Hierarchical Surrogate-Assisted Evolutionary Optimization Framework

Zongzhao Zhou
School of Computer Engineering
Nanyang Technological University
Singapore
Email: zhouzongzhao@pmail.ntu.edu.sg

Yew Soon Ong
School of Computer Engineering
Nanyang Technological University
Singapore
Email: asysong@ntu.edu.sg

Prasanth B. Nair
School of Engineering Sciences
University of Southampton
Southampton SO17 1BJ, U.K
Email: P.B.Nair@soton.ac.uk

Abstract— **This paper presents enhancements to a surrogate-assisted evolutionary optimization framework proposed earlier in the literature for solving computationally expensive design problems on a limited computational budget [1]. The main idea of our former framework was to couple evolutionary algorithms with a feasible sequential quadratic programming solver in the spirit of Lamarckian learning, including a trust-region approach for interleaving the true tness function with computationally cheap local surrogate models during gradient-based search. In this paper, we propose a hierarchical surrogate-assisted evolutionary optimization framework for accelerating the convergence rate of the original surrogate-assisted evolutionary optimization framework. Instead of using the exact high- delity tness function during evolutionary search, a Kriging global surrogate model is used to screen the population for individuals that will undergo Lamarckian learning. Numerical results are presented for two multi-modal benchmark test functions to show that the proposed approach leads to a further acceleration of the evolutionary search process.**

I. INTRODUCTION

In recent years, Evolutionary Algorithms (EAs) have been successfully applied to many complex engineering design optimization problems. Its popularity lies in the ease of implementation and the ability to arrive close to the global optimum design. However, high computational costs associated with the use of high- delity simulation models pose a serious impediment to the successful application of Evolutionary Algorithms (EAs) to engineering design optimization. This is primarily because a single function evaluation (involving the analysis of a complex engineering system) often consumes many minutes to hours of computer time and EAs typically require thousands of tness function evaluations to locate a near optimal solution. One promising way to signi cantly reduce the computational cost of EAs is to employ computationally cheap surrogate models in place of computationally expensive tness evaluations. By introducing surrogate models, the computational burden can be greatly reduced since the efforts involved in building the surrogate model and optimization using it are much smaller than the standard approach of direct coupling the simulation code with the optimizer.

Many existing evolutionary frameworks for solving computationally intractable problems consist of building a statistical global model of the tness landscape from a small number of data points that represents the search space. These data points are usually obtained during one or more generations of a classical evolutionary search. The statistical global model is then exploited by the evolutionary algorithm as an auxiliary tness function in order to get the maximum amount of information out of these initial data points. Subsequently, the statistical models are updated online based on new data points obtained during the surrogate-assisted evolutionary search [2], [3], [4]. Alternatively, the hybrid surrogate-assisted evolutionary optimization framework may employ a trust-region approach in the evolutionary search for interleaving use of the exact models for the objective and constraint functions with computationally cheap local surrogate models during Lamarckian learning [1]. Lamarckian learning forces the genotype to re ect the result of improvement by placing the locally improved individual back into the population to compete for reproductive opportunities.

In this paper, we present a preliminary investigation into a hierarchical surrogate-assisted evolutionary optimization framework for computationally expensive problems. The hierarchical optimization framework considered in this study brings together a number of ideas proposed in the literature [1], [2], [5]. We would like to highlight that the present work is mainly motivated by the lack of studies in the literature on combining global and local surrogate models as well as suitable hierarchical surrogate-assisted frameworks for evolutionary optimization of computationally expensive problems. In other words, we show how global and local surrogate models can be synergistically combined to accelerate EAs.

The remainder of this paper is organized as follows. In the next section, we brie y describe the structure of a general surrogate-assisted evolutionary optimization framework. Section III presents the proposed hierarchical surrogate-assisted evolutionary optimization framework which employs both global and local surrogate models to accelerate convergence. The results of numerical studies conducted on two benchmark test functions are presented and discussed in Section IV while Section V summarizes the major conclusions.

II. SURROGATE-ASSISTED EVOLUTIONARY OPTIMIZATION FRAMEWORK

In this section, we give a brief overview of the general surrogate-assisted evolutionary optimization framework for computationally expensive problem. Several efforts in the area have been made over the recent years, particularly using Genetic Algorithms (GA) and Evolutionary Strategies (ES) [1], [2], [4], [5].

