

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 the knowledge gained in previous iterations. Using surrogate assisted (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 possibly address the issue. By comparing the behaviour of the two, we propose a step size adaptation mechanism with an emergency trigger and systematically evaluate the strategy for 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 $((\mu/\mu, \lambda) - (\mu/\mu, \lambda))$ -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 reduce 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. Cumulative step size adaptation (CSA) [13] 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 adapted 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 [10]. There are a range of surrogate models and a survey of the development can be found by Jin [8] and Loshchilov [12]. 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 sophisticated algorithm where surrogates are combined or the model is updated online according to some heuristic. Comparison 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 surrogate 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 [11]. The experimental result is followed by a discussion 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 [5]. Some successful surrogated models include but are not limited to Polynomial Regression (PR, response surface methodology) [7], Gaussian Process (GP, Kriging models) [16], 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 Kriging was examined by Ratle [14]. Another ES using global surrogate model based on Artificial neural networks was constructed by Jin [9] 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 local surrogate models

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

[20] can be constructed using methods like radial basis function (RBF) [6] 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 [20] 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 [ARASH]. Secondly, it does not include a step size adaptation mechanism. Besides that, Ulmer et al [19] 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 [11] proposed a surrogated-assisted (1+1)-ES that investigates the acceleration and single 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 robust 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 robustness 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 $y_i \in \mathbb{R}^N, i = 1, \dots, \lambda$ are generated from μ parents $x_i \in \mathbb{R}^N, i = 1, \dots, \mu$, where $\lambda > \mu$. The parental population are replaced by the best μ candidate solutions $y_{i:\lambda}, i = 1, 2, \dots, \mu$ 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

offsprings generate λ candidate solutions that are evaluated by the surrogate model. the candidate solution generated using the surrogate models. We first sample a random direction and generate two points y_{\pm} by adding the positive and negative directions to the parent. The optimization is done by first fitting a quadratic model using the parent (evaluated by the true objective function) and the other two points (evaluated by the surrogate models where the inaccurate estimate can be achieved at vanishing cost), then choosing the lowest point of the quadratic model as the offspring of this iteration.

For simplicity, we assume the error of using the GP estimate is a Gaussian random variable with mean coincides with the true objective function value of the candidate solution and some variance introduced as the noise for the estimate. In this case, we can apply the analysis of evolution strategies with the presence of Gaussian noise (see [2] and references theorem). The analysis could be extended to biased surrogate models where the mean of the estimates' distribution is different from the exact objective function value. It is likely that analyzing the effect of non-Gaussian noise for the performance of ES could be achieved by using models with error distribution that is skew [1].

Considering the minimization of quadratic sphere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with $f(x) = x^T x$ where the local surrogate model assisted (1+1)-ES is applied, this section will use the surrogate model described above to replace the candidate solution generated in each iteration. We first consider a simple iteration of the strategy. In each iteration the strategy generates a single candidate solution $y = x + \sigma z$ as is described in Section 3 where σ is the step size parameter and x the parent (best candidate solutions obtained so far). Then we randomly sample standard normally distributed random vector $z \in \mathbb{R}^n$ referred to as the direction vector and generate two points based on the direction vector $y_{\pm} = x \pm z$. We want to fit a quadratic model that optimizes y in direction z where the minimization of the quadratic model can be written as:

$$\text{Given } x \in \mathbb{R}^n, z \in \mathbb{R}^n \text{ find } \alpha_{\text{opt}} \in \mathbb{R} \text{ s.t. } \min\{f(x + \alpha z)\}$$

We denote the quadratic model $g(\alpha) = a\alpha^2 + b\alpha + c = f(x + \alpha z)$ where the surrogate step size α essentially illustrates the signed distance we added to the given candidate solution y in direction z . For the three points described above, the strategy uses the surrogate model estimate $f_{\epsilon}(y_{\pm})$ to replace the true objective function value of y_{\pm} that goes like the following:

$$\begin{cases} g(1) = a + b + c = f_{\epsilon}(y_{+}) \\ g(-1) = a - b + c = f_{\epsilon}(y_{-}) \\ g(0) = c = f(x) \end{cases}$$

Solving the above system gives $\alpha_{\text{opt}} = -\frac{b}{2a} = -\frac{f_{\epsilon}(y_{+}) - f_{\epsilon}(y_{-})}{f_{\epsilon}(y_{+}) - f_{\epsilon}(y_{-}) + 2f(x)}$ where $y = x + \alpha_{\text{opt}} z$ is the offspring obtained in this iteration.

By assumption, $f_{\epsilon}(y_{\pm})$ is a random variable with mean $f(y_{\pm})$ and some standard deviation $\sigma_{\epsilon} > 0$ that can be rewritten as $y_{\epsilon}(y_{\pm}) = f(y_{\pm}) + \sigma_{\epsilon}$. Better surrogate models with small value of σ_{ϵ} can

result more precise estimation of the true objective function. In the case where $\sigma_\epsilon = 0$, we obtain the estimation without Gaussian noise, which is essentially the same as we evaluate using the true objective function.

