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

Jingyun Yang and Dirk V. Arnold
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia B3H 4R2
jingyun.yang@dal.ca,dirk@cs.dal.ca

ABSTRACT

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.

KEYWORDS

$(\mu/\mu, \lambda)$-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. In *Proceedings of , , ,* 4 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) [16] 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 [13]. There are a range of surrogate models and a survey of the development can be found by Jin [11] and Loshchilov [15]. 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 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 [14]. The experimental result is followed by a discuession and future work in Section 5.

ACM ISBN .
https://doi.org/

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) [10], Gaussian Process (GP, Kriging models) [19], Artificial neural networks [20]. 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 [17]. Another ES using global surrogate model based on Artificial neural networks was constructed by Jin [12] which gives an imperial criterion on using the true objective function or the surrogate model to evaluate the offspring. Ulmer et al [21] 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.* Online local surrogate models [23] 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.

Recent works in surrogated assisted EAs uses a combination of different surrogate models to estimate the fitness strength of the candidate solutions. Zhou et al [23] proposed a hierarchical surrogate-assisted ES where a global surrogate model and a local surrogate model are integrated. The Global surrogate model uses GP and PR to estimate the global fitness of ES's search space, filtering the unpromising candidate solutions. Then, a local surrogate-assisted Lamarckian learning based on RBF is performed to search the promising candidate solutions.

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 investigated the surrogated-assisted 1+1-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 [14]. Secondly, it does not include a step size adaptation mechanism. Besides that, Ulmer et al [22] 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 [14] 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 degress the choice of population can benefit the ES in terms of rubustness and acceleration.

2.2 Step size adaptation

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[14].....

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 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* [1]). 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 [18] 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 [$dirk'sslides$]). 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

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

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 $\|z_{\text{step}}\|^n/n = \mu$?????μor1??? assuming $\mu \ll n$. 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 [1]). 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

$$\delta_\epsilon = \delta + \sigma_\epsilon^* z_\epsilon$$

The expected value of the normalized change in objective function value

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

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 $N \to \infty$, by substituting λ with 1 in equation (7) from [1] is

$$\eta = \frac{1}{1} E[\Delta] = \frac{\sigma^* c_{\mu/\mu,\lambda}}{\sqrt{1+\vartheta^2}} - \frac{(\sigma^*)^2}{2\mu}, \tag{3}$$

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 [2] that follows

$$c_{\mu/\mu,\lambda} = \frac{\lambda-\mu}{2\pi} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} e^{-x^2} \left[\Phi(x)\right]^{\lambda-\mu-1} \left[1-\Phi(x)\right]^{\mu-1} \mathrm{d}x, \tag{4}$$

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 (3) over σ^* and get the following

$$\eta_{opt} = \frac{\sigma_{opt}^* c_{\mu\ \mu,\lambda}}{\sqrt{1+\vartheta^2}} - \frac{(\sigma_{opt}^*)^2}{2\mu} \tag{5}$$

$$\sigma_{opt}^* = \frac{\mu c_{\mu\ \mu,\lambda}}{\sqrt{1+\vartheta^2}} \tag{6}$$

The expected fitness gian of against the normalized step size is plotted in Fig. 1. The line shows the result obtained from equation (xx) (xx) (xx). The dots represent the experimental result for unbiased Gaussin surroagte error for $n \in \{10, 100\}$ obtained by averaging 100 runs.

For a fixd dimensionality N the normalized progress rate [4] is

$$\eta \approx \frac{c_{\mu/\mu,\lambda}\sigma^*(1+\sigma^{*2}/2\mu N)}{\sqrt{1+\sigma^{*2}/2\mu N)}\sqrt{1+\vartheta^2+\sigma^{*2}/2N}} - N\left[\sqrt{1+\frac{\sigma^{*2}}{\mu N}} - 1\right]$$

4 STEP SIZE ADAPTATION

Use CSA tune some parameters -> show that the step size is reasonable

In this section we propose a step size mechanism for the suroagted model assisted $((\mu/\mu,\lambda)+(\mu/\mu,\lambda))$-ES. 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.

The step size of $(\mu/\mu, \lambda)$-ES is commonly adapated using Cumulative step size adaptation (CSA) [16] and Covariance Matrix Adaptation (CMA) [9]. The former uses an evolution path and a trial path

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

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. -. Beyer. 2004. Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49, 4, (Apr. 2004), 617–622.

[2] Dirk V. Arnold and Hans-Georg 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.

[3] Dirk V Arnold and Hans-Georg Beyer. 2002. *Noisy optimization with evolution strategies*. Vol. 8. Springer Science & Business Media.

[4] Dirk V. Arnold and Hans-Georg Beyer. 2002. Performance analysis of evolution strategies with multi-recombination in high-dimensional rn-search spaces disturbed by noise. *Theoretical Computer Science*, 289, 1, 629–647. http://www.sciencedirect.com/science/article/pii/S030439750100384X.

[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.

[6] Yu Chen and Xiufen Zou. 2014. Performance analysis of a (1+1) surrogate-assisted evolutionary algorithm. In *Intelligent Computing Theory*. De-Shuang Huang, Vitoantonio Bevilacqua, and Prashan Premaratne, (Eds.) Springer International Publishing, Cham, 32–40.

[7] B. Dunham, D. Fridshal, R. Fridshal, and J. H. North. 1963. Design by natural selection. *Synthese*, 15, 1, (Jan. 1963), 254–259. https://doi.org/10.1007/BF00484855.

[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] Nikolaus Hansen, Sibylle D. Müller, and Petros 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.

[10] William J. Hill and William 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.

[11] Yaochu 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.

[12] Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. 2002. A framework for evolutionary optimization with approximate fitness functions. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 6, 481–494.

[13] Yaochu Jin and Bernhard 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.

[14] Arash Kayhani and Dirk 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.

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

[16] Andreas Ostermeier, Andreas Gawelczyk, and Nikolaus Hansen. 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.

[17] Alain 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.

[18] Ingo Rechenberg. 1973. Evolutionsstrategie–optimierung technisher systeme nach prinzipien der biologischen evolution.

[19] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. 1989. Design and analysis of computer experiments. *Statist. Sci.*, 4, 4, (Nov. 1989), 409–423. https://doi.org/10.1214/ss/1177012413.

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

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

[22] Yaochu 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.

[23] 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.