Design of a Surrogate Model Assisted $(\mu/\mu, \lambda)$-ES

JINGYUN YANG AND DIRK V. ARNOLD, Faculty of Computer Science, Dalhousie University

Surrogate models have been widely used to assist evolutionary algorithms (EAs) to avoid unnecessary objective function evaluations. The cost is reduced by substituting the true objective function evaluation with a cheap but inaccurate estimate using the surrogate model. The surrogate model is built on th knowledge gained in previous iterations. Using surrogate assised (1+1)-ES for simple model and single steps have been studied, but the effect of actual inferior parent resulted from an inaccurate surrogate estimation and the corresponding poor step size are not well understood. We study the behaviour using a surrogate model assisted $(\mu/\mu, \lambda)$-ES using a population instead of a single offspring with the hope to possible address the issue. The bahaviours of the two strategies are compared using several test functions.

Additional Key Words and Phrases: $(\mu/\mu, \lambda)$-ES, Surrogate Model, Evolutionary algorithms(EAs), Gaussian Process

ACM Reference format:
Jingyun Yang and Dirk V. Arnold. 2018. Design of a Surrogate Model Assisted $(\mu/\mu, \lambda)$-ES. 1, 1, Article 1 (November 2018), 10 pages. https://doi.org/

1 INTRODUCTION

Evolution strategies (ESs) have been widely utilized to solve optimization problems where the true objective function evaluation is computationally-intensive. Various attempts have been made to reducte the cost by extracting the information obtained from points evaluated in previous iterations. Such information yields insights into better mutation and recombination that help generate and select promising offspring. Cummulative step size adaptation (CSA) [20] builds an evolution path based on the history step size (mutation) of ESs, the population in the next iteration is generated based on the mutation adpated by the evolution path.

The history information could be used to construct a surrogate model, referred either as a local approximation or a global approximation to the true objective function [16]. There are a range of surrogate models and a survey of the development can be found by Jin [14] and Loshchilov [18]. Those algorithms are usually heuristic by nature and the behaviour of each step is likely not well interpreted. Recent work in surrogate assisted EAs tend to use sophosticated algroithm where surrogates are combined or the model is updated online according to some heustic. Comparision is often made by comparing the performance using the algorithm with and without model assistance where the behaviour of the surrogate is not well simulated. In this context, an approach that could simulate the surrogate would be helpful in understanding the surrogate behaviour, leading to potential modification for surrogate update or parameter-setting. A surrogate that models the objective function with desired precise gains benefit especially for algorithms that requires

Manuscript submitted to ACM

a large population size for good performance.The computational saving largely lies in the saved evaluations outshine the potential poor step resulted from relative inaccurate estimation of candidate solutions.

This paper intend to improve the understanding of the impact of population size on surrogate-assisted ESs' by analyzing using simple test functions with strong theoretical basis and established baselines. The paper is organized as follows: In Section 2 we give a brief review of related background, in Section 3 we propose a local surroagte model-assisted $((\mu/\mu, \lambda) + (\mu/\mu, \lambda))$-ES and study its behaviour on sphere functions. Based on the existing knowledge and step behaviour, in Section 4, we then propose a combined step size adaptation mechanism for the this algorithm, analyze the performance using several test functions and compare the result with a surrogate model-assisted (1+1)-ES [17]. The experimental result is followed by a discuession and future work in Section 5.

2 RELATED WORK

2.1 Surrogate Model

Using an approximate model to reduce computational cost can be traced back to 1960s [7]. Some successful surrogated models include but are not limitted to Polynomial Regression (PR, response surface methodology) [11], Gaussian Process (GP, Kriging models) [13], Artificial neural networks [24]. There are two types of surrogate models, global surrogate model and local surrogate model, . ES using global surrogate model based on Kring was examined by Ratle [21]. Another ES using global surrogate model based on Artificial neural networks was constructed by Jin [15] which gives an imperial criterion on using the true objective function or the surrogate model to evaluate the offspring. Ulmer et al [25] and Buche et al [5] also applied GP as surrogate models in ES. But the performance of global surrogate models degrade as the dimension of the data increases, known as *curse of dimensionality*. Since the performance of ES is straightly affected by the surrogate model accurancy, online surrogates has been introduced by using a surrogate-adaptation mechanism that updated the model according to some heustic. Loshchilov et al [12] uses . Online local surrogate models [26] can be constructed using methods like radial basis function (RBF) [8] to replace the global surrogate model, where the surrogate model is updated online, giving a more accurate estimation compared with the global surrogate model.

