IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 30, 2018, accepted December 22, 2018, date of publication January 1, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890593

INVITED PAPER

Finding High-Dimensional D-Optimal Designs for Logistic Models via Differential Evolution

WEINAN XU[1], WENG KEE WONG[2], KAY CHEN TAN[3], (Fellow, IEEE), AND JIAN-XIN XU[4], (Fellow, IEEE)

[1]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 119077
[2]Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA 90095, USA
[3]Department of Computer Science, City University of Hong Kong, Hong Kong
[4]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077

Corresponding authors: Weinan Xu (e0011559@u.nus.edu) and Kay Chen Tan (kaytan@cityu.edu.hk)

This work was supported in part by the National Natural Science Foundation of China under Grant 61876162/F060604, and in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project CityU11202418. The work of W. K. Wong was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award R01GM107639. The contents in this paper are solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ABSTRACT D-optimal designs are frequently used in controlled experiments to obtain the most accurate estimate of model parameters at minimal cost. Finding them can be a challenging task, especially when there are many factors in a nonlinear model. As the number of factors becomes large and interacts with one another, there are many more variables to optimize and the D-optimal design problem becomes high-dimensional and non-separable. Consequently, premature convergence issues arise. Candidate solutions get trapped in the local optima, and the classical gradient-based optimization approaches to search for the D-optimal designs rarely succeed. We propose a specially designed version of differential evolution (DE), which is a representative gradient-free optimization approach to solve such high-dimensional optimization problems. The proposed specially designed DE uses a new novelty-based mutation strategy to explore the various regions in the search space. The exploration of the regions will be carried out differently from the previously explored regions, and the diversity of the population can be preserved. The proposed novelty-based mutation strategy is collaborated with two common DE mutation strategies to balance exploration and exploitation at the early or medium stage of the evolution. Additionally, we adapt the control parameters of DE as the evolution proceeds. Using the logistic models with several factors on various design spaces as examples, our simulation results show that our algorithm can find the D-optimal designs efficiently and the algorithm outperforms its competitors. As an application, we apply our algorithm and re-design a 10-factor car refueling experiment with discrete and continuous factors and selected pairwise interactions. Our proposed algorithm was able to consistently outperform the other algorithms and find a more efficient D-optimal design for the problem.

INDEX TERMS Approximate design, design efficiency, generalized linear model, high-dimensional, non-separable, sensitivity function.

I. INTRODUCTION

Optimal design problems frequently arise in scientific investigations when we want to obtain the most accurate statistical inference at minimal cost. For example, D-optimal designs are commonly used to estimate parameters in the statistical model by minimizing the volume of the confidence ellipsoid of the parameters. When the model is nonlinear, the design criterion contains the unknown model parameters, which we want to estimate. Nominal values for the parameters are required to replace the unknown parameters before optimization and the resulting optimal design is termed locally optimal [1], [2] because it depends on the nominal values. Nominal values for the parameters may come from an expert's opinion or from a pilot study. The locally D-optimal design is then implemented to generate data to estimate the model parameters and the estimated parameters become the nominal values in the next step. The expectation is that after a couple of iterations, the estimates will become stable.

In the statistical literature, the optimal design is usually found from theory and when the model is nonlinear, there is usually only one or two factors. The theoretical approach encounters mathematical difficulties when the nonlinear model has several factors or the design criterion becomes complicated. Under such situations, our experience is that the classical optimization numerical techniques fail to find the locally optimal design or they become very inefficient. This is because as the number of factors in the model increases, the number of parameters in the model also increases. Consequently, the number of design points for the optimal design increases, resulting in having substantially many more variables to optimize. Thus, the design problem becomes quickly high-dimensional and also non-separable when factors interact with one another. Premature convergence can become a severe issue since solutions can easily get trapped in local optima.

Nature-inspired metaheuristic algorithms are now increasingly applied to solve a large variety of complicated optimization problems [3], [4]. Particle Swarm Optimization (PSO) [5] is one such algorithm [5]–[7], which has been recently used to solve various optimal design problems in the statistical literature [8]–[10]. However, the D-optimal design problems in these papers have only 3 or fewer factors in the statistical model and so premature convergence may not be an issue. Since PSO exerts the selective pressure onto some current best solutions termed as *gbest* and *pbest*, our experience is that models with 4 or more factors can cause PSO to experience premature convergence and make PSO less effective [11], [12].

Differential Evolution (DE) is an algorithm from a family of gradient-free algorithms-evolutionary algorithms. Mutation, crossover and selection are three fundamental operations in DE [13], [14]. One advantage that DE has over other evolutionary algorithms is that it has fewer control parameters [15]–[17], and works well in handling numerical optimization problems [18]–[21]. Compared with PSO, DE can alleviate the premature convergence issue moderately [13] since most of the mutation strategies of DE do not exert the selective pressure onto the current best solution [22]–[26]. However, based on the studies of DE variants for solving high-dimensional problems, there is no specially designed mechanism to explore various but novelty regions in the search space and to preserve the diversity of the population.

To circumvent the above issues and also motivated by novelty search methods [27], [28] which are capable of escaping from local optima by trying some novelty solutions for efficient exploration, we propose a new novelty-based mutation strategy. At the start of the evolution, a portion of individuals are randomly selected as the novelty-based individuals, and their aim is to explore various individuals which are potentially to be novelty individuals. For each novelty-based individual, we sample some difference vectors to be added to the current individual. Among these sampled difference vectors, we select the one which has the largest angle differences from the difference vector used in the previous generation. Each novelty-based individual explores the region of the search space different from the region explored in the previous generation so that novelty solutions can be obtained. As evolution proceeds, various regions of the search space would be explored and the diversity of the population is enhanced. The novelty-based mutation strategy is combined with two common mutation strategies, 'DE/rand/2' and 'DE/current-to-rand/1'. These two mutation strategies can balance the exploration and exploitation well at the early or medium stage of evolution as compared with other mutation strategies [29]. When the individuals obtained from these two mutation strategies converge, the novelty-based individuals can provide information of these recently explored regions in the search space so that these convergent individuals can both exploit in their current region and explore more regions in the search space.

We apply the proposed algorithm to generate locally D-optimal designs for logistic models with several factors with and without interactions on various design spaces. Logistic regression models have a binary response with one or more factors and is among the most frequently used in scientific investigations across many disciplines. Using a broad simulation study, we show our proposed algorithm consistently outperforms several of its top competitors. As an application, we implement our DE based algorithm to re-design a 10-factor car refueling experiment with both discrete and continuous factors, with and with factor interactions.

The remainder of this paper is organized as follows. Section II introduces statistical background and locally D-optimal designs for logistic regression models. It also reviews previous applications of using PSO to solve optimal design problems and a literature review of DE algorithms. In Section III, we propose a new DE algorithm NovDE and in Section IV, we apply it to construct and study properties of D-optimal designs on various design spaces. In Section V, we apply the proposed algorithm to generate D-optimal designs for a ten-factor car refueling experiment with and without factor interactions and there are mixed factors in the experiment. Section VI concludes with a summary of our work.

II. BACKGROUND

A. LOCALLY D-OPTIMAL DESIGNS FOR LOGISTIC REGRESSION MODELS

A generalized linear model is commonly used to study the mean of a response variable Y as a function of n independent variables [1]. We focus on models with a binary response variable even though the methodology proposed herein applies more generally. Let $E(Y_l) = \mu_l$ and let $\eta_l = r^T(x)\beta$ be the linear predictor, where $r(x)$ is a user-selected regression function that depends on the n factors. Additionally, let g(.) be a monotonic function such that $g(\mu_l) = \eta_l$ [30]. Some common choices for the regression function are $r(x)^T = \{1, x_1, \cdots, x_n\}$ (additive model) or $r(x)^T = \{1, x_1, \cdots, x_n, x_1x_2, \cdots, x_{n-1}x_n\}$ (model with all pairwise interaction terms). We assume Y_k is independent of Y_l if $l \neq k$ and the design space is user-selected compact set and contains all allowable combination levels of the factors to observe the response.

For the logistic model, we have

$$g(\mu_l) = \log(\frac{\mu_l}{1-\mu_l}) = \eta_l \tag{1}$$

Our goal is to find an optimal set of factor levels $x_1, \cdots, x_L$ to estimate the vector of parameters β in the linear predictor [2], [31] when we are given resources to take N observations. This means that we determine the optimal number of support points required, i.e. the value of L, the best choices of the support points $x_1, \cdots, x_L$ from a given design space and the optimal number of replicates at n_i at x_i, $i = 1, \cdots, L$ subject to $n_1 + \cdots + n_L = N$. The upshot is we have a constrained optimization problem where some of the variables to be optimized are positive integers and constrained to sum to N.

Following [31], the worth of a L-point design ξ with n_l replicates at x_l is determined by its Fisher information matrix defined by

$$I_\xi = \sum_{l=1}^{L} n_l \Upsilon(\eta_l) r(x_l) r(x_l)^T, \tag{2}$$

where $\Upsilon(\eta_l) = \frac{(du_l/d_{\eta_l})^2}{u_l(1-u_l)}$. For the logistic regression model, the link function is the logit function in (1) and

$$\Upsilon(\eta_l) = \frac{1}{2+e^{\eta_l}+e^{-\eta_l}} = \frac{e^{\eta_l}}{(1+e^{\eta_l})^2}. \tag{3}$$

A locally D-optimal design maximizes the log-determinant of the Fisher information matrix I_ξ in (2), or equivalently minimizes the generalized variance of the estimates of the parameters. Thus, D-optimal designs provide the most accurate estimates of all the model parameters in β. Clearly, I_ξ depends on β and so nominal values for β are required before optimization. Frequently, the nominal values for β come from prior experiences or a pilot study [32].

We focus on approximate designs obtained by replacing each n_l by $w_l = n_l/N$, the proportion of the total observations to be taken at x_l. More generally, we allow w_l to take on any value between 0 and 1 and doing so turns the problem into a convex optimization problem where convex optimization tools can be used to find and verify optimality of a design. Designs with weights $w_i's$ that sum to unity are called approximate designs.

