Evolution by Adapting Surrogates

Minh Nghia Le mnle@ntu.edu.sg
School of Computer Engineering, Nanyang Technological University, 639798, Singapore

Yew Soon Ong asysong@ntu.edu.sg
School of Computer Engineering, Nanyang Technological University, 639798, Singapore

Stefan Menzel stefan.menzel@honda-ri.de
Honda Research Institute Europe GmbH, 63073 Offenbach/Main, Germany

Yaochu Jin yaochu.jin@surrey.ac.uk
Department of Computing, University of Surrey Guildford, Surrey, GU27XH United Kingdom

Bernhard Sendhoff bernhard.sendhoff@honda-ri.de
Honda Research Institute Europe GmbH, 63073 Offenbach/Main, Germany

Abstract

To deal with complex optimization problems plagued with computationally expensive fitness functions, the use of surrogates to replace the original functions within the evolutionary framework is becoming a common practice. Despite the significant research effort spent on optimizing computational expensive problems more efficiently, the performance of existing surrogate-assisted evolutionary frameworks may be fundamental bounded by the use of fitness prediction errors as the key performance indicator used to assess and select surrogates. Further, the suitability of data-centric approximation methodology used to construct the surrogate model, also depends greatly on the nature of the optimization problems of interest, namely, problem fitness landscape, state of the evolutionary search, characteristics of the search algorithm, and others. This explains why a plethora of surrogate-assisted evolutionary frameworks with ad-hoc approximation/surrogate modelling methodologies have emerged widely in the literature. In contrast to earlier works, this paper presents a novel evolutionary framework with the *Evolvability Learning* of Surrogates (EvoLS) by operating with *multiple diverse* approximation methodologies within the search. As opposed to the common use of fitness prediction error, the concept of "*evolvability*" to indicate the productivity or suitability of an approximation methodology that brings about fitness improvement in the evolutionary search is introduced as the basis for adaptation. The backbone of the proposed EvoLS is a statistical learning scheme to determine the *evolvability* of each approximation methodology while the evolutionary search progresses online. For a given individual solution, using the most productive approximation methodology inferred, i.e., the method with highest *evolvability* measure, fitness improving surrogates are subsequently constructed for use within a trust-region enabled local search strategy, leading to the self-configuration of surrogate-assisted memetic algorithm for optimizing computationally expensive problems. Numerical study of EvoLS on representative benchmark problems and a real-world computationally expensive problem of aerodynamic car rear design highlights the competitiveness of EvoLS in attaining reliable, high quality and efficient performance under limited computational budget.

1 Introduction

Engineering reliable and high quality products is now becoming an important practice of many industries to stay competitive in today's increasingly global economy, which is constantly exposed to high commercial pressures. A strong engineering design know-how results in lower time to market and better quality at lower cost. In the field of computer science and mathematical programming, the search process of seeking for improved or best alternative solution from a number of possible solutions to a problem is generally referred to as optimization, which is now part and parcel of problem-solving in many areas of arts, business & finance, design, science and engineering, including those that are directly applicable in our daily life. Over the years, optimization methods have evolved considerably, with many algorithms and implementations now available. Among them, modern stochastic optimization, evolutionary algorithms (EA) in particular, has gained increasing popularity and success for the past decades, due to their ease of implementation and abilities to locate high quality design solutions fast.

In recent years, advancement in science, engineering and the availability of massive computational power have led to the increasing high-fidelity approaches introduced for precise studies of complex systems *in silico*. Modern Computational Structural Mechanics, Computational Electro-Magnetics, Computational Fluid Dynamics and Quantum mechanical calculations represent some of the approaches that have been shown to be highly accurate [Jin, 2002, Zienkiewicz and Taylor, 2006, Hirsch, 2007]. These techniques play a central role in the modelling, simulation and design process since they serve as efficient and convenient alternatives for conducting trails on the original real-world complex system that are otherwise deemed to be too costly or hazardous to construct. Typically, when high-fidelity analysis codes are used, it is not uncommon for the single simulation process to take minutes, hours to days of supercomputer time to compute. A motivating example at Honda Research is aerodynamic car rear design, where one function evaluation involving a Computational Fluid Dynamics (CFD) simulation to calculate the fitness performance of a potential design can take many hours of wall clock time. Since the design cycle time of a product is directly proportional to the number of calls made to the costly analysis solvers, researchers are now seeking for novel stochastic optimization approaches, including evolutionary frameworks, that handles these forms of problems elegantly. Besides parallelism, which is an obvious choice to achieving near linear order improvement in evolutionary search, researchers are gearing towards surrogate-assisted or meta-model assisted evolutionary frameworks when handling optimization problems imbued with costly non-linear objective and constraint functions [Liang et al., 1999, Jin et al., 2002, Song, 2002, Ong et al., 2003, Jin, 2005a, Lim et al., 2007, Lim et al., 2010, Tenne, 2009, Shi and Rasheed, 2010, Voutchkov and Keane, 2010, Santana-Quintero et al., 2010].