The basic evolutionary algorithm will not be described here, the reader is referred to the literature for a detailed exposition; see, for example, [6]. The outline of a typical surrogate-assisted evolutionary optimization framework is shown in Figure 1. In the rst step, a database is initialized using a population of designs, which are generated either randomly or using design of experiments techniques such as Latin hypercube sampling, orthogonal arrays, orthogonal array-based Latin hypercube sampling. All the design points thus generated and the associated exact values of the objective and constraint functions are then archived in the database that will be used later for constructing global or local surrogate models. Alternatively, one could also use a database containing the results of a previous search on the problem or a combination of the two or the database builds dynamically up as the search progresses. Subsequently, with ample design points in the database (i.e., this is often after some pre-de ned number of EA search generations using the exact tness function), the search proceeds according to the strategies employed in the surrogate-assisted evolutionary optimization framework.

Ratle[7] examined a strategy for integrating GAs with Kriging models. It uses a heuristic convergence criterion to determine when an approximate model must be updated. The same problem was revisited by El-Beltagy et al. [3], where it was argued that the issue of balancing the concerns of optimization with those of design of experiments should be addressed. Jin et al. [4] coupled ES with neural network-based surrogate models. In their approach, an empirical criterion was proposed to decide the frequency at which the expensive and approximate models should be used throughout the search. In Song et al. [8], a real-coded GA was coupled with Kriging for rtree structural optimization. A strategy for coupling ES with local search along a quadratic response surface model was proposed in Liang et al. [10]. A parallel hybrid EA framework that leverages local surrogate models for solving computationally expensive design problems with general constraints was proposed in [1] and further extended in [11] to incorporate gradient information in the approximation process.

In general, most existing strategies employed in surrogate-assisted evolutionary search typically involve considering one or more of the following issues listed below:

* steps to control the switch between surrogate model and exact tness function, for example, individual or generation levels or using a trust-region approach,
* the choice of approximation techniques,
* de ning the sample points and data used for surrogate modeling,
* creating global or local surrogate models,
* working with non-linear equality/inequality constraints,
* parallelism, and last but not least,
* convergence guarantee schemes.

For a review of existing surrogate-assisted evolutionary optimization frameworks for high- delity engineering design problems, the reader may refer to [12].

BEGIN
Initialize: Generate a database containing a population of designs.
(*Optional: upload a historical database if one exists*)
While (*computational budget not exhausted*)

* Evaluate all individuals in the population using the exact models for a pre-de ned number of EA generations.
* Proceeds according to the strategies employed in the surrogate-assisted evolutionary optimization framework.
* Apply standard EA operators to create a new population.

End While
END

Fig. 1. Outline of a General Surrogate-Assisted Evolutionary Optimization Framework.

III. Hierarchical Surrogate-Assisted Evolutionary Optimization Framework

The hierarchical framework begins with the initialization of a population of design points, which are evaluated using the exact tness function. These design points then form the training dataset used later for constructing surrogate models. Figure 2 summarizes a hierarchical surrogate-assisted evolutionary optimization framework that employs both global and local surrogate models.

In the present investigation, Radial Basis Function (RBF) and Kriging are employed as the local and global surrogate models, respectively. Since local surrogate models will probably be built thousands of times during search, computational ef ciency is a major concern. This consideration motivates the use of radial basis function networks, which can be ef ciently applied to approximate multiple-input multiple-output data, particularly when a few hundred data points are used for training.

Incidentally, to be ef cient and effective for complex engineering design optimization, the selected global surrogate model should retain the characteristics of modest computational complexity and accurate representation of the global trends of the tness landscape [2]. A statistically rigorous alternative to RBF approximation is the idea of Bayesian interpolation or regression which is also referred to as Gaussian process regression in the neural networks literature and Kriging in the geostatistics literature. It is generally recognized as a powerful tool for modelling and estimation. Hence, Kriging interpolation is chosen since it retains the aforementioned features. Besides, the Kriging method is statistically more meaningful and also allows the possibility of computing error estimates for the predicted outputs.

A. Global Surrogate Modelling

In the present hierarchical optimization framework, a global Kriging model is built using all the design data points in the database. The computationally cheap global surrogate model is then used in place of the exact tness function for evaluating all individuals in the EA population. The role of the global surrogate model consider here is to identify potentially promising areas in the search space. An obvious and commonly used metric is the absolute tness value pre-evaluated using the global surrogate [5]. Alternatively, the estimated error of the predicted tness may be taken into consideration. Jones et al. [13] proposed the expected improvement approach, which attempts to achieve a balance between seeking promising areas of the design space based on both the absolute tness criterion and the uncertainty in the surrogate model.