For D-optimality, the design criterion is $-log|I(\xi, \theta)|$ and this is a convex function over the space of all approximate designs on the given and compact design space of interest [1]. Following [33], the approximate design ξ^* is locally D-optimal among all designs if and only if for all x in the design space, the following checking condition is satisfied:

$$\frac{e^{\beta^T r(x)}}{(1+e^{\beta^T r(x)})^2} r(x)^T I_{\xi^*}^{-1} r(x) - k \leq 0 \tag{4}$$

with equality at each support point of ξ^*. Here k is the dimension of β and the left-hand side of (4) is sometimes called the sensitivity function.

Often, the worth of a design ξ is measured by its efficiency relative to the optimal design ξ^* [1]. For D-optimality, the D-efficiency of a design ξ is

$$\left(\frac{\det(I_\xi)}{\det(I_{\xi^*})}\right)^{1/k}.$$

If its D-efficiency is near 1, ξ is close to ξ^*. If the theoretical optimal design ξ^* is unknown, the proximity of a design ξ to ξ^* can be determined from convex analysis theory. Specifically, its D-efficiency is at least $exp(-\theta/k)$, where θ is the maximum positive value of the sensitivity function across the entire design space [34]. If the D-efficiency lower bound is close to 1, the design ξ is close to the D-optimal design ξ^*.

B. FUNDAMENTALS OF DIFFERENTIAL EVOLUTION

Differential Evolution (DE) was proposed by Storn and Price [13] in 1995. It is a population-based optimization algorithm that searches for the optimum iteratively. DE is simple to implement and has good performance for solving various types of optimization problems. Compared with other evolutionary algorithms (EA), the space complexity of DE is low [14] and number of control parameters in DE is small [15]–[17]. There are two control parameters in DE; a scaling factor F for mutation and a crossover rate CR for the crossover operation. The parameter F controls the convergence speed and the parameter CR affects both the convergence and the diversity of the population [13], [35], [36].

To fix ideas, suppose $f(X)$ is the given objective function and we want to minimize it over a user-selected D-dimensional space comprising the decision variables. DE has three main operations: mutation, crossover and selection. Each solution of generation g is represented by $X_{i,g}$, where i is the index of the corresponding solution. Sometimes $X_{i,g}$ is referred to as the target vector, which needs to be updated for the next generation $g+1$. Mutation generates a mutant vector $V_{i,g}$, followed by a crossover which then generates a trial vector $U_{i,g}$ based on both $V_{i,g}$ and $X_{i,g}$. The next step is Selection, where a decision is made whether to update the solution $X_{i,g+1}$ from $U_{i,g}$ or $X_{i,g}$ based on their objective function values. Some details for the three operations follow.

1) MUTATION

Each target vector $X_{i,g}$ generates a new individual, called the mutant vector $V_{i,g}$ and some frequently used mutation strategies are listed below.

"DE/rand/1":

$$V_{i,g} = X_{r1,g} + F \cdot (X_{r2,g} - X_{r3,g}) \tag{5}$$

"DE/rand-to-best/2":

$$V_{i,g} = X_{i,g} + F \cdot (X_{best,g} - X_{i,g}) + F \cdot (X_{r1,g} - X_{r2,g}) + F \cdot (X_{r3,g} - X_{r4,g}) \tag{6}$$

"DE/rand/2":

$$V_{i,g} = X_{r1,g} + F \cdot (X_{r2,g} - X_{r3,g}) + F \cdot (X_{r4,g} - X_{r5,g}) \tag{7}$$

"DE/current-to-rand/1":

$$V_{i,g} = X_{i,g} + K \cdot (X_{r1,g} - X_{i,g}) + F \cdot (X_{r2,g} - X_{r3,g}) \quad (8)$$

In (8), K is randomly generated from [0, 1]. The subscripts $r1$ to $r5$ of X in (5)−(8) represent the random individuals selected from the population pool.

2) CROSSOVER

Crossover operation is employed after mutation. In crossover, the mutant vector $V_{i,g}$ is recombined with the original individual $X_{i,g}$ to form the trial vector $U_{i,g}$. Two types of crossover schemes of DE are binomial crossover and exponential crossover. Binomial crossover is commonly used in DE to determine the trial vector as follows [24]:

$$u_{i,g}^{j} = \begin{cases} v_{i,g}^{j}, & rand(0,1) \le Cr \quad j = j_{rand} \\ x_{i,g}^{j}, & otherwise \end{cases} \quad (9)$$

where $j_{rand} \in \{1, 2, 3, \cdots, D\}$ is a randomly selected index to ensure that the trial vector $U_{i,g}$ can get at least one variable from the mutant vector $V_{i,g}$. The notation rand(0,1) is a uniform random number from the interval [0,1] and Cr is the pre-specified crossover rate.

An exponential crossover is another way to implement a crossover [37]. An integer z is randomly generated from [1,D]. Another integer L, i.e. the length of decision variables to be mutated, is determined as follows:

L=0
WHILE(rand(0,1)≤ Cr AND L≤ D)
DO(L=L+1)

If $L \geq 1$, the trial vector $U_{i,g}$ is generated as follows:

$$u_{i,g}^{j} = \begin{cases} v_{i,g}^{j}, & for \ j = z, z+1, z+2, \cdots z+L-1 \\ x_{i,g}^{j}, & otherwise \end{cases} \quad (10)$$

If $L = 0$, then $U_{i,g}$ is identical to $X_{i,g}$.

3) SELECTION

Selection is the last step to determine whether the trial vector $U_{i,g}$ survives to enter the next generation based on the objective function value $f(U_{i,g})$.

The selection operation in DE is described below:

$$X_{i,g+1} = \begin{cases} U_{i,g}, & if \ f(U_{i,g}) \le f(X_{i,g}) \\ X_{i,g}, & otherwise \end{cases} \quad (11)$$

C. LITERATURE REVIEW OF DIFFERENTIAL EVOLUTION

1) THE ADAPTATION SCHEME OF CONTROL PARAMETERS

The success of DE in solving a specific problem crucially depends on the appropriate choice of mutation strategies and the associated control parameter values. Many DE studies have proposed new mutation strategies that stayed constant for the entire evolution process but a few such as SaDE [29] have proposed an adaptive approach to select appropriate mutation strategies based on the successful experiences in the previous generations.

In terms of the control parameter adaptation schemes, most DE studies adapt the control parameters F and CR based on a pre-defined distribution. The mean of this distribution depends on the successful F or CR values in the previous generations. In [38], a new crossover method Multiple Exponential Recombination (MER) that combines the advantages of binomial crossover and exponential crossover was proposed to solve the non-separable problems, where the decision variables are dependent on each other [39]. It has been shown both theoretically and empirically that for the same value of CR, MER can result in improved performances. Hence, it is promising to embed MER into the control parameter adaptation schemes when we solve non-separable optimization problems.

2) HIGH-DIMENSIONAL PROBLEMS

For solving high-dimensional problems, DE algorithms have a cooperative coevolution (CC) framework and a noncooperative coevolution framework [40]. CC-based framework partitions either the entire population into subpopulations, or partitions the entire decision variables into subcomponents. The optimization process is both parallel within each subgroup and centralized for the entire group. In [41], DECC-DML adopted CC framework, and the new partition strategy called delta grouping was proposed. To emphasize the interaction between variables in the same group, the improvement interval of interacting variables in different group would be limited. DECC-DML was efficient in solving non-separable problems with one group of rotated variables but not so when there are multiple groups. In [42], DCDE applied the CC framework and a ring connection to enhance the interactions among variables between different groups so that the search behavior of exploration and exploitation can be balanced. DCDE was capable of solving some non-separable and multimodal high-dimensional problems. In [43], DDE-AMS was proposed to solve the high-dimensional problems by a distributed differential evolution with adaptive mergence and split on subpopulations. The mergence and split operators made full use of the population resource to efficiently solve the problems in a cooperative and efficient way. For non-CC frameworks, most of the DE studies focused on adding adaptive mechanisms into the algorithms or proposed new mutation strategies. In [44], both F, CR, population size and mutation strategies were adapted and in [40], a new triangular mutation strategy was proposed.

III. PROPOSED ALGORITHM: NOVDE

A. OVERVIEW

Since the D-optimal design problems in this paper are high-dimensional and non-separable, premature convergence can be a severe issue with solutions easily getting trapped as local optima. Compounding the problem is that most of the state-of-the-art DE methods do not have a special mechanism to preserve diversity of the solutions and so the issue of premature convergence is not completely solved. For the mutation strategies such as 'DE/rand-to-best/2', solutions tend to be close to the current best region thus limiting the exploration capability at the early stage. For the mutation strategies such as 'DE/rand/2' or 'DE/current-to-rand/1', solutions tend to be close to each other at the early or medium stage of the evolution. Thus, to circumvent the issue of premature convergence of DE-based algorithms for solving high-dimensional and non-separable optimization problems, a mechanism of exploring various regions of the search space should be specially designed and combined with other DE mutation strategies.

Assume that there are n factors in the model and we denote a L-point design by $\xi = ([x_{11}x_{12}\cdots x_{1n}p_1], \cdots, [x_{l1}x_{l2}\cdots x_{ln}p_l], \cdots [x_{L1}x_{L2}\cdots x_{Ln}p_L])$, where p_l is the proportion of the total observations to be taken at the l-th design point $[x_{l1}x_{l2}\cdots x_{ln}]$. It follows that each individual $X_{i,g}$ in the current generation g with population index i is constructed as $X_{i,g} = (x_{11}x_{12}\cdots x_{1n}p_1\cdots x_{l1}x_{l2}\cdots x_{ln}p_l\cdots x_{L1}x_{L2}\cdots x_{Ln}p_L)$. For an additive model with no interactions among the factors, the dimension D of $X_{i,g}$ is $(n+1)L$.