The general consensus on surrogate-assisted evolutionary frameworks is that the efficiency of the search process can be improved by replacing as often as possible, calls to the costly high-fidelity analysis solvers with surrogates that are deem to be less costly to build and compute. In this manner, the overall computational burden of the evolutionary search can be greatly reduced since the efforts required to build the surrogates and to use them are much lower than those in the traditional approach that directly couples the evolutionary algorithm (EA) with the costly solvers [Serafini, 1998, Booker et al., 1999, Jin et al., 2000, Kim and Cho, 2002, Emmerich et al., 2002, Jin and Sendhoff, 2004, Ulmer et al., 2004, Branke and Schmidt, 2005, Zhou et al., 2005]. Among many data-centric approximation methodologies used to construct surrogates to date, polynomial regression or response surface methodology [Lesh, 1959], support vector machine [Cortes and Vapnik, 1995, Vapnik, 1998], artificial neural networks [Zurada, 1992], radial basis function [Powell, 1987], Gaussian process referred or Kriging or design and analysis of computer experiment models [Mackay, 1998] and ensemble of surrogates [Zerpa et al., 2005, Goel et al., 2007, Sanchez et al., 2008, Acar and Rais-Rohani, 2009] are among the most prominently investigated [Jin, 2005a, Zhou et al., 2005, Shi and Rasheed, 2010]. Early proposed approaches have considered using surrogates that target to model the whole solution space or fitness landscape of the costly exact objective or fitness function. However, due to the limited and sparseness of data points collected along the evolutionary search, the construction of accurate global surrogates [Ulmer et al., 2004, Buche et al., 2005] that mimics the entire problem landscape well is impeded by the effect of "curse of dimensionality" [Donoho, 2000]. To enhance the accuracies of the surrogates used, researchers have turned to localized [Giannakoglou, 2002, Emmerich et al., 2002, Ong et al., 2003, Regis and Shoemaker, 2004] as opposed to globalized models or their synergies [Zhou et al., 2005]. Others have also considered the use of gradient information [Ong et al., 2004] to enhance the prediction accuracy of the constructed surrogate models or physics-based models that are deem to be more trustworthy than the purely data-centric counterparts [Keane and Petruzzelli, 2000, Keane and Petruzzelli, 2000, Lim et al., 2008].

In the context of surrogate-assisted optimization [Jin et al., 2001, Shi and Rasheed, 2010], present performance or assessment metrics and schemes used for surrogate model selection and validation involve many prominent approaches that take roots in the field of statistical and machine learning [Fielding and Bell, 1997, Shi and Rasheed, 2010]. Particularly, existing focus have been placed on attaining surrogate model that has minimal apparent error or training error on some optimization data collected during the evolutionary search, as an estimation of the true error when used to replace the original costly high-fidelity analysis solver. Maximum/Mean Absolute Error, Root Mean Square Error (RMSE) and Correlation Measure denote some of the performance metrics that are commonly used [Jin et al., 2001]. Typical model selection schemes that stem from the field of statistical and machine learning, including the split sample (holdout) approach, cross-validation and bootstrapping, are subsequently used to choose surrogate models that have low estimation of apparent and true errors [Queipo et al., 2005, Tenne and Armfield, 2008a]. [Tenne and Armfield, 2008b] used the multiple cross-validation scheme for the selection of low-error surrogates that replace the original costly high-fidelity analysis solver to avoid convergence at false optima of poor accuracy models.

Despite the significant research effort spent on optimizing computational expensive problems more efficiently, existing surrogate-assisted evolutionary frameworks may be fundamental bounded by the use of apparent or training error as the performance metric used to assess and select surrogates. Clearly, a surrogate having zero prediction error although contributes beneficially as a cost-efficient replacement of the original computational expensive problem in the optimization process, it however does not guarantee fitness improvement in the evolutionary search [Lim et al., 2010]. In contrast, it would be more worthwhile to assess and select surrogates that enhances search improvement in the context of optimization, as opposed to the usual practice of choosing surrogate model with estimated minimal true error. Further, the influence of the data-centric approximation methodology employed is deemed to have a major impact on surrogate-assisted evolutionary search performance. The varied suitability of approximation methodology to different fitness landscape, state of the search, and characteristics of the search algorithm suggests why a variety of surrogate-assisted evolutionary frameworks in the literature have emerged with ad-hoc approximation methodologies. To the best of our knowledge, little work has been done to mitigate this issue since only limited knowledge of the "black-box" optimization problem is available before one starts.

Falling back on the basics of Darwin's grand idea of "Natural Selection" or "Survival of the fittest" as the criterion for the choice of surrogates that brings about fitness improvement to the search, this paper describes a novel evolutionary search process that evolves with *fitness improving* surrogates. Here, we focus our study on the *Evolvability Learning* of Surrogates (EvoLS), particularly the adaptive choice of data-centric approximation methodologies that build fitness improving surrogates in place of the original computationally expensive "black-box" problem, during the evolutionary search. EvoLS infers the fitness improvement contribution of each approximation methodology towards the search, which is here referred to as *evolvability* measure. Hence for each individual or design solution in the evolutionary search, the *evolvability* of each approximation methodology is determined statistically, according to the current state of the search, properties of the search operators and characteristics of the fitness landscape, while the search progress online. Based on the *evolvability* measures derived, the search adapts by using the most productive approximation methodology inferred for the construction of surrogates within a trust-region enabled local search strategy, leading to the self-configurations of surrogate-assisted memetic algorithm that deals with complex optimization of computationally expensive problems effectively.

The paper is outlined as follows: Section 2 discusses the roles of surrogate models in evolutionary optimization and briefly describes the data-centric approximation methodologies considered in the present study. Subsequently, the notion of *evolvability* as a performance or assessment measure that expresses the suitability of an approximation methodology in guiding towards improved evolutionary search and the essential ingredients of our proposed EvoLS, are introduced in Section 3. Section 4 presents a numerical study of EvoLS on representative benchmark problems. Detailed analyses on the suitability and cooperation of surrogates in search, as well as the correlation between the surrogate models' estimated *fitness prediction error* and *evolvability* in EvoLS, are also presented in the section. Section 5 then showcases the real world application of EvoLS on an aerodynamic car rear design that involves highly computationally expensive CFD simulations. Finally, Section 6 summarizes the present study with a brief discussion and concludes the paper.

