Design of a Surrogate Model Assisted $((\mu/\mu, \lambda) + (\mu/\mu, \lambda))$-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 $((\mu/\mu, \lambda) + (\mu/\mu, \lambda))$-ES using a population instead of a single offspring with the hope to possible address the issue. By comparing the bahaviour of the two, we propose a step size adaptation mechinisum with an emergency tigger and systematically evaluate the strategy for several test functiions.

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 $((\mu/\mu, \lambda) + (\mu/\mu, \lambda))$-ES. In *Proceedings of , , ,* 3 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) [15] 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 [12]. There are a range of surrogate models and a survey of the development can be found by Jin [10] and Loshchilov [14]. 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 benefitted from large population size.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 [13]. 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 [**Dunham1963**]. Some successful surrogated models include but are not limitted to Polynomial Regression (PR, response surface methodology) [8], Gaussian Process (GP, Kriging models) [9], Artificial neural networks [17]. 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 [16]. Another ES using global surrogate model based on Artificial neural networks was constructed by Jin [11] which gives an imperial criterion on using the true objective function or the surrogate model to evaluate the offspring. Ulmer et al [18] and Buche et al [3] 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

ACM ISBN .
https://doi.org/

local surrogate models [19] can be constructed using methods like radial basis function (RBF) [5] 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 [19] 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 [4] 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 [13]. Secondly, it does not include a step size adaptation mechanism. Besides that, Ulmer et al [6] 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 [13] 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 a population. Specifically, we propose an EA that, in each iteration, a population size of λ new candidate solutions $y_i \in \mathbb{R}^N, i = 1, \ldots, \lambda$ are generated from μ parents $x_i \in \mathbb{R}^N, i = 1, \ldots, \mu$, where $\lambda > \mu$. The parental population are replaced by the best μ candidate solutions $y_{i;\lambda}, i = 1, 2, \ldots, \mu$ evaluated by the surrogate model with fitness estimate $\hat{f}(y_{i;\lambda}) \leq \hat{f}(y_{j,\lambda}), 1 \leq i < j \leq \lambda$. For each of the λ candidate solution $y_i = x_{\text{opt}} + \sigma z$, x_{opt} is the best centroid of the parental population so far referred to as the parent, . $x = \sum_{i=1}^{n} x_i/\mu$ are the. This step is referred to as intermediate recombination. candidate solutions are ranked according to the fitness estimated by the surrogate with vanishing cost, follow.

the each iteration either in emergency, discard the offspring generated iteration in each iteration, a complete population of size λ is evaluated by the surrogate. Then a size of μ candidate solutions ranking

For the dimensionality $N \to \infty$ the normalized progress rate [1] is

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

By taking derivative we can find the opt.expected progress rate is $\eta_{opt} = \frac{\sigma^*_{opt} c_{\mu\ \mu,\lambda}}{\sqrt{1+\upsilon^2}} - \frac{(\sigma^*_{opt})^2}{2\mu}$ and the coresponding opt. normalized step size is $\sigma^*_{opt} = \frac{\mu c_{\mu\ \mu,\lambda}}{\sqrt{1+\upsilon^2}}$.

For a fixd dimensionality N the normalized progress rate [2] 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

In this section we propose a step size mechanism for the surroagted 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) [15] and Covariance Matrix Adaptation (CMA) [7]. The former uses an evolution path and a trial path

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

$c \leftarrow \frac{\mu+2}{n+\mu+5}$
$d \leftarrow 1 + 2\max(0, \sqrt{\frac{\mu-1}{n+1}} - 1)$
$p \leftarrow 0$
$D \leftarrow 0.68$
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

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

[2] D. V. Arnold and H.-G. 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] N. Hansen, S. D. Müller, 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.