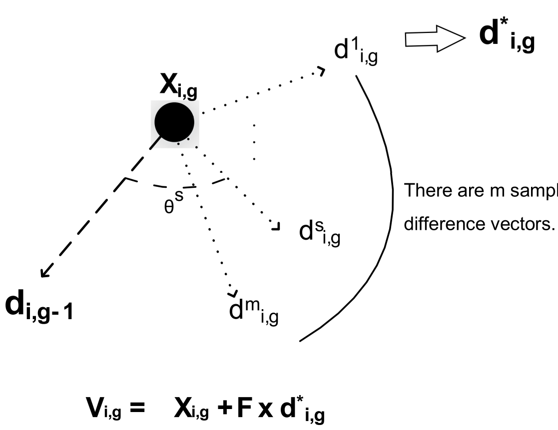

FIGURE 1. The operation of novelty-based mutation strategy. The target vector is $X_{i,g}$, and the difference vector from the previous generation is $d_{i,g-1}$. In the current generation, the m sampled difference vectors are $d^1_{i,g}\cdots d^m_{i,g}$ and θ^s is the computed angle between $d^s_{i,g}$ and $d_{i,g-1}$, where $s = 1, \cdots, m$. The $d^s_{i,g}$ with the largest angle differences θ^s is selected to be $d^*_{i,g}$ and the mutant vector $V_{i,g}$ is generated from $X_{i,g}$ and $d^*_{i,g}$.

We propose a new novelty-based DE-based algorithm and denote it by NovDE to solve our complex optimization problems using a novelty-based mutation strategy. At the start of the evolution process, a group of individuals are randomly selected to be novelty-based individuals. To preserve the diversity of solutions, various regions of the search space are explored by these novelty-based individuals. Fig. 1 shows the difference vector $d_{i,g-1}$ which is the difference between the trial vector $u_{i,g-1}$ and the target vector $x_{i,g-1}$ in the previous generation $g-1$. For the current generation g and a user-selected value of m, the number of m difference vectors $d^1_{i,g}\cdots d^m_{i,g}$ are sampled. Fig. 1 displays the computed angle θ^s between the sampled difference vector $d^s_{i,g}$ in the current generation g and the difference vector $d_{i,g-1}$ in the previous generation $g-1$ where $s = 1, 2, \cdots, m$. We then add the difference vector $d^*_{i,g}$, which has the largest angle difference between $d^s_{i,g}$ and $d_{i,g-1}$ among the m samples, to the target vector $x_{i,g}$ to generate the mutant vector $v_{i,g}$. This is because the largest angle differences between $d^*_{i,g}$ and $d_{i,g-1}$ would enhance each novelty-based individual to explore region in the search space entirely different than what was explored in the previous generation $g-1$. As the evolution proceeds, the novelty-based individuals can gradually explore various and novelty regions in the search space because of the efficient exploration and the diversity of solutions can be preserved. The proposed novelty-based mutation strategy is combined with 'DE/rand/2' and 'DE/current-to-rand/1' since they can balance exploration and exploitation at the early or medium stage of evolution [29]. If the individuals obtained based on 'DE/rand/2' and 'DE/current-to-rand/1' are close to each other, the novelty-based individuals can provide the information of the recent explored regions of the search space to those convergent individuals. The convergent individuals can either exploit in their current region of the search space or explore more regions of the search space.

The D-optimality criterion is a function of the information matrix in (2), where x_l is part of the consecutive component in the decision variables. The term $x_l x_l^T$ in equation (2) is the consecutive variables multiplied by each other, so physically proximate variables have stronger correlation and the problem is non-separable. According to [38], the crossover method MER can solve non-separable problems more efficiently than the binomial or exponential crossover method if the CR rate is the same. Further, MER updates the consecutive variables altogether which is more suitable for the structure of the decision variables in our problem. Thus, MER is selected to be the crossover method for our problem.

We adapt the control parameters F and CR to find locally D-optimal designs. The adaptation of F is the same as the state-of-the-art adaptive DE algorithm SaDE [29] where the F value for each individual is generated from $F = N(0.5, 0.3)$. In this way, the value of F falls in the range [−0.4,1.4] with probability of 0.997, which covers exploration capability when F is large and exploitation capability when F is small [29]. Because the novelty-based individuals explore various regions in the search space, F is not required to be adaptive so it is fixed at 0.5. Our adaptation method of CR in NovDE is new. A First-in-First-out (FIFO) memory $CRpool_k$ with a fixed size is applied, and the memory size for strategy k is proportion to the number of individuals involved in strategy k. The $CRmean_k$ is the mean value of the successful CR values stored in $CRpool_k$ memory. The mean value of $CRmean_k$ for each strategy k is adapted based on the success values of CR stored in $CRpool_k$ for strategy k. This adaptation method updates the distribution of CR more frequently based on the solutions in the current evolution stage. In NovDE, CR value for each individual for strategy k is generated from $CR = N(CRmean_k, 0.1)$. The initial value of $CRmean_k$ is selected to be 0.7 since if the value of CR is larger, the exploration would be encouraged. At the start of the evolution, exploration should be encouraged.

Algorithm 1 NovDE

Require: Target Vector $X_{i,g} = (x^1_{i,g}, x^2_{i,g}, \cdots, x^D_{i,g})$, population size N, p1=0.45, p2=0.9, sample size m, $CRpool_k$ with size LP_k, where k represents the $k-$th mutation strategy.

Ensure: Trial Vector $U_{i,g} = (u^1_{i,g}, u^2_{i,g}, \cdots, u^D_{i,g})$.

1: **if** $i \leq p1 * N$ **then**
2: F is generated from $N(0.5, 0.3)$.
3: $X_{i,g}$ performs 'DE/rand/2' to generate mutant vector $V_{i,g}$.
4: **end if**
5: **if** $p1 * N < i \leq p2 * N$ **then**
6: F is generated from $N(0.5, 0.3)$.
7: $X_{i,g}$ performs 'DE/current-to-rand/1' to generate mutant vector $V_{i,g}$.
8: **end if**
9: **if** $i > p2 * N$ **then**
10: F is fixed to be 0.5.
11: $d_{i,g-1}$ is the differences between trial vector $U_{i,g-1}$ and target vector $X_{i,g-1}$ in the previous generation $g-1$.
12: Sample number of m difference vectors as $d^1_{i,g} \cdots d^m_{i,g}$.
13: Compute the angle θ^s between $d^s_{i,g}$ and $d_{i,g-1}$ where $s = 1, 2, \cdots, m$.
14: $d^*_{i,g}$ is the one with the largest θ^s.
15: The mutant vector $V_{i,g} = X_{i,g} + F * d^*_{i,g}$.
16: **end if**
17: CR is generated from $(CRmean_k, 0.1)$ for different mutation strategy k.
18: Trial vector $U_{i,g}$ is generated based on MER and the CR rate.
19: **if** $i > p2 * N$ **then**
20: $d_{i,g} = U_{i,g} - X_{i,g}$.
21: **end if**
22: **if** $f(U_{i,g}) < f(X_{i,g})$ **then**
23: Record CR value into the corresponding $CRpool_k$.
24: Perform first-in-first-out operation once the size of $CRpool_k$ exceeds LP_k.
25: Update $CRmean_k$ to compute the mean value of elements in $CRpool_k$.
26: **end if**

B. ALGORITHM STRUCTURE

The proposed algorithm NovDE is displayed in Algorithm 1. In NovDE, three mutation strategies 'DE/rand/2', 'DE/current-to-rand/1' and the proposed 'novelty-based DE' are employed to generate the mutant vector $V_{i,g}$. Population are assigned to these three groups to employ one of the mutation strategies based on the pre-defined ratios $p1$ and $p2$. From step 9 to step 16, the proposed novelty-based DE mutation is presented. For each novelty-based individual $X_{i,g}$, F is fixed to be 0.5. The number of m difference vectors are sampled as $d^1_{i,g} \cdots d^m_{i,g}$, and the value of m is user-selected. For each $d^s_{i,g}$ in the samples for $s = 1, \cdots, m$, the angle between $d^s_{i,g}$ and the difference vector from last generation $d_{i,g-1}$ is computed and denoted as θ^s. The $d^s_{i,g}$ with the largest θ_s is denoted as $d^*_{i,g}$. Then in step 15, the mutant vector $V_{i,g}$ can be generated based on target vector $X_{i,g}$ and difference vector $d^*_{i,g}$.

For the adaptation of CR, a first-in-first-out memory for each mutation strategy k is established as $CRpool_k$ with size LP_k. $CRpool_k$ is to store the values of CR that make the trial vector $U_{i,g}$ successfully replace the target vector $X_{i,g}$ for strategy k. The $CRmean_k$ is computed as the mean value of elements in $CRpool_k$, and CR for each individual is generated from $N(CRmean_k, 0.1)$. The crossover method is MER. After the crossover operation, the novelty-based individuals should update $d_{i,g}$ to be used in the next generation as $d_{i,g+1}$.

IV. EMPIRICAL STUDY

In this section, we evaluate the performance of the proposed algorithm NovDE for finding locally D-optimal designs for logistic models on various design spaces with several factors. Specifically, we compare NovDE with six state-of-the-art variants of the DE algorithms. 'DE/rand/2/bin' [13] and SaDE [29] are effective in handling general numerical optimization problems; SaDE+MER [38] is effective in solving non-separable optimization problems; JADE [23] is an effective DE variant for its control parameter adaptation scheme; ANDE [40] and DDE-AMS [43] are effective in solving high-dimensional optimization problems. In order to validate the effectiveness of novelty-based mutation, we also compare the novelty-based mutation combined with the conventional crossover (i.e. binomial crossover), which is termed as NovDE-Bin. We compare using logistic models on various design spaces with seven continuous factors and five sets of nominal values. The design space of each factor is first selected to be on the prototype interval $[-1, 1]$ before we vary the design space to $[-3, 3]$, followed by the interval $[0, 3]$. We next describe the details of our experimental setup for comparing the four algorithms.

A. EXPERIMENTAL SETUP

1) Population size is 100.

2) The preset upper bound on the number of support points L is 100.

3) The dimension D of the problem to be optimized for seven factors without interactions is 800 $(=(7+1) \times 100)$. The dimension for each support point is 8, which includes the number of factors (i.e. 7) and the dimension of the corresponding portion of observations taken at each support point (i.e. 1).