2 Surrogate Modelling

Data-centric surrogates are (statistical) models that are built to approximate computationally expensive simulation codes or the exact fitness evaluations. They are orders of magnitude cheaper to run and can be used in lieu of the exact analysis during evolutionary search. Let $\{(\mathbf{x}_i, t_i)\}_{i=1}^m$ where $t_i = f(\mathbf{x}_i)$ denote the training dataset, where $\mathbf{x} \in R^n$ is the input vector of scalars or design parameters, and $f(\mathbf{x}) \in R$ is the output or exact fitness value. Based on the approximation methodology M, the constructed surrogate model $\widehat{f}_M(\mathbf{x})$ is an approximation of $f(\mathbf{x})$ such that the estimation error $|f(\mathbf{x}) - \widehat{f}_M(\mathbf{x})|$ is minimal. Further, the surrogate model can also yield insights into the functional relationship between the input $\mathbf{x}$ and the objective function $f(\mathbf{x})$.

There exist a variety of approximation methodologies for constructing surrogate models that take its roots from the field of statistical and machine learning field [Jin, 2005b, Shi and Rasheed, 2010]. One popular approach in the design optimization literature is polynomial regression or response surface methodology [Lesh, 1959]. Neural networks which have shown to be effective tools for function approximation have also been employed as surrogates extensively in evolutionary optimization. These include support vector machine [Cortes and Vapnik, 1995, Vapnik, 1998], artificial neural networks [Zurada, 1992] and radial basis function [Powell, 1987]. A statistically sound alternative is Gaussian process or Kriging [Mackay, 1998], referred to as design and analysis of computer experiment models. Next, we provide a brief overview on three different approximation methods used in the paper, namely, polynomial regression (PR), radial basic function (RBF) and Gaussian process (GP).

2.1 Polynomial Regression

The most widely used polynomial regression model is the quadratic model which takes the form

$$\widehat{f}_M(\mathbf{x}) = \beta_0 + \sum_{i=1}^{n} \beta_i \mathbf{x}^{(i)} + \sum_{1 \le i \le j \le n} \beta_{n-1+i+j} \mathbf{x}^{(i)} \mathbf{x}^{(j)} \quad (1)$$

where n is the number of input variables, $\mathbf{x}^{(i)}$ is the i-th component of $\mathbf{x}$, and β_i are the coefficients to be estimated. As the number of terms in the quadratic model is $n_t = (n+1)(n+2)/2$

in total, the number of training sample points should be at least n_t for proper estimation of the unknown coefficients, by means of either least square or gradient-based methods [Jin, 2005a].

2.2 Radial Basic Function

The surrogate models in this category are interpolating radial basic function (RBF) networks of the form

$$\widehat{f}_M(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i K(\| \mathbf{x} - \mathbf{x}_i \|) \tag{2}$$

where $K(\| \mathbf{x} - \mathbf{x}_i \|) : \mathbb{R}^d \to \mathbb{R}$ is a RBF and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T \in \mathbb{R}^m$ denotes the vector of weights. The number of hidden nodes used in interpolating RBF is often assumed as equal to the number of training vector points.

Typical choices for the kernel function include linear splines, cubic splines, multiquadrics, thin-plate splines and Gaussian functions [Bishop, 1996]. Given a suitable kernel, the weight vector $\boldsymbol{\alpha}$ can be computed by solving the linear algebraic system of equations

$$\mathbf{K}\boldsymbol{\alpha} = \mathbf{t}$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_m]^T \in \mathbb{R}^m$ denotes the vector of outputs and $\mathbf{K} \in \mathbb{R}^{m \times m}$ denotes the Gram matrix formed using the training inputs (i.e., the *ij*-th element of $\mathbf{K}$ is computed as $K(\| \mathbf{x}_i - \mathbf{x}_j \|)$).

2.3 Kriging/Gaussian Process

The Kriging model or Gaussian Process (GP) assumes the presence of a global model $g(\mathbf{x})$ and an additive noise term $Z(\mathbf{x})$ in the original function.

$$f(\mathbf{x}) = g(\mathbf{x}) + Z(\mathbf{x})$$

where $g(\mathbf{x})$ is a known function of $\mathbf{x}$ as a global model of the original function, and $Z(\mathbf{x})$ is a Gaussian random function with zero mean and non-zero covariance that represents a localized noise or deviation from the global model. Usually, $g(\mathbf{x})$ is a polynomial and in many cases, it is reduced to a constant β. The approximation model of $f(\mathbf{x})$, given the m samples and the current input $\mathbf{x}$, is defined as:

$$\widehat{f}_M(\mathbf{x}) = \widehat{\beta} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{t} - \widehat{\beta}\mathbf{I}) \tag{3}$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_m]^T$, $\mathbf{I}$ is a unit vector of length m, and $\mathbf{R}$ is the correlation matrix which can be obtained by computing the correlation function between any two samples, i.e., $\mathbf{R}_{i,j} = R(\mathbf{x}_i, \mathbf{x}_j)$. While the correlation function can be specified by the user, Gaussian exponential correlation function, defined by correlation parameters $\{\theta_k\}_{k=1}^{n}$, has often been used:

$$R(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\sum_{k=1}^{n} \theta_k |\mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)}|^2 \right]$$

where $\mathbf{x}_i^{(k)}$ and $\mathbf{x}_j^{(k)}$ are the k-th component of sample points $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. $\mathbf{r}$ is the correlation vector of length m between the given input $\mathbf{x}$ and the samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, i.e., $\mathbf{r} = [R(\mathbf{x}, \mathbf{x}_1), R(\mathbf{x}, \mathbf{x}_2), \ldots, R(\mathbf{x}, \mathbf{x}_m)]^T$.

