Swarm and Evolutionary Computation 12 (2013) 18–23

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

How can surrogates influence the convergence of evolutionary algorithms?

Yu Chen [a,*], Weicheng Xie [b], Xiufen Zou [b]

[a] School of Science, Wuhan University of Technology, Wuhan 430070, China
[b] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 20 October 2012
Received in revised form
25 March 2013
Accepted 8 April 2013
Available online 15 April 2013

Keywords:
Theoretical analysis
Surrogate-assisted evolutionary algorithm
First-order polynomial model
Exploitation ability
Exploration ability

ABSTRACT

Surrogate-assisted evolutionary algorithms have been widely utilized in science and engineering fields, while rare theoretical results were reported on how surrogates influence the performances of evolutionary algorithms (EAs). This paper focuses on theoretical analysis of a (1+1) surrogate-assisted evolutionary algorithm ((1+1)SAEA), which consists of one individual and pre-evaluates a newly generated candidate using a first-order polynomial model (FOPM) before it is precisely evaluated at each generation. By performing comparisons between a unimodal problem and a multi-modal problem, we rigorously estimate the variation of exploitation ability and exploration ability introduced via the FOPM. Theoretical results show that the FOPM employed to pre-evaluate the candidates sometimes accelerate the convergence of evolutionary algorithms, while sometimes prevents the individuals from converging to the global optimal solution. Thus, appropriate adaptive strategies of candidate generation and surrogate control are needed to accelerate the convergence of the (1+1)EA. Then, the accelerating effect of FOPM decreases monotonically with p, the probability of performing precise function evaluation when a candidate is pre-evaluated worse than the present individual.

1. Introduction

Nowadays evolutionary algorithms (EAs) have been widely utilized to solve various optimization problems in the science and engineering fields. However, when the computational simulation for individual evaluations is highly time-consuming, EAs hardly perform well because the evaluation process of individuals costs too much computation time. To reduce the heavy complexity of fitness evaluations in real applications, the surrogate-assisted evolutionary algorithms (SAEAs) approximately evaluate solutions by surrogates, which greatly improves the performances of EAs when they are utilized to solve various complicated optimization problems [11–13,19,15,26,8,22]. Meanwhile, the algorithms employing surrogate models can also outperform the original algorithms when noise exists [21].

Because surrogates save the computational cost at the cost of evaluation precision, appropriate surrogates and strategies of evolution control are needed to construct efficient SAEAs for problems with different mathematical properties [3,7,10,20,9,23,14,6,2428,17,27,16,5]. Some SAEAs employ low-quality models filtering out poor solutions to reduce the requirement on the quality of surrogates [2,25,1,18], while these results are all based on the numerical results, and no theoretical work has been reported to show how surrogate models reduce the number of expensive evaluations [13].

This paper tries to reveal how surrogates influence the convergence of SAEAs in a rigorous way. By comparing the convergence rates and expected improvements of fitness value (EIFV) of a (1+1) surrogate-assisted evolutionary algorithm ((1+1)SAEA) and its counterpart, a (1+1) evolutionary algorithm ((1+1)EA), we theoretically estimate the effects of a first-order polynomial model (FOPM) on the exploration and exploitation abilities, and consequently, some hints on the design of SAEAs are obtained. The remainder of the paper is structured as follows. Section 2 presents a (1+1)SAEA based on a FOPM, and its performances on a unimodal problem and a multi-modal problem are analyzed in Section 3. Then, Section 4 concludes the paper.

2. A (1+1) surrogate-assisted evolutionary algorithm

2.1. The first-order polynomial model

When a first-order polynomial model (FOPM)

$$f_A(\mathbf{x}) = f_A(x_1, \ldots, x_n) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (1)$$

is utilized to approximate the n-variable function $f(x_1, \ldots, x_n)$, N ($\geq n+1$) sample points are needed to estimate the unknown coefficients of the polynomial models by the least squared method (LSM) [11]. Then, the N samples $y^{(i)} = f(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)})$, $i = 1, \ldots, N$ constitute a linear system $\mathbf{y} = \mathbf{X}\Theta$, where

$$\mathbf{y} = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}]^T, \quad \Theta = [\beta_0, \beta_1, \ldots, \beta_n]^T,$$

* Corresponding author. Tel.: +86 15972176539.
E-mail address: chymath@gmail.com (Y. Chen).

2210-6502/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.swevo.2013.04.005

and

$$\mathbf{X}=\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} & \cdots & x_n^{(N)} \end{bmatrix}.$$

When the rows of $\mathbf{X}$ are linearly independent, LSM gives the estimation

$$\hat{\Theta}=(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_n)=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

To reduce the complexity of model construction as much as possible, only n+1 samples are drawn to construct the first-order polynomial model. For this case, the LSM estimation of Θ is $\mathbf{X}^{-1}\mathbf{y}$, and the constructed surrogate model $f_A(\mathbf{x})$ coincides with $f(\mathbf{x})$ at the n+1 sample points $\mathbf{x}^{(i)}=(x_1^{(i)},x_2^{(i)},\ldots,x_n^{(i)}),\ i=1,2,\ldots,n+1$.