There are various surrogate-assisted EAs integrating global and local surrogate models or using a combination of heuristics. These methods tend to be sophisticated for good performance, while few literatures have *systematically investigated*??? the surrogated-assisted $(\mu/\mu, \lambda)$-ES. One exception is what Chen and Zou [6] proposed but yet incomplete in terms of two aspects. Firstly, it uses a linear surrogate that cannot give a precise estimate when coordinate transform is applied, the precondition to solve a generalized optimization problem [17]. Secondly, it does not include a step size adaptation mechanism. Besides that, Ulmer et al [9] proposed a Model Assisted Steady-State Evolution Strategy (MASS-ES), which is a $(\mu + \lambda)$-ES that is a (1+1)-ES when we set $\mu = \lambda = 1$. But the behavior of step size adaptation is unclear given the proposed conditions.

There is a wealth of literatures for solving black box optimization using (1+1)-ES on unimodal test problems given the convergence property of convex functions. Kayhani and Arnold [17] proposed a surrogated-assisted (1+1)-ES that investigates the acceleration and signgle step behaviour of the algorithm using GP based local surrogate. In this algorithm, the local surrogate acts as a filter and is updated every time when a true objective function is made. Since (1+1)-ES generate a single offspring per iteration and is not as rubust as $(\mu/\mu, \lambda)$ especially in the presence of surrogate (bias due to choice of points), we argue that it is natural to ask to what degree the choice of population can benefit the ES in terms of rubustness and acceleration.

2.2 Step size adaptation

The step size of $(\mu/\mu, \lambda)$-ES is commonly adapated using cumulative step size adaptation (CSA) proposed by Ostermeier et al [20]. In each iteration, $(\mu/\mu, \lambda)$-ES generate λ candidate solutions $y_i \in \mathbb{R}^N$, $i = 1, \ldots, \lambda$ from a parental population $x_i \in \mathbb{R}^N i = 1, \ldots, \mu$ and the centroid of the parent population is $x = 1/\mu \sum_{i=1}^{\mu} x_i$ where $\mu < \lambda$. The parental population is replaced by the best μ candidate solutions gennerated by $y_i = x + \sigma z$ where $\sigma \in \mathbb{R}$ is a scalar referred to as the step size and $z \in \mathbb{R}^N$ as the mutation. For a strategy with idelaly adapted step size, eahch step should be uncorrelated. If the connective are negatively correlated, the step size should be decreased. In contrast, if the connective steps are positively correlated, the steps are pointing to the same direction. Then a number of small steps can be replaced by fewer large steps and therefore, the step sie should increase.

To decide the correlation, information from previous steps and mutations are cummulated. By comparing the step size with its expected length under random selection, the step size is adapted according to its expedcted length. It increases if the length is less than expected and decrease otherwise.

Define the search path as

$$p_{k+1} \leftarrow (1-c)p_k + \sqrt{\mu c(2-c)}z, \tag{1}$$

where $0 < c \geq 1$ is the proportion of history information retaiend and passed to the evolution path in the next iteration, $\sqrt{\mu c(2-c)}$ is a normalization constant that updates the evolution path from the mutation of this iteration and z the mutation obatined by averaging the best μ candidate solutions generated.

The step size is adapated

$$\sigma \leftarrow \sigma \exp\left(\frac{c}{d}\left(\frac{\|p\|}{E\|N(O,I)\|}\right)\right), \tag{2}$$

where $E\|N(0,I)\|$ is the expected length of the search path p and can be approximated as $E\|N(0,I)\| \approx \sqrt{n}(1 - 1/4n + 1/21n^2)$. In Section 4, the detail will be discussed.

(*not step size adaptation COULD IN FUTURE WORK*) Covariance Matrix Adaptation (CMA) [19].

3 ANALYSIS

To understand the potential implications of using surrogate models in EAs with varying population size, in this section, we use a simple model that applis a surrogate on the population. Specifically, we propose an EA that, in each iteration, a population of new λ candidate solutions are generated and then evaluated by the surrogate instead of true obejctive function calls and a selection based on the inaccurate surrogate estimate is done followed by a true objective function evaluation for the centroid of the selected referred to as the parent for next iteration. We assume that the inaccurate estimate of the surrogate model is a Gaussin random variable with mean equals the true objective function value of the candidate solution with some variance that describes the accuracy of the surrogate model. So, we can apply the technique of analyzing ESs's behaviours in the presence of Gaussian noise [3]. The analysis could be extend to biased surrogate modlels where the distribution mean is different from the exact objective function value[17].....