The estimation of the unknown parameters β and $\{\theta_k\}_{k=1}^{n}$ can be carried out using the maximum likelihood method [Shi and Rasheed, 2010]. Aside from the approximation values, Kriging model or Gaussian process can also provide a confidence interval without much additional computational cost incurred. However, one main disadvantage of Gaussian process is the significant increasing of computational expense when the dimensionality becomes high, due to the matrix inversions in the estimation of parameters.

3 Proposed Framework: Evolvability Learning of Surrogates

In this section we present the essential ingredients of the proposed *Evolvability Learning* of Surrogates (EvoLS) for handling computationally expensive optimization problems. In particular, we concentrate on the general nonlinear programming problem of the following form:

$$\begin{aligned} \textbf{Minimize}: &\quad f(\mathbf{x}) \\ \textbf{Subject to}: &\quad \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of design variables, and $\mathbf{x}_l$, $\mathbf{x}_u$ are vectors of lower and upper bounds, respectively. In this paper, we are interested in cases where the evaluation of $f(\mathbf{x})$ is computationally expensive, and it is desired to obtain a near-optimal solution on a limited computational budget, using a novel evolutionary process that adapts fitness improving surrogates, $\widehat{f}_M(\mathbf{x})$ as a replacement of $f(\mathbf{x})$, in the search.

In what follows, we begin with a formal introduction on the notion of *evolvability* as a performance or assessment measure to indicate the productivity and suitability of an approximation methodology for constructing surrogate that brings about fitness improvement to the evolutionary search (see Section 3.1). The essential backbone of our proposed *Evolvability Learning* of surrogates framework is a evolutionary algorithm coupled with a trust-region enabled local search strategy with adaptive surrogates, in the spirit of Lamarckian learning. In contrast to existing works, we adapt the choice of approximation methodology for the construction of fitness improving data-centric surrogates in place of the original computationally expensive "black-box" problem when conducting the computationally intensive local search in the context of memetic optimization [Hart et al., 2004, Krasnogor and Smith, 2005] (see Section 3.2.).

3.1 Evolvability of Surrogate

Conventionally, surrogate models are assessed and chosen according to their estimated true error, $|f(\mathbf{x}) - \widehat{f}_M(\mathbf{x})|$, where $\widehat{f}_M(\mathbf{x})$ denotes the predicted fitness value of input vector $\mathbf{x}$ by a surrogate constructed using approximation method M. In contrast to existing surrogate-assisted evolutionary search, the surrogate model employed for each individual design solution in the present study, favors fitness improvement as the choice of merit to assess the usefulness of surrogates in enhancing search improvement, as opposed to minimal estimated true error.

In this subsection, we introduce the concept on "*Evolvability*" of an approximation methodology as the basis for adaptation. Since the term "*Evolvability*" has been used in different contexts[1], it is worth highlighting that here our concept of *evolvability* generalizes from that of learnability in machine learning [Valiant, 2009] where an evolutionary process is regarded as "*evolvable*" on a given optimization problem if the progress in search performance is observed for some moderate number of generations. Hence *evolvability* of an approximation methodology is referred here to the propensity of the method in constructing surrogate model that guides towards viable, or "potentially favorable" individuals that leads towards the global optimum.

In particular, the *evolvability* measure of an approximation methodology M for the construction of fitness improving data-centric surrogate on individual solution $\mathbf{x}$ at generation t, assuming a minimization problem, is denoted here as $Ev_M(\mathbf{x})$ and derived in the form of

$$\begin{aligned} Ev_M(\mathbf{x}) &= Exp[f(\mathbf{x}) - f(\mathbf{y}^{opt})|\mathbf{P}^t, \mathbf{x}] \\ &= f(\mathbf{x}) - \int_y f(\varphi_M(\mathbf{y})) \times P(\mathbf{y}|\mathbf{P}^t, \mathbf{x}) d\mathbf{y} \end{aligned} \quad (4)$$

[1]In [Wagner and Altenberg, 1996], "evolvability" is defined as the genome's ability to produce adaptive variants when acted upon by the genetic system. Others have generally referred the term to the ability of stochastic or random variations to produce improvement for adaptation to happen [Ong et al., 2006a].

Here $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$ denotes the density function of the stochastic variation operators applied on parent $\mathbf{x}$ to arrive at solution $\mathbf{y}$ at generation t and $\varphi_M(\mathbf{y})$ represents the local search strategy operating on the surrogate constructed by approximation method M. The *evolvability* measure of an approximation methodology indicates the *expectation* of fitness improvement which the refined offspring, denoted here as $\mathbf{y}^{opt} = \varphi_M(\mathbf{y})$, has gained over its parent, upon undergoing local search on the respective constructed surrogate. A high *evolvability* measure encapsulates two core essences of existing surrogate-assisted evolutionary search: 1) When a surrogate exhibits low true error estimates, fitness improvement on the refined offspring $\mathbf{y}^{opt}$ over initial parent $\mathbf{x}$ can be expected and 2) When a surrogate exhibits high true error estimates, the discovery of offspring solutions with improved fitness $\mathbf{y}^{opt}$ of $\mathbf{x}$ can be still attained due to the effect of "bless of uncertainty" [Ong et al., 2006b] (see Fig. 1 for an example illustration).