Let $\{\mathbf{x}^i, y_i, i = 1, 2, \ldots, n\}$ denote the training dataset, where $\mathbf{x} \in \mathbb{R}^d$ is the input vector and $y \in \mathbb{R}$ is the output. The Kriging model can be expressed as:

$$y(\mathbf{x}) = \beta + Z(\mathbf{x}), \tag{1}$$

where β represents a constant term of the model, and $Z(\mathbf{x})$ is a zero mean Gaussian stochastic process. The covariance matrix of $Z(\mathbf{x})$ is given by

$$Cov(Z(\mathbf{x}^i), Z(\mathbf{x}^j)) = \sigma^2 R(\mathbf{x}^i, \mathbf{x}^j), \tag{2}$$

where σ^2 is the so called process variance and $R(...)$ is the correlation function. Different types of correlation functions can be employed. A commonly used type of correlation function is the Gaussian kernel

$$R(\mathbf{x}^i, \mathbf{x}^j) = \prod_{k=1}^{n} exp(-\theta_k |x_k^i - x_k^j|^{p_k}), \tag{3}$$

where $\theta_k \geq 0$ and $0 < p_k \leq 2$ are the hyperparameters. Note that the above equation asserts that there is a complete correlation of a point with itself and this correlation deteriorates rapidly as the two points move away from each other in the parameter space. The choice of $p_k = 2$ would provide enough exibility for modelling smooth and highly non-linear functions for most cases.

The hyperparameters in the Kriging model can be estimated by maximizing a likelihood function using numerical optimization techniques. Once the hyperparameters are obtained from the training data, the function value at a new point can be predicted by

$$\hat{y}(\mathbf{x}^*) = \hat{\beta} + \mathbf{r}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\beta}), \tag{4}$$

where $\mathbf{R} \in \mathbb{R}^{n \times n}$ is the correlation matrix and $\mathbf{1} = \{1, 1, \ldots 1\}^T \in \mathbb{R}^n$.

The posterior variance of the prediction $s^2(\mathbf{x}^*)$ is given by

$$s^2(\mathbf{x}^*) = \sigma^2 [1 - \mathbf{r}\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}^T \mathbf{R}^{-1} \mathbf{r})^2)}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{r}}] \tag{5}$$

where $\mathbf{r}(\mathbf{x}) = \{R(\mathbf{x}, \mathbf{x}^1), ..., R(\mathbf{x}, \mathbf{x}^n)\}$ is the correlation vector between the new point $\mathbf{x}$ and the points in the training dataset.

The main computational cost involved in constructing Kriging models occur in the maximum likelihood estimation phase. Here, a nonlinear optimization technique has to be employed to estimate the hyperparameters by maximizing a likelihood function. Evaluation of the likelihood function requires factorization of the correlation matrix $\mathbf{R}$ which scales as $\mathcal{O}(n^3)$. For this reason, constructing a Kriging model for cases with more than two thousand points can be computationally very expensive. As an aside, we would like to point out here that using a data parallel approach [14], it is possible to apply Kriging to dataset with tens of thousands of points.

B. Local Surrogate Modelling

Subsequently, the members in the population are pre-evaluated and rank sorted based on the global surrogate model. The rst mechanism of the hierarchical framework lies in the use of the global surrogate model as a screening operation such that only the top σ percentage individuals in the EA population undergo the Lamarckian learning process, where $0 < \sigma < 100\%$ is a percentage value which is speci ed by the user. The local strategy or local improvement procedure used in the Lamarckian learning process is a trust-region approach for interleaving the exact tness functions with computationally cheap local surrogate models during local search in the spirit of Lamarckian learning [1]. The local surrogate models are built dynamically using only portions of the design points in the database and interpolating radial basis function networks of the form

$$\hat{y} = \sum_{i=1}^{n} \alpha_i K(||\mathbf{x} - \mathbf{x}^i||), \tag{6}$$

where $K(||\mathbf{x} - \mathbf{x}^i||) : \mathbb{R}^d \rightarrow R$ is a radial basis kernel and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \in \mathbb{R}^n$ denotes the vector of weights.

Typical choices for the kernel include linear splines, cubic splines, multiquadrics, thin-plate splines, and Gaussian functions [15]. Given a suitable kernel, the weight vector can be computed by solving the linear algebraic system of equations $\mathbf{K}\boldsymbol{\alpha} = \mathbf{y}$, where $\mathbf{y} = \{y_1, y_2, \ldots, y_n\} \in \mathbb{R}^n$ denotes the vector of outputs and $\mathbf{K} \in \mathbb{R}^{n \times n}$ denotes the Gram matrix formed using the training inputs (i.e., the ijth element of $\mathbf{K}$ is computed as $K(||\mathbf{x}^i - \mathbf{x}^j||)$).