2.2. A (1+1) surrogate-assisted evolutionary algorithm

In this paper, we investigate a (1+1) surrogate-assisted evolutionary algorithm ((1+1)SAEA) (described by Algorithm 1) solving the minimization problem

$$\min\ \ f(\mathbf{x}),\quad \mathbf{x}=(x_1,\ldots,x_n)\in S_x\subset\mathbb{R}^n, \tag{2}$$

where S_x is the feasible region. At the beginning, an archive of size n+1 is randomly generated, and a member with minimum function value is selected to be the individual $\mathbf{x}$ in the population. At each generation, the surrogate model $f_A(\mathbf{x})$ is generated by n+1 members of $\mathcal{A}$, and a newly generated candidate $\mathbf{x}'$ is preevaluated by the surrogate $f_A(\cdot)$. If $\{f_A(\mathbf{x}')<f(\mathbf{x})\}\vee\{f_A(\mathbf{x}')\geq f(\mathbf{x})\wedge\ rand()<p\}$ holds for a predetermined probability p, evaluate $\mathbf{x}'$ by the objective function $f(\cdot)$, and replace $\mathbf{x}$ with $\mathbf{x}'$ when $f(\mathbf{x}')<f(\mathbf{x})$. Simultaneously, replace the worst member $\mathbf{x}_w$ in the archive $\mathcal{A}$ by $\mathbf{x}'$, where $\mathbf{x}_w=\arg\max_{\mathbf{x}\in\mathcal{A}}f(\mathbf{x})$.

Algorithm 1. A (1+1) surrogate-assisted; evolutionary algorithm.

Randomly generate an archive $\mathcal{A}$ of n+1 feasible solutions;
Set $\mathbf{x}=\arg\min_{\mathbf{w}\in\mathcal{A}}f(\mathbf{w})$;
Set the function evaluations counter t≔1;
while the stop criterion is not satisfied **do**
  Generate $f_A(\mathbf{x})$ by n+1 members of the archive $\mathcal{A}$;
  $\mathbf{x}'=mutate(\mathbf{x})$;
  if $\{f_A(\mathbf{x}')<f(\mathbf{x})\}\vee\{f_A(\mathbf{x}')\geq f(\mathbf{x})\wedge rand()<p\}$ **then**
    if $f(\mathbf{x}')<f(\mathbf{x})$**then**
      $\mathbf{x}=\mathbf{x}'$;
      Update $\mathcal{A}$ by replacing $\mathbf{x}_w=\arg\max_{\mathbf{x}\in\mathcal{A}}f(\mathbf{x})$ with $\mathbf{x}'$;
    end if
    t≔$t+1$;
  end if
end while
output the results.

To show the functions of the FOPM, we perform a comparative analysis between the (1+1)SAEA (Algorithm 1) and its counterpart, a (1+1)EA described by Algorithm 2. In both algorithms, the Gaussian mutation is performed to generate new candidate $\mathbf{x}'$, i.e.,

$$\mathbf{x}'=\mathbf{x}+\boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is a random vector obeying the Gaussian distribution $\mathbf{N}(\mathbf{0},\boldsymbol{\sigma})$.

Algorithm 2. A (1+1) evolutionary algorithm.

Randomly generate an individual $\mathbf{x}\in S_x$;
Set the function evaluations counter t≔1;
while the stop criterion is not satisfied **do**
  $\mathbf{x}'$≔$mutate(\mathbf{x})$;
  if $f(\mathbf{x}')<f(\mathbf{x})$ **then**
    $\mathbf{x}=\mathbf{x}'$;
  end if
  t≔$t+1$;
end while
output the results.

3. Comparative analysis between the (1+1)SAEA and the (1+1) EA

3.1. Comparative analysis on the exploitation ability of the (1+1) SAEA and the (1+1)EA

To investigate how the first-order polynomial model influences the exploitation ability of the (1+1)EA, we do the analysis via the minimization problem of the quadratic problem

$$\min_{x\in\mathbb{R}}\ \ f(x)=x^2. \tag{3}$$

Then, the first-order polynomial surrogate is

$$f_A=\beta_0+\beta_1 x. \tag{4}$$

The FOPM is confirmed by two samples in the archive $\mathcal{A}$, one of which is denoted as (s,s^2), and another is (x,x^2) coinciding with the present individual of the (1+1)SAEA.[1]

Definition 1. Let $I(x)=|x|$ be the fitness function of feasible solutions, and $x^{(t)}(t=1,2,\ldots)$ be the individual of Algorithm 1 (or Algorithm 2) after the tth function evaluation. The expectation

$$R_C(x)=\mathbf{E}\left[\frac{I(x^{(t+1)})-I(x^*)}{I(x^{(t)})-I(x^*)}\,|x^{(t)}=x\right] \tag{5}$$

is called the convergence rate of Algorithm 1 (or Algorithm 2) at x, where x^* is the global optimal solution of (3).

