Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

AN EMPIRICAL REVIEW OF MODEL-BASED ADAPTIVE SAMPLING FOR GLOBAL OPTIMIZATION OF EXPENSIVE BLACK-BOX FUNCTIONS

Nazanin Nezami
Hadis Anahideh

Department of Mechanical and Industrial Engineering
University of Illinois at Chicago
842 W Taylor St
Chicago, IL 60607, USA

ABSTRACT

This paper reviews the state-of-the-art model-based adaptive sampling approaches for single-objective black-box optimization (BBO). While BBO literature includes various promising sampling techniques, there is still a lack of comprehensive investigations of the existing research across the vast scope of BBO problems. We first classify BBO problems into two categories of engineering design and algorithm design optimization and discuss their challenges. We then critically discuss and analyze the adaptive model-based sampling techniques focusing on key acquisition functions. We elaborate on the shortcomings of the variance-based sampling techniques for engineering design problems. Moreover, we provide in-depth insights into the impact of the discretization schemes on the performance of acquisition functions. We emphasize the importance of dynamic discretization for distance-based exploration and introduce *EEPA*$^+$, an improved variant of a previously proposed Pareto-based sampling technique. Our empirical analyses reveal the effectiveness of variance-based techniques for algorithm design and distance-based methods for engineering design optimization problems.

1 INTRODUCTION

The task of optimizing a set of design parameters of a black-box system is of great importance whether the parameters are for a real engineering process (Chen et al. 2019), a computer simulation (April et al. 2003), a real scientific experiment (Angermueller et al. 2020), or a complex algorithm (Feurer and Hutter 2019). The term black-box is referred to the underlying structure (derivatives or their approximations) of the objective function that cannot be explicitly defined as it is unknown, complex, or does not exist. In most cases, the black-box function is expensive to compute, as the evaluation time and process to obtain the responses of sample points are prohibitive. The unconstrained black-box optimization (BBO) problem is defined as minimizing or maximizing a black-box function $f(\mathbf{x})$ over a box-constrained input space $\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$, where $\mathbf{x}$ is a d-dimensional input vector of variables, $\mathbf{x}^L$ and $\mathbf{x}^U$ are the lower bound and upper-bound of the range of the variables, and $f(\mathbf{x})$ denotes the unknown expensive objective function. Although we use minimization for mathematical derivations, we consider $-f(\mathbf{x})$ for maximization problems.

Derivative-free optimization (Rios and Sahinidis 2013) is the most widely used group of techniques in BBO literature. Due to the characteristics of the black-box functions, classical gradient-based optimization methods are not applicable in the BBO context. Direct-search methods (Alarie et al. 2021) and Model-based methods (Jiang et al. 2020) are two well-known classes of techniques in derivative-free optimization. The former evaluates function values in a subset of sample points and selects points based on direct comparison of function values without any derivative approximation. The latter, however, considers a surrogate model to estimate the unknown black-box function. The fitted surrogate is either directly optimized (Jones 2001)

or utilized to guide a sampling core through an adaptive approach to select informative points in the design space towards the global optimum.

The model-based adaptive sampling strategy is one of the most effective approaches in BBO literature. Adaptive sampling selects sample points from regions of interest where the *interest* is measured by two critical elements of exploration and exploitation. Exploitation attempts to sample points in regions that may contain global optimum, which can be detected utilizing the properties of the fitted surrogate model in explored regions. The exploration, on the other hand, attempts to find undiscovered regions. Designing a sampling core depends on defining acquisition functions to incorporate exploration and exploitation criteria. In this paper, we categorize acquisition functions into two main groups of functions that have a single explicit mathematical formula including Expected Improvement (EI) (Jones et al. 1998), Probability Improvement (PI) (Kushner 1964), Entropy Search (ES) (Hennig and Schuler 2012), Knowledge Gradient (KG) (Frazier et al. 2008), and Upper Confidence Bound (UCB) (Srinivas et al. 2009), Dynamic Coordinate Search using Response Surfaces (DYCORS) (Regis and Shoemaker 2013), and the multi-criteria Pareto-based including Surrogate Optimization with Pareto Selection (SOP) (Krityakierne et al. 2016) and Exploration Exploitation Pareto Approach (EEPA) (Dickson 2014). In the former, a control hyperparameter linearly aggregates the exploration and exploitation metrics. In the latter, however, a non-dominated set is obtained from a Pareto front construction on the surrogate estimates as the exploitation dimension and a distance metric (comparing evaluated samples and new candidates) as the exploration metric. The Pareto construction provides a more flexible trade-off for exploration and exploitation compared to the restricted linear aggregation and does not require tuning a control parameter.

