A Study on Polynomial Regression and Gaussian Process Global Surrogate Model in Hierarchical Surrogate-Assisted Evolutionary Algorithm

Zongzhao Zhou, Yew Soon Ong
School of Computer Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798
{zhou0018, asysong}@ntu.edu.sg

My Hanh Nguyen, Dudy Lim
School of Computer Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798
{myhanh, dudy}@pmail.ntu.edu.sg

Abstract- This paper presents a study on Hierarchical Surrogate-Assisted Evolutionary Algorithm (HSAEA) using different global surrogate models for solving computationally expensive optimization problems. In particular, we consider the use of Gaussian Process (GP) and Polynomial Regression (PR) methods for approximating the global fitness landscape in the surrogate-assisted evolutionary search. The global surrogate model serves to pre-screen the EA population for promising individuals. Subsequently, these potential individuals undergo a local search in the form of Lamarckian learning using online local surrogate models. Numerical results are presented on two multi-modal benchmark test functions. The results obtained show that both PR-HSAEA and GP-HSAEA converge to good designs on a limited computational budget. Further, our study also shows that the GP model is suitable for global surrogate modeling in HSAEA if the evaluation function is very expensive in computations. On moderately expensive problems, the PR method may serve to generate better efficiency than using GP.

1 Introduction

Evolutionary Algorithms (EAs) have emerged as a powerful paradigm for global optimization. Over the last decades, this approach has gained significant interest in complex engineering design problems. However, in many complex engineering design problems where high-fidelity analysis codes are used, each objective function evaluation requires the simulation of the high-fidelity analysis codes, such as Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD) or Computational Electro Magnetics (CEM) etc., may cost from minutes to days of supercomputer time. Since EAs typically require thousands of function evaluations to locate a near optimal solution, the use of EAs often becomes computationally prohibitive for this class of problems. To address this problem, several efforts have been made to reduce the computational cost and accelerate the convergence of EAs. One promising way to significantly reduce the computational cost of EAs is to employ computationally cheap surrogate models in place of computationally expensive exact function evaluations [1], [2], [3], [4], [5]. By leveraging surrogate models, the computational burden can be greatly reduced since the cost of building the surrogate model and optimization using it are much lower than the standard approach of directly coupling the simulation codes with the optimizer.

A variety of techniques for the constructions of surrogate model, often also referred to as meta-models or approximation models, have been used in engineering design optimization. Among these techniques, Polynomial Regression (PR, also known as response surface method), Artificial Neural Network (ANN), Radial Basis Function (RBF), and Gaussian Process (GP) (also referred to as Kriging or Design and Analysis of Computer Experiments (DACE) models) are the most prominent and commonly used techniques [6], [7], [8].

Over the recent years, there has been a growing body of research focusing on the development of new EA frameworks for solving computationally expensive problems on a limited computational budget. Hence, there are now various ways to integrate surrogate models into evolutionary optimization. In general, researchers either use global or local surrogate models in place of the exact simulation codes that are constructed to predict the fitness of promising individuals more efficiently. Ratle [3] and El-Beltagy et al. [9] examined strategies for integrating evolutionary search with global surrogate models based on Kriging. Jin et al. [10] also employs an artificial neural network to construct global surrogate models and an empirical criterion is proposed to switch between the expensive exact fitness function and surrogate model during the search. Various strategies using GP global surrogate models have also been considered in Ulmer et al. [11] and D. Büche [12]. However, since the idea of constructing accurate global surrogate models might be fundamentally flawed due to the curse of dimensionality, online local surrogate models using RBF were proposed in Ong et al. [1] and Giannakoglou [4] et al. in place of global surrogate models.

Hybrid evolutionary approaches, also sometimes known as memetic strategies in surrogate-assisted EA have also been considered recently to accelerate evolutionary search. Liang and Yao proposed a strategy for coupling EAs with local search based on a PR global models in [13]. Ong et al. used a trust-region approach in the hybrid evolutionary search to interleave use of the exact fitness functions with computationally cheap online RBF local surrogate models in the Lamarckian learning process [1]. The use of gradient information to improve the approximation accuracy of surrogate-assisted EAs was also considered in Ong et al. [14]. A recent survey paper that outlines some of the typical fitness approximation methods and data sampling techniques used in evolutionary computation can be found in Jin

[15].

Existing surrogate-assisted evolutionary algorithm generally uses the global or local surrogate models. Our recent studies on surrogate-assisted evolutionary algorithms have combined both global and local surrogate models for solving computationally expensive optimization problems. A novel Hierarchical Surrogate-Assisted Evolutionary Algorithm (HSAEA) that employs computationally cheap online global and local surrogate models to replace the exact computationally expensive exact fitness evaluations during evolutionary search was reported in [16]. At the first level, the algorithm employs a GP global surrogate model to filter the EA population of promising individuals. These potential individuals then undergo a memetic search in the form of Lamarckian learning at the next level. The Lamarckian evolution involves a trust-region enabled gradient-based search strategy that employs RBF local surrogate models to accelerate convergence. It was shown to accelerate the evolutionary search much more efficiently than using the global or local surrogate model alone [16].