4) The evolution process terminates if and when a design with at least 99.99% D-efficiency is found. Otherwise, the process terminates when the maximum number of generations we specify is met. For all our experiments, we set the maximum number of generation to be 20000.

5) The maximum number of run is 30.

6) For 'DE/rand/2/bin', we set $F = 0.5$ and $CR = 0.9$ based on the suggested settings in [13],.

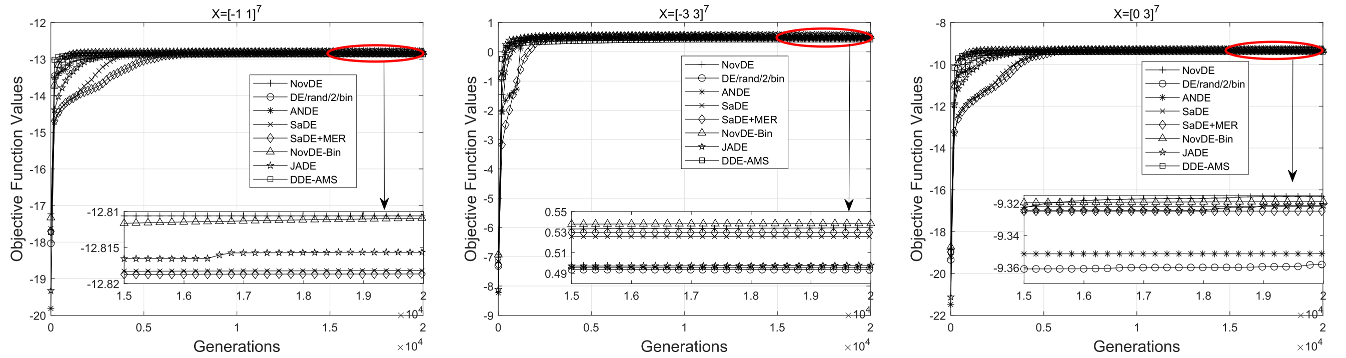

FIGURE 2. Average best-of-run objective function values of 30 independent runs over generations for $X = [-1\ 1]^7$, $X = [-3\ 3]^7$ and $X = [0\ 3]^7$, respectively. The nominal parameter is $\beta 3$.

TABLE 1. Performances of NovDE, NovDE-Bin and six competitors for finding locally D-optimal designs on $[-1, 1]^7$ using 5 sets of nominal values. In each cell, the numbers in the upper line are the mean and standard deviation of the values of the objective function over 30 runs, and the number at the bottom line is its success rate. For each set of nominal values, the best values of the mean and success rates are in bold. The entries with an $*$ means that NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum test.

Algorithm	$\beta 1$	$\beta 2$	$\beta 3$	$\beta 4$	$\beta 5$
NovDE	**-13.2018 (0.0079)** **86.67%**	-13.5845 (0.0153) 90%	**-12.8106 (0.0054)** **73.33%**	-12.6012 (0.0041) 83.33%	**-13.0221 (0.0093)** **90%**
NovDE-Bin	-13.2028 (0.0077)* **86.67%**	-13.5958 (0.0526)* 90%	-12.8116 (0.0049) 70%	**-12.5946 (0.0044)** **96.67%**	-13.0224 (0.0242) 86.67%
DE/rand/2/bin	-13.2274 (0.0153)* 3.33%	-13.5973 (0.0092)* 0%	-12.8459 (0.0128)* 3.33%	-12.6412 (0.0253)* 3.33%	-13.0650 (0.0205)* 0%
ANDE	-13.2340 (0.0146)* 3.33%	-13.6009 (0.0197)* 6.67%	-12.8460 (0.0095)* 0%	-12.6330 (0.0144)* 3.33%	-13.0543 (0.0163)* 3.33%
SaDE	-13.2030 (0.0060)* 73.33%	-13.5778 (0.0052) 93.33%	-12.8195 (0.0106)* 23.33%	-12.6006 (0.0078) **96.67%**	-13.0505 (0.0651)* 73.33%
SaDE+MER	-13.2032 (0.0071)* 80%	**-13.5761 (0.0012)** **96.67%**	-12.8186 (0.0057)* 53.33%	-12.6022 (0.0052) 90%	-13.0233 (0.0059)* 76.67%
JADE	-13.2035 (0.0021)* 30%	-13.5799 (0.0021) 73.33%	-12.8156 (0.0033)* 53.33%	-12.5958 (0.0072) 40%	-13.0248 (0.0039)* 33.33%
DDE-AMS	-13.2390 (0.0096)* 3.33%	-13.5961 (0.0186)* 3.33%	-12.8690 (0.0154)* 0%	-12.6258 (0.0169)* 6.67%	-13.0549 (0.0204)* 0%

TABLE 2. Performances of NovDE, NovDE-Bin and six competitors for finding locally D-optimal designs on $[-3, 3]^7$ using 5 sets of nominal values. In each cell, the numbers in the upper line are the mean and standard deviation of the values of the objective function over 30 multiple runs, and the number at the bottom line is its success rate. For each set of nominal values, the best values of the mean and success rates are in bold. The entries with $*$ represent NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum test.

Algorithm	$\beta 1$	$\beta 2$	$\beta 3$	$\beta 4$	$\beta 5$
NovDE	-0.1052 (0.0048) **100%**	-0.4441 (0.0038) 80%	0.5343 (0.0128) **46.67%**	**0.7487 (0.0076)** **93.33%**	**0.3678 (0.0028)** **90%**
NovDE-Bin	-0.1054 (0.0049) 93.33%	-0.4438 (0.0036) **90%**	**0.5384 (0.0177)** 40%	0.7476 (0.0051) **93.33%**	0.3645 (0.0055) **90%**
DE/rand/2/bin	-0.1209 (0.0135)* 43.33%	-0.4581 (0.0221)* 23.33%	0.4879 (0.0283)* 0%	0.7140 (0.0127)* 6.67%	0.3396 (0.0061)* 16.67%
ANDE	-0.1131 (0.0080) * 53.33%	-0.4561 (0.0113)* 40%	0.4940 (0.0207)* 3.33%	0.7097 (0.0295)* 3.33%	0.3412 (0.0138)* 6.67%
SaDE	-0.1033 (0.0027) 80%	**-0.4435 (0.0040)** **90%**	0.5234 (0.0155)* 40%	0.7457 (0.0109) 86.67%	0.3641 (0.0070) 73.33%
SaDE+MER	**-0.1022 (0.0032)** **100%**	-0.4440 (0.0030) 86.67%	0.5240 (0.0143) * 36.67%	0.7474 (0.0067) 90%	0.3661 (0.0061) 83.33%
JADE	-0.1031 (0.0025) 50%	-0.4436 (0.0038) 46.67%	0.4979 (0.0274) 10%	0.7467 (0.0058) 83.33%	0.3667 (0.0031) 83.33%
DDE-AMS	-0.1241 (0.0320)* 30%	-0.4590 (0.0304)* 16.67%	0.4278 (0.0350)* 0%	0.7130 (0.0197)* 3.33%	0.3547 (0.0098)* 33.33%

7) For ANDE, we follow recommendations in [40] and generate $F1$, $F2$, and $F3$ from the uniform distribution on [0, 1]. We select CR accordingly to [40] and set LP to 10% of the maximum generation.

8) For SaDE and SaDE+MER, we set $LP = 20$ and initial value of $p_k = 0.25$ for each strategy. The initial CRm for each strategy is 0.5 and generate the value of F from the normal distribution [0.5,0.3] based on [29]. For SaDE+MER, we set $T = 10$ based on [38].

9) For JADE, we set $p = 0.05$, $c = 0.1$ and set $\mu_{CR} = 0.5$, $\mu_F = 0.5$ as their initial values indicated in [23].

10) For DDE-AMS, we use 4 sub-populations, and set $U_p = 25$, $T = 80$, $D_r = 0.3$, and $\phi = 0.05$, $F = 0.5$ and $CR = 0.9$ based on [43].

11) For NovDE and NovDE-Bin, we set $p1 = 0.45$, $p2 = 0.9$ and the sample size is $m = 10$. Initial $CRmean_k$ for each strategy is 0.7 to encourage the exploration at the start of evolution. The upper bound of $CRmean_k$ is 0.9, and the lower bound of $CRmean_k$ is 0.1. For both 'DE/rand/2' and 'DE/current-to-rand/1', we set $LP = 50$, and for the novelty-based mutation strategy, we set $LP = 10$. We generate values of F for both 'DE/rand/2' and 'DE/current-to-rand/1' from the normal distribution [0.5, 0.3], and set $F = 0.5$ for the novelty-based mutation strategy.

12) We generate each of the nominal values in the vector of 8 coefficients $\beta^T = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)$ in an additive 7-factor logistic regression model randomly from the

TABLE 3. Performances of NovDE, NovDE-Bin and six competitors for finding locally D-optimal designs on $[0, 3]^7$ using 5 sets of nominal values. In each cell, the numbers in the upper line are the mean and standard deviation of the values of the objective function over 30 multiple runs, and the number at the bottom line is its success rate. For each set of nominal values, the best values of the mean and success rates are in bold. The entries with $*$ represent NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum test.