Comparision based surrogate model

Consider the minimization of the quadratic sphere $f : \mathbb{R}^N \to \mathbb{R}$ with $f(x) = x^T x$ where the surrogate model assisted $(\mu/\mu.\lambda)$-ES is applied. This section will use the surrogate model described above to replace the true objective function calls of candidate solutions in each iteration, inaccurate but at vanishing cost. We first consider a simple iteration of the

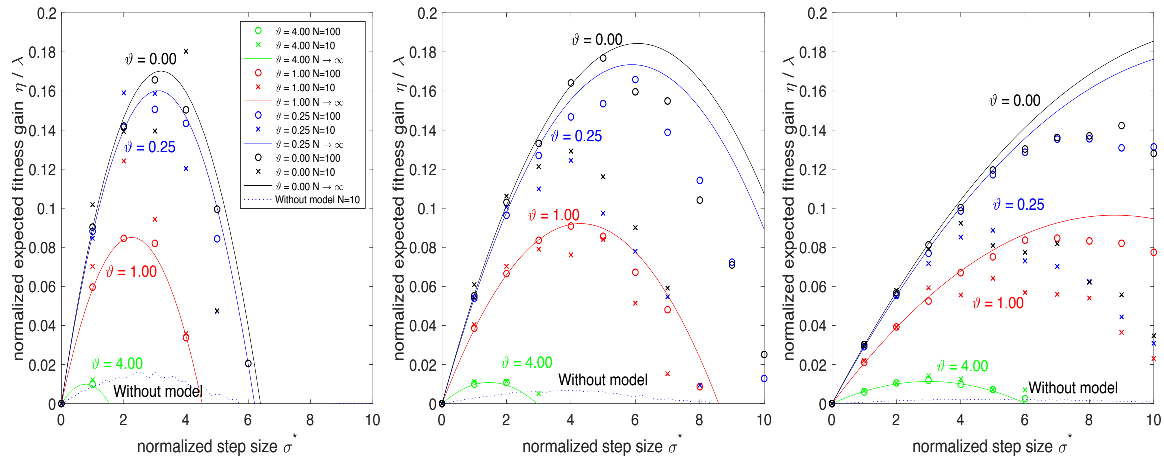

Fig. 1. The figures from left to right shows the expected signle step behaviour of the surrogate model assisted $(\mu/\mu, \lambda)$-ES with unbiased Gaussin distributed surrogate error with $\lambda = 10, 20, 40$ respectively where $\mu = \lceil \lambda/4 \rceil$. The solid lines are the results obtained analytically when $n \longrightarrow \infty$, while the dotted line below illustrates the corresponding performance ($n = 10$) of the $(\mu/\mu, \lambda)$-ES without model assistance. The dots represents the experimental result for $n = 10$ (croesses) and $n = 100$ (circles).

strategy. In each iteration, a population size of λ new candidate solutions $y_i \in \mathbb{R}^N, i = 1, ..., \lambda$ are generated from μ parents $x_i \in \mathbb{R}^N, i = 1, ..., \mu$, where $\lambda > \mu$. The parental population with size μ are replaced by the best μ candidate solutions $y_{i;\lambda}, i = 1, 2, ..., \mu$ evaluated by the surrogate model with fitness estimate $f_\epsilon(y_{i;\lambda}) \leq f_\epsilon(y_{j,\lambda}), 1 \leq i < j \leq \lambda$ at vanishing cost. For each of the λ candidate solutions $y_i = x + \sigma z$ where the parent $x = \sum_{i=1}^{n} x_i/\mu$, the centroid of the parental population is obtained through intermediate recombination, $z \in \mathbb{R}^N$ is a standard normally distributed random vector, $\sigma > 0$ is the step size of the strategy that the adaptation is discussed in Section 4. The strategy uses the surrogate model to obtain a fitness estimate of the candidate solution $f_\epsilon(y_i), 1 \leq i \geq \lambda$ and by the assumption the estimate has mean $f(y_i)$ with some standard deviation $\sigma_\epsilon > 0$ (*surroagte model error* also as *fitness noise* [2]). Better surrogate model results in smaller model error σ_ϵ. For the λ new candidate solutions $f_\epsilon(y_i) < f_\epsilon(y_j), 1 \leq i < j \leq \lambda$ indiccates the estimated objective function value of y_i is superior to y_j and therefore the best μ candidate solutions are selected,replacing the old parental population of size μ (used for offspring generation in next iteration), while the other inferior candidate solutions are discarded. Therefore, in each iteration only one objective function call is made in evaluating the fitness of the parent (centroid of parental population). The surrogate essentially does a pre-selection for $(\mu/\mu, \lambda)$-ES over candidate solutions, avoiding the unecessary objective function calls determined by the surroagte model.