Denote $R_C^1(x)$ and $R_C^2(x)$ to be the convergence rates of Algorithms 1 and 2 at x, respectively. In Algorithm 1, the precise function evaluation is performed only when $A=\{f_A(x')<f(x)\}\cup\{f_A(x')\geq f(x)\cap rand()<p\}$ occurs, where x' is the new candidate generated by mutation. Let $\mathbf{P}(A)$ be the probability in which A occurs. Then,

$$R_C^1(x)=\frac{1}{\mathbf{P}(A)}\int_{A\bigcap S_x}\frac{I(x^{(t+1)})-I(x^*)}{I(x)-I(x^*)}\,dP$$
$$=\frac{1}{\mathbf{P}(A)}(Int_p(x)+pInt_s(x)),$$

where

$$\mathbf{P}(A)=\mathbf{P}(\{f_A(x')<f(x)\})+p\mathbf{P}\{f_A(x')\geq f(x)\},$$

$$Int_p(x)=\int_{\{f_A(x')<f(x)\}\bigcap\{f(x')\geq f(x)\}}dP+\int_{\{f_A(x')<f(x)\}\bigcap\{f(x')<f(x)\}}\frac{I(x')-I(x^*)}{I(x)-I(x^*)}\,dP,$$

and

$$Int_s(x)=\int_{\{f_A(x')\geq f(x)\}\bigcap\{f(x')\geq f(x)\}}dP+\int_{\{f_A(x')\geq f(x)\}\bigcap\{f(x')<f(x)\}}\frac{I(x')-I(x^*)}{I(x)-I(x^*)}\,dP.$$

[1] To apparently show the different positions of the two samples, in the following (s,s^2) is called the sample point, and (x,x^2) is termed as the present individual.

In Algorithm 2, a precise function evaluation is performed once a new candidate x' is generated. Then,

$$R_C^2(x) = \int_{S_x} \frac{I(x^{(t+1)}) - I(x^*)}{I(x) - I(x^*)} = Int_p(x) + Int_s(x).$$

Thus, the ratio of convergence rate

$$\frac{R_C^1(x)}{R_C^2(x)} = \frac{1}{\mathbf{P}(A)} \frac{Int_p(x) + pInt_s(x)}{Int_p(x) + Int_s(x)}.$$

For the unconstrained optimization problem (2), it holds that

$$\mathbf{P}(\{f_A(x') < f(x)\}) = \mathbf{P}(\{f_A(x') \ge f(x)\}) = \frac{1}{2}.$$

Consequently,

$$\frac{R_C^1(x)}{R_C^2(x)} = \frac{2}{1+p} \frac{Int_p(x) + pInt_s(x)}{Int_p(x) + Int_s(x)}, \quad (6)$$

and the following theorem holds.

Theorem 1. *When the first-order polynomial model is introduced in the (1+1)EA illustrated by Algorithm2, the relative change of convergence rate for problem (3) cannot exceed* $(1-p)(1-\mathcal{E}(x,\sigma))/(1+p)(1+\mathcal{E}(x,\sigma))$*, that is,*

$$1 + \frac{(1-p)(1-\mathcal{E}(x,\sigma))}{(1+p)(1+\mathcal{E}(x,\sigma))} \ge \frac{R_C^1(x)}{R_C^2(x)} \ge 1 - \frac{(1-p)(1-\mathcal{E}(x,\sigma))}{(1+p)(1+\mathcal{E}(x,\sigma))}, \quad (7)$$

where $\mathcal{E}(x,\sigma) = \frac{1}{2} e^{-8x^2/\sigma^2}$.

Proof. According to the positions of x and the sample point s, there are two different cases to be distinguished. For the two different cases, the respective convergence rate can be improved and reduced, however, the changes of convergence rate can always be upper bounded.

1. When the present individual x and the sample point s are on different sides of the global optimal solution x^* (Fig. 1a), it holds that

$$Int_p(x) = \int_{\{f_A(x') < f(x)\} \cap \{f(x') \ge f(x)\}} dP = \int_{\{f_A(x') < f(x)\}} dP = \frac{1}{2}.$$

Moreover,

$$Int_s(x) = \int_{\{f_A(x') \ge f(x)\} \cap \{f(x') \ge f(x)\}} dP + \int_{\{f_A(x') \ge f(x)\} \cap \{f(x') < f(x)\}} \frac{|x'|}{|x|} dP$$

$$= \int_{f_A(x') \ge f(x)} dP - \int_{\{f_A(x') \ge f(x)\} \cap \{f(x') < f(x)\}} dP + \int_{\{f_A(x') \ge f(x)\} \cap \{f(x') < f(x)\}} \frac{|x'|}{|x|} dP.$$