In the present study, we use linear splines to construct surrogate models since experimental studies in the literature [9] suggest that this kernel is capable of providing models with good generalization capability at a low computational cost. The local improvement procedure embeds a Feasible Sequential Quadratic Programming optimizer (FSQP) within the trust-region framework, which ensures convergence to a stationary point or local optimum of the exact computationally expensive tness function [1], [16]. More speci cally, for each non-duplicated individuals among the top ranking σ percentage in the population, the local strategy proceeds with a sequence of trust-region subproblems of the form

$$\begin{array}{ll} \textbf{Minimize:} & \hat{f}^k(\mathbf{x} + \mathbf{x}_c^k) \\ \textbf{Subject to:} & \hat{g}_i^k(\mathbf{x} + \mathbf{x}_c^k) \leq 0, i = 1, 2, \ldots, p \\ & ||\mathbf{x}|| \leq \Omega^k \end{array} \tag{7}$$

where $k = 0, 1, 2, \ldots, k_{max}$, $\hat{f}(x)$ and $\hat{g}(x)$ are the approximation functions corresponding to the objective function $f(x)$ and constraint function $g(x)$ respectively, $\mathbf{x}_c^k$ and Ω^k are the initial guess and the trust-region radius used for local search at iteration k, respectively. In practice, the L_∞ norm can be employed to impose the second constraint in Eqn. (7). Hence, this constraint can be transformed into appropriate bounds on the design variables, which is updated at each trust-region iteration based on the value of Ω^k.

For each subproblem (or during each trust-region iteration), surrogate models of the objective and constraint functions, viz., $\hat{f}^k(\mathbf{x})$ and $\hat{g}_i^k(\mathbf{x})$ are created dynamically. The m nearest neighbors of the initial guess, $\mathbf{x}_c^k$, are extracted from the archived database of design points evaluated so far using the exact analysis code. The criterion used to determine the similarity between design points is the simple Euclidean distance metric. These points are then used to construct local surrogate models of the objective and constraint functions.

The surrogate models thus created are used to facilitate the necessary objective and constraint function estimations in the local searches. During local search, we initialize the trust-region Ω using the minimum and maximum values of the design points used to construct the surrogate model. After each iteration, the trust-region radius Ω^k is updated based on a measure which indicates the accuracy of the surrogate model at the kth local optimum, $\mathbf{x}_{lo}^k$. After computing the exact values of the objective and constraint functions at this point, the gure of merit, ρ^k, is calculated as

$$\rho^k = \min(\rho_f^k, \rho_{g_i}^k), \text{ for } i = 1, 2, \ldots, p, \tag{8}$$

where

$$\rho_f^k = \frac{f(\mathbf{x}_c^k) - f(\mathbf{x}_{lo}^k)}{\hat{f}(\mathbf{x}_c^k) - \hat{f}(\mathbf{x}_{lo}^k)} \text{ and } \rho_{g_i}^k = \frac{g_i(\mathbf{x}_c^k) - g_i(\mathbf{x}_{lo}^k)}{\hat{g}_i(\mathbf{x}_c^k) - \hat{g}_i(\mathbf{x}_{lo}^k)} \tag{9}$$

The above equations provide a measure of the actual versus predicted change in the objective and constraint function values at the kth local optimum. The value of ρ^k is then used to update the trust-region radius as follows [17]:

$$\begin{aligned} \Omega^{k+1} &= 0.25\Omega^k, && \text{if } \rho^k \le 0.25, \\ &= \Omega^k, && \text{if } 0.25 < \rho^k \le 0.75, \\ &= \xi\Omega^k, && \text{if } \rho^k \ge 0.75, \end{aligned} \tag{10}$$

where $\xi = 2$, if $||\mathbf{x}_{lo}^k - \mathbf{x}_c^k||_\infty = \Omega^k$ or $\xi = 1$, if $||\mathbf{x}_{lo}^k - \mathbf{x}_c^k||_\infty < \Omega^k$.

The trust-region radius, Ω^k, is reduced if the accuracy of the surrogate, measured by ρ^k is low. Ω^k is doubled if the surrogate is found to be accurate and the kth local optimum, $\mathbf{x}_{lo}^k$, lies on the trust-region bounds. Otherwise the trust-region radius remains unchanged.

The exact solutions of the tness functions at the kth local optimum are combined with the existing neighboring data points to generate new surrogate models in the subsequent trust-region iterations. The initial guess for the k+1 iteration is de ned by

$$\begin{aligned} \mathbf{x}_c^{k+1} &= \mathbf{x}_{lo}^k, && \text{if } \rho^k > 0 \\ &= \mathbf{x}_c^k, && \text{if } \rho^k \le 0. \end{aligned} \tag{11}$$

The trust-region process for an individual terminates when the maximum number of trust-region iterations permissible, k_{max}, and con gurable by the user is reached. Lamarckian learning then proceeds if the k_{max} local optimum solution obtained is an improvement over that of the initial individual.