Decomposition of z, first proposed by Rechenberg [22] can be used to study the expected step size of the strategy. We can decompose the vector z as a vector sum $z = z_1 + z_2$, where z_1 is in the direction of the negative gradient of the objecive function $\nabla f(x)$, while z_2 orthogonal to z_1. We have z_1 standard normally distributed, while $\|z_2\|^2$ χ-distributed with $N - 1$ degree of freedom and $\|z_2\|^2/N \overset{N\to\infty}{=} 0$ (see reference theorem [*dirk's slides*]). Denote $\delta = N(f(x) - f(y))/(2R^2)$ where $R = \|x\|$ is the distance to the optimal, we further introduce normalized step size $\sigma^* = N\sigma/R$ and $z_{\text{step}} = \sum_{i=1}^{\mu} z_{i;\lambda}$ (the averaged z taken by the best *mu* candidate solutions). The normalized fitness advantage of y over x follows

$$\begin{aligned}\delta &= \frac{N}{2R^2}(x^T x - (x+\sigma z_{\text{step}})^T(x+\sigma z_{\text{step}})) \\ &= \frac{N}{2R^2}(-2\sigma x^T z_{\text{step}} - \sigma^2 \|z_{\text{step}}\|^2) \\ &\overset{N\to\infty}{=} \sigma^* z_{\text{step},1} - \frac{\sigma^{*2}}{2},\end{aligned} \tag{3}$$

where $z_{\text{step},1}$, the component of z_{step} pointing to the negative graident of $f(x)$, is normally distributed and $\overset{N\to\infty}{=}$ denotes the convergence of the distribution $\|z_{\text{step}}\|^N/N = 1$. We further indtroudce $\sigma_\epsilon^* = N\sigma_\epsilon/(2R^2)$, the normalized surrogate model error (also referred to as the normalized fitness noise in Noise Sphere from Arnold and Beyer [2]). The estimate of true objective function value of y_i is $f_\epsilon(y_i) = f(y_i) + \sigma_\epsilon z_\epsilon, z_\epsilon \in \mathbb{R}$ is standard normally distributed.

The actual normalized fitness advantage of y using the surrogate model is

$$\delta_\epsilon = \delta + \sigma_\epsilon^* z_\epsilon \tag{4}$$

The expected value of the normalized change in objective function value

$$\begin{aligned}\Delta &= -\frac{N}{2}E\left[\log f(y) - \log f(x)\right] \\ &= -\frac{N}{2}E\left[\log\frac{f(x^{t+1})}{f(x^t)}\right],\end{aligned} \tag{5}$$

where y^t is the centroid of parental population in timestamp t, the equation is normalized in terms of dimensionality.

Since the fitness of λ offspring generated are evaluated by the surrogate model with vanishing cost. The objective function evaluation per iteration is 1 instead of λ (for $(\mu/\mu,\lambda)$-ES), therefore the normalized progress rate when dimensionality $N \longrightarrow \infty$, by substituting λ with 1 in equation (7) from [4] is

$$\eta = \frac{1}{1}E[\Delta] = \frac{\sigma^* c_{\mu/\mu,\lambda}}{\sqrt{1+\vartheta^2}} - \frac{(\sigma^*)^2}{2\mu}, \tag{6}$$

where $\vartheta = \sigma_\epsilon^*/\sigma^*$ is the noise-to-signal ratio, defined to measure the quality of surrogate model relative to the algorithm's step size, $c_{\mu/\mu,\lambda}$ is the $(\mu/\mu,\lambda)$-progess coefficient derived by Arnold and Beyer [1] that follows

$$c_{\mu/\mu,\lambda} = \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} e^{-x^2}\left[\Phi(x)\right]^{\lambda-\mu-1}\left[1-\Phi(x)\right]^{\mu-1}\mathrm{d}x, \tag{7}$$

where Φ^{-1} is the inverse function of Φ, the normal cumulative distribution function. The integral can be solve numerically.