Because $|x'|/|x|$ is always less than one when $f(x') < f(x)$, it holds that

$$\frac{1}{2} \ge Int_s(x) \ge \frac{1}{2} - \int_0^{2|x|} dP,$$

and from (6) we know that

$$1 \le \frac{R_C^1(x)}{R_C^2(x)} \le 1 + \frac{(1-p)\int_0^{2|x|} dP}{(1+p)(1-\int_0^{2|x|} dP)}. \quad (8)$$

2. When the present individual x and the sample point s are on the same side of the global optimal solution x^* (Fig. 1b), it holds that

$$Int_s(x) = \int_{\{f_A(x') \ge f(x)\} \cap \{f(x') \ge f(x)\}} dP = \frac{1}{2}.$$

Meanwhile,

$$Int_p(x) = \int_{\{f_A(x') < f(x)\} \cap \{f(x') \ge f(x)\}} dP + \int_{\{f_A(x') < f(x)\} \cap \{f(x') < f(x)\}} \frac{|x'|}{|x|} dP$$

$$= \int_{f_A(x') < f(x)} dP - \int_{\{f_A(x') < f(x)\} \cap \{f(x') < f(x)\}} dP + \int_{\{f_A(x') < f(x)\} \cap \{f(x') < f(x)\}} \frac{|x'|}{|x|} dP.$$

Because $|x'|/|x|$ is always less than one when $f(x') < f(x)$, for this case it holds that

$$\frac{1}{2} \ge Int_p(x) \ge \frac{1}{2} - \int_0^{2|x|} dP,$$

and from (6) we know that

$$1 \ge \frac{R_C^1(x)}{R_C^2(x)} \ge 1 - \frac{(1-p)\int_0^{2|x|} dP}{(1+p)(1-\int_0^{2|x|} dP)}. \quad (9)$$

From (8) and (9), we know that

$$1 + \frac{(1-p)\int_0^{2|x|} dP}{(1+p)(1-\int_0^{2|x|} dP)} \ge \frac{R_C^1(x)}{R_C^2(x)} \ge 1 - \frac{(1-p)\int_0^{2|x|} dP}{(1+p)(1-\int_0^{2|x|} dP)}.$$

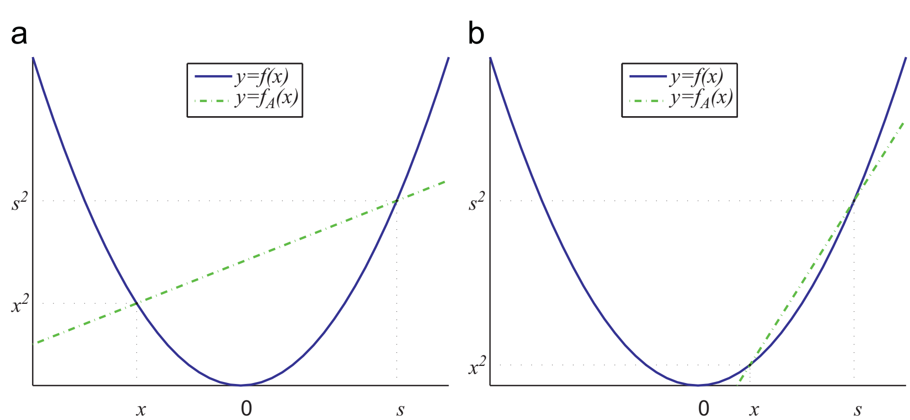

Fig. 1. Two different cases for positions of the present individual x and the sample point s in relation to the global optimal solution x^* of problem (3). (a) Case 1: the present individual x and the sample point s are on different sides of the global optimal solution zero; (b) Case 2: the present individual x and the sample point s are on same side of the global optimal solution zero.

Moreover, because

$$\int_0^{2|x|} dP = \frac{1}{\sqrt{2\pi}\sigma}\int_0^{2|x|} e^{-y^2/\sigma^2}\,dy \le \frac{\sqrt{1-e^{-8x^2/\sigma^2}}}{2} \le \frac{1}{2}\left(1-\frac{1}{2}e^{-8x^2/\sigma^2}\right),$$

we come to the conclusion that

$$1+\frac{(1-p)(1-\mathcal{E}(x,\sigma))}{(1+p)(1+\mathcal{E}(x,\sigma))} \ge \frac{R_C^1(x)}{R_C^2(x)} \ge 1-\frac{(1-p)(1-\mathcal{E}(x,\sigma))}{(1+p)(1+\mathcal{E}(x,\sigma))},$$