C. Convergence Criteria

Using the hierarchical surrogate-assisted evolutionary optimization framework in Figure 2, a convergence criteria has to be de ned to decide when the global surrogate model must be updated with new sample points obtained using the exact tness function. The convergence criteria introduced in [2] is used in the present study and convergence is assumed to occur when there are zero improvements in the solution over the last Δ generations, where Δ is another user-de ned value. Hence, if no improvements in the solution can be found after Δ generations, the subsequent population of individuals will be evaluated based on the exact tness function and the global surrogate model is updated with these new design points. Otherwise, the search continues with the former global surrogate model.

Note that apart from the parameters used in the surrogate-assisted evolutionary optimization framework in [1], the proposed hierarchical optimization framework under study has two additional user-speci ed parameters: σ and Δ.

IV. Empirical Results

In this section, we present numerical studies on two multi-modal benchmark test functions to investigate the convergence properties of the Hierarchical Surrogate-Assisted Evolutionary Optimization Framework (HSAGA). We employed a standard GA with population size of 50, uniform crossover and mutation operators at probabilities 0.6 and 0.001, respectively. A linear ranking algorithm is used for selection. Besides the standard GA, a performance comparison to our original Surrogate-Assisted Evolutionary Optimization Framework (SAGA) [1] is also considered in the present study. The SAGA adopts the same parameter con gurations as the standard GA. However, apart from the standard GA settings, the two user-speci ed parameters of the SAGA are, 1) number of nearest neighboring data points used to construct the local surrogate model, m and 2) maximum trust region iterations k_{max} are con gured to be 100 and 3, respectively.

A. Rastrigin Test Function

The rst benchmark problem adopted in this study involves minimizing the Rastrigin function [18]:

$$\begin{gathered} f(x) = 10n + \textstyle\sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i)) \\ -5.12 \le x_i \le 5.12, i = 1, 2, \ldots, n. \end{gathered} \tag{12}$$

Figure 3 shows a plot of the Rastrigin test function for 2 dimension. The function is highly multi-modal, having

BEGIN
Initialize: Generate a database containing a population of designs.
Construct global Kriging model using all available design points in the database.
Set global tness function := Global surrogate model
While(*computational budget not exhausted*)

- Evaluate all individuals in the population using the global tness function.
- **For** each non-duplicated top ranking σ percent individuals in the population,
  - ∗ Apply trust-region enabled FSQP solver to the individual by interleaving the exact tness function and RBF local surrogate model for the local tness function.
  - ∗ Update the database with any new design points generated during the trust-region iterations together with their exact tness values.
  - ∗ 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.
- **If** (global tness function := Exact tness function)
  - * Update database with any new designs generated using the exact model.
  - * Update the global surrogate model using all design points in the database (note that at this point, the database contains both the previous and new design points).

  End If
- **If** (convergence over global surrogate model)
  global tness function := Exact tness function
  Else
  global tness function := Global surrogate model
  End If

End While
END

Fig. 2. Hierarchical Surrogate-Assisted Evolutionary Optimization Framework.

many local minima surrounding the global minimum. It is a separable function. A twenty dimensional (n = 20) version of the test function is used here in the numerical studies.

B. Ackley Test Function

The second benchmark problem adopted in this study is minimization of the Ackley function [19]:

$$f(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n} \cos 2\pi x_i} \quad (13)$$
$$-32.768 \leq x_i \leq 32.768, i = 1, 2, \ldots, n.$$