To obtain the opt. expected fitness gain η_{opt} and its coresponding opt. normalized step size σ_{opt}^*, we take derivative of equation (6) over σ^* and get the following

$$\sigma_{opt}^* = \frac{\mu c_{\mu/\mu,\lambda}}{\sqrt{1+\vartheta^2}} \tag{8}$$

$$\eta_{opt} = \frac{\sigma_{opt}^* c_{\mu/\mu,\lambda}}{\sqrt{1+\vartheta^2}} - \frac{(\sigma_{opt}^*)^2}{2\mu} \tag{9}$$

The expected fitness gian is normalized in terms of the population size λ for easy comparision. The normalzied fitness gain against the normalized step size for $(\mu/\mu,\lambda)$-ES with population size $\lambda = 10, 20, 40$ corresponding $\mu = 3, 5, 10$ are ploted in 1 from left to right respectively. The line shows the result obtained from Eqs. (6) (7). The dots represent the

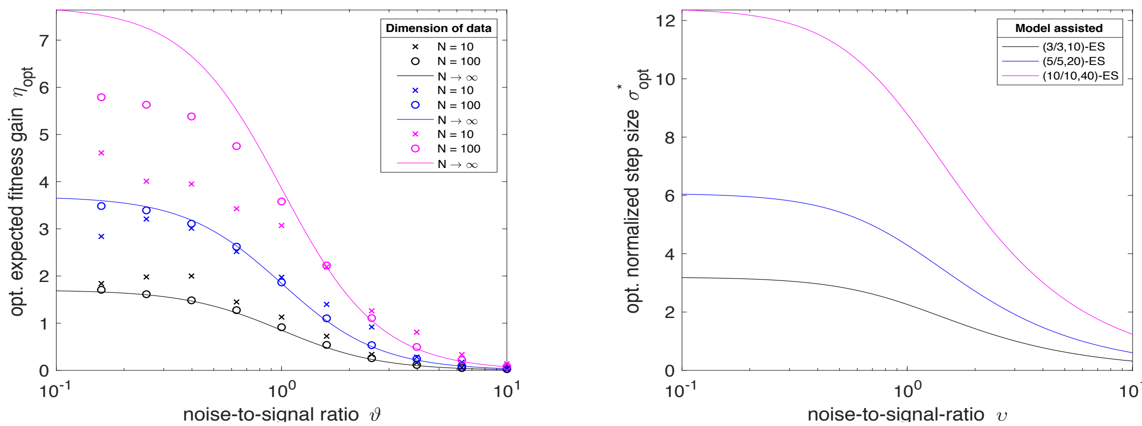

Fig. 2. Opt. expected fitness gain and corsponding opt. normalized step size of the surrogate model assisted $(\mu/\mu, \lambda)$-ES plotted against the noise-to-signal ratio. The line and dots with colour black, blue green represent (3/3, 10)-ES, (3/3, 10)-ES, (3/3, 10)-ES The solid line representa the results obtained analytically when $n \to \infty$.

experimental result for unbiased Gaussin surroagte error for $n \in \{10, 100\}$ obtained by averaging 100 runs. The result obtained for $n \longrightarrow \infty$ are considered to be cases with a large normalized step size with very small noise to signal ratio.

It can be inferred from 1, for a fixed population size, the expected fitness gain decreases with an increasing noise-to-signal-ratio. When $\vartheta \longrightarrow \infty$, the surrogate model becomes useless and the strategy becomes a random search. For moderate noise-to-signal ratio ϑ, the surrogate model assisted algorithm can acieve much larger value for expected fitness gain at a larger normalized step size. When $\vartheta = 1$, the maximal expected fitness gain achievable for (3/3, 10)-ES,(5/5, 20)-ES and (10/10, 40)-ES are 0.8507, 1.841, 3.808 with $\sigma^* = 2.254, 4.251, 8.738$ respectively. Compared with the result of the surrogate assisted (1+1)-ES [17] where maximal fitness gain is 0.548 achieved at $\sigma^* = 1.905$, $(\mu/\mu, \lambda)$-ES does benefit from using a larger population from the analysis. For $\vartheta = 0$ (the surrogate models the objective function exactly), from equation 8 we can obatin the maximal expected fitness gain is achieved at $\sigma^*_{opt} = \mu c_{\mu/\mu,\lambda}$ with value $\eta_{opt} = \mu(c_{\mu/\mu,\lambda})^2/2$. Even if this indicates the potential benefit the strategy may gain with a growing population, it is important to note the analytical results derived when $n \longrightarrow \infty$ is an approximation for the finite-dimensional case. 2 shows the relation of optimal expected fitness gain and the coresponding optimial normalized step size over noise-to-signal ratio derived analytically in the limit of $n \longrightarrow \infty$ for three different population sizes. The optimal expected fitness gain is also measured experimentally for $n \in \{10, 100\}$. how to calculate the speed up? For a finite-dimension, the speed-up achieved with surrogate model assistance for small noise-to-signal ratio appears to be between ...