where $\mathcal{E}(x,\sigma)=\frac{1}{2}e^{-8x^2/\sigma^2}$. □

$R_C(x)$, the convergence rate of the(1+1)EA (or the (1+1)SAEA) for problem (3), indicates its exploitation ability for local search. The proof of Theorem 1 reveals that for Case 1, the surrogate will hinder the convergence of (1+1)SAEA (Algorithm 2), and it can accelerate its convergence when Case 2 occurs. Thus, negative influence of Case 1 and the positive influence of Case 2 will weaken each other. Meanwhile, Theorem 1 also shows that by introducing the surrogate in Algorithm 1, the variation of convergence rate cannot exceed $((1-p)(1-\mathcal{E}(x,\sigma))/(1+p)(1+\mathcal{E}\,(x,\sigma)))R_C^2(x)$, which decreases monotonically with the predetermined probability p. Thus, to make the variation as great as possible, p is preferred to be zero, that is, the newly generated candidate x' have to be refused when it is evaluated by surrogate worse than the present x. Although the surrogate can maximize the convergence of Algorithm 1 in Case 2 (discussed in the proof of Theorem 1), but, in this way the (1+1)SAEA cannot outperform its counterpart as greatly as possible, since the pre-evaluation result of model (4) will eliminate the promoting candidates generated in Case 1 (discussed in the proof of Theorem 1). Thus, to maximize the promoting effect of the FOPM on the (1+1)EA, an appropriate value of the parameter σ is needed, and then, the maximum improvement of convergence rate can be obtained when p is endowed with a small value. Ideally, when Case 1 discussed in the proof of Theorem 1 is eliminated, the maximum improvement of convergence rate could be $(1-\frac{1}{2}e^{-8x^2/\sigma^2}/1+\frac{1}{2}e^{-8x^2}\quad/\sigma^2)R_C^{(2)}(x)$ by setting $p=0$.

3.2. Comparative analysis on the exploration ability of the (1+1) SAEA and the (1+1)EA

To investigate how the first-order polynomial model influences the exploration ability of the (1+1)EA, we study the minimization problem of one-dimensional Rastrigin's function

$$\min_{x\in\mathbb{R}} \quad f(x)=x^2-10\cos(2\pi x)+10. \tag{10}$$

by estimating the expected improvement of fitness value (EIFV) for one function evaluation, where

$$EIFV(x)=\mathbf{E}[|x^{(t)}|-|x^{(t+1)}|\,|x^{(t)}=x]. \tag{11}$$

In Algorithm 1, the EIFV for one fitness evaluation is

$$EIFV^{(1)}(x)=\frac{1}{\mathbf{P}(A)}\int_{A\cap S_x}\left|x^{(t)}\right|-\left|x^{(t+1)}\right|\,dP = \frac{1}{\mathbf{P}(A)}(\Delta I^{(e)}(A,x)+p\Delta I^{(s)}(x)),$$

where

$$\mathbf{P}(A)=\mathbf{P}(\{f_A(x')<f(x)\})+p\mathbf{P}\{f_A(x')\ge f(x)\} = \frac{1}{2}+\frac{p}{2},$$

$$\Delta I^{(e)}(A,x)=\int_{\{f_A(x')<f(x)\}\cap\{f(\mathbf{x}')<f(\mathbf{x})\}}|x|-|x'|\,dP,$$

and

$$\Delta I^{(s)}(A,x)=\int_{\{f_A(x')\ge f(x)\}\cap\{f(x')<f(x)\}}|x|-|x'|\,dP.$$

In Algorithm 1, the EIFV for one function evaluation is

$$EIFV^{(2)}(x)=\Delta I^{(e)}(A,x)+\Delta I^{(s)}(x).$$

Thus, the difference between EIFVs of Algorithms 1 and 2

$$\left|EIFV^{(1)}(x)-EIFV^{(2)}(x)\right| = \frac{1-p}{1+p}|\Delta I^{(e)}(A,x)-\Delta I^{(s)}(A,x)|. \tag{12}$$

Theorem 2. *When the first-order polynomial model is introduced in the (1+1)EA illustrated by Algorithm2, $|EIFV^{(1)}(x)-EIFV^{(2)}(x)|$, the absolute change of EIFVs, is in the order of $\Omega((\sqrt{[x]}/\sigma)e^{-2/\sigma^2})$, where $[x]$ is the nearest integer from x.*

Proof. For any $x\in\mathbb{R}$, denote $C=\{x'\in\mathbb{R}|I(x')<I(x)\,\&\,f(x')<f(x)\}$. The region $\{x'\in\mathbb{R}|I(x')\ge I(x)\&f(x')<f(x)\}$ can be divided into several disjoint parts, denoted as A_i and B_i, where A_i and B_i, $(i=1,2,\ldots)$ are symmetric about zero (See Fig. 2). Then,

$$\{x'\in\mathbb{R}|f(x')<f(x)\}=\bigcup_i\{A_i\cup B_i\}\cup C.$$

Denote the global optimal solution to be x^*. There are two different cases to be distinguished. For the two different cases, the respective EIFVs can be improved and reduced, however, the absolute changes of EIFVs can always be lower bounded.

