-2}.$$ 

Then $\sum_{t=1}^{+\infty} \gamma_i^{(t)} < +\infty$. Applying Theorem 1(2), we finish the proof. 

Here is a mutation example which satisfies the conditions of Theorem 2: $\sigma^{(t)} = t^{-0.9}$ and $\varphi(z)$ is a Gaussian probability density function.

4 Conclusions and future work

This paper has investigated the convergence properties of a self-adaptive $(\mu + 1)$-MOEA based on grids [9,11]. Theorem 1 establishes general convergence conditions; Theorem 2 proves that the MOEA using self-adaptive mutations is convergent. Although the discussion of this paper is oriented towards a $(\mu + 1)$ MOEA, it is possible to generalize Theorem 1 and 2 to other types of MOEAs. An future research issue is to specify the time complexity for the algorithms to converge to the Pareto-optimal set [11].
References