Compared the result obtained by the surrogate assisted (1+1)-ES (Fig 1. [17]), there is indeed a

there is a significant increase in expected fitness gain as the population size λ increases, the

4 STEP SIZE ADAPTATION

4.1 Cummulative step size adaptation

Even though the analysis in Section 3 suggests a potential better performance for the surrogate-assisted $(\mu/\mu, \lambda)$-ES. There is no gaurantee the step size of the strategy can be properly adapted and further the analysis is very inaccurate in terms of finite dimension. In this section we experiemnt the surroagted model assisted $(\mu/\mu, \lambda)$-ES using the cumulative step size adaptation described in Section 2.2 and exploit the potential insight that it may offer. The strategy is evaluated by using a Gaussian Process based surrogate model replacing the simple model that simulates the surrogate behaviour in Section 3. Several test functions are used for testing the strategy.

Here, we use the well established parameters from Hansen's CMA tutorial [10] that follows

$$\begin{cases} c = (\mu + 2)/(N + \mu + 5) \\ d = 1 + 2\max\left(0, \sqrt{(\mu - 1)/(N + 1) - 1}\right) + c. \end{cases} \tag{10}$$

Then one signle iteration fo the surrogate model assisted $(\mu/\mu, \lambda)$-ES is shown in 1.

Algorithm 1 A Surrogate Assisted $(\mu/\mu, \lambda)$-ES

$c \leftarrow \frac{\mu+2}{n+\mu+5}$
$d \leftarrow 1 + 2\max(0, \sqrt{\frac{\mu-1}{n+1}} - 1)$
$p \leftarrow 0$
while not terminate() **do**
  for $i = 1, 2, \ldots, \lambda$ **do**
    Generate standard normally distributed $z_i \in \mathbb{R}^N$
    $y_i \leftarrow x + \sigma z_i$
    Evaluate y_i using the surrogate model, yieding $f_\epsilon(y_i)$
  end for
  $z = \frac{1}{\mu}\sum_{i=1}^{\mu} z_{i;\lambda}$
  $y = x + \sigma x$
  Evaluate y using true objective function, yieding $f(y)$
  Update surrogate modle
  $s \leftarrow (1 - c)s + \sqrt{c(2 - c)\mu} z$
  $\sigma \leftarrow \sigma \times \exp\left(\frac{c}{d}\frac{\|X\|}{E\|N(0,I)\|} - 1\right)$
end while

Five ten-dimensional test problems are used to test if the step size of the strategy has been appropriately adapted, namely sphere functions $f(x) = (x^T x)^{\alpha/2}$ for $\alpha = \{1, 2, 3\}$ referred to as linear, quadratic and cubic spheres, $f(x) = \sum_{i=1}^{n}(\sum_{j=1}^{i} x_j)^2$ (i.e. a convex quadratic function with condition number of the Hessian approximately equal to 175.1) referred to as SchwefelŚ Problem 1.2 [23]) and quartic function [17] defined as $f(x) = \sum_{i=1}^{n-1}\left[\beta(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right]$ where $\beta = 1$. The quartic function becomes the Rosenbrock function when the condition number of the Hessian at the optimier exceeds 3,500, making it very hard to find the global optima without adapting the shape of mutation distribution. We use the quartic function in the context with $\beta = 1$ and condition number of the Hessian at the optimizer equals to 49.0. The value of global optima for all test functiosn is zero. For each test problem, 1000 runs are conducted both for surrogate assisted $(1 + 1)$-ES and surrogate assisted $(\mu/\mu, \lambda)$-ES where a parental population size $\lambda = 10, 20, 40$ with $\mu = \lceil\lambda/4\rceil$. For surrogate model, we use Gaussin process. We use squard exponential kernel and the length scale paramter in the kernel is set proportional to the square of the dimension and the step size of the Evolution strategy. For simplicity, the length scale is $8\sigma\sqrt{n}$. The Gaussian process kernel is constructed using a trainijng size of 40. The training set consists of the most recent 40 candidate so that the surrogate model approximates the local landscape of the objective function. All runs are initialized with starting point sampled from a Gaussin distribution with zero mean

Table 1. Median test results.