1. When $x^*\in\{x'\in\mathbb{R}|f_A(x')\ge f(x)\}$ (Fig. 2a),

$$\Delta I^{(e)}(A,x)-\Delta I^{(s)}(A,x)$$

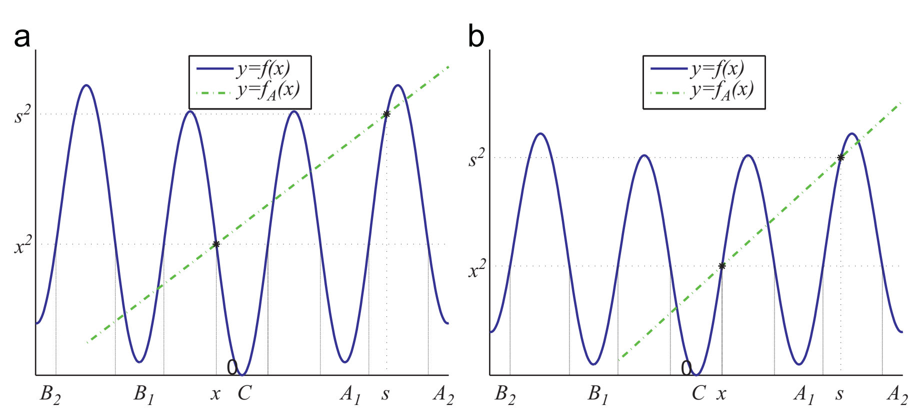

Fig. 2. Two different cases for positions of the present individual x and the sample point s in relation to the global optimal solution x^* of problem (10). (a) Case 1: $x^*\in\{x'\in\mathbb{R}|f_A(x')\ge f(x)\}$; (b) Case 2: $x^*\in\{x'\in\mathbb{R}|f_A(x')<f(x)\}$.

$$= \left(\sum_i \int_{B_i} - \sum_i \int_{A_i}\right) |x| - |x'| \, dP - \int_C |x| - |x'| \, dP$$
$$\le -\int_C |x| - |x'| \, dP. \tag{13}$$

2. When $x^* \in \{x' \in \mathbb{R} | f_A(x') < f(x)\}$ (Fig. 2b),

$$\Delta I^{(e)}(A, x) - \Delta I^{(s)}(A, x)$$
$$= \left(\sum_i \int_{B_i} - \sum_i \int_{A_i}\right) |x| - |x'| \, dP + \int_C |x| - |x'| \, dP$$
$$\ge \int_C |x| - |x'| \, dP. \tag{14}$$

Thus,

$$\left| EIFV^{(1)}(x) - EIFV^{(2)}(x) \right| \ge \frac{1-p}{1+p} \int_C |x| - |x'| \, dP.$$

When x is in the absorbing region S_k of the local optimum x_k,

$$\int_C |x_k| - |x'| \, dP = \min_{x \in S_k} \int_C |x| - |x'| \, dP.$$

To compute the lower bound of $\int_C |x_k| - |x'| \, dP$, let us consider the absorbing region of x_{k-1}, just as described in Fig. 3. Denote

$$b_{k-1} = \min\{x > k-1;\ f(x) = f(x_k)\}$$

and

$$a_{k-1} = \max\{x < k-1;\ f(x) = f(x_k)\}.$$

Then, it holds that [4],[2]

$$x_k \approx k - \frac{k}{20\pi^2 + 1},$$

$$a_{k-1} \approx k-1 + \frac{-(k-1) - \sqrt{(k-1)^2 + (20\pi^2+1)(2k-1)}}{20\pi^2 + 1}$$

and

$$b_{k-1} \approx k-1 + \frac{-(k-1) + \sqrt{(k-1)^2 + (20\pi^2+1)(2k-1)}}{20\pi^2 + 1}.$$

Thus,

$$b_k - a_k \approx \frac{2\sqrt{(k-1)^2 + (20\pi^2+1)(2k-1)}}{20\pi^2 + 1} \ge 2\sqrt{\frac{2k-1}{20\pi^2 + 1}}, \tag{15}$$

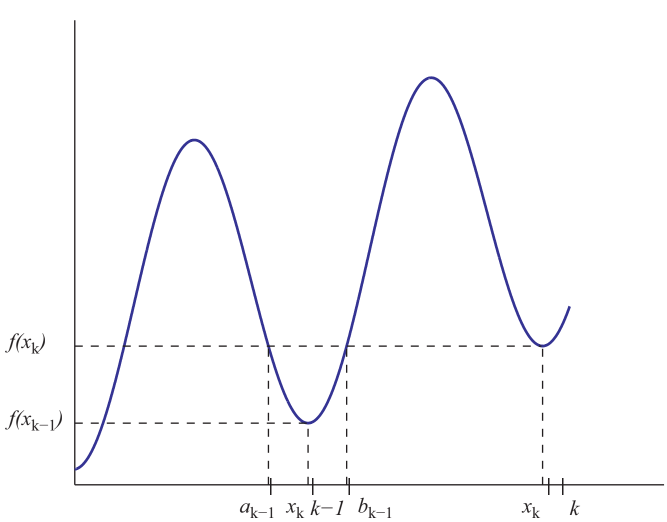

Fig. 3. Local absorbing region of x_{k-1}.