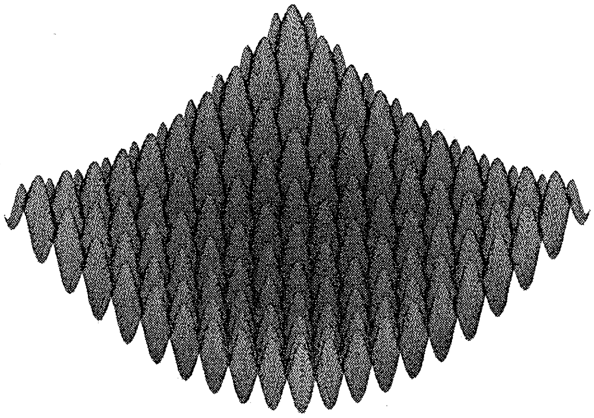

Fig. 3. Rastrigin benchmark test function of 2 dimension

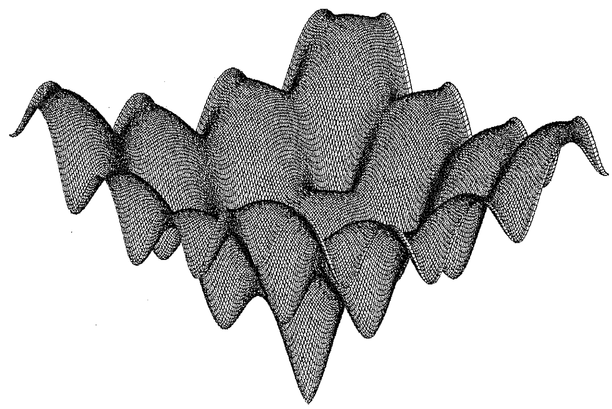

Fig. 4. Ackley benchmark test function of 2 dimension

Figure 4 shows a plot of the Ackley test function for 2 dimension. It is also a highly multi-modal function with many local minima and a global minimum located at (0,...,0). A 20 dimensional (n = 20) version of the function is used in the present study. It has a very rugged landscape and is dif cult to search for most optimizers.

The effects of (σ) and (Δ) parameters on the convergence trends of HSAGA are presented in Figures 5-10. Figures 5 and 6 present the convergence trends of the HSAGA framework on the 20-variable Rastrigin and Ackley test functions, respectively, for three different con gurations on the top percentage ranking individuals with σN_{pop}=1 (i.e., representing the elitist strategy), σ of 40% and 80%, but with the convergence criterion, Δ kept constant at 4. N_{pop} is the EA population size. The results indicate that the setting of σ signi cantly affects the performance of the HSAGA. It is shown that the HSAGA tends to stall much earlier on both test problems when only the elite individual is permitted to undergo the Lamarckian learning process (see trace HSAGA-Elitist in Figure 5 and Figure 6).

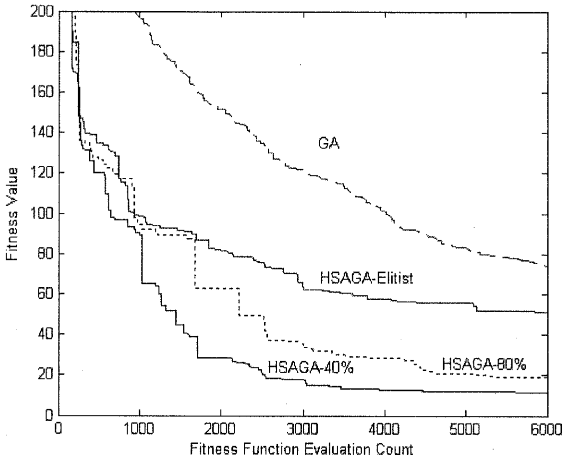

Fig. 5. Convergence trends of GA and HSAGA with various percent individuals σ = Elitist, 40%, 80% and $\Delta = 4$ for Rastrigin function

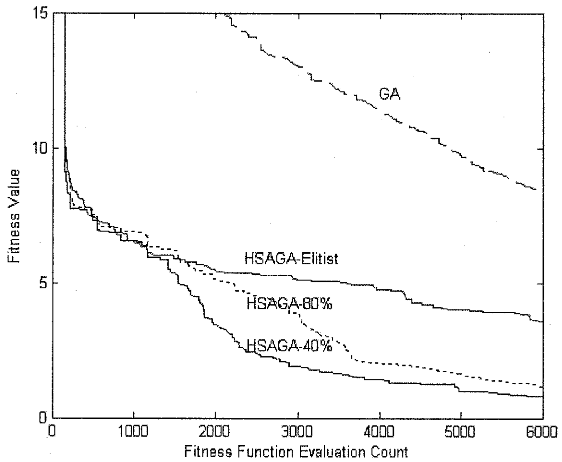

Fig. 6. Convergence trends of GA and HSAGA with various percent individuals σ = Elitist, 40%, 80% and $\Delta = 4$ for Ackley function

A generous σ value of 80% would tend to reduce the stalling effect but may however lead to a lower search ef ciency. Note that if σ is set at 100%, the HSAGA becomes similar to the original SAGA proposed in [1]. In our numerical studies, we found that the HSAGA improves the search ef ciency and converges to good designs on both multi-modal functions can be obtained when σ is con gured to either 40% or 80%.