Furthermore, acquisition functions either use variance-based exploration (EI, PI, ES, KG, and UCB), or distance-based exploration metrics (SOP, DYCORS, and EEPA). Although evaluating points with high uncertainty using variance-based acquisition functions is valuable for identifying undiscovered regions, it requires accurate estimates of the error distribution. Thus, it lacks a global exploration metric. Variance-based methods often converge to local optima because of their limitation in exploring the entire search space, effectively using finite samples. As a result, model-independent exploration metrics seem to be more competent for global exploration, either by defining a weighted score acquisition function (e.g., Regis and Shoemaker (2013)) or by constructing a Pareto front (e.g, Dickson (2014)).

In this paper, we aim to investigate the performance of the most commonly used acquisition functions in model-based adaptive sampling approaches for black-box functions on a wide range of applications and conduct a comprehensive comparison to provide valuable guidelines for BBO researchers and practitioners. We specifically separate the applications by the nature of the problems as **engineering design** optimization and **algorithmic design** optimization. The first group's ultimate goal is to find a globally optimal design by optimizing the performance objective of an unknown complex system, whereas the second group's goal is to optimize the performance of a complex learning function. Examples of **engineering design** problems exist in *airfoil design* (Chen et al. 2019), *circuit design* (Lockwood 2022), *groundwater resources* (Müller et al. 2021), *fluid mechanics* (Forrester et al. 2006), *chemical processes* (Griffiths and Hernández-Lobato 2020), and other science problems such as *biological design* settings (Angermueller et al. 2020). Another class of BBO problems deals with **optimizing algorithmic parameters**. These algorithms can be machine learning models or optimization solvers. Optimization solvers (e.g., CPLEX and GUROBI) consist of a large number of combinations of possible solver options that largely affect the performance of the solver (Liu et al. 2019). Machine learning models have hyperparameters that must be optimized (aka *Hyperparameter Tuning*) where training time is computationally expensive and the solution space of hyperparameters is often intractable. Instances of BBO for hyperparameter tuning are widespread (Wistuba et al. 2016; Feurer and Hutter 2019). Our results support the hypothesis that balancing the exploration-exploitation trade-off is more critical for engineering design and global optimization test problems with highly complex and several local optima compared to algorithmic design problems (e.g., hyperparameter tuning) with less oscillated mappings. The results confirm that the Pareto-based sampling approaches or acquisition functions with explicit emphasis on exploration and exploitation (e.g., UCB) are more suitable for such complex settings.

Other variance-based acquisition functions with an uncontrolled emphasis on exploration and exploitation are more successful for less complex settings such as hyperparameter tuning.

Furthermore, the above-mentioned common acquisition functions are used to sample from a discretized set of generated sample points. Many applications are naturally defined over discrete solution space (e.g., the number of hidden units in Neural networks (Luong et al. 2019)). For problems with continuous variables ($\mathbf{x} \in \mathbb{R}^n$), the assumption of natural discretized solution space existence is violated. There are common approaches in the literature to discretize the representation of the continuous solution space. Design of Experiment space-filling methods (Pronzato and Müller 2012) can be employed in a one-shot manner to predesign the solution space, or in a sequential manner to generate discretized candidate points in each iteration. Smarter discretization strategies such as the Dynamic Dimensioned Search (DDS) approach has been proposed by Tolson and Shoemaker (2007) and adopted in DYCORS by Regis and Shoemaker (2013) to generate high fidelity candidate points that are not necessarily space-filling but rather generated around interesting regions. DDS approach has been shown to be effective in practice compared to one-shot or sequential space-filling approaches. EEPA (Dickson 2014) constructs the Pareto front on a randomly generated discretized set of candidates identifying the non-dominated points. Since the performance of EEPA on a fixed predesigned random set is neither practical nor promising, inspired by DYCORS, we design a dynamic discretization scheme for EEPA in this paper, which significantly improves its performance on a variety of problems. We refer to the advanced EEPA as EEPA$^+$. As explained in Section 1, DYCORS is a score-based sampling with a single explicit formula, whereas EEPA is a multi-criteria Pareto-based sampling. We simply use the dynamic discretization approach that DYCORS employs.