and

$$x_k - x' \approx k - \frac{k}{20\pi^2 + 1} - \left(k-1 + \frac{-(k-1) + \sqrt{(k-1)^2 + (20\pi^2+1)(2k-1)}}{20\pi^2 + 1}\right)$$
$$= 1 - \frac{k + \sqrt{(k-1)^2 + (20\pi^2+1)(2k-1)}}{20\pi^2 + 1}$$
$$\ge 1 - \frac{k + (k-1) + \sqrt{(20\pi^2+1)(2k-1)}}{20\pi^2 + 1}$$
$$= 1 + \frac{1}{20\pi^2 + 1} - \sqrt{\frac{2k-1}{20\pi^2 + 1}}$$
$$\ge \frac{1}{20\pi^2 + 1}. \tag{16}$$

Consequently,

$$\int_C |x| - |x'| \, dP \ge \frac{1}{\sqrt{2\pi}\sigma} \int_{a_k}^{b_k} |x| - |x'| \, dP$$
$$\ge \frac{2}{\sqrt{2\pi}\sigma} \sqrt{\frac{2k-1}{20\pi^2 + 1}} \cdot \frac{1}{20\pi^2 + 1} \cdot e^{-(x-a_k)^2/2\sigma^2}$$
$$\ge \frac{2}{\sqrt{2\pi}\sigma} \sqrt{\frac{2k-1}{20\pi^2 + 1}} \cdot \frac{1}{20\pi^2 + 1} \cdot e^{-2/\sigma^2}$$
$$= \Omega\left(\frac{\sqrt{[x]}}{\sigma} e^{-2/\sigma^2}\right), \tag{17}$$

and we can conclude that the difference between EIFVs of Algorithms 1 and 2 is $\Omega((\sqrt{[x]}/\sigma)e^{-2/\sigma^2})$, where $[x] = k$ is the nearest integer from x. □

Theorem 2 tells us that the difference of EIFVs between Algorithms 1 and 2 is approximately bounded below by $\Omega((\sqrt{[x]}/\sigma)e^{-2/\sigma^2})$. This shows that to magnify the influence of the introduced surrogate, a relatively small value of σ is preferred. Meanwhile, a σ which is large enough can provide a satisfactory convergence speed of the (1+1)SAEA. Thus, at the early stage of the evolution process, a relatively large value of σ should be adopted to endow the (1+1)EA with strong exploration ability. Consequently, the (1+1)SAEA can also explore search space efficiently, although the surrogate model cannot greatly accelerate the convergence of (1+1)SAEA with a big value of σ. However, at the terminal stage, small values of σ are needed to refine the exploitation, and small values of σ can also accelerate the convergence of (1+1)SAEA greatly. Therefore, to maximize the convergence of (1+1)SAEA applied to problem (10), we should endow the standard deviation σ with a relatively big value at the beginning, and let it decrease in the process of population evolution. Meanwhile, the probability p should be set as small as possible if the first case discussed in the proof of Theorem 2 can be eliminated.

4. Conclusions and future work

By theoretically comparing a (1+1)EA and a (1+1)SAEA, this paper endeavors to reveal how the surrogates function on EAs by prescreening newly generated candidates. Theoretical investigations for a unimodal problem and a multi-modal problem show that the (1+1) SAEA can accelerate the convergence of (1+1)EA if the generation process of candidates is carefully designed. First, the negative functions of surrogates decline with the increase of the probability parameter p. Second, to obtain strong exploitation and exploration

[2] According to the derivations in [4], the approximate errors are in lower orders, which will not influence the validity of the following argument.

abilities, the value of σ should be carefully selected to accommodate landscape of the objective function. Future work will focus on how to design an efficient adaptive strategy of mutation parameter σ and the surrogate control method to get a high-performance surrogate-assisted evolutionary algorithm.

Acknowledgment

This work was supported by the Fundamental Research Funds for the Central Universities of China and the Key Program of the National Natural Science Foundation of China under Grant 51039005.