Algorithm	$\beta1$	$\beta2$	$\beta3$	$\beta4$	$\beta5$
NovDE	**-8.2056 (0.0091)**	**-11.0117 (0.0023)**	**-9.3156 (0.0191)**	-7.6625 (0.0183)	**-9.1025 (0.0124)**
	83.33%	**100%**	**26.67%**	83.33%	76.67%
NovDE-Bin	-8.2064 (0.0047)	-11.0134 (0.0027)	-9.3188 (0.0314)	**-7.6508 (0.0038)**	-9.1076 (0.0171)
	80%	**100%**	23.33%	**96.67%**	**80%**
DE/rand/2/bin	-8.2344 (0.0196)*	-11.0293 (0.0121)*	-9.3562 (0.0193)*	-7.6977 (0.0184)*	-9.1549 (0.0194)*
	3.33%	16.67%	3.33%	6.67%	3.33%
ANDE	-8.2430 (0.0194)*	-11.0249 (0.0055)*	-9.3500 (0.0256)*	-7.6861 (0.0163)*	-9.1510 (0.0135)*
	3.33%	46.67%	0%	0%	3.33%
SaDE	-8.2147 (0.0145)*	-11.0141 (0.0019)*	-9.3282 (0.0213)*	-7.6571 (0.0063)	-9.1066 (0.0128)*
	66.67%	90%	10%	86.67%	46.67%
SaDE+MER	-8.2067 (0.0052)	-11.0139 (0.0030)*	-9.3348 (0.0270)*	-7.6594 (0.0121)	-9.1098 (0.0147)*
	70%	86.67%	16.67%	76.67%	56.67%
JADE	-8.2085 (0.0019)*	-11.0330 (0.0093)*	-9.3213 (0.0207)*	-7.6517 (0.0040)	-9.1095 (0.0032)*
	20%	20%	6.67%	33.33%	36.67%
DDE-AMS	-8.2212 (0.0145)*	-11.0239 (0.0114)*	-9.3939 (0.0279)*	-7.7216 (0.0250)*	-9.1529 (0.0377)*
	3.33%	26.67%	0%	0%	3.33%

interval [−1,1] without loss of generality. In this experiment, we generate five parameter sets and they are as follows: $\beta1 = (0.6294, 0.8116, -0.7460, 0.8268, 0.2647, -0.8049, -0.4430, 0.0938)$, $\beta2 = (-0.6710, 0.8256, -0.9221, 0.8348, 0.0538, 0.8664, 0.9186, 0.7741)$, $\beta3 = (-0.4926, -0.6280, -0.3283, 0.4378, 0.5283, -0.6120, -0.6837, -0.2061)$, $\beta4 = (-0.4336, 0.3501, -0.8301, 0.3295, 0.0853, 0.5650, 0.0870, 0.1688)$, $\beta5 = (0.8379, -0.5372, 0.1537, -0.1094, -0.2925, 0.2599, -0.8201, -0.8402)$.

13) The program is implemented in MATLAB R2017b.

14) In this paper, the D-efficiency lower bound criterion is applied to evaluate the optimality of the generated design ξ and "DE/rand/1/bin" with F=0.5 and CR=0.9 is used to find the maximum positive value of the sensitivity function θ. We recall this value is used to compute the D-efficiency lower bound of the design ξ, which is $\exp(-\theta/k)$ where k is the dimension of β. In what is to follow, if a design has at least 95% D-efficiency, we accept the design as close enough to the optimum.

B. RESULTS AND DISCUSSIONS

We compare the performance of the proposed NovDE with NovDE-Bin and six competitive DE-based algorithms using 3 different design spaces to validate that NovDE is an effective DE variant in solving the high-dimensional optimal design problems. Since the optimal designs of the logistic model under various sets of nominal values and design spaces are unknown, the average of the objective function values obtained in 30 runs is considered as one performance indicator. Since the aim is to maximize the log-determinant, the larger the objective function values are, the better is the performance of the algorithm. Another performance indicator is the success rate, which is the percentage of runs where the generated design has at least 95% D-efficiency. To judge whether the proposed NovDE algorithm outperforms each of the other seven DE-based algorithms in a statistically significant way, we employ a nonparametric statistical test called Wilcoxon rank-sum test [45] and the 5% significance level. For each algorithm, the numbers in the upper line in each cell represent the mean and standard deviation of the objective values. The numbers in the bottom line represent the success rate of the algorithm. The best values of the mean and success rates are in bold, and entries with * represent NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum test at the 0.05 significance level.

For each design space, there are 5 different settings with nominal values $\beta1$ to $\beta5$. Hence, for the three different design spaces, there are 15 different settings in total. For the 15 different settings, when NovDE compares with the other seven DE algorithms, NovDE ranks first 9 out of 15 in terms of the mean of the objective function values. Furthermore, in these 9 cases, excluding NovDE-Bin, NovDE significantly outperforms the other six DE algorithms in 6 out of 9. NovDE also ranks first 10 out of 15 in terms of the success rate.

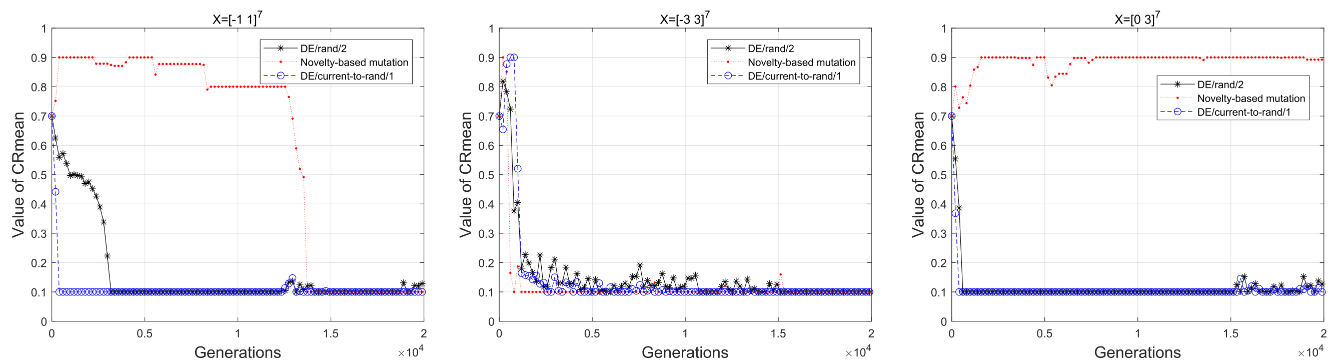

FIGURE 3. Adaptation behaviors of the median run among the 30 multiple runs of the CRmean values in NovDE for $X = [-1\ 1]^7$, $X = [-3\ 3]^7$ and $X = [0\ 3]^7$, respectively. The nominal parameter is $\beta3$.

These empirical results suggest that since the novelty-based mutation strategy combined with the MER crossover has the advantages of superior capability of exploration [27] and maintaining the dependent variables structure [38], NovDE can work well in handling non-separable problems and can avoid trapping into local optimum with higher chances. Thus, NovDE is more effective in generating locally D-optimal designs compared with the other seven algorithms.

To validate the effectiveness of novelty-based mutation alone, NovDE-Bin which is the novelty-based mutation with the conventional crossover-binomial crossover, is involved for the comparisons as well. For the 15 different settings, NovDE-Bin ranks first 3 out of 15, and second 8 out of 15 in terms of the mean of the objective function values. NovDE-Bin also ranks first 8 out of 15 in terms of the success rate. These empirical results suggest that the novelty-based mutation strategy presents better exploration capability and can prevent solutions from trapping into local optimum, which is consistent with the advantages of novelty search methods as illustrated in [27].

To give a clearer picture of the performance difference between NovDE and the other four DE algorithms, Fig. 2 plots the change of best-of-run objective function values over generations for each DE algorithm. The plots in Fig. 2 are based on nominal parameter $\beta 3$ and plots based on the other 4 sets of nominal values showed a similar pattern. We observe from Tables 1-3 and Fig. 2, both NovDE and NovDE-Bin clearly outperform 'DE/rand/2/bin' for all of the settings. Although 'DE/rand/2/bin' converges faster than NovDE, 'DE/rand/2/bin' has the issue of premature convergence so that the solutions tend to become close to each other and its exploration capability is deteriorated. The better performance of both NovDE-Bin and NovDE validate that exploration is important to solve our high-dimensional non-seperable problem which have local optimums. Furthermore,the novel information collected from exploration can be provided to the individuals generated from 'DE/rand/2' and 'DE/current-to-rand/1' to enhance both exploration and exploitation. As shown in Fig. 2, NovDE has the best converged objective function values close to the global optimum on various design spaces.

Since *CRmean* can represent the overall *CR* values of the individuals under different strategies, it is instructive to plot the CRmean values versus generations for each design space. Fig. 3 plots the *CRmean* values based on the median run using β_3 as nominal values for the same reason explained earlier. In Fig. 3, we observe that the *CRmean* values for 'DE/rand/2' and 'DE/current-to-rand/1' would converge to 0.1, which is the lower bound of the *CRmean* in NovDE. The variation of the *CRmean* values for novelty-based strategy presents distinct patterns under different design spaces. When $X = [-1\ 1]^7$ and $X = [-3\ 3]^7$, the *CRmean* would converge to 0.1, which is the lower bound of the *CRmean* in NovDE. Under these two design spaces, the decrease of *CRmean* values indicates their exploration capability tends to be restricted as evolution proceeds; for $X = [-3\ 3]^7$, the *CRmean* values would converge faster. When $X = [0\ 3]$, the *CRmean* converges to around 0.9, which is the upper bound of the *CRmean* in NovDE. The increase of *CRmean* values indicates their exploration capability tends to be enhanced as evolution proceeds. For different design spaces, the *CRmean* for the novelty-based strategy presents its adaptation to the exploration capability.

TABLE 4. NovDE-generated locally D-optimal design for the logistic model with seven variables when the vector of nominal values for the parameters is $\beta 3 = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T = (-0.4926, -0.6280, -0.3283, 0.4378, 0.5283, -0.6120, -0.6837, -0.2061)^T$, and $X = [-1, 1]^7$.