The results provide deep insights into the choice of acquisition function for the problem of interest for practitioners and shed light on the challenges of each group of techniques for BBO researchers to address them appropriately. Our empirically justified observations are as follows:

- Variance-based are mostly outperformed by distance-based acquisition functions for global optimization test problems. For high-dimensional cases, the outperformance is even more noticeable.
- Distance-based acquisition functions are effective in real-world engineering design problems.
- Variance-based acquisition functions are more suitable for hyperparameter tuning problems, although most acquisition functions have close performances.
- Overall, we observed competitive performance for our proposed EEPA$^+$ across all test problems.

In Section 2, we review the generic model-based BBO and the well-known acquisition functions. In Section 3, we provide background on the complexities associated with different BBO problems. Section 4 mainly discusses the critical role of the discretization scheme for sampling. In Section 5, we perform an extensive experimental evaluation for three main categories of problems utilizing different acquisition functions. Section 6 and Section 7 provide discussion on our findings and general guidelines on the performance of each group of techniques for different types of problems.

2 ACQUISITION FUNCTIONS FOR MODEL-BASED ADAPTIVE SAMPLING

A generic model-based adaptive sampling approach follows an iterative sequential sampling. Let $\mathscr{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be the initial set of evaluated points, commonly obtained from a DOE method and $\mathscr{F} = \{f(\mathbf{x}_i) | \mathbf{x}_i \in \mathscr{D}\}$ be the set of function values corresponding to the evaluated points. The adaptive sampling starts with fitting a cheap to evaluate surrogate model, $\hat{f}$, on the evaluated data points. Next, a sample generation scheme (e.g., random) is utilized to discretize the solution space, $\mathscr{X}$. Finally, an acquisition function determines the promising set of candidate points S with the maximum acquisition value to be evaluated next. Note that $\mathscr{D}$ is updated by appending the newly evaluated points in each iteration. The algorithm iterates until the termination criterion is satisfied, which is commonly an evaluation budget.

Gaussian processes (GP) (aka Kriging) (Rasmussen 2004) is one of the most popular surrogate models in Black-Box Optimization that considers the objective function $f(.)$ as the realization of a Gaussian Process. A Gaussian process (GP), $\hat{f}$, models the predicted mean $\hat{\mu}_t(\mathbf{x})$ and the degree of uncertainty $\hat{\sigma}_t(\mathbf{x})$ at any

point $\mathbf{x}$ in the input space, given a set of observations $\mathscr{D}$, where $\mathbf{x}_t$ is the input vector, and $f(\mathbf{x}_t)$ is the corresponding output collected at iteration t. Let X represent the matrix of already evaluated points $\mathscr{D}$, and $\mathbf{f}$ be the vector containing elements of $\mathscr{F}$, the estimated mean $\hat{\mu}(\mathbf{x}) = K(X,\mathbf{x})\Sigma^{-1}\mathbf{f}$ and variance $\hat{\sigma}(\mathbf{x}) = K(\mathbf{x},\mathbf{x}) - K(X,\mathbf{x})\Sigma^{-1}K(\mathbf{x},X)$ are updated according to their equations where Σ represents the covariance matrix and K is a positive semidefinite (PSD) similarity kernel. In this paper, we consider investigating the most commonly used acquisition functions for model-based adaptive sampling of black-box optimization. Let $f_n^o = \max\{f(\mathbf{x}_i)|\mathbf{x}_i \in \mathscr{D}_n\}$ be the maximum observed value after n evaluations. Note that based on GP, at any given point, $\mathbf{x}$, the function is sampled from a normal distribution, $f(\mathbf{x}) \sim \mathscr{N}(\hat{\mu}(\mathbf{x}), \hat{\sigma}^2(\mathbf{x}))$. Proposed by Kushner (1964), **Probability of Improvement (PI)** directly measures the probability of inducing improvement $P(I(\mathbf{x})) = P(f(\mathbf{x}) > f_n^o) = \Phi(\frac{\hat{\mu}(\mathbf{x})-f_n^o}{\hat{\sigma}(\mathbf{x})})$, where $\Phi(\mathbf{x})$ is the standard normal cumulative distribution function. The **Expected Improvement (EI)** criterion gained considerable attention after Jones et al. (1998) with the closed-form expression of $E(I(\mathbf{x})) = (\hat{\mu}(\mathbf{x}) - f_n^o)\Phi(\frac{\hat{\mu}(\mathbf{x})-f_n^o}{\hat{\sigma}(\mathbf{x})}) + \hat{\sigma}(\mathbf{x})\phi(\frac{\hat{\mu}(\mathbf{x})-f_n^o}{\hat{\sigma}(\mathbf{x})})$, where $I(\mathbf{x}) = \max(f(\mathbf{x}) - f_n^o, 0)$ denotes improvement over the current best-known solution at a given point $\mathbf{x}$ and $\phi(\mathbf{x})$ is the standard normal density function. **Upper Confidence Bound (UCB)** is among the most successful acquisition functions, which was developed by Srinivas et al. (2009). The UCB equation combines explicit objectives for exploration $\hat{\sigma}(\mathbf{x})$ and exploitation $\hat{\mu}(\mathbf{x})$ in the format $\alpha_{UCB}(\mathbf{x},\beta_t) = \hat{\mu}(\mathbf{x}) + \beta_t^{0.5}\hat{\sigma}(\mathbf{x})$ through the trade-off hyperparameter β_t, where t denotes the number of function evaluations.

Following the development of information-based acquisition functions such as Entropy Search (ES) in Hennig and Schuler (2012) and Predictive Entropy Search (PES) in Hernández-Lobato et al. (2014), a **Max-value Entropy Search (MES)** approach has been proposed in Wang and Jegelka (2017) which is a relatively cheaper and more robust acquisition objective than its previous counterparts. MES computes the information gain about the maximum value $y^* = f(\mathbf{x}^*)$ instead of the optimal location $\mathbf{x}^*$ through the expression $MES(\mathbf{x}) = H(P(y|\mathbf{x},\mathscr{D}_n,\mathscr{F}_n)) - E[H(P(y|y^*,\mathscr{D}_n,\mathbf{x})))]$. Proposed by Frazier et al. (2008), **Knowledge Gradients (KG)** quantifies the knowledge gained from each candidate point as the conditional expectation of the resulting increase in the posterior mean. Let $\hat{\mu}_n^o = \max_{\forall \mathbf{x}\in\mathscr{X}} \hat{\mu}_n(\mathbf{x})$ be the maximum of posterior mean after n evaluations and $\hat{\mu}_{n+1}^* = \max_{\forall \mathbf{x}\in\mathscr{X}} \hat{\mu}_{n+1}(\mathbf{x})$ be the updated posterior mean after evaluation of $n+1$ point. The knowledge gained from this sampling could be interpreted as the resulting increase in the prediction mean for the maximization problem $(\hat{\mu}_{n+1}^* - \hat{\mu}_n^o)$. As a result, the KG acquisition expression for sampling each point $\mathbf{x}$ for the $n+1$ evaluation could be calculated as the $KG_n(\mathbf{x}) = E_n[\hat{\mu}_{n+1}^* - \hat{\mu}_n^o | x_{n+1} = \mathbf{x}]$.

Most analytical acquisition functions above require evaluating an expectation term over a discrete finite set of candidate points to return the best candidate based on the acquisition values. As a result, they are not directly applicable in a batch setting where joint consideration of sample points is of interest. **Monte Carlo (MC) Batch Sampling** can serve as an appropriate alternative to approximate the batch sampling process. In this paper, we adopt an efficient MC approach (Balandat et al. 2020), which uses a reparameterized representation to derive approximations for the analytical acquisition functions to sample a batch of sample points. We refer to the associated batched strategies as "q-acquisition name" (e.g., qEI) and denote the sequential version where $q = 1$ as the "s-acquisition name" (e.g., sEI).

Originally introduced by Regis and Shoemaker (2007), the **Weighted Score Acquisition** function assigns scores to each candidate point based on a linearly aggregated function over a distance metric $\Delta(\mathbf{x})_{x\in\mathscr{X}} = \min_{x\in\mathscr{X}} \|\mathbf{x} - x\|$ and the estimated surrogate value $\hat{f}(\mathbf{x})$ as explicit exploration and exploitation criteria. Let w_r^t and w_d^t be the predefined weight pattern for $\hat{f}(\mathbf{x})$ and $\Delta(\mathbf{x})_{x\in\mathscr{X}}$ at iteration t ($w_d^t = 1 - w_r^t$). Let $V_d^t(\mathbf{x}) = \frac{\Delta(\mathbf{x})_{max} - \Delta(\mathbf{x})}{\Delta(\mathbf{x})_{max} - \Delta(\mathbf{x})_{min}}$ and $V_r^t(\mathbf{x}) = \frac{\hat{f}(\mathbf{x}) - \hat{f}(\mathbf{x})_{min}}{\hat{f}(\mathbf{x})_{max} - \hat{f}(\mathbf{x})_{min}}$ be the functions to return a normalized score for the surrogate value and the distance criterion. The weighted score acquisition for the minimization problem can be written as, $W^t(\mathbf{x}) = w_d^t \times V_d^t(\mathbf{x}) + w_r^t \times V_r^t(\mathbf{x})$.

Proposed by Dickson (2014), **Exploration Exploitation Pareto Approach (EEPA)** is an algorithm that constructs a Pareto frontier over the unevaluated candidate points using two explicit criteria of their Maximin distance from the already evaluated points for exploration ($\mathbf{x} \in \mathscr{X}$) and the estimated function values for exploitation ($\min_{\mathbf{x}\in\mathscr{X}} \hat{f}(\mathbf{x})$). EEPA utilizes a fixed random set of unevaluated points to represent

the solution space. Lastly, proposed by Krityakierne et al. (2016), **Surrogate Optimization with Pareto Selection (SOP)** constructs a Pareto front based on exploration and exploitation metrics on the already evaluated points, $\mathscr{D}$, instead of unevaluated candidate points (in EEPA) to select a few reference points (i.e., *centers*). The algorithm generates a dynamic discretization of unevaluated sample points through random perturbation of the selected centers. Next, a surrogate-assisted candidate search is performed to select the candidate point with the minimum estimated function value. A multi-layer heuristic approach is proposed to overcome the shortcomings of single Pareto front construction to generate quality sample points. SOP creates a Tabu list to assist with not selecting points that did not significantly improve the hypervolume of the exploration-exploitation trade-off in previous iterations.

3 BlACK-BOX OPTIMIZATION PROBLEMS

Model-based adaptive sampling strategies for Black-box Optimization (BBO) is a well-motivated approach for optimizing expensive to evaluate functions without making specific assumptions about the function behavior. Single objective expensive BBO literature consists of two primary classes of problems; Engineering Design and Algorithm Design optimization. Although the aforementioned problems are black-box with expensive evaluations, the topology of the unknown underlying response surface may vary around the optimal region. To elaborate on this fact, we provide an example by creating the response surface of the mapping between hyperparameters of the Deep Neural Network (DNN) for the MNIST image classification task through an exhaustive search over a dense grid of two main hyperparameters *(number of dense units, learning rate)* in Figure 1(a). Now, comparing the surface of the DNN tuning with a global optimization test problem (Rastrigin) in Figure 1(b), on the surface of DNN tuning, we can observe that there are several regions with objective levels comparable to the global optimizer (noted by the red star). This, however, is not the case for the Rastrigin function, which has numerous identifiable local optima and a single unique minimizer. As a result, regret (i.e., $|f(\mathbf{x}^*) - f(\mathbf{x})|$) is reduced for hyperparameter tuning due to the modest objective differences between local and global optima, resulting in faster convergence. However, the complex multimodal nature of Rastrigin yields a distinct convergence pattern in which regret minimization can not be easily achieved in areas near global optima. As a result, converging to a global optimum is slower for complex global optimization test problems with identifiable local optima.

4 DISCRETIZATION

The generalized assumption of continuous input space in BBO is violated in many cases (Luong et al. 2019). Moreover, most of the acquisition functions are only tractable over a finite discretized solution space representation. Hence, an effective space representation yields a robust search for complex high-dimensional black-box settings. In contrast, a weak representation leads to poor acquisition performance, especially when the black-box is complex and/or higher-dimensional.

Despite the primary role in the model-based adaptive approach, solution space discretization (aka candidate generation) for BBO has only been investigated in a few research articles. Using random distributions (Regis and Shoemaker 2007; Dickson 2014), a DOE method (Golzari et al. 2015), dividing the search space (Alarie et al. 2021), and adaptive grid construction (Bardenet and Kégl 2010) to generate a large set of unevaluated points are common practice yet ineffective in most BBO applications. As a result, more practical discretization schemes such as Monte Carlo (Snoek et al. 2012) and dynamic (Regis and Shoemaker 2013; Müller and Day 2019) discretization have been proposed in the literature. Dynamic discretization methods smartly generate sample points in promising regions and thus, are more effective for high-dimensional complex problems. For example, in Regis and Shoemaker (2013), a dynamic coordinate search strategy, **DYCORS**, has been proposed, which uses a decreased number of the input coordinates to perturb the current best-known solution and generate promising candidate points. Similarly, Müller and Day (2019) utilized a perturbation of the current best-known solution where the perturbation range is drawn from a random normal distribution. Anahideh et al. (2022) introduced a new discretization strategy in

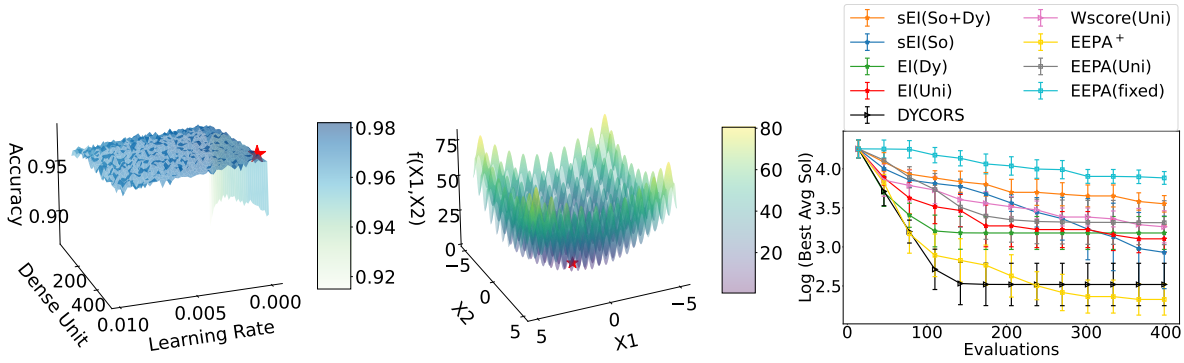
which a set of promising candidates, centroids obtained from Decision Tree partitioning, are dynamically added to a randomly generated set of unevaluated points to improve the representation of the solution space with high-quality sample points.

EEPA$^+$: An Improved Exploration-Exploitation Pareto Approach. Since the original **EEPA** algorithm (Dickson 2014) utilizes a fixed predesigned discretization of unevaluated sample points, in this paper, we first propose an improved version of **EEPA** utilizing a dynamic coordinate discretization scheme similar to **DYCORS** (Regis and Shoemaker 2013), which we refer to it as **EEPA$^+$**. We then empirically investigate the impact of the solution space discretization scheme on the quality of the best-known solution obtained from different acquisition functions. Algorithm 1 represents the main steps of our proposed **EEPA$^+$**. Similar to most BBO strategies, **EEPA$^+$** starts with fitting a surrogate model (e.g., GP) on the already evaluated data points $\mathscr{D}$ followed by a candidate generation (discretization) step to create an effective representation of the solution space. In this step, we dynamically select a subset of coordinates of the best-known solution and perturb to generate new candidate points similar to **DYCORS**. Next, we construct a Pareto frontier on the candidate points based on a *distance* $\Delta(\mathbf{x})_{x\in\mathscr{X}} = \min_{x\in\mathscr{X}} \|\mathbf{x}-x\|$ and *estimated function value* $\hat{f}$ criterion and obtain a non-dominated set of points. To select a batch of points from the non-dominated set, **EEPA$^+$** includes the point with minimum $\hat{f}(\mathbf{x})$ and eliminates the close points by applying a *maximin* measure to identify the most diverse candidates. In general, **EEPA$^+$** can adapt itself to both large and small batch sizes. When the non-dominated set is smaller than the batch size $|\mathscr{C}| \leq q$, it selects all available non-dominated points as the selected subset S and saves the remaining evaluation budget $B-|S|$ for future iterations.

Algorithm 1 EEPA$^+$ Adaptive Sampling

B :Total Budget, q :Batch Size, R :Remained Budget, S : Selected Subset
$\mathscr{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\mathscr{F} = \{f(\mathbf{x}_i) | \mathbf{x}_i \in \mathscr{D}\}$, $R = B$, $S = \varnothing$
while $R > 0$ **do**
  Modeling: Fit a surrogate model $\hat{f}$ (e.g., GP) on $(\mathscr{D}, \mathscr{F})$
  Coordinate selection: Randomly select a decreasing subset of dimensions
  Candidate generation: Perturb the selected coordinates of $\mathbf{x}^o = \arg\min_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x})$ to obtain $\mathscr{X}$
  Pareto frontier: Construct the non-dominated set $\mathscr{C}$ over the $\mathscr{X}$, using $\hat{f}$ and $\Delta(\mathbf{x})_{x\in\mathscr{X}}$
  Sampling: Select a subset of q candidate points S from $\mathscr{C}$
  Evaluation: $\mathscr{F}_S = \{f(\mathbf{x}_i) | \mathbf{x}_i \in S\}$; $\mathscr{D} = \mathscr{D} \cup \mathscr{S}$; $\mathscr{F} = \mathscr{F} \cup \mathscr{F}_S$
end while
Return $\mathbf{x}^o \in \arg\min_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x})$

Figure 1(c) shows the performance of **EI**, **EEPA**, and **Wscore** with different discretization approaches (uniform random (**Uni**, Sobol (**So**) (Owen 1998), and Dynamic coordinate search (**Dy**)) for the 6-dimensional Rastrigin test problem. It is observable that the space discretization schemes have a significant impact on both the convergence pattern and the quality of the final solution obtained by distance-based acquisition functions (**EEPA** and **Wscore**). We refer to **Wscore** acquisition with **Dy** discretization as **DYCORS** in Figure 1 and throughout Section 5 and Section 6. On the other hand, the most widely-used variance-based acquisition function **EI** is less sensitive to the choice of candidate generation strategy. Note that the dynamic discretization approaches may not be necessarily as helpful for variance-based approaches (e.g., **EI**). One justification for this observation is the random perturbation of coordinates of the current best-known solution denoted by **Dy** (proposed in Regis and Shoemaker (2013)), which leads to convergence to local optima with **EI** acquisition function (green line). As discussed earlier, the variance-based acquisition functions without explicit trade-offs between exploration and exploitation have weaker exploration strength, and the reduced subspace search of **Dy** restricts the exploration even further. In this paper, we consider the best observed combination of acquisition functions and discretization.



(a): MNIST-DNN ($d = 2$). (b): Rastrigin $d = 6$. (c): Rastrigin $d = 6$.

Figure 1: Examples of response surfaces and discretization schemes' evaluation

5 EXPERIMENTAL EVALUATION

In this section, we present a comprehensive comparison of the model-based adaptive sampling strategies for Black-Box Optimization. As discussed in Section 2, our baselines include sequential as well as the batch version of popular acquisitions for global optimization test problems. Hence, we consider **sEI** and **