Test functions	Median number of objective function calls (with model assistance)			
	$(1+1)$-ES	$(3/3, 10)$-ES	$(5/5, 20)$-ES	$(10/10, 40)$-ES
`linear sphere`	505	754	689	755
`quadratic sphere`	214	310	245	228
`cubic sphere`	202	274	250	254
`Schwefelś function`	1496	$+\infty$	$+\infty$	$+\infty$
`quartic function`	1244	1006	750	662

and unit covariance matrix and initial step size $\sigma_0 = 1$. The termination criteria is denfined as one solution achieves objective function value below 10^{-8}.

The number of objective function calls needed to solve the test problems within the required accurancy are represented in the histogram first row in Fig. 3, the median objective funcation calls for each test problem is shown in Table 1. The result of surrogate assisted (1+1)-ES is also included for comparision.

The step size of $(\mu/\mu, \lambda)$-ES is commonly adapted using cumulative step size adaptation (CSA)

4.2 Alternative

From Figure x, the success rate for all test functions are approxiamtely 0.48, the strategy makes a bad step every other step. It comes natural to ask, how much we are to benefit if we can avoid or simplily not take those bad steps.

Recent papers in surroagte model assisted ES consider (1+1)-ES, the step size of the stratgey is successfully adapted using the 1/5-rule []. Then we probbaly can apply a similar criteria when encounter a bad step size.

Propose step size adaptation in terms of emergency.

$table(test\ functions)$ Table for median of test results for surrogate model assisted $(\mu/\mu, \lambda)$-ES using CSA with emergency

Figure for success rate for surroagte assisted $(\mu/\mu, \lambda)$-ES with $\lambda = 10, 20, 40$

5 CONCLUSIONS

In this paper, We proposed a local surrogate-assisted $(\mu/\mu, \lambda) + (\mu/\mu, \lambda)$-ES. The strategy uses a lcoal surrogate model to optimize the candidate solution obtained in each iteration. The performance is analyzed by adding different levels of Gaussian distributed noise and applying the strategy to sphere functions.

REFERENCES

[1] D. V. Arnold and H. -G. Beyer. 2000. Efficiency and mutation strength adaptation of the (mu, mui, lambda)-es in a noisy environment. In *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature* (PPSN VI). Springer-Verlag, London, UK, UK, 39–48. http://dl.acm.org/citation.cfm?id=645825.669117.

[2] D. V. Arnold and H. -G. Beyer. 2004. Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49, 4, (Apr. 2004), 617–622.

[3] D. V. Arnold and H.-G. Beyer. 2002. *Noisy optimization with evolution strategies*. Vol. 8. Springer Science & Business Media.

[4] Dirk V. Arnold and Hans-Georg Beyer. 2001. Local performance of the (mu/mu, lambda)-es in a noisy environment. In *Foundations of Genetic Algorithms 6*. W. N. Martin and W. M. Spears, (Eds.) Morgan Kaufmann, San Francisco, 127–141. http://www.sciencedirect.com/science/article/pii/B9781558607347500901.

[5] D. Buche, N. N. Schraudolph, and P. Koumoutsakos. 2005. Accelerating evolutionary algorithms with gaussian process fitness function models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35, 2, (May 2005), 183–194.

Algorithm 2 Cummulative Step Size Adaptation with Emergency

$c \leftarrow \frac{\mu+2}{n+\mu+5}$
$d \leftarrow 1 + 2\max(0, \sqrt{\frac{\mu-1}{n+1}} - 1)$
$p \leftarrow 0$
$D \leftarrow 0.68$
while not terminate() **do**
  for $i = 1, 2, \ldots, \lambda$ **do**
    Generate standard normally distributed $z_i \in \mathbb{R}^N$
    $y_i \leftarrow x + \sigma z_i$
    Evaluate y_i using the surrogate model, yieding $\hat{f}(y_i)$
  end for
  $z = \frac{1}{\mu} \sum_{i=1}^{\mu} z_{i;\lambda}$
  $y = x + \sigma x$
  Evaluate y using true objective function, yieding $f(y)$
  Update surrogate modle
  if $f(x) < f(y)$ (Emergency) **then**
    $\sigma \leftarrow \sigma D$
  else
    $s \leftarrow (1-c)s + \sqrt{c(2-c)\mu z}$
    $\sigma \leftarrow \sigma \times \exp\left(\frac{c}{d} \frac{\|X\|}{E\|N(0,I)\|} - 1\right)$
  end if
end while

[6] Y. Chen and X. Zou. 2014. Performance analysis of a (1+1) surrogate-assisted evolutionary algorithm. In *Intelligent Computing Theory*. V. Bevilacqua D. Huang and P. Premaratne, (Eds.) Springer International Publishing, Cham, 32–40.