The choice of global surrogate model in the HSAEA should be one that is capable of modeling any complex global trends of the exact fitness landscape accurately. Since the GP interpolation method is commonly regarded to provide accurate approximation of complex global fitness landscape, it was proposed in [16] for solving complex engineering design problem. Nevertheless, model construction based on GP can be very time-consuming when compared to other approximation methods which scales exponentially with the training data size and problem dimensionality.

Since the purpose of the GP global surrogate model is to pre-screen the EA population, the accuracy of the predicted fitness may not be very critical to the HSAEA algorithm. Rather, the correct selection of promising individuals would be of upmost importance from the evolutionary search point of view. Such a observation was also reported in Jin et al. [17] where it was shown that the qualitative fitness approximation of the surrogate model may be sufficient in surrogate-assisted evolutionary frameworks, even though the approximation error may be quite large. Taking these cues, in this paper we consider using the PR method in place of the GP model for building global surrogate model in the HSAEA algorithm since the PR method is computationally more efficient. The trade-off is of course a lower accuracy in the global surrogate model obtained by using PR.

The remainder of this paper is organized as follows. In the next section, we outline the HSAEA for solving computationally expensive optimization problems. The basic theory of the GP and PR modeling is briefly introduced. Empirical studies on HSAEA using GP or PR global surrogate models were conducted on two commonly used multimodal benchmark functions. The numerical results obtained are presented and discussed in Section 3 while Section 4 summarizes our main conclusions.

BEGIN

Initialize: Generate a database containing a population of designs.

Construct a global surrogate model.

While(*computational budget is not exhausted*)

- Evaluate all individuals in the population using the global surrogate model.
- **For** each non-duplicated pre-selected η percent promising individuals in the population
  - $*$ Apply trust-region enabled gradient-based local search strategy to the individual which interleaves the exact fitness function with a local surrogate model.
  - $*$ Update the database with any new evaluated design points.
  - $*$ Replace the individuals in the population with the locally improved solution in the spirit of Lamarckian learning.

  End For
- Apply standard EA operators to create a new population.

End While

END

Figure 1: Hierarchical Surrogate-Assisted Evolutionary Algorithm (HSAEA)

2 Hierarchical Surrogate-Assisted Evolutionary Algorithm

Here, we are interested in cases where the evaluation of the objective/fitness function is computationally expensive and it desired to obtain a near-optimal solution on a limited computational budget. The basic steps of the proposed Hierarchical Surrogate-Assisted Evolutionary Algorithm (HSAEA) are outlined in figure 1.

For the sake of readability, we present the proposed HSAEA optimization framework as four main phases:

Phase 0 $\{$*Initialization*$\}$: At the first step, a population of design points is initialized either randomly or using design of experiments techniques such as Latin hypercube sampling. These design points are evaluated using the exact objective function. The exact fitness values obtained are then archived in a central database together with the design vectors. After some initial period of time, a global surrogate model is constructed to represent the global trends of the entire fitness landscape.

Phase 1 $\{$*Global Search Strategy*$\}$: Subsequently, the global surrogate model is used to pre-evaluate all individuals of the population. The predictions produced by using the global surrogate model are used to pre-screen subsequent EA populations such that only the $\eta\%$ $(0 < \eta < 100)$

promising individuals undergo Lamarckian learning. This eliminates any unnecessary local searches from being conducted on individuals whose actual fitness is anticipated to be poor.

Phase 2 {*Local Search Strategy*}: A Lamarckian evolution process involving a trust-region framework devised for interleaving exact objective functions with computationally cheap local surrogate models is used during the gradient-based search. For each non-duplicated $\eta\%$ individuals, a local surrogate model is built dynamically to represent the local fitness landscape in the vicinity of an individual and is hence termed a local surrogate model. If an improved solution is found in the Lamarckian learning process, the genotype is forced to reflect the result of improvement by placing the locally improved individual back into the population to compete for reproductive opportunities. Subsequently, results of any new exact fitness obtained during the Lamarckian learning process are added into the central database, facilitating possible updating of surrogate models through online learning.

Phase 3 {*Standard EA Operations*}: The population then proceeds with the standard EA operators of crossover, mutation, etc. This process of hierarchical surrogate-assisted EA search is continued until the computational budget is exhausted or a user specified termination criterion is met.

Since local surrogate models will probably be built thousands of times during local searches, computational efficiency is a major concern. This consideration motivates the use of RBF method, which can be efficiently applied to approximate multiple-input multiple-output data, particularly when a few hundred data points are used for training. In the HSAEA algorithm, the RBF local surrogate model is constructed using only the m neighboring data points in the database nearest to the design point of interest because the neighboring points are likely to have more impact than remote ones [4]. In this way, the RBF model offers reasonable accuracy as well as fast training.

The global surrogate model is constructed by using the top ranking q archived design points of the database. Since the training data points spread across the entire search space to represent the whole fitness landscape, hence the model is termed a global surrogate model. In this paper, we study the effect on using GP or PR global surrogate models in HSAEA for solving computationally expensive optimization problems. For greater details on the HSAEA framework, the reader is referred to [16].

2.1 Gaussian Process Modeling

Here, we give a brief overview of the GP modeling technique for global surrogate model construction.