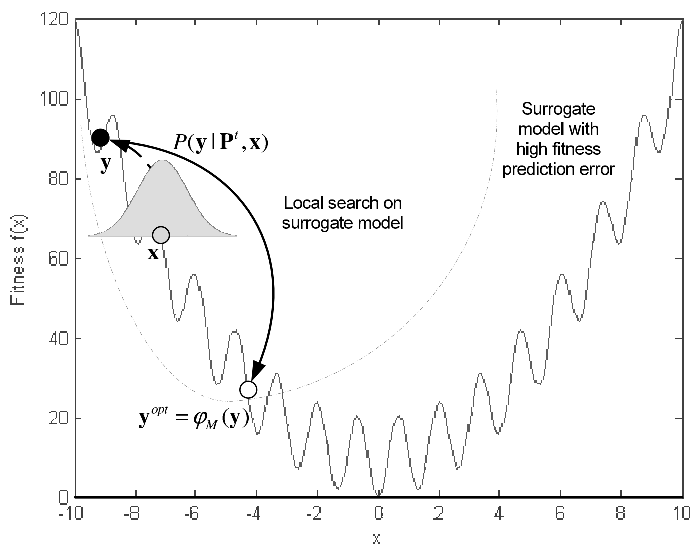

Figure 1: Illustration of *evolvability* under the effect of "bless of uncertainty"

Taking into account the current state of the evolutionary optimization search, properties of the search operators, and characteristics of the fitness landscape, a statistical learning approach to estimate the *evolvability* measure $Ev_M(\mathbf{x})$ of each approximation methodology (as defined in Eqn. (4)) for use on a given individual solution $\mathbf{x}$ at generation t, is proposed. Let $\Phi_M = \{(\mathbf{y}_i, \varphi_M(\mathbf{y}_i))\}_{i=1}^K$ denote the database of distinct samples archived along the search which represents the historical contribution of the approximation methodology on the problem considered. Through a weighted sampling approach, the weight $w_i(\mathbf{x})$ that defines the probability of choosing a sample $(\mathbf{y}_i, \varphi_M(\mathbf{y}_i))$ for the estimation of $Ev_M(\mathbf{x})$, or the relevancy of the sample for the *evolvability* learning process is first derived. Considering $\{(\mathbf{y}_i, \varphi_M(\mathbf{y}_i))\}_{i=1}^K$ as distinct samples from current distribution $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$, the weights $w_i(\mathbf{x})$ associated with samples $(\mathbf{y}_i, \varphi_M(\mathbf{y}_i))$ satisfy the equations: $\sum w_i(\mathbf{x}) = 1$ and $w_i(\mathbf{x}) \sim \int_{V(\mathbf{y}_i)} P(\mathbf{y}_i|\mathbf{P}^t, \mathbf{x})d\mathbf{y}$ (i.e., proportional to), in which $V(\mathbf{y}_i)$ denotes the arbitrarily small bin around solution $\mathbf{y}_i$. Note that the conditional density function $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$ is modeled probabilistically based on the properties of the evolutionary variation operators used to reflect the current state of the search. Since the integration $\int_{V(\mathbf{y}_i)} P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})d\mathbf{y}$ is computationally intensive, weight $w_i(\mathbf{x})$ is efficiently estimated here as follows:

$$w_i(\mathbf{x}) = \frac{P(\mathbf{y}_i|\mathbf{P}^t, \mathbf{x})}{\sum P(\mathbf{y}_i|\mathbf{P}^t, \mathbf{x})} \tag{5}$$

Using the archived samples in $\Phi_M = \{(\mathbf{y}_i, \varphi_M(\mathbf{y}_i))\}_{i=1}^{K}$ and weights w_i obtained using Eqn. (5), $Ev_M(\mathbf{x})$ is estimated as follows:

$$Ev_M(\mathbf{x}) = f(\mathbf{x}) - \sum f(\varphi_M(\mathbf{y}_i)) \times w_i(\mathbf{x}) \tag{6}$$

3.2 Evolution with Adapting Surrogates

The proposed *Evolvability Learning* of Surrogates (EvoLS) for solving computationally expensive optimization problems is presented and outlined in Fig. 2. The essential ingredients of our proposed EvoLS framework, are composed of *multiple* data-centric approximation methodologies having diverse characteristics[2], denoted here as $\{M_{id}\}_{id=1}^{ID}$. In the first step, a population of N individuals is initialized either randomly or using design of experiment techniques such as Latin hypercube sampling. The cost or fitness value of each individuals in the population is then determined using $f(\mathbf{x})$. The evaluated population then undergoes natural selection, for instance, via fitness-proportional or tournament selection. Each individual $\mathbf{x}$ is evolved to arrive at the offspring $\mathbf{y}$ using stochastic variation operators including crossover and mutation. Subsequently, with ample design points in the database Ψ or after some predefined database building phase of generations G_{db}, the trust-region enabled local search with adaptive surrogates kicks in for each non-duplicated design point or individuals in the population. For a given individual solution $\mathbf{x}$ at generation t, the local search strategy proposed here embeds the *evolvability* learning formulations derived in Section 3.1, where the *evolvability* of each data-centric approximation methodology $Ev_{M_{id}}(\mathbf{x})$ is estimated statistically by taking into account the current state of the search, properties of search operators and characteristics of the fitness landscape via the historical contribution by the respective constructed surrogates, while the search progresses online. Without loss of generality, in the event of a minimization problem, the most productive data-centric approximation methodology, which is deemed as one that has the highest estimated *evolvability* measure $\arg max Ev_{M_{id}}(\mathbf{x})$, is then chosen to construct a surrogate that will be used by the local search to bring about fitness improvement on individual $\mathbf{x}$.