[7] B Dunham, D Fridshal, R Fridshal, and JH North. 1963. Design by natural selection. *Synthese*, 15, 1, 254–259.

[8] K.C. Giannakoglou. 2002. Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence. *Progress in Aerospace Sciences*, 38, 1, 43–76. http://www.sciencedirect.com/science/article/pii/S0376042101000197.

[9] Y. Jin, (Ed.) 2005. *Model assisted evolution strategies. Knowledge Incorporation in Evolutionary Computation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 333–355. https://doi.org/10.1007/978-3-540-44511-1_16.

[10] N. Hansen. 2016. The cma evolution strategy: a tutorial. *arXiv preprint arXiv:1604.00772*.

[11] W. J. Hill and W. G. Hunter. 1966. A review of response surface methodology: a literature survey. *Technometrics*, 8, 4, 571–590. eprint: https://amstat.tandfonline.com/doi/pdf/10.1080/00401706.1966.10490404. https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1966.10490404.

[12] M. Schoenauer I. Loshchilov and M. Sebag. 2012. Self-adaptive surrogate-assisted covariance matrix adaptation evolution strategy. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation*. ACM, 321–328.

[13] W. J. Welch J. Sacks, T. J. Mitchell, and H. P. Wynn. 1989. Design and analysis of computer experiments. *Statist. Sci.*, 4, 4, (Nov. 1989), 409–423. https://doi.org/10.1214/ss/1177012413.

[14] Y. Jin. 2011. Surrogate-assisted evolutionary computation: recent advances and future challenges. *Swarm and Evolutionary Computation*, 1, 2, 61–70. http://www.sciencedirect.com/science/article/pii/S2210650211000198.

[15] Y. Jin, M. Olhofer, and B. Sendhoff. 2002. A framework for evolutionary optimization with approximate fitness functions. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 6, 481–494.

[16] Y. Jin and B. Sendhoff. 2002. Fitness approximation in evolutionary computation - a survey. In *Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation* (GECCO'02). Morgan Kaufmann Publishers Inc., New York City, New York, 1105–1112. http://dl.acm.org/citation.cfm?id=2955491.2955686.

[17] A. Kayhani and D. V. Arnold. 2018. Design of a surrogate model assisted (1 + 1)-es. In *Parallel Problem Solving from Nature - PPSN XV - 15th International Conference, Coimbra, Portugal, September 8-12, 2018, Proceedings, Part I*, 16–28. https://doi.org/10.1007/978-3-319-99253-2%5C_2.

[18] I. Loshchilov. 2016. LM-CMA: an Alternative to L-BFGS for Large Scale Black-box Optimization. *Evolutionary Computation*, to appear.

[19] S. D. Müller N. Hansen and P. Koumoutsakos. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evol. Comput.*, 11, 1, (Mar. 2003), 1–18. http://dx.doi.org/10.1162/106365603321828970.

[20] G. Andreas O. Andreas and H. Nikolaus. 1994. A derandomized approach to self-adaptation of evolution strategies. *Evol. Comput.*, 2, 4, (Dec. 1994), 369–380. http://dx.doi.org/10.1162/evco.1994.2.4.369.

[21] A. Ratle. 2001. Kriging as a surrogate fitness landscape in evolutionary optimization. *Artif. Intell. Eng. Des. Anal. Manuf.*, 15, 1, (Jan. 2001), 37–49. http://dx.doi.org/10.1017/S0890060401151024.

[22] I. Rechenberg. 1973. Evolutionsstrategie–optimierung technisher systeme nach prinzipien der biologischen evolution.

[23] H. -P. Schwefel. 1981. *Numerical Optimization of Computer Models*. John Wiley & Sons, Inc., New York, NY, USA.

[24] M. Smith. 1993. *Neural Networks for Statistical Modeling*. (1st ed.). Thomson Learning.

[25] H. Ulmer, F. Streichert, and A. Zell. 2003. Evolution strategies assisted by gaussian processes with improved pre-selection criterion. In *in IEEE Congress on Evolutionary Computation,CEC 2003*, 692–699.

[26] Z. Zhou, Y. S. Ong, P. B. Nair, A. J. Keane, and K. Y. Lum. 2007. Combining global and local surrogate models to accelerate evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37, 1, (Jan. 2007), 66–76.