Let $\mathcal{D} = \{\mathbf{x}_i, t_i\}, i = 1 \ldots n$ denote the training dataset, where $\mathbf{x}_i \in \mathbb{R}^d$ is the input design vector and $t_i \in \mathbb{R}$ is the corresponding target value. The GP surrogate model assumes the presence of an unknown true modeling function $f(\mathbf{x})$ and an additive noise term v to account for anomalies in the observed data. Thus:

$$t(\mathbf{x}) = f(\mathbf{x}) + v \tag{1}$$

The standard analysis requires the specification of prior probabilities on the modeling function and the noise model. From a stochastic process viewpoint, the collection $\mathbf{t} = \{t_1, t_2, ..., t_n\}$ is called a Gaussian process if every subset of $\mathbf{t}$ has a joint Gaussian distribution. More specifically,

$$P(\mathbf{t}|\mathbf{C}, \{x_n\}) = \frac{1}{Z}\exp\left(-\frac{1}{2}(\mathbf{t}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{t}-\boldsymbol{\mu})\right) \tag{2}$$

where $\mathbf{C}$ is a covariance matrix parameterized in terms of hyperparameters $\boldsymbol{\theta}$, i.e., $\mathbf{C}_{ij} = k(x_i, x_j; \theta)$ and $\boldsymbol{\mu}$ is the process mean. The Gaussian process is characterized by this covariance structure since it incorporates prior beliefs both about the true underlying function as well as the noise model. In the present study, we use the following exponential covariance model

$$k(x_i, x_j) = e^{-(x_i - x_j)^T\Theta(x_i - x_j)} + \theta_{d+1} \tag{3}$$

where $\Theta = \mathrm{diag}\{\theta_1, \theta_2, ..., \theta_d\} \in \mathbb{R}^{d\times d}$ is a diagonal matrix of undetermined hyperparameters, and $\theta_{d+1} \in \mathbb{R}$ is an additional hyperparameter arising from the assumption that noise in the dataset is Gaussian (and output dependent). We shall henceforth use the symbol $\boldsymbol{\theta}$ to denote the vector of undetermined hyperparameters, i.e., $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_{d+1}\}$.

In practice, the undetermined hyperparameters are tuned to the data using the evidence maximization framework. Once the hyperparameters have been estimated from the data, predictions can be readily made for a new testing point. To illustrate this, assume that $\mathbf{t}_n$ represents the set of n targets, $\mathbf{C}_n$ the corresponding covariance matrix and that the process to be modeled has zero mean, i.e., $\boldsymbol{\mu} = 0$. Given a new point $\mathbf{x}_{n+1}$, it can be shown that the prediction t_{n+1} has a conditional probability distribution given by :

$$P(t_{n+1}|\mathcal{D}, \mathbf{C}, \mathbf{x}_{n+1}) = \frac{1}{Z}exp\left(-\frac{(t_{n+1} - \hat{t}_{n+1})^2}{2\hat{\sigma}^2}\right) \tag{4}$$

where,

$$\hat{t}_{n+1} = \mathbf{k}_{n+1}^T(\mathbf{x})\mathbf{C}_n^{-1}\mathbf{t}_n \tag{5}$$

and

$$\sigma_{n+1}^2 = \mathbf{k}(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}; \boldsymbol{\theta})\mathbf{k}_{n+1}^T(\mathbf{x})\mathbf{C}_n^{-1}\mathbf{k}_{n+1} \tag{6}$$

where, $\hat{t}_{n+1}$ and σ_{n+1}^2 are the prediction for the posterior mean and variance, respectively, and $\mathbf{k}_{n+1} = \{k(x_{n+1}, x_1), k(x_{n+1}, x_2), \ldots, k(x_{n+1}, x_n)\} \in \mathbb{R}^n$.

From a computational perspective, the search for an optimal GP under the evidence maximization framework [18] involves solving the following nonlinear maximum likelihood estimation (MLE) problem to determine the most probable hyperparameters $\boldsymbol{\theta}_{MP}$ for the given data:

$$\boldsymbol{\theta}_{MP} = \min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}), \tag{7}$$

where

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\log\det\mathbf{C}_n - \frac{1}{2}\mathbf{t}_n^T\mathbf{C}_n^{-1}\mathbf{t}_N - \frac{n}{2}\log 2\pi \tag{8}$$

is the negative log likelihood function.

Since computing $L(\boldsymbol{\theta})$ and its gradient generally involves computing and inverting a dense $n \times n$ covariance matrix (requiring $\mathcal{O}(n^3)$ resources) at each iteration, training the GP model can be prohibitively expensive even for moderately sized data (e.g., say a few thousand data points).

Here we alleviate the computational bottleneck in standard GP modeling by employing a data parallel approach, which makes it possible to deal with datasets containing tens of thousands of points at modest computational cost. The proposed Data Parallel Gaussian Process (DPGP) model achieves parallelism by employing a specialized compactly supported covariance function defined over spatially localized clusters. It use a compactly supported covariance function as a building block for GP models is shown to decompose the maximum likelihood estimation problem into smaller decoupled subproblems. The attendant benefits which include a significant reduction in training complexity, as well as sparse predictive models for the posterior mean and variance make this scheme extremely attractive. Greater details of the DPGP approach is available in [19].