The outline of the approximation methodology selection process is detailed in Algorithm 1. Nearest sampled points to $\mathbf{y}$ in the database Ψ are selected as training dataset $T = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^{m}$ for building fitness improving surrogate $\widehat{f}_{M_{id}}$. The improved solution found using the respective constructed surrogate, denoted here as $\mathbf{y}^{opt} = \varphi_{M_{id}}(\mathbf{y})$, is subsequently evaluated using the original computational expensive fitness function $f(\mathbf{x})$ and replaces the parent individual in the population, in the spirit of Lamarckian learning. Exact evaluations of all newly found individuals $\{(\mathbf{y}^{opt}, f(\mathbf{y}^{opt}))\}$, together with $\{(\mathbf{y}, \varphi_{M_{id}}(\mathbf{y}))\}$ are then archived into the database Ψ and $\Phi_{M_{id}}$, respectively. The entire process repeats until the specified stopping criteria are satisfied.

3.3 Complexity Analysis and Parameters of EvoLS Framework

Here we comment on the computational complexity of present conventional surrogate selection schemes that take roots in the fields of statistical and machine learning [Fielding and Bell, 1997, Queipo et al., 2005, Tenne and Armfield, 2008b]. In conventional surrogate selection schemes, multiple sets of sample data are generally segregated, typically into training and test sets. For each approximation methodology, the respective surrogate model is commonly constructed

[2]Typical approximation techniques are Radial Basic Function (RBF), Kriging or Gaussian process (GP) and Polynomial Regression (PR).

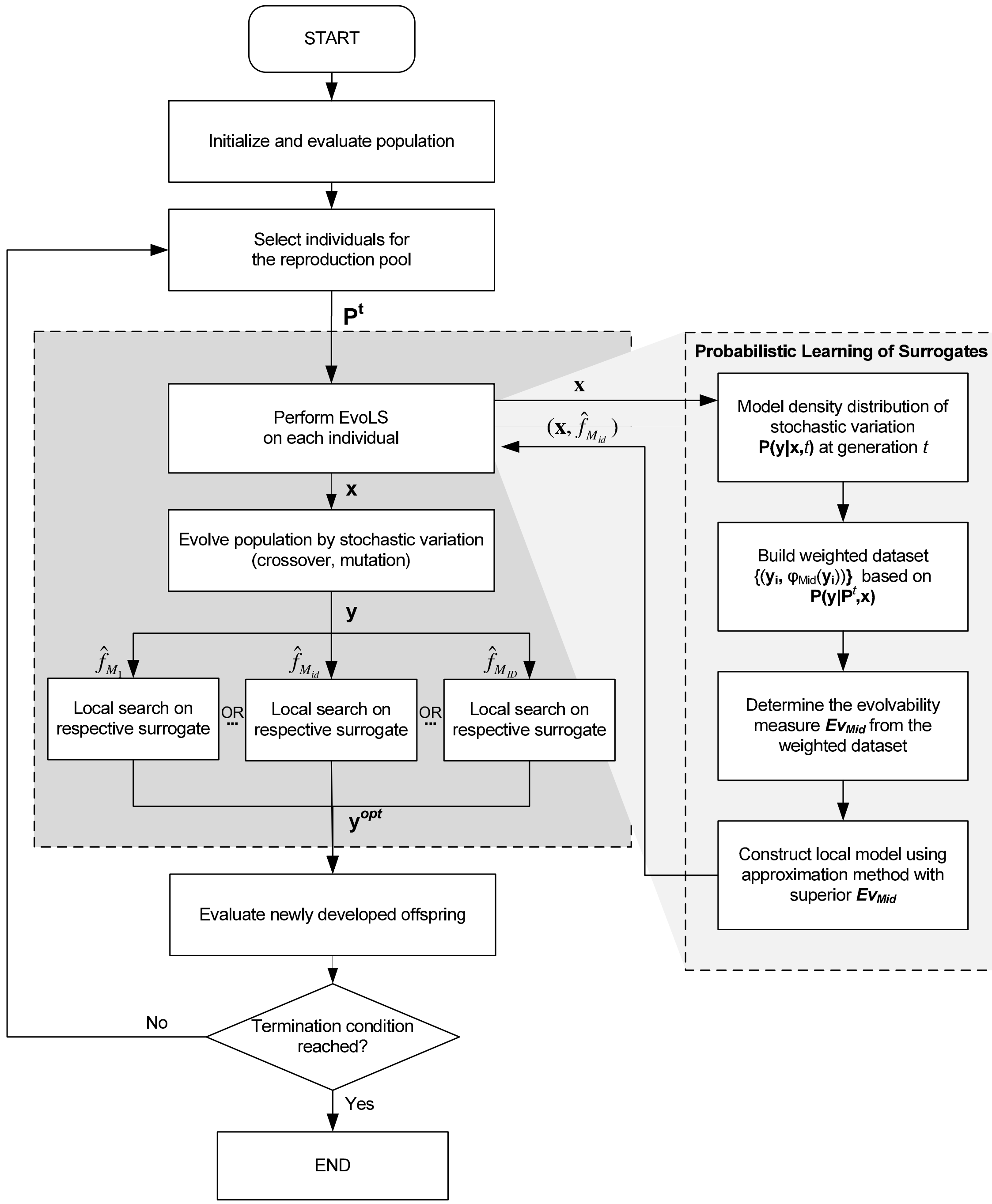

Figure 2: Evolution by Adapting Surrogates

Algorithm 1 Probabilistic Evolvability Learning of Surrogates

Construct density distribution $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$ of variation operators
for each approximation methodology M_{id} **do**
    Query archived data $\Phi_{M_{id}} = \{(\mathbf{y}_j, \varphi_{M_{id}}(\mathbf{y}_j))\}$ for M_{id}
    Calculate weight $w_i(\mathbf{x}) = P(\mathbf{y}_i|\mathbf{P}^t, \mathbf{x})$ for each sample $\mathbf{y}_i$
    if $\sum w_i(\mathbf{x}) < \epsilon$ **then**
        $w_i(\mathbf{x}) = 0$ {No relevant data available}
        $Ev_{M_{id}}(\mathbf{x}) = -\infty$
    else
        Normalize $w_i = w_i(\mathbf{x}) / \sum w_i(\mathbf{x})$
        $Ev_{M_{id}}(\mathbf{x}) = f(\mathbf{x}) - \sum f(\varphi_{M_{id}}(\mathbf{y}_i)) \times w_i(\mathbf{x})$ (Eqn. (6))
    end if
end for
if $Ev_{M_{id}}(\mathbf{x}) < 0 \; \forall \; M_{id}$ **then**
    Select approximation methodology randomly
else
    Select approximation methodology with highest $Ev_{M_{id}}(\mathbf{x})$ for $\mathbf{x}$
end if