We use the decomposing of z proposed by Rechenberg [15] to analyze the expected step size of the strategy. Vector z could be decomposed as a vector sum $z = z_1 + z_2$, where z_1 is in the direction of the negative gradient of z , while z_2 orthogonal to z_A . We have z_1 standard normally distributed while $\|z_2\|^2 \chi$ -distributed with $n - 1$ degree of freedom and $\frac{\|z_2\|^2}{n} \xrightarrow{n \rightarrow \infty} 0$. The estimate and optimal α_{opt} of this iteration using surrogate model results the following where z_ϵ^\pm is standard normally distributed.

$$f_\epsilon(y_\pm) = (R \mp \sigma z_1) + \sigma \|z_2\|^2 + \sigma_\epsilon^\pm$$

$$\alpha_{opt} = \frac{4\sigma z_1 + \sigma_\epsilon(z_\epsilon^+ - z_\epsilon^-)}{2(2\sigma^2\|z\|^2 + \sigma_\epsilon(z_\epsilon^+ + z_\epsilon^-))}$$

The normalized fitness advantage of the offspring over its parent is denoted as $\delta = n(f(x) - f(y))/(2R^2)$ where $R = \|x\|$ is referred to as the distance from the parent to the global minima. By introducing the normalized step size $\sigma^* = n\sigma/n$ and estimated normalized fitness advantage $\sigma_\epsilon^* = n\sigma_\epsilon/(2R^2)$ (the normalized fitness advantage using the surrogate model) we can simplify α_{opt}

$$\alpha_{opt} = -\frac{-2\sigma^* z_1 + \sigma_\epsilon^*(z_\epsilon^+ - z_\epsilon^-)}{2(\sigma^*)^2\|z\|^2/n + 2\sigma_\epsilon^*(z_\epsilon^+ - z_\epsilon^-)}$$

$$\xrightarrow{n \rightarrow \infty} -\frac{\sigma^* z_1 - \sigma_\epsilon^*/2(z_\epsilon^+ - z_\epsilon^-)}{(\sigma^*)^2 + \sigma_\epsilon^*(z_\epsilon^+ + z_\epsilon^-)}$$

Simplify the expected fitness advantage of the offspring over its parent

$$\delta = \frac{n}{2R^2} (x^T x - (x + \alpha_{opt} \sigma z)^T (x - (x + \alpha_{opt} \sigma z)))$$

$$\xrightarrow{n \rightarrow \infty} \alpha_{opt} \sigma^* z_1 - \frac{(\alpha_{opt} \sigma^*)^2}{2}$$

where $z_1 = -x^T z$ is standard normally distributed random variable that coincides with z in the negative gradient direction and $\xrightarrow{n \rightarrow \infty}$ denotes the convergence of the distribution. The equation could be further written as

$$\delta(x, y) \xrightarrow{n \rightarrow \infty} x \sigma^* y - \frac{(x \sigma^*)^2}{2}$$

The probability density of α_{opt} conditioned on z_1 obtained by Dirk is

$$p_{\alpha|z_1}(x|y)$$

$$= 2De^{\frac{-E^2}{2(1+D^2)}} \left[\frac{(\sigma_\epsilon^* + 2y + 2\theta \frac{E}{1+D^2})(1 - 2\text{normcdf}(\beta)) + 4\theta \frac{De^{-\beta^2/2}}{\sqrt{2\pi}\sqrt{1+D^2}}}{(1 - 2x)^2 \sqrt{2\pi}\sqrt{1+D^2}} \right]$$

where we have the following

$$\begin{cases} \theta = \frac{\sigma_\epsilon^*}{\sigma^*} \\ E = \frac{-2(y + x \sigma_\epsilon^*)}{1+2x} / \theta \\ D = \frac{1-2x}{1+2x} \\ \beta = \frac{(\sigma_\epsilon^*/2 + y) * \sqrt{1+D^2} + \theta E / \sqrt{1+D^2}}{D} \\ \text{normcdf}(\beta) : \text{cdf of } N(0,1) \text{ evaluated at } \beta \end{cases}$$

In each iteration, one objective function call is made after the candidate solution is optimized using the quadratic model. The offspring y replacing its parent x if and only if $\delta > 0$. We write $p_{step} = \text{Prob}[\delta > 0]$ for the probability of the offspring replacing its parent

In each iteration, it could be computed as

$$E[\Delta] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-y^2/2} p_{\alpha|z_1}(x|y) \delta(x, y) dx dy$$

We define $\theta = \sigma_\epsilon^*/\sigma^*$ the noise-to-signal ratio to measure the quality of surrogate model relative to the algorithm's step size and the function to be integrated is denoted as

$$g(x, y, \sigma_\epsilon^*, \sigma^*) = 1/\sqrt{2\pi} e^{-y^2/2} p_{\alpha|z_1}(x|y) \delta(x, y)$$