Using the output mean prediction $\hat{t}(\mathbf{x})$ and standard deviation $\sigma(\mathbf{x})$ of DPGP model, a variety of pre-selection criteria for the selection of promising individuals may be formulated to accelerate evolutionary optimization search. Here we use the Probability of Improvement (PoI) as the pre-selection criterion to filter out promising individuals of population. Previous work in [11], [16] have also shown that that the PoI pre-selection criterion performs well.

Let t^- denote the smallest value of all the outputs in the training dataset used to construct the GP surrogate model. Subsequently, it is intended to use the surrogate model to predict a new point $\mathbf{x}^*$ at which the output is likely to be lower than t^-. The PoI at the point $\mathbf{x}^*$ (i.e., the probability that the surrogate prediction at $\mathbf{x}^*$ is lower than t^-) can be readily computed from the posterior mean $\hat{t}(\mathbf{x}^*)$ and standard deviation $\sigma(\mathbf{x}^*)$ as follows:

$$PoI(\mathbf{x}^*) = \Phi\left(\frac{t^- - \hat{t}(\mathbf{x}^*)}{\sigma(\mathbf{x}^*)}\right) \tag{9}$$

where $\Phi(.)$ is the normal cumulative distribution function.

2.2 Polynomial Regression Modeling

Next, we briefly describe the Polynomial Regression (PR) method for approximating multi-dimensional input data of any order [20].

Consider the input-output pair $(\mathbf{x}_i, t_i), i = 1 \ldots n$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the input design vector, $t_i \in \mathbb{R}$ is the corresponding target value, and $\mathbf{x_i} = (x_{i_1}, x_{i_2}, \ldots, x_{i_d})$, d denotes the dimension of the problem, we have:

$$t_i = t_i(\mathbf{x_i}) = t(x_{i_1}, \ldots, x_{i_d}) \tag{10}$$

Define an exponent vector ε containing n positive integers $(\pi_1, \pi_2, \ldots, \pi_d)$ and define $\mathbf{x_i}^{\varepsilon}$ as an exponent input vector $({x_{i_1}}^{\pi_1}, {x_{i_2}}^{\pi_2}, \ldots, {x_{i_d}}^{\pi_d})$.

Given a set of exponent vectors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ and the set of data $(\mathbf{x}_i, t_i)$, where $i = 1, 2, \ldots, n$, the polynomial model of $(m-1)^{th}$ order has the form:

$$P(\mathbf{x_i}) = C_1 {\mathbf{x_i}}^{\varepsilon_1} + C_2 {\mathbf{x_i}}^{\varepsilon_2} + \ldots + C_m {\mathbf{x_i}}^{\varepsilon_m} \tag{11}$$

where $C_1, C_2, \ldots, C_m$ are the coefficient vectors to be estimated, and $C_j = (c_{j_1}, c_{j_2}, \ldots, c_{j_d}), j = 1, 2, \ldots, m$.

The least square method is then used to estimate the coefficients of the polynomial model. By definition, the least square error E to be minimized is:

$$E = \sum_{i=1}^{n} [t_i - P(\mathbf{x}_i)]^2 \tag{12}$$

It may be easily shown that $t_i = P(\mathbf{x}_i)$, and by multiplying both sides of equation (11) with $\mathbf{x}_i^{\varepsilon_j}$ and taking the sum of n pairs of input-output data, we arrive at

$$C_1 \sum_i \mathbf{x}_i^{\varepsilon_1 + \varepsilon_j} + \ldots + C_m \sum_i \mathbf{x}_i^{\varepsilon_m + \varepsilon_j} = \sum_i t_i \mathbf{x}_i^{\varepsilon_j} \tag{13}$$

For $j = 1, 2, \ldots, m$, the polynomial model can be represented in the matrix notation as follow

$$A\gamma^T = \boldsymbol{b}^T \tag{14}$$

where

$$A = \begin{bmatrix} \sum_i \mathbf{x}_i^{\varepsilon_1 + \varepsilon_1} & \ldots & \sum_i \mathbf{x}_i^{\varepsilon_1 + \varepsilon_m} \\ \vdots & & \vdots \\ \sum_i \mathbf{x}_i^{\varepsilon_m + \varepsilon_1} & \ldots & \sum_i \mathbf{x}_i^{\varepsilon_m + \varepsilon_m} \end{bmatrix} \tag{15}$$

$$\boldsymbol{b} = (\sum t_i \mathbf{x}_i^{\varepsilon_1}, \ \ldots \ , \sum t_i \mathbf{x}_i^{\varepsilon_m}) \tag{16}$$

$$\gamma = (C_1, C_2, \ldots, C_m) \tag{17}$$

Then the coefficient matrix of the polynomial is:

$$\gamma = (A^{-1}\boldsymbol{b}^T)^T \tag{18}$$

Let $B_i = (\mathbf{x}_i^{\varepsilon_1}, \ldots, \mathbf{x}_i^{\varepsilon_m})$, the following equations may be derived:

- $A = \sum_i B_i^T B_i$
- $\boldsymbol{b} = \sum_i t_i B_i$
- $P(\mathbf{x}_i) = \gamma . B_i^T$

The predicted output for a new input pattern is then given by $P(\mathbf{x}_i) = \gamma . B_i^T$.