We also studied the effects of varying Δ on the convergence behavior of the HSAGA framework. Figures 7 and 8 show the convergence trends of the HSAGA framework for three con gurations of Δ at 1, 2 and 3, but this time σ is kept constant at 40% on both the Rastrigin and Ackley functions. From these results, it appears that appropriate settings of Δ would affect signi cantly on the abilities of the HSAGA in attaining high search ef ciency towards good quality designs, even though the Ackley function is less affected. A larger value of Δ tends to reduce the number of calls to the computationally expensive exact tness function and Kriging approximation, and hence faster convergence. However, the search could stall easily due to the inaccuracy of the global surrogate model, leading to convergence to false optima. On the other hand, a small value of Δ may result in performing the global approximation too frequently. This may cause a problem as the computational cost of the Kriging approximation scales as $\mathcal{O}(n^3)$. Nevertheless, our empirical studies shows that the HSAGA was able to generate better search performance when Δ is con gured as 2 on both multi-modal test problems.

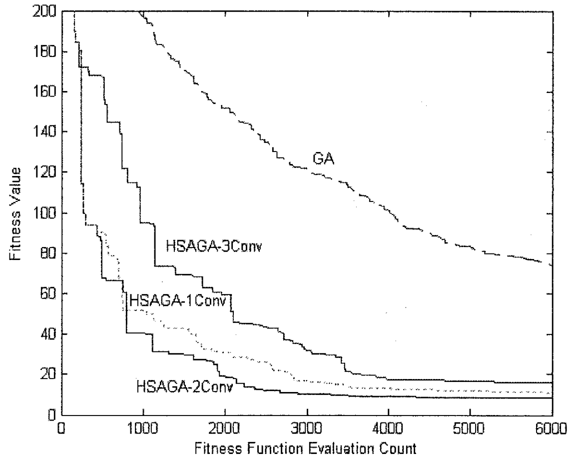

Fig. 7. Convergence trends of GA and HSAGA with convergence criteria Δ = 1, 2, 3 and σ = 40% for Rastrigin function

With respect to the standard GA, the results clearly show that the HSAGA is capable of converging to good designs at a signi cantly lower computational budget. In addition, the convergence histories of the HSAGA framework in comparison to the original SAGA framework in [1] are summarized in Figures 9 and 10 for both test functions.The results indicate that HSAGA can bring about signi cant improvements over the SAGA when the two additional user-parameters of the HSAGA are properly con gured. Since Lamarckian learning is conducted only on the promising GA individuals among the entire population in the HSAGA framework, it is reasonable to expect that the HSAGA can perform better than SAGA. In other words, the HSAGA attempts to accelerate the evolutionary optimization process by reducing the total number of exact tness function calls.

V. Conclusion

In this paper, we presented a Hierarchical Surrogate-Assisted Evolutionary Optimization Framework(HSAGA) that integrates both global and local surrogate models to accelerate evolutionary search. Experimental studies are presented

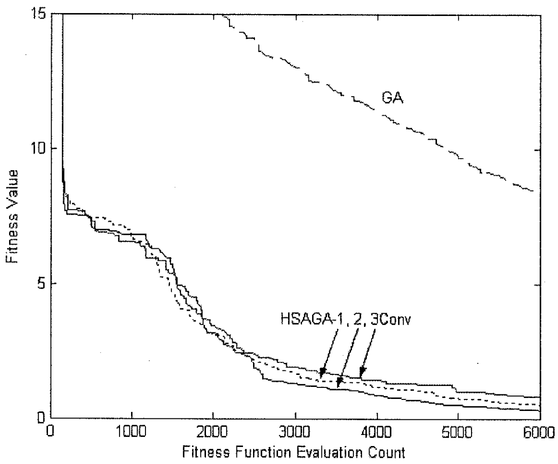

Fig. 8. Convergence trends of GA and HSAGA with convergence criteria Δ = 1, 2, 3 and σ = 40% for Ackley function

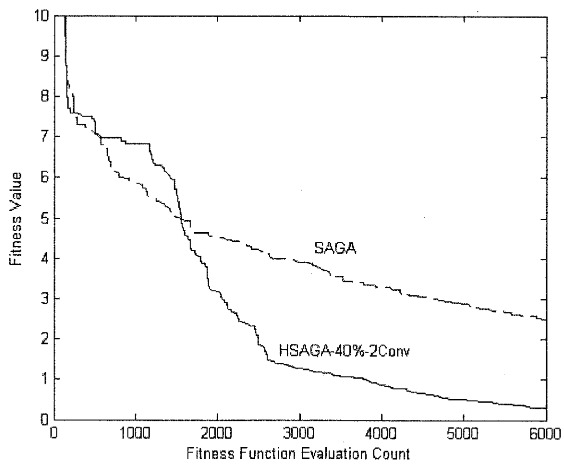

Fig. 10. Convergence trends of the SAGA and HSAGA framework with σ = 40 and Δ = 2 for Ackley function