Support point	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	P_i
1	1	-1	1	-1	-1	-1	-1	0.0230
2	1	-1	1	-1	-1	-1	1	0.0160
3	1	1	1	1	1	-1	1	0.0255
4	1	1	-1	1	1	-1	-1	0.0223
5	1	-1	1	-1	-1	1	1	0.0152
6	-1	-1	1	1	1	1	1	0.0212
7	-1	1	1	-1	1	-1	1	0.0269
8	1	-1	1	-1	1	-1	-1	0.0101
9	1	-1	1	-1	-1	1	-1	0.0269
10	-1	1	1	-1	-1	1	-1	0.0117
11	1	-1	-1	1	1	-1	1	0.0269
12	-1	-1	1	1	1	1	1	0.0142
13	-1	1	1	1	1	-1	1	0.0219
14	1	1	1	-1	-1	-1	1	0.0182
15	1	1	-1	1	-1	-1	1	0.0183
16	-1	-1	1	1	1	1	-1	0.0199
17	-1	-1	-1	-1	-1	-1	1	0.0269
18	-1	-1	-1	-1	1	-1	-1	0.0101
19	-1	-1	1	-1	1	1	-1	0.0269
20	-1	1	1	-1	-1	-1	-1	0.0163
21	1	-1	1	1	-1	-1	1	0.0102
22	-1	1	-1	-1	1	-1	-1	0.0269
23	-1	1	-1	1	-1	-1	1	0.0241
24	-1	-1	-1	-1	-1	-1	1	0.0213
25	-1	-1	-1	-1	-1	1	1	0.0269
26	-1	-1	1	-1	1	-1	-1	0.0269
27	1	-1	-1	1	-1	-1	-1	0.0269
28	-1	1	-1	-1	-1	1	-1	0.0184
29	-1	-1	-1	1	-1	1	-1	0.0269
30	1	-1	-1	1	-1	1	-1	0.0161
31	1	-1	1	1	-1	1	-1	0.0165
32	1	-1	-1	-1	-1	-1	1	0.0143
33	-1	1	-1	-1	1	-1	-1	0.0269
34	-1	-1	1	1	1	1	-1	0.0124
35	-1	1	1	1	1	-1	-1	0.0269
36	-1	1	1	-1	-1	1	1	0.0204
37	1	1	1	1	-1	1	-1	0.0269
38	-1	1	-1	1	1	-1	-1	0.0103
39	1	-1	1	1	1	-1	1	0.0269
40	-1	1	-1	1	-1	1	1	0.0152
41	1	1	-1	1	-1	-1	-1	0.0150
42	1	1	1	-1	-1	-1	-1	0.0260
43	-1	-1	-1	-1	-1	-1	1	0.0144
44	-1	1	-1	1	-1	1	-1	0.0260
45	-1	-1	-1	1	1	1	1	0.0269
46	1	1	1	1	-1	1	1	0.0203
47	-1	1	1	1	-1	1	1	0.0245
48	1	-1	-1	1	1	-1	-1	0.0269

Table 4 to Table 6 present the support points of the locally D-optimal designs when $\beta 3$ is the set of nominal values. Interestingly, each support point of these locally D-optimal

TABLE 5. NovDE-generated locally D-optimal design for the logistic model with seven variables when the vector of nominal values for the parameters is $\beta 3 = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T = (-0.4926, -0.6280, -0.3283, 0.4378, 0.5283, -0.6120, -0.6837, -0.2061)^T$, and $X = [-3, 3]^7$.

Support point	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	P_i
1	-3	-3	-3	-3	3	-3	-3	0.0311
2	3	3	3	3	-3	3	3	0.0100
3	-3	-3	-3	-3	3	-3	3	0.0347
4	3	-3	3	3	3	-2.9971	3	0.0480
5	-3	-3	3	-3	3	2.9217	-3	0.0292
6	3	-3	-3	-3	-3	-3	-3	0.0100
7	-3	3	-3	-3	2.0891	-3	-3	0.0107
8	-3	-3	3	-3	-2.7545	3	3	0.0468
9	3	3	3	3	3	-3	-3	0.0187
10	3	3	3	3	-3	3	-3	0.0298
11	-3	-3	3	3	3	3	3	0.0409
12	3	-3	-3	3	-3	3	-3	0.0403
13	3	3	3	-3	-3	-3	3	0.0100
14	3	3	-3	3	-3	-3	3	0.0100
15	-3	3	3	3	3	3	-3	0.0385
16	3	-3	-3	-3	-3	-3	-3	0.0260
17	3	3	3	-3	-3	-3	-3	0.0517
18	-3	-3	-3	3	3	3	-3	0.0338
19	3	-3	3	-3	3	-3	-3	0.0375
20	-3	3	-3	-3	-3	-3	3	0.0303
21	3	3	-3	3	-3	-3	3	0.0258
22	-2.9190	3	-3	3	-3	3	-3	0.0451
23	-3	3	3	-3	-3	3	3	0.0431
24	3	-3	-3	3	3	-3	-3	0.0100
25	3	-3	-3	-3	-3	-3	3	0.0316
26	-3	-3	-3	-3	-3	3	-3	0.0136
27	-3	3	3	-3	3	-3	-3	0.0356
28	-3	3	-3	3	-3	3	3	0.0162
29	3	-3	3	3	-3	3	3	0.0438
30	-3	3	3	-3	-3	3	3	0.0305
31	3	-3	-3	3	-3	3	-3	0.0140
32	-3	3	-3	3	3	-3	3	0.0517
33	3	3	3	3	3	-3	3	0.0512

designs has at most one factor level supported at its non-extreme values. This observation may provide an impetus for further study using analytical tools.

V. CAR REFUELING EXPERIMENT

We now apply the proposed NovDE algorithm to re-design a ten factor experiment to test a vision-based car refueling system [46]. The investigators were interested in finding whether a computer-controlled nozzle was able to insert into gas pipe correctly or not implying that the response variable in the study is binary. Table 7 lists the ten factors. Four factors are discrete, each with two levels -1 or $+1$, and six factors are continuous. Table 7 shows that the range of values for each continuous factor and they do vary considerably. The proposed NovDE algorithm is applied to find a locally D-optimal design for this high-dimensional nonlinear model with mixed factors using the vector of nominal values $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})^T = (3, 0.5, 0.75, 1.25, 0.8, 0.5, 0.8, -0.4, -1, 2.65, 0.65)$ from literature [46].

Design issues for this ten-factor experiment were also considered in [47] but without interaction terms. In practice, the binary response is likely dependent on the joint changes in two or more of the factors, suggesting that interaction terms should be in the model. To fix ideas, we include five pairwise interactions into the model and believe that this is the first

TABLE 6. NovDE-generated locally D-optimal design for the logistic model with seven variables when the vector of nominal values for the parameters is $\beta 3 = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T = (-0.4926, -0.6280, -0.3283, 0.4378, 0.5283, -0.6120, -0.6837, -0.2061)^T$, and $X = [0, 3]^7$.

Support point	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	P_i
1	3	0	3	3	3	0	0	0.0232
2	0	3	3	3	3	0	0	0.0262
3	0	3	3	0	0	0	3	0.0151
4	0	3	3	3	0	3	0	0.0270
5	3	3	3	3	0	0	0	0.0191
6	3	0	3	0	0	0	0	0.0113
7	0	0	3	3	0	3	0	0.0103
8	3	0	3	0	0	0	0	0.0215
9	0	3	3	0	0	0	3	0.0220
10	3	0	0	3	0	0	0	0.0106
11	3	0	3	3	0	0	3	0.0366
12	0	0	0	3	3	0	3	0.0389
13	0	3	3	3	0	0	0	0.0318
14	0	3	0	3	0	0	3	0.0309
15	0	0	3	3	0	3	3	0.0171
16	0	0	3	3	3	0	3	0.0315
17	0	3	3	0	0	0	3	0.0385
18	0	0	0	0	0	0	0	0.0367
19	0	0	3	0	3	0	0	0.0319
20	0	0	3	3	3	0	0	0.0323
21	0	0	0	3	3	0	0	0.0128
22	0	0	0	3	0	0	3	0.0217
23	3	3	3	3	0	0	0	0.0333
24	0	0	3	3	3	0	3	0.0114
25	0	0	0	3	0	3	0	0.0389
26	0	0	0	0	0	0	0	0.0107
27	0	0	3	0	0	3	0	0.0271
28	0	0	3	0	0	0	0	0.0380
29	0	3	3	3	3	0	0	0.0263
30	3	0	3	3	0	0	3	0.0298
31	0	0	3	3	0	3	0	0.0346
32	3	0	0	3	0	0	0	0.0316
33	0	3	0	3	0	0	0	0.0312
34	0	0	3	0	0	0	0	0.0131
35	0	0	0	3	0	0	3	0.0225
36	0	3	0	3	0	0	0	0.0138
37	0	3	3	3	0	3	3	0.0300
38	3	0	3	3	0	0	0	0.0100
39	0	0	3	3	3	3	0	0.0118
40	0	0	3	0	0	0	3	0.0146
41	0	0	3	3	0	3	3	0.0242

design work for such a high-dimensional logistic model. Previous attempts using common algorithms, like multiplicative and modified Fedorov-Wynn algorithms did not converge. The vector of nominal values for the model with the five selected pairwise interactions is $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{1,9}, \beta_{2,5}, \beta_{3,4}, \beta_{6,7}, \beta_{8,10})^T = (3, 0.5, 0.75, 1.25, 0.8, 0.5, 0.8, -0.4, -1, 2.65, 0.65, 0.01, -0.02, 0.03, -0.04, 0.05)^T$ based on literature [46].

Some of the tuning parameters used to find the locally D-optimal designs are the population size, maximum number of generations and maximum number of support points. For the model without factor interactions, the population size is 100, and the maximum number of generations is 10000. The maximum number of support points L is set to 100 so the dimension D of the problem is 1100 $(=(10 + 1) \times 100)$. The dimension for each support point is 11, which includes the number of factors (i.e. 10) and the dimension of the corresponding portion of observations taken at each support

TABLE 7. Factor types and levels for the car refueling experiment.

Type	Factor	Level	
		Low	High
Discrete	Ring Type	White paper	Reflective
	Lighting	Room lighting	2 flood lights and room lights
	Sharpen	No	Yes
	Smooth	No	Yes
Continuous	Lighting angle	from 50 degrees to 90 degrees	
	Gas-cap angle (Z axis)	from 30 degrees to 55 degrees	
	Gas-cap angle (Y axis skew)	from 0 degrees to 10 degrees	
	Car distance	from 18 in. to 48 in.	
	Reflective ring thickness	from 0.125 in. to 0.425 in.	
	Threshold step value	from 5 to 15	

TABLE 8. Comparisons of the performance of NovDE, NovDE-Bin and six competitors for the car refueling experiments without factor interactions. The best values of the mean and success rates are in bold. The entries with * represent NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum test.