3 Empirical Results

In this section, we present a numerical study on two multi-modal benchmark test functions (i.e., the Ackley and Griewank functions) so as to investigate the convergence properties of the HSAEA using PR and GP global surrogate models. Since the Genetic Algorithm (GA) is used in our present study, the Hierarchical Surrogate-Assisted Genetic Algorithm with PR or GP global surrogate model is referred in short as PR-HSAGA and GP-HSAGA, respectively.

- **Ackley Test Function**:

$$f(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos 2\pi x_i}$$
$$-32.768 \leq x_i \leq 32.768, i = 1, 2, \ldots, n.$$

- **Griewank Test Function**:

$$f(x) = 1 + \sum_{i=1}^{n} x_i^2/4000 - \prod_{i=1}^{n} \cos(x_i/\sqrt{i})$$
$$-600 \leq x_i \leq 600, i = 1, 2, \ldots, n.$$

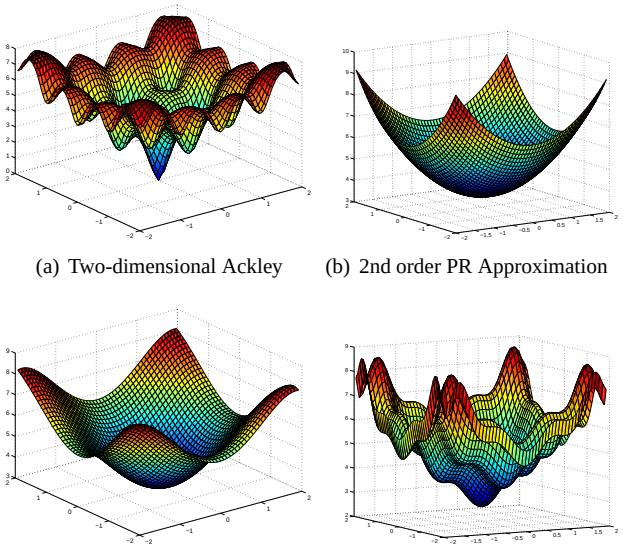
In our study, we employed a standard GA with population size of 50, crossover and mutation operators at probabilities 0.6 and 0.001, respectively. A stochastic universal sampling algorithm is used for selection. Besides the HSAGA, the Surrogate-Assisted Genetic Algorithm with Local search Strategy (SAGA-LS) proposed in [1] is also reported in this paper to facilitate possible comparison in search performances. In contrast to the HSAGA, SAGA-LS does not employ any global surrogate models, instead all individuals of the GA population undergoes local improvements in the form of Lamarckian learning. Note that in our numerical study, SAGA-LS adopts the same parameter configurations as the standard GA. However, apart from the standard GA settings, the two user-specified parameters of the SAGA-LS 1) number of nearest neighboring data points used to construct the local surrogate model, m and 2) maximum trust region iterations k, are configured to be 100 and 3, respectively.

Furthermore, the maximum number of training design points (i.e., q_{max}) and clusters for constructing the global surrogate model using DPGP (referred here as GP) for HSAGA are configured as 2000 and 2, respectively. While η is configured as 40%. It is worth noting that all configurations used in this study were based on those suggested in earlier studies [1], [16]. All results presented are the average of 20 independent simulations conducted on a limited computational budget of (6×10^3) exact objective function evaluations.

3.1 Effect of PR Order

We begin by presenting the effect of PR orders for approximating the multi-modal Ackley function.

The two dimensional Ackley function depicted in Figure 2(a) is approximated using 2nd, 5th and 10th order polynomial regression and 300 training data points, which are plotted in Figures 2(b) to 2(d), respectively. The Ackley function is highly multi-modal with many local minima and a global minimum located at (0,...,0). From the figures, it may be observed the 2nd order PR model is unable to approximate the complex landscape of the Ackley function very well. The approximated Ackley becomes a quadratic function. Higher order PR models, on the other hand, approximates the multi-modal landscape of the Ackley function more closely. However, a higher order PR incurs greater computational cost to construct the global model. The main computational cost of PR modeling is $\mathcal{O}((dm)^3)$ for n pairs of data, and $(m-1)^{th}$ order.



(a) Two-dimensional Ackley (b) 2nd order PR Approximation

(c) 5th order PR Approximation (d) 10th order PR Approximation

Figure 2: Two-dimensional Ackley function

We also study the convergence behavior of HSAGA framework using different orders of PR for global surrogate modeling. Figures 3 and 4 depict the search traces of PR-HSAGA search on the two multi-modal test functions. The PR^{2nd}-HSAGA fares poorer than both the 5th and 10th order counterparts on the test functions. This may be attributed to the inferior approximation accuracy of the 2nd order PR global model which results in poor identification of promising individuals during the pre-screening process.

The results in figures 3 and 4 also show little improvements in using the PR^{10th}-HSAGA over the 5th order counterpart. In effect, we use the PR^{5th}-HSAGA in our subsequent experimental study since it is more computationally efficient and competitive than the PR^{10th}-HSAGA.