based on the training set and the true error estimated using the test set in the prediction process. This procedure of computation cost C_M is typically repeated for k times on different training and test sets to arrive at a statistically sound estimation of the approximation error that is then used in the selection scheme. Although many error estimation approaches are in abundance, the major differences lie mainly on how the training and test sets are generated, which vary from random subsampling (holdout), k-fold cross-validation and bootstraping as described in [Kohavi, 1995]. For ID number of approximation methodologies considered, the overall computational complexity of the conventional selection scheme in estimating the error of the surrogates can thus be derived as $O(ID \times k \times C_M)$.

Next, a complexity analysis of the EvoLS framework is detailed. Apart from the standard parameters of a typical surrogate-assisted evolutionary algorithm [Lim et al., 2010], EvoLS has two additional parameters: database $\Phi_{M_{id}}$ and density function $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$. Typically, the form of $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$ can be explicitly defined according to the stochastic operator used (as illustrated in Section 4), while databases $\Phi_{M_{id}}$ naturally follows a first-in-first-out queue structure to favor more recently archived optimization data $\{(\mathbf{y}, \varphi_{M_{id}}(\mathbf{y}))\}$. The complexity for *evolvability* learning of surrogates can be derived as $O(ID \times |\Phi_{M_{id}}| \times C_E)$ where $|\Phi_{M_{id}}|$ denotes the database size, ID denotes the number of approximation methodologies used, and C_E is the computational effort incurred to determine $P(\mathbf{y}_i|\mathbf{P}^t, \mathbf{x})$ for each $\mathbf{y}_i$. For each individual, since only the most productive approximation methodology inferred is used to construct a new surrogate at a computational requirement of C_M, the complexity of EvoLS can be derived as $O(ID \times |\Phi_{M_{id}}| \times C_E + C_M)$. Nevertheless, as $(ID \times |\Phi_{M_{id}}| \times C_E) << C_M$ in practice, the computational complexity of EvoLS becomes $O(C_M)$. Thus, EvoLS offers an alternative to the conventional selection scheme with a significantly lower complexity of $O(C_M)$ that is independent of the number of approximation methodologies considered in the framework.

4 Empirical Study

In this section, we present the numerical results obtained by the proposed EvoLS using three commonly used approximation methodologies, namely: 1) interpolating linear spline Radial Basic Function (RBF), 2) 2^{nd} order Polynomial Regression (PR) and 3) interpolating Kriging/ Gaussian Process (GP). For the details on GP, PR and RBF, the reader is referred to Section 2. Representative *30* dimensional benchmark functions considered in the present study are summarized in Table 1 while the algorithmic parameters of EvoLS are summarized in Table 2.

The stochastic variation operators considered in the present study are uniform crossover and mutation, which have been widely used in real-coded genetic evolution [Herrera et al., 1998, Herrera et al., 2003]. The stochastic variations thus impose the resultant offspring $\mathbf{y}$ to be bounded by $\min_{j=1...N}\{\mathbf{x}_j^{(i)}\}$ and $\max_{j=1...N}\{\mathbf{x}_j^{(i)}\}$ for each dimension[3], i.e., $\forall i = 1\dots n$. Hence, the density distribution can be modeled as a uniform distribution with

$$P(\mathbf{y}|\mathbf{P}^t, \mathbf{x}) = UniformDist(\mathbf{R}) = \begin{cases} \frac{1}{Vol(\mathbf{R})} & \textit{if } \mathbf{y} \in \mathbf{R} \\ 0 & \textit{otherwise} \end{cases} \tag{8}$$

where $Vol(\mathbf{R})$ denotes the hyper-volume of hyper-rectangle of bounds $\mathbf{R}$, defined as

$$\mathbf{R} = [\min_{j=1...N}\{\mathbf{x}_j^{(i)}\}, \max_{j=1...N}\{\mathbf{x}_j^{(i)}\}]_{i=1...n}$$

Note that since hyper-rectangle $\mathbf{R}$ reduces as the search progresses, the probabilistic model of the variation operators reflects well on the refinement of the search space throughout the evolution. Without loss of generality, Eqn. (8) models the density distribution $P(\mathbf{y}|\mathbf{P}^t, \mathbf{x})$ of the stochastic variations considered in the present study.