References

[1] K. Abboud, M. Schoenauer, Surrogate deterministic mutation: preliminary results, in: Artificial Evolution, Lecture Notes in Computer Sciences, 2002, pp. 919–954.
[2] K. Anderson, Y. Hsu, Genetic crossover strategy using an approximation concept, in: IEEE Congress on Evolutionary Computation, 1999, pp. 527–533.
[3] A.J. Brooker, J. Dennis, P.D. Frank, D.B. Serafini, V. Torczon, M. Trosset, A rigorous framework for optimization of expensive functions by surrogates, Structural Optimization 17 (1998) 1–13.
[4] Y. Chen, X. Zou, J. He, Drift conditions for estimating the first hitting times of evolutionary algorithms, International Journal of Computer Mathematics 88 (1) (2011) 37–50.
[5] G. Chen, X. Han, G. Liu, C. Jiang, Z. Zhao, An efficient multi-objective optimization method for black-box functions using sequential approximate technique, Applied Soft Computing 12 (1) (2012) 14–27.
[6] M. Emmerich, K.C. Giannakoglou, B. Naujoks, Single- and multiobjecitve evolutionary optimization assisted by Gaussian random field metamodels, IEEE Transactions on Evolutionary Computation 10 (4) (2006) 421–439.
[7] M. Farina, A minimal cost hybrid strategy for Pareto optimal front approximation, Evolutionary Optimization 3 (1) (2001) 41–52.
[8] M. Hauschild, M. Pelikan, An introduction and survey of estimation of distribution algorithms, Swarm and Evolutionary Computation 1 (2011) 111–128.
[9] Y.-S. Hong, H. Lee, M.-J. Tahk, Acceleration of the convergence speed of evolutionary algorithms using multi-layer neural networks, Engineering Optimization 35 (1) (2003) 91–102.
[10] Y. Jin, M. Olhofer, B. Sendhoff, A framework for evolutionary optimization with approximate fitness functions, IEEE Transactions on Evolutionary Computation 6 (5) (2002) 481–494.
[11] Y. Jin, A comprehensive survey of fitness approximation in evolutionary computation, Soft Computing 9 (1) (2005) 3–12.
[12] Y. Jin, J. Branke, Evolutionary optimization in uncertain environments—a survey, IEEE Transactions on Evolutionary Computation 9 (3) (2005) 303–317.
[13] Y. Jin, Surrogate-assisted evolutionary computation: recent advances and future challenges, Swarm and Evolutionary Computation 1 (2011) 61–70.
[14] J. Knowles, Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, IEEE Transactions on Evolutionary Computation 10 (1) (2006) 50–66.
[15] Y. Lian, A. Oyama, M.S. Liou, Progress in design optimization using evolutionary algorithms for aerodynamic problems, Progress in Aerospace Sciences 46 (2010) 199–223.
[16] Z. Liang, K.K. Choi, I. Lee, Metamodeling method using dynamic kriging for design optimizaiton, The American Institute of Aeronautics and Astronautics Journal 49 (9) (2011) 2034–2046.
[17] D. Lim, Y. Jin, Y.S. Ong, B. Sendhoff, Generalizing surrogate-assisted evolutionary computation, IEEE Transactions on Evolutionary Computation 14 (3) (2010) 329–355.
[18] I. Loshchilov, M. Schoenauer, S. Sebag, A mono surrogate for multiobjecitve optimiation, in: Genetic and Evolutionary Computation Conference, 2010, pp. 471–478.
[19] S. Mitra, Y. Hayashi, Bioinformatics with soft computing, IEEE Transactions on Systems, Man and Cybernetics, Part C 36 (5) (2006) 616–635.
[20] P. Nain, K. Deb, A computationally effective multi-objective search and optimization techniques using coarse-to-fine grain modeling, in: 2002 PPSN Workshop on Evolutionary Multiobjective Optimization.
[21] F. Neri, X. del Toro Garcia, G.L. Cascella, N. Salvatore, Surrogate assisted local search in PMSM drive design, COMPEL: The International Journal of Computation and Mathematics in Electrical and Electronic Engineering 27 (3) (2008) 573–592.
[22] F. Neri, C. Cotta, Memetic algorithms and memetic computing optimization: a literature review, Swarm and Evolutionary Computation 2 (2012) 1–14.
[23] Y.S. Ong, P.B. Nair, A.J. Keane, Evolutionary optimiztion of computationally expensive problems via surrogate modeling, The American Institute of Aeronautics and Astronautics Journal 41 (4) (2003) 687–696.
[24] Y.S. Ong, P.B. Nair, K.Y. Lum, Max–min surrogate-assisted evolutionary algorithm for robust design, IEEE Transactions on Evolutionary Computation 10 (4) (2006) 392–404.
[25] K. Rasheed, H. Hirsh, Informed operators: speeding up genetic-algorithm-based design optimization using reduced models, in: Genetic and Evolutionary Computation Conference, Morgan Kaufmann, 2000, pp. 628–635.
[26] I. Voutchkov, A. Keane, Multi-objective optimization using surrogates, in: Computational Intelligence in Optimization (ALO 7), Springer, 2010, pp. 155–175.
[27] Q. Zhang, W. Liu, E. Tsang, B. Virginas, Expensive multiobjective optimization by MOEA/D with Gaussian process model, IEEE Transactions on Evolutionary Computation 14 (3) (2010) 456–474.
[28] A. Zhou, Y.-S. Ong, M.-H. Lim, B.-S. Lee, Memetic algorithm using multi-surrogate for computationally expensive optimization problems, Soft Computing 11 (10) (2007) 957–971.