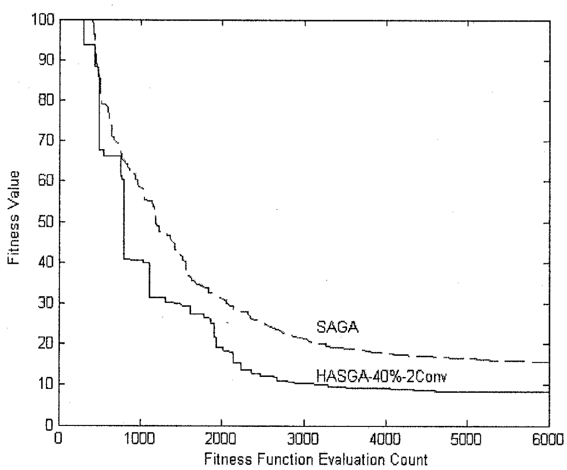

Fig. 9. Convergence trends of SAGA and HSAGA with σ = 40 and Δ = 2 for Rastrigin function

for the Rastrigin and Ackley multi-modal benchmark test functions with various con gurations of the two additional user-speci ed parameters introduced in the hierarchical optimization framework. The empirical results were compared with those obtained using a standard GA and a Surrogate-Assisted GA (SAGA) proposed earlier in the literature [1]. The results obtained suggest that the proposed hierarchical framework is capable of solving computationally expensive optimization problems more ef ciently than both the standard GA and SAGA under a limited computational budget.

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.

REFERENCES

[1] Y. S. Ong, P. B. Nair, 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] A. Ratle, "Accelerating Evolutionary Algorithms Using Fitness Function Models," *Parallel Problem Solving from Nature*, Vol. 5, pp. 87-96, 1998.

[3] M. A. El-Beltagy, P. B. Nair, A. J. Keane, "Metamodelling Techniques For Evolutionary Optimization of Computationally Expensive Problems: Promise and Limitations," *Proceeding of the Genetic and Evolutionary Compution Conference (GECCO99)*, pp. 196-203, 1999.

[4] 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.

[5] K. C. Giannakoglou "Design of Optimal Aerodynamic Shapes using Stochastic Optimization Methods and Computational Intelligence," *International Review Journal Progress in Aerospace Sciences*, Vol. 38, pp. 43-76, 2002.

[6] T. Back and H.-P. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation*, Vol. 1, No. 1, pp. 1-23, 1993.

[7] A. Ratle, "Kriging as a Surrogate Fitness Landscape in Evolutionary Optimization," *Arti cial Intelligence for Engineering Design Analysis and Manufacturing*, Vol. 15, No. 1, pp. 37-49, 2001.

[8] W. B. Song, "Shape Optimisation Of Turbine Blade Firtrees," *PhD Thesis*, University of Southampton, 2002.

[9] 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.

[10] 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.

[11] 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," *Congress on Evolutionary Computation, Special Session on Design Optimisation with Evolutionary Computation*, Canberra, Australia, 2003.

[12] Y. S. Ong, P. B. Nair, A. J. Keane and K. W. Wong, "Surrogate-Assisted Evolutionary Optimization Frameworks for High-Fidelity Engineering Design Problems", to appear in Knowledge Incorporation in Evolutionary Computation, editor: Y. Jin, *Studies in Fuzziness and Soft Computing Series*, Springer Verlag, 2004.

[13] D. R. Jones, M. Schonlau, W. J. Welch, "Ef cient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, pp. 455-492, 1998.
[14] A. Choudhury, P. B. Nair, and A. J. Keane, "A Data Parallel Approach for Large-Scale Gaussian Process Modeling," *Proceedings of the Second SIAM International Conference on Data Mining*, Arlington, VA, 2002.
[15] C. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995.
[16] J. F. Rodriguez, J. E. Renaud,and L. T. Watson, "Convergence of Trust Region Augmented Lagrangian Methods Using Variable Fidelity Approximation Data," *Structural Optimization*, Vol. 15, No. 3-4, pp. 141-156, 1998.
[17] N. Alexandrov, J. E. Dennis, R. M. Lewis, and V. Torczon, "A Trust Region Framework for Managing the use of Approximation Models in Optimization," *Structural Optimization*, Vol. 15, No. 1, pp. 16-23, 1998.
[18] H. Miihlenbein, M. Schomisch, and J. Born, "The parallel genetic algorithm as function optimizer," *Parallel Computing*, Vol 17, pp. 619-632, 1991.
[19] T. Back, "Evolutionary algorithms in theory and practice", Oxford University Press, 1996.