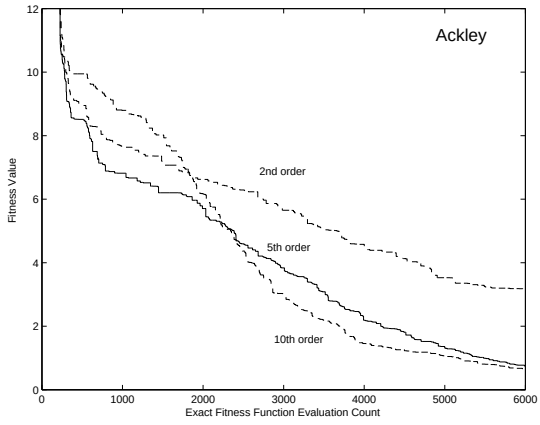


Figure 3: Averaged search trends of PR-HSAGA for different orders on 20-dimensional Ackley function

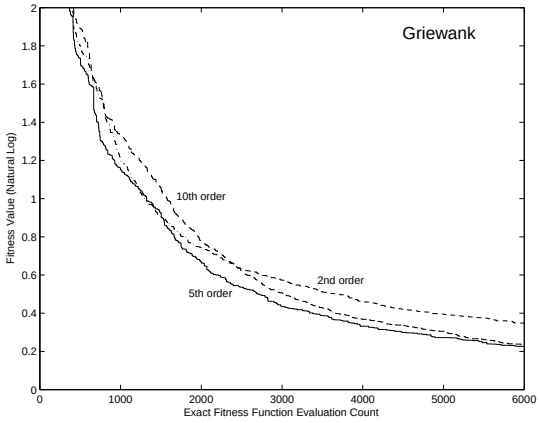


Figure 4: Averaged search trends of PR-HSAGA for different orders on 20-dimensional Griewank function

3.2 Performance Comparison of PR-HSAGA and GP-HSAGA

Figures 5 and 6 depicts the search performance traces of the $5th$ order PR-HSAGA and GP-HSAGA on the Ackley and Griewank functions, respectively.

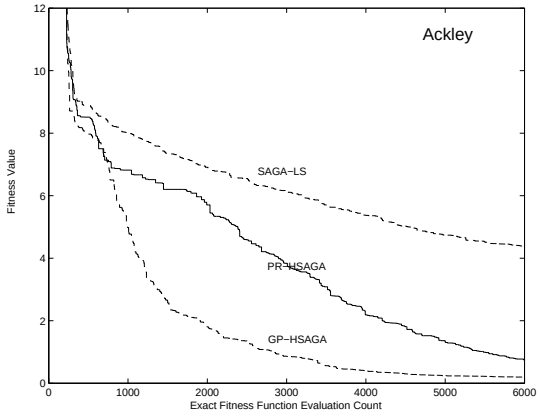


Figure 5: Averaged search trends of SAGA-LS, GP-HSAGA, and PR-HSAGA on 20-dimensional Ackley function

It may be observed in these figures that both PR-HSAGA and GP-HSAGA were able to filter off unnecessary local searches, thus accelerating convergence and providing significant computational cost savings over the SAGA-LS. Overall, GP-HSAGA converges to better designs faster than PR-HSAGA. We observed that this is due to the inability of the PR method to approximate the complex high dimensional landscape accurately, resulting in poor qualitative fitness approximation. Consequently, the PR global surrogate model fails to accurately identify the $\eta\%$ top ranking promising individuals in the EA population that will undergo Lamarckian learning. Figure 7 illustrates one such case where vector ***y*** is preferred over ***x*** due to the poor qualitative fitness approximation of the PR model. Nevertheless, it is worth noting that PR-HSAGA is still capable of providing significant computational cost savings over the SAGA-LS. This is most likely due to the second level surrogate-assisted local search which helps maintain good accuracy in the evolutionary search.

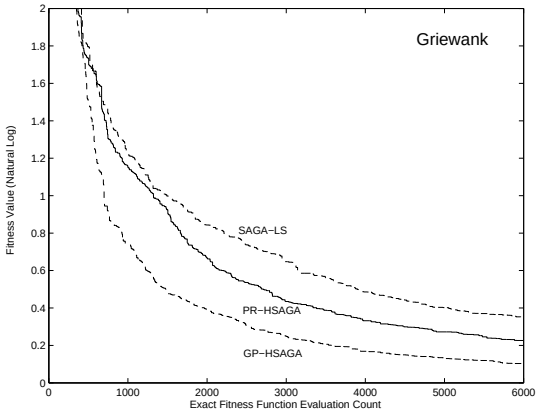


Figure 6: Averaged search trends of SAGA-LS, GP-HSAGA, and PR-HSAGA on 20-dimensional Griewank function

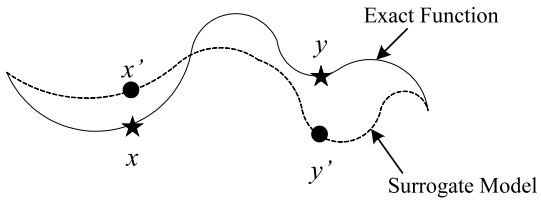


Figure 7: Poor qualitative approximation of surrogate model

3.3 Wall Clock Time of PR-HSAGA and GP-HSAGA

Many existing works on surrogate-assisted evolutionary algorithms only report the fitness values obtained against the number of exact evaluation calls made to the computationally expensive optimization problem. However, it is obvious there is significant difference in the computationally efforts incurred by GP and PR for surrogate modeling. Hence, it would be interesting to consider the actual wall clock time of the PR-HSAGA and GP-HSAGA, inclusive of the modeling time. The computational efforts for a single generation of the HSAGA search may be formulated as

$$T_{gen} = t_g + t_{ls} + k \cdot \eta \cdot n_{pop} \cdot t_{obj} + t_{ga} \qquad (19)$$

where t_g, t_{ls}, t_{obj} and t_{ga} denotes the wall clock time to construct the global surrogate model, complete a trust-region local search with on-line RBF local surrogate models, evaluate a single exact objective function and perform standard GA operators, respectively.