Algorithm	Success Rate	Mean (std)
NovDE	**90%**	-35.9390 (0.1060)
NovDE-Bin	**90%**	-35.9180 (0.0035)
DE/rand/2/bin	13.33%	-37.4963 (0.8482)*
ANDE	26.67%	-36.7708 (0.8923)*
SaDE	70%	-35.9480 (0.0485)*
SaDE+MER	46.67%	-36.2242 (0.7181)*
JADE	70%	-35.9645 (0.0911)*
DDE-AMS	10%	-39.5935 (0.5128)*

point (i.e. 1). For the model with factor interactions, the population size is 100, and the maximum number of generations is 20000. The maximum number of support points L is 100 so the dimension D of the problem is 1600 (=(10+1+5)×100).

Due to the number of factors in this study, it is hard to construct and visually appreciate the high-dimensional sensitivity function of the generated design to confirm its optimality. An option is to generate 1000000 random vectors within the design spaces and check whether the sensitivity function is positive at these points. One may repeat this procedure and if none is found and the sensitivity function is zero at the support points of the generated design, then we conclude we have found an optimal design. Otherwise, we apply "DE/rand/1/bin" with F=0.5 and CR=0.9 to find the maximum positive value of the function and compute its D-efficiency lower bound. The lower bound D-efficiency is defined as exp $(-\theta/k)$ where k is the dimension of the model parameter β. Since the variables of this problems are mixed, the variation of lower bound D-efficiency is very large. In what is to follow, if a design has at least 90% D-efficiency, we accept the design as close enough to the optimum.

A. WITHOUT FACTOR INTERACTIONS

Table 8 compares the mean of locally D-optimal objective value and success rate of NovDE with NovDE-Bin and the other six differential evolution algorithms. Wilcoxon rank-sum test [45] is also conducted at the 5% significance level. In Table 8, both the mean of the objective value and the success rate of NovDE-Bin are the highest. NovDE ranks the second. Both NovDE-Bin and NovDE significantly outperform all the other six algorithms. Thus, our empirical results validate the effectiveness of novelty exploration in solving the car refueling experiment. By extension, our work suggests that the NovDE and NovDE-Bin are effective for searching locally D-optimal designs for high-dimensional non-separable problems with mixed variables on various design spaces. In problems with mixed factors, the design space is less complex than the design space of the problems with continuous factors. The solutions obtained from D-optimal design with mixed factors are more likely to be close to one another at early evolution stage. Thus, it is more crucial to handle the premature issue especially for the problems with mixed factors. NovDE and NovDE-Bin can present the advantages in handling the premature issue and preserve

TABLE 9. NovDE-generated locally D-optimal design for car refueling experiment without factor interactions.

Support point	Ring Type	Lighting	Sharpen	Smooth	Lighting Angle	Z axis	Y axis skew	Car Dist.	Ring Thick.	Threshold Step-size	P_i
1	-1	-1	-1	1	50	30	10	48	0.1250	5	0.0909
2	-1	-1	-1	-1	50	30	4.1991	48	0.1250	5	0.0909
3	-1	-1	-1	-1	50	30	10	48	0.1250	8.5698	0.0909
4	-1	-1	-1	-1	50	30	10	48	0.1250	5	0.0807
5	-1	-1	-1	-1	54.6407	30	10	48	0.1250	5	0.0909
6	-1	1	-1	-1	50	30	10	48	0.1250	5	0.0909
7	1	-1	-1	-1	50	30	10	48	0.1250	5	0.0752
8	1	-1	-1	-1	50	30	10	48	0.4250	5	0.0397
9	-1	-1	-1	-1	50	32.9005	10	48	0.1250	5	0.0909
10	-1	-1	-1	-1	50	30	10	45.6796	0.1250	5	0.0909
11	-1	-1	-1	-1	50	30	10	48	0.4250	5	0.0772
12	-1	-1	1	-1	50	30	10	48	0.1250	5	0.0909

TABLE 10. Comparisons of the performance of NovDE, NovDE-Bin and six competitors for the car refueling experiment with factor interactions. The best values of the mean and success rates are in bold. The entries with * represent NovDE significantly outperforms the other algorithm based on Wilcoxon rank-sum tests.

Algorithm	Success Rate	Mean (std)
NovDE	**80%**	**-71.5401 (0.3365)**
NovDE-Bin	70%	-71.5640 (0.3453)
DE/rand/2/bin	0%	-74.4495 (1.4082)*
ANDE	23.33%	-72.0390 (0.6751)*
SaDE	3.33%	-71.7135 (0.3182)*
SaDE+MER	20%	-71.6072 (0.1860)*
JADE	30%	-71.5843 (0.3562)*
DDE-AMS	6.67%	-71.7058 (0.3136)*

the diversity of the solutions. As a result, NovDE performs even better than it performs on handling D-optimal design with continuous factors. Table 9 lists the locally D-optimal design for the car refueling experiment, and 12 support points are generated. The design criteria value is -35.9178. A direct calculation shows that the D-efficiency lower bound for the generated design is 94.60%. This is not surprising even though we set the lower bound to be 90% for this problem. The reason is because the algorithm is not monotonic in the sense that it does not necessarily produce increasingly more efficient designs with each iteration. Another reason is that the higher D-efficiency optimal design may exist in the continuous design spaces instead of the mixed design spaces. Based on Table 9, the common rule of the mining knowledge still satisfies. For each support point, there is at most one factor value not at the boundary of the design space. This is consistent with the observation in Section IV.

B. WITH FACTOR INTERACTIONS

It seems realistic that there are factor interactions between Ring type and Reflective ring thickness, Lighting and Lighting angle, Sharpen and Smooth, Gas-cap angle (Z axis) and Gas-cap angle (Y axis skew) and Car distance and Threshold step value, respectively. The former two interactions are between a discrete factor and a continuous factor; the third interaction is between a discrete factor and a discrete factor and the latter two interactions are between a continuous factor and a continuous factor. In practice, the researcher uses content information to specify interaction terms in the model and implements a parsimonious model. Our conjecture that interaction terms were ignored in earlier design work for such a model is to simplify the design construction.

Table 10 compares the mean of locally D-optimal objective value and success rate of NovDE with NovDE-Bin and the other six differential evolution algorithms. Wilcoxon rank-sum test [45] is also conducted at the 5% significance level. In Table 10, both the mean of the objective value and the success rate of NovDE is the highest, and NovDE-Bin is the second highest. NovDE significantly outperforms all the other six algorithms. We observe that the overall outperformance of the NovDE and NovDE-Bin algorithms relative to the other six algorithms are less dramatic than when the model has no interaction terms, our results still show it is effective in solving non-separable high-dimensional locally D-optimal design problems with mixed factors on various design spaces. In particular, it shows the NovDE is able to handle premature convergence and non-separable issues well in complex optimization problems. The proposed NovDE algorithm can produce optimal designs for a more realistic situation and so represents an advancement. Table 11 shows

TABLE 11. NovDE-generated locally D-optimal design for the car refueling experiment with five pairwise factor interactions.

Support point	Ring Type	Lighting	Sharpen	Smooth	Lighting Angle	Z axis	Y axis skew	Car Dist.	Ring Thick.	Threshold Step-size	P_i
1	-1	1	-1	-1	50	35.5152	10	48	0.1250	5	0.0625
2	-1	1	1	1	50	30	10	48	0.1250	5	0.0625
3	-1	-1	-1	-1	50	30	10	48	0.1250	5	0.0466
4	-1	1	-1	-1	50	30	10	48	0.1250	5	0.0511
5	-1	1	-1	-1	50	34.5716	8.8571	48	0.1250	5	0.0625
6	-1	1	-1	-1	50	30	8.6212	48	0.1250	5	0.0625
7	-1	1	-1	-1	50	30	10	48	0.4250	5	0.0486
8	-1	1	-1	-1	50	30	10	45.1640	0.1250	5.6974	0.0625
9	1	1	-1	-1	50	30	10	48	0.4250	5	0.0625
10	-1	1	-1	-1	50	30	10	48	0.1250	5.7233	0.0625
11	-1	1	1	-1	50	30	10	48	0.1250	5	0.0625
12	-1	1	-1	-1	54.5960	30	10	48	0.1250	5	0.0625
13	1	-1	-1	-1	50	30	10	48	0.1250	5	0.0242
14	1	1	-1	-1	50	30	10	48	0.1250	5	0.0499
15	-1	-1	-1	-1	54.1003	30	10	48	0.1250	5	0.0625
16	-1	-1	-1	-1	50	30	10	48	0.4250	5	0.0296
17	-1	1	-1	1	50	30	10	48	0.1250	5	0.0625
18	-1	1	-1	-1	50	30	10	45.0586	0.1250	5	0.0625

the optimal design for the car refueling experiment with five pairwise factor interactions. There are 18 support points, and the design criteria value is -71.4284. The design has a D-efficiency of 95% or higher. An interesting note is that Table 11 shows each support point can have one or more factors supported other than at its extreme values. This violates the common rule mentioned earlier and serves to show that as the model gets more complicated, the structure of the optimal design also becomes harder to characterize and understand.

VI. CONCLUSION

We propose a DE-based algorithm NovDE to search for locally D-optimal designs for logistic models with multiple factors and the factors may or may not interact with one another. We employ a new novelty-based mutation strategy to explore various regions of the search space so that the diversity of the population would be preserved. The new novelty-based mutation strategy is collaborated with 'DE/rand/2' and 'DE/current-to-rand/1' which can balance exploration and exploitation at early or medium stage of the evolution so that both convergence and diversity of the population are enhanced, and premature convergence issues are alleviated. We have demonstrated that NovDE provides the best objective function values and success rates compared with four other DE-related evolutionary algorithms. NovDE also outperforms the others in terms of finding a highly efficient D-optimal design for the ten-factor car refueling study where there are discrete and continuous factors in the logistic model and some of them interact with one another. Our empirical results also show that the distribution of the support points for optimal designs for models with interaction terms are more complex than those for models without interaction terms.

We focus on logistic models which are the most commonly used in practice to model binary responses. We expect our proposed algorithm works for other link functions as well, including cases when the response is continuous and there are many mixed factors. Our future study includes testing the capability of our proposed algorithm for tackling these problems and multiple-objective optimal design problems.