The expected fitness gain relative to normalized step size is plotted in figure 2. The dots show the corresponding values observed in experiments with unbiased Gaussian surrogate error for $n \in \{10, 100\}$ obtained by averaging 40 runs. We tried to compute the expected fitness gain ($n \rightarrow \infty$) over the normalized step size analytically, but failed because the function to integrate does not seem to converge over y . We experimented with different values of σ_ϵ^* and θ for $g(x, y, \sigma_\epsilon^*, \sigma^*)$ but the value obtained is a negative constant along the y -axes when y exceeds a certain value for a range of x . An example of the plot for $g(x, y, \sigma_\epsilon^*, \sigma^*)$ over x and y with $\sigma_\epsilon^* = 1$ and $\sigma^* = 1$ is shown in Fig.3 where two half-cylinder-shape along y -axes can be observed. $n \rightarrow \infty$ can be interpreted as a large step size with small noise-to-signal ratio.

It can be seen from Fig.2 that for a small noise-to-signal ratio, the evaluation rate increases with an increasing step size. While for a large noise-to-signal ratio, evaluation rate decreases with an increasing step size. The strategy makes progress one out of two steps for very small step size. The strategy becomes more "wise" in generating offsprings with a large step size and a relative small noise-to-signal ratio in a sense that the candidate solution optimized in this iteration is more close to the true lowest point of the quadratic model with a relative large space for candidate generation. In the case of zero noise-to-signal ratio, the candidate solution is deemed superior to its parent by choosing the best candidate solution in the direction sampled.

The expected fitness gain decreases as the step size increases (observed in Fig. 2 $\theta = 4$ green dots) which is contrary to the case where $\theta = 0.25$ and $\theta = 1$. One interpretation is that the strategy samples one direction in each iteration and the candidate solution is optimized in that direction using a quadratic model. The fitness gain of the candidate solution largely depends on the direction sampled and the accuracy of the estimation for the points fitting the quadratic model. The noise-to-signal ratio θ controls the precision of the optimal candidate solution estimated in the quadratic model relative to step size. A large θ means the estimation

of the optimal candidate solution in the direction sampled is less accurate and therefore the evaluation rate decreases as the step size increases (the increasing step size also adds inaccuracy to the estimation of the optimal candidate solution).

The direction sampled also affects the model performance. We would like to optimize the candidate solution in a direction with a potentially large fitness gain. But random sampling makes the process uncertain and according to experimental results, chances are that we do not sample a good direction in most cases i.e. the improvement for the parent is limited. So that a selection mechanism for the (direction of) candidate solutions in each iteration should be further considered. One possible approach is to consider the difference (distance) of the candidate solution before and after optimization using the quadratic model. Or we could possibly try unit vectors in each dimension, optimize using the quadratic model and choose the direction with the largest absolute $-2a/b$ (surrogate step size) in the optimization using the quadratic model.

4 STEP SIZE ADAPTATION

5 CONCLUSIONS

In this paper, We proposed a local surrogate-assisted (1+1)-ES on a quadratic sphere function. The strategy uses a local surrogate model to optimize the candidate solution obtained in each iteration. The performance is analyzed by adding different levels of Gaussian noise in the strategy.

For future work, we will work on a selection mechanism for candidate solutions deciding what candidate solutions to choose based on the fitness gain it may bring. Further, a step size adaptation mechanism for the surrogate model assisted (1 + 1)-ES should be considered.

REFERENCES

- [1] D. V. Arnold and H. - Beyer. 2006. A general noise model and its effects on evolution strategy performance. *IEEE Transactions on Evolutionary Computation*, 10, 4, (Aug. 2006), 380–391.
- [2] Dirk V. Arnold and Hans-georg Beyer. 2002. Noisy Optimization with Evolution Strategies. Tech. rep. SIAM Journal on Optimization.
- [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] 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.
- [5] 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>.
- [6] 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>.
- [7] 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>.
- [8] 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>.
- [9] 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.
- [10] 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>.
- [11] 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.
- [12] I. Loshchilov. 2016. LM-CMA: an Alternative to L-BFGS for Large Scale Black-box Optimization. *Evolutionary Computation*, to appear.
- [13] Andreas Ostermeier, Andreas Gawelczyk, and Nikolaus Hansen. 1994. A de-randomized 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>.
- [14] 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>.
- [15] 1973. *Evolutionsstrategie: optimierung technischer systeme nach prinzipien der biologischen evolution*. Problemata, 15. Frommann-Holzboog. <https://books.google.ca/books?id=-WAQAQAAMAAJ>.
- [16] 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>.
- [17] Murray Smith. 1993. *Neural Networks for Statistical Modeling*. (1st ed.). Thomson Learning.
- [18] Holger Ulmer, Felix Streichert, and Andreas Zell. 2003. Evolution strategies assisted by gaussian processes with improved pre-selection criterion. In *IEEE Congress on Evolutionary Computation, CEC 2003*, 692–699.
- [19] 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.
- [20] 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.