In solving optimization problems, the terms t_{ls}, k, η, n_{pop} and t_{ga} in equation (19) are the same for both PR-HSAGA and GP-HSAGA, but differs in the time taken to constructing the global surrogate model, t_g. The main computational cost involved in constructing GP surrogate models occurs in the maximum likelihood estimation phase so as to determine the most probable hyperparameters [18]. The evaluation of the likelihood function requires factorization

Model Type	Using training data size of	
	500	2000
the 5th order PR	0.333165	1.20895
GP(2 cluster DPGP)	11.912108	1255.4707

Table 1: Average Wall Clock Time (seconds) to Construct Surrogate Model

of the correlation matrix and scales as $\mathcal{O}(n^3)$. In comparison, the main computational cost in constructing PR model is performing matrix inversion which scales as $\mathcal{O}((dm)^3)$ but is independent on the number of training data, i.e. n in GP model.

Since the number of training data, n, used to built the models is often large in size and $dm \ll n$, the time to construct PR modeling is significantly lower compared to using GP. Table 1 tabulates the average wall clock time to construct the GP and PR global surrogate models for different sizes of training dataset.

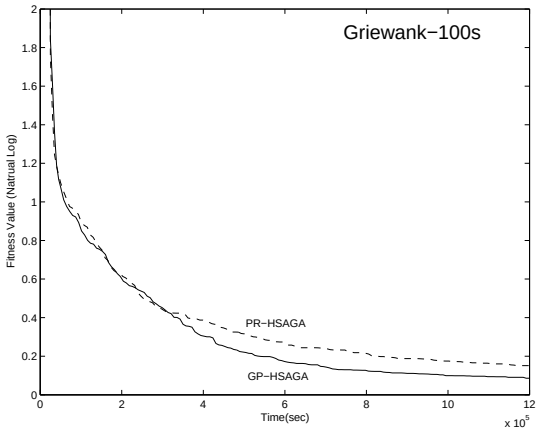


Figure 8: Wall Clock Time of GP-HSAGA and PR-HSAGA on the 20-Dimensional $100s$ Griewank Function

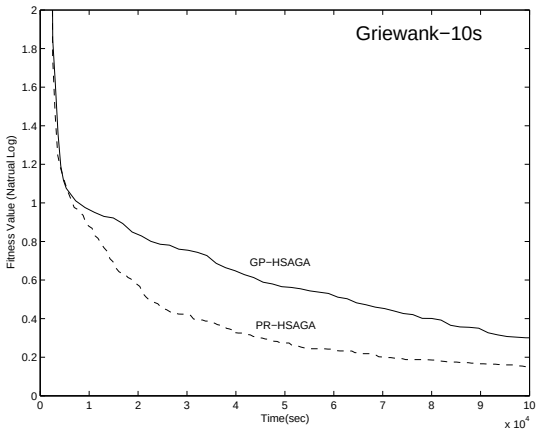


Figure 9: Wall Clock Time of GP-HSAGA and PR-HSAGA on the 20-Dimensional $10s$ Griewank Function

With the great difference in the time taken to construct models using PR and GP, it is apparent some approximation methods may not be efficient if the the computational expense of the optimization problem is not very high. Hence, it would be interesting to investigate a suitable approximation method for the optimization problem in hand. Technically, when using HSAGA to solve an optimization problem having a computational expense of t_{obj}, one should use a GP global model over PR if the following condition is satisfied:

$$t_{obj} \gg \frac{t_{gp}}{k \cdot \eta \cdot n_{pop}} \tag{20}$$

where t_{gp} is the wall clock time to construct a GP model in each EA generation, and n_{pop} is the population size.

On the other hand, the PR method is preferred over GP when:

$$\frac{t_{pr}}{k \cdot \eta \cdot n_{pop}} \ll t_{obj} < \frac{t_{gp}}{k \cdot \eta \cdot n_{pop}} \tag{21}$$

where t_{pr} and t_{gp} are the wall clock time to construct a PR or GP model in each EA generation, respectively.

To validate this, we investigate the search performances of HSAGA for variable-fidelity problems by using the Griewank function with time delays. Note that in the present experiment, all parameter configurations are kept the same as before. By substituting the values of k, η, n_{pop}, t_{pr} and t_{gp} to equations (20) and (21), we can arrive at $t_{obj} \gg 20.924512$ and $0.0201492 \ll t_{obj} < 20.924512$, respectively.

Figures 8 and 9 show the search traces of GP-HSAGA and PR-HSAGA on the $100s$ and $10s$ Griewank function. On the $100s$ Griewank function, i.e., $t_{obj} \gg 20.924512$, GP-HSAGA is observed to perform better than PR-HSAGA. On the other hand, PR-HSAGA is superior to GP-HSAGA on the $10s$ Griewank function when $0.0201492 \ll t_{obj} < 20.924512$. These results thus prove our hypothesis in equations (20) and (21).