ACKNOWLEDGMENT

Wong wishes to acknowledge the support from the Institute for Mathematical Sciences at the National University of Singapore for hosting the Particle Swarm and Evolutionary Computation Workshop in February 2018 and co-chaired by the K. C. Tan and W. K. Wong, where parts of this paper were discussed and written.

REFERENCES

[1] S. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation*, vol. 1. Springer, 2013.

[2] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Statist.*, vol. 24, no. 4, pp. 586–602, 1953.

[3] J. M. Whitacre, "Recent trends indicate rapid growth of nature-inspired optimization in academia and industry," *Computing*, vol. 93, nos. 2–4, pp. 121–133, 2011.

[4] J. M. Whitacre, "Survival of the flexible: Explaining the recent popularity of nature-inspired optimization within a rapidly evolving world," *Computing*, vol. 93, nos. 2–4, pp. 135–146, 2011.

[5] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*. Boston, MA, USA: Springer, 2011, pp. 760–766.

[6] M. Clerc, *Particle Swarm Optimization*, vol. 93. Hoboken, NJ, USA: Wiley, 2010.

[7] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, Jun. 2007.

[8] J. Qiu, R.-B. Chen, W. Wang, and W. K. Wong, "Using animal instincts to design efficient biomedical studies via particle swarm optimization," *Swarm Evol. Comput.*, vol. 18, pp. 1–10, Oct. 2014.

[9] W. K. Wong, R.-B. Chen, C.-C. Huang, and W. Wang, "A modified particle swarm optimization technique for finding optimal designs for mixture models," *PLoS ONE*, vol. 10, no. 6, p. e0124720, 2015.

[10] R.-B. Chen, S.-P. Chang, W. Wang, H.-C. Tung, and W. K. Wong, "Minimax optimal designs via particle swarm optimization methods," *Statist. Comput.*, vol. 25, no. 5, pp. 975–988, 2015.

[11] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. ICML*, vol. 97, 1997, pp. 412–420.

[12] W.-N. Chen *et al.*, "Particle swarm optimization with an aging leader and challengers," *IEEE Trans. Evol. Comput.*, vol. 17, no. 2, pp. 241–258, Apr. 2013.

[13] R. Storn and K. Price, "Differential evolution–A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[14] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 4–31, Feb. 2011.

[15] M. Keshk, H. Singh, and H. Abbass, "Automatic estimation of differential evolution parameters using hidden Markov models," *Evol. Intell.*, vol. 10, nos. 3–4, pp. 77–93, 2018.

[16] O. Kramer, "Evolutionary self-adaptation: A survey of operators and strategy parameters," *Evol. Intell.*, vol. 3, no. 2, pp. 51–65, 2010.

[17] N. Boukhari, F. Debbat, N. Monmarché, and M. Slimane, "A study on self-adaptation in the evolutionary strategy algorithm," in *Proc. IFIP Int. Conf. Comput. Intell. Appl. (CIIA)*. Oran, Algeria: Springer, 2018, pp. 150–160.

[18] A. Qing, "Dynamic differential evolution strategy and applications in electromagnetic inverse scattering problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 116–125, Jan. 2006.

[19] A. K. M. K. A. Talukder, M. Kirley, and R. Buyya, "Multiobjective differential evolution for scheduling workflow applications on global grids," *Concurrency Comput., Pract. Exper.*, vol. 21, no. 13, pp. 1742–1756, 2009.

[20] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.

[21] C. Zhang, K. C. Tan, and R. Ren, "Training cost-sensitive deep belief networks on imbalance data problems," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 4362–4367.

[22] J. Brest, S. Greiner, B. Boskovic, M. Mernik, and V. Zumer, "Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems," *IEEE Trans. Evol. Comput.*, vol. 10, no. 6, pp. 646–657, Dec. 2006.

[23] J. Zhang and A. C. Sanderson, "JADE: Adaptive differential evolution with optional external archive," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 945–958, Oct. 2009.

[24] S. Das, A. Abraham, U. K. Chakraborty, and A. Konar, "Differential evolution using a neighborhood-based mutation operator," *IEEE Trans. Evol. Comput.*, vol. 13, no. 3, pp. 526–553, Jun. 2009.

[25] M. G. Epitropakis, D. K. Tasoulis, N. G. Pavlidis, V. P. Plagianakos, and M. N. Vrahatis, "Enhancing differential evolution utilizing proximity-based mutation operators," *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 99–119, Feb. 2011.

[26] S. M. Islam, S. Das, S. Ghosh, S. Roy, and P. N. Suganthan, "An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 482–500, Apr. 2012.

[27] S. Risi, S. D. Vanderbleek, C. E. Hughes, and K. O. Stanley, "How novelty search escapes the deceptive trap of learning to learn," in *Proc. 11th Annu. Conf. Genetic Evol. Comput.*, 2009, pp. 153–160.

[28] J. Lehman and K. O. Stanley, "Evolving a diversity of virtual creatures through novelty search and local competition," in *Proc. 13th Annu. Conf. Genetic Evol. Comput.*, 2011, pp. 211–218.

[29] A. K. Qin and P. N. Suganthan, "Self-adaptive differential evolution algorithm for numerical optimization," in *Proc. IEEE Congr. Evol. Comput.*, vol. 2. Sep. 2005, pp. 1785–1791.
[30] C. E. McCulloch, "Generalized linear models," *J. Amer. Stat. Assoc.*, vol. 95, no. 452, pp. 1320–1324, 2000.
[31] I. Ford, B. Torsney, and C. F. J. Wu, "The use of a canonical form in the construction of locally optimal designs for non-linear problems," *J. Roy. Stat. Soc. Ser. B, Methodol.*, vol. 54, no. 2, pp. 569–583, 1992.
[32] J. Stufken and M. Yang, "Optimal designs for generalized linear models," in *Design and Analysis of Experiments: Special Designs and Applications*, vol. 3. 2012, p. 137.
[33] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *Ann. Math. Statist.*, vol. 30, no. 2, pp. 271–294, 1959.
[34] A. Pázman, *Foundations of Optimum Experimental Design*, vol. 14. Cham, Switzerland: Springer, 1986.
[35] D. Zaharie, "Control of population diversity and adaptation in differential evolution algorithms," in *Proc. MENDEL*, vol. 9, 2003, pp. 41–46.
[36] H. A. Abbass, "The self-adaptive Pareto differential evolution algorithm," in *Proc. Congr. Evol. Comput. (CEC)*, vol. 1, May 2002, pp. 831–836.
[37] S.-Z. Zhao and P. N. Suganthan, "Empirical investigations into the exponential crossover of differential evolutions," *Swarm Evol. Comput.*, vol. 9, pp. 27–36, Apr. 2013.
[38] X. Qiu, K. C. Tan, and J.-X. Xu, "Multiple exponential recombination for differential evolution," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 995–1006, Apr. 2017.
[39] N. Hansen, R. Ros, N. Mauny, M. Schoenauer, and A. Auger, "Impacts of invariance in search: When CMA-ES and PSO face ill-conditioned and non-separable problems," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5755–5769, 2011.
[40] A. W. Mohamed and A. S. Almazyad, "Differential evolution with novel mutation and adaptive crossover strategies for solving large scale global optimization problems," *Appl. Comput. Intell. Soft Comput.*, vol. 2017, Mar. 2017, Art. no. 7974218.
[41] M. N. Omidvar, X. Li, and X. Yao, "Cooperative co-evolution with delta grouping for large scale non-separable function optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2010, pp. 1–8.
[42] R. Tang, "Decentralizing and coevolving differential evolution for large-scale global optimization problems," *Appl. Intell.*, vol. 47, no. 4, pp. 1208–1223, 2017.
[43] Y.-F. Ge *et al.*, "Distributed differential evolution based on adaptive mergence and split for large-scale optimization," *IEEE Trans. Cybern.*, vol. 48, no. 7, pp. 2166–2180, Jul. 2018.
[44] J. Brest, A. Zamuda, I. Fister, and M. S. Maučec, "Large scale global optimization using self-adaptive differential evolution algorithm," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2010, pp. 1–8.
[45] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.
[46] S. D. Grimshaw, B. J. Collings, W. A. Larsen, and C. R. Hurt, "Eliciting factor importance in a designed experiment," *Technometrics*, vol. 43, no. 2, pp. 133–146, 2001.
[47] J. Lukemire, A. Mandal, and W. K. Wong, "*d*-QPSO: A quantum-behaved particle swarm technique for finding *D*-optimal designs with discrete and continuous factors and a binary response," *Technometrics*, to be published.

WEINAN XU received the B.E. degree in electrical engineering from the National University of Singapore, in 2015, where she is currently pursuing the Ph.D. degree. Her current research interests include many objective optimizations, differential evolution, constraint optimization problems, and applications of evolutionary computation.

WENG KEE WONG is currently a Full Professor with the Department of Biostatistics, Fielding School of Public Health, University of California at Los Angeles. He has published more than 200 refereed articles in medical and statistics journals, including two edited books and a design monograph. He is an elected member of the Delta Omega Public Health Honor Society and the International Statistical Institute and a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science. He is an Associate Editor of *Statistical Methods in Medical Research*, *Communication in Statistics*, *TEST*, and the *Journal of Statistical Planning and Inference*.

KAY CHEN TAN (SM'08–F'14) is currently a Full Professor with the Department of Computer Science, City University of Hong Kong. He has published more than 200 refereed articles and six books. He is an elected member of the IEEE CIS AdCom (2017–2019). He was the Editor-in-Chief of the *IEEE Computational Intelligence Magazine*, from 2010 to 2013. He is the Editor-in-Chief of the IEEE Transactions on Evolutionary Computation. He serves as a member of the Editorial Board for over 20 journals.

JIAN-XIN XU (M'92–SM'98–F'11) received the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 1989. In 1991, he joined the National University of Singapore. He has published more than 200 journal papers and six books in the field of system and control and has supervised 30 Ph.D. students and 15 research associates. Over the years, his cumulative research is supported by nearly 30 research grants. His current research interests include the fields of learning theory, intelligent system and control, nonlinear and robust control, robotics, and precision motion control.