On the other hand, the local search strategy in EvoLS involves a trust-region framework [Ong et al., 2006b] with Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) method [Zhu et al., 1997] that ensures convergence to the local optimum of the exact objective function under mild assumptions [Ong et al., 2003]. More specifically, for each individual $\mathbf{y}_i$ in the population, the local search method L-BFGS-B proceeds on the inferred fitness improving surrogate model $\widehat{f}_M(\mathbf{x})$ with a sequence of trust-region sub-problems of the form:

$$\begin{aligned} \textbf{Minimize}: &\quad \widehat{f}_M(\mathbf{y} + \mathbf{y}_i^k) \\ \textbf{Subject to}: &\quad ||\mathbf{y}|| \leq \Omega^k \end{aligned}$$

where $k = 0, 1, 2, \dots k_{max}$, $\widehat{f}_M(\mathbf{y})$ denotes the approximate function corresponding to the original fitness function $f(\mathbf{y})$, $\mathbf{y}_i^k$ and Ω^k are the starting point and trust-region radius at iteration k, respectively. For each individual $\mathbf{y}_i$, the surrogate model $\widehat{f}_M(\mathbf{y})$ of the exact fitness function is created dynamically based on training data points in archived database $\Psi = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^Q$ to estimate the fitness during local search. Note that $\mathbf{y}_i^{k+1} = \arg min \widehat{f}_M(\mathbf{y} + \mathbf{y}_i^k)$ denotes the local optimum of the trust-region sub-problem at iteration k. At each kth iteration, $\mathbf{y}_i^{k+1}$ and the trust-region radius Ω^k are updated accordingly. In the present study, the resultant individual denoted here as $\mathbf{y}_i^{k_{max}} = \varphi_M(\mathbf{y}_i)$, represents the improved solution attained based on L-BFGS-B and surrogate model $\widehat{f}_M(\mathbf{x})$.

[3] If $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{y}$ denote the parents and the offspring then each locus of the offspring $\mathbf{y}$ satisfies the inequality

$$\min\{\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}\} \leq \mathbf{y}^{(i)} \leq \max\{\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}\}, \forall i = 1 \dots n \tag{7}$$

Table 1: Benchmark problems considered in the empirical study. On shifted rotated problems, note that $\mathbf{z} = \mathbf{M} \times (\mathbf{x} - \mathbf{o})$ where $\mathbf{M}$ is the rotation matrix and $\mathbf{o}$ is the shifted global optimum. Otherwise, $\mathbf{z} = \mathbf{x}$.

Function	**Benchmark test functions**	**Range of $\mathbf{x}$**	**Multi***	**Non-sep***
Ackley (F1)	$F_{Ackley} = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} z_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi z_i)}$	$[-32, 32]^n$	Yes	Yes
Griewank (F2)	$F_{Griewank} = 1 + \sum_{i=1}^{n} z_i^2/4000 - \prod_{i=1}^{n} \cos(z_i/\sqrt{i})$	$[-600, 600]^n$	Yes	Yes
Rosenbrock (F3)	$F_{Rosenbrock} = \sum_{i=1}^{n-1}(100 \times (z_{i+1} - z_i^2)^2 + (1 - z_i)^2)$	$[-2.048, 2.048]^n$	Yes	Yes
Shifted rotated Rastrigin (F4)	$F_{Rastrigin-SR} = 10n + \sum_{i=1}^{n}(z_i^2 - 10\cos(2\pi z_i))$	$[-5, 5]^n$	Yes	Yes
Shifted rotated Weierstrass (F5)	$F_{Weierstrass-SR} = \sum_{i=1}^{n}(\sum_{k=0}^{k_{max}}(a^k \cos(2\pi b^k (z_i + 0.5)))) - n\sum_{k=0}^{k_{max}}(a^k \cos(\pi b^k))$	$[-0.5, 0.5]^n$	Yes	Yes
Expanded Griewank plus Rosenbrock (F6)	$F_{Grie+Rosen} = \sum_{i=1}^{D} F_{Griewank}(F_{Rosenbrock}(z_i, z_{i+1})),\ z_{D+1} = z_1$	$[-3, 1]^n$	Yes	Yes

Table 2: Algorithm parameters setting

General parameters	
Population size	100
Selection scheme	Roulette wheel
Stopping criteria	8000 evaluations
Local search method	L-BFGS-B
Number of trust region iteration	3
Crossover probability	1
Mutation probability	0.01
Variation operator	Uniform crossover and mutation
Database building phase	2000 evaluations

4.1 Solution Quality of EvoLS

The averaged convergence trends obtained by EvoLS on the benchmark problems as a function of the total number of exact fitness function evaluations are summarized in Figs. 3(a)-3(f). The results presented here are averaged of 20 independent runs for each test problem. Also shown in the figures are the averaged convergence trends obtained using the canonical SAEAs with single approximation methodology (i.e., EA-RBF, EA-PR and EA-GP) as described in Algorithm 2. In addition, EA-Perfect, refers to a canonical surrogate-assisted EA that employs an imaginary approximation method that generates error-free surrogates [4], i.e., $RMSE = 0$, is also considered for the comparison between the use of *evolvability* measure versus approximation error.

Algorithm 2 Canonical SAEA

Generate and *evaluate* an initial population
while computational budget is not exhausted **do**
    Select individuals for the reproduction pool
    if generation count $<$ database building phase (G_{db}) **then**
        Evolve population by evolutionary operators (crossover, mutation)
        Evaluate new population using exact fitness function evaluation
        Archive all exact evaluations into the database Ψ
    else
        Evolve population by evolutionary operators (crossover, mutation)
        for each individual $\mathbf{x}$ in the population **do**
            /*** Local Search Phase on Surrogate Model***/
            Find m nearest points to $\mathbf{y}$ in the database Ψ as training points for surrogate model
            Build surrogate model $\widehat{f}_M(\mathbf{x})$ based on training points
            Apply local search strategy $\varphi_M(\mathbf{y})$ to arrive at $\mathbf{y}_{opt}$
            Replace $\mathbf{y}$ with $\mathbf{y}^{opt}$ (Lamarckian learning)
        end for
        Archive exact evaluations into the database Ψ
    end if
end while

[4] An error-free surrogate model is realized by using exact fitness function for evaluation inside the local search strategy where a surrogate model should be used. Note that the incurred fitness evaluation here is not counted as part of the computational budget.