4 Conclusion

This paper has elaborated the use of combining both global and local surrogate models for solving different complex engineering design problems. The PR and GP global surrogate models are studied to guide the hierarchical evolutionary search. Numerical study show that both PR-HSAEA and GP-HSAEA converge to good designs on a limited computational budget. Further, it may be inferred that the GP method should generally be used in the HSAEA for global surrogate modeling if the evaluation function is very expensive in computations. On moderately expensive problems, the PR method may serve to be a better choice.

Acknowledgment

The authors would also like to thank the two research centres, Parallel and Distributed Computing Centre, Centre for Multimedia and Network Technology of Nanyang Technological University for their support in this work.

Bibliography

[1] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *American Institute of Aeronautics and Astronautics Journal*, vol. 41, no. 4, pp. 687–696, 2003.

[2] Z. Zhou, Y. S. Ong, and P. B. Nair, "Hierarchical surrogate-assisted evolutionary optimization framework," in *Proc. IEEE Congress on Evolutionary Computation CEC'04*, Portland, USA, June 2004, pp. 1586–1593.

[3] A. Ratle, "Kriging as a surrogate fitness landscape in evolutionary optimization," *Artificial Intelligence for Engineering Design Analysis and Manufacturing*, vol. 15, no. 1, pp. 37–49, 2001.

[4] K. C. Giannakoglou, "Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence," *International Review Journal Progress in Aerospace Sciences*, vol. 38, no. 5, pp. 43–76, 2002.

[5] M. K. Karakasis, A. P. Giotis, and K. C. Giannakoglou, "Inexact information aided, low-cost, distributed genetic algorithms for aerodynamic shape optimization," *International Journal for Numerical Methods in Fluids*, vol. 43, pp. 1149–1166, 2003.

[6] A. A. Guinta and L. Watson, "A comparison of approximation modelling techniques: polynomial versus interpolating models," in *Proc. the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA-98-4758*, St. Louis, MO, USA, Sept. 1998.

[7] D. D. Daberkow and D. N. Marris, "New approaches to conceptual and preliminary aircraft design: A comparative assessment of a neural network formulation and a response surface methodology," in *Proc. 1998 World Aviation Conference, AIAA*, Florida , USA, Sept. 1998.

[8] R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodeling techniques under multiple modeling criteria," *Structural and Multidisciplinary Optimization*, vol. 23, no. 1, pp. 1–13, 2001.

[9] M. A. El-Beltagy, P. B. Nair, and A. J. Keane, "Metamodelling techniques for evolutionary optimization of computationally expensive problems: Promise and limitations," in *Proc. IEEE the Genetic and Evolutionary Computation Conference (GECCO'99)*, Florida , USA, July 1999, pp. 196–203.

[10] Y. Jin, M. Olhofer, and B. Sendho, "A framework for evolutionary optimization with approximate fitness functions," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 5, pp. 481–494, 2002.

[11] H. Ulmer, F. Streichert, and A. Zell, "Evolution startegies assisted by gaussian processes with improved preselection criterion," in *Proc. IEEE Congress on Evolutionary Computation CEC'03*, Canberra, Australia, Dec. 2003, pp. 1741–1751.

[12] D. B*ü*che, N. N. Schraudolph, and P. Koumoutsakos, "Accelerating evolutionary algorithms with gaussian process fitness function models," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 2004.

[13] K. H. Liang, X. Yao, and C. Newton, "Evolutionary search of approximated n-dimensional landscapes," *International Journal of Knowledge-Based Intelligent Engineering Systems*, vol. 4, no. 3, pp. 172–183, 2000.

[14] Y. S. Ong, K. Y. Lum, P. B. Nair, *et al.*, "Global convergence unconstrained and bound constrained surrogate-assisted evolutionary search in aerodynamic shape design," in *Proc. IEEE Congress on Evolutionary Computation CEC'03*, vol. 3, Canberra , Australia, Dec. 2003, pp. 1856–1863.

[15] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," *Soft Computing*, vol. 9, no. 1, pp. 3–12, 2005.

[16] Z. Zhou, Y. S. Ong, P. B. Nair, A. J. Keane, and K. Y. Lum, "Combining global and local surrogate models to accelerate evolutionary optimization," *IEEE Transactions on Systems, Man and Cybernetics (SMC), part C*, 2005, In press.

[17] Y. Jin, M. Huesken, and B. Sendhoff, "Quality measures for approximate models in evolutionary computation," in *Proc. GECCO Workshops: Workshop on Adaptation, Learning and Approximation in Evolutionary Computation*.

[18] D. Mackay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992. [Online]. Available: citeseer.ist.psu.edu/article/mackay91bayesian.html

[19] A. Choudhury, P. B. Nair, and A. J. Keane, "A data parallel approach for large-scale gaussian process modeling," in *Proc. the Second SIAM International Conference on Data Mining*, Arlington, VA, Apr. 2002.

[20] F. H. Lesh, "Multi-dimensional least-square polynomial curve fitting," *Communications of ACM*, vol. 2, no. 9, pp. 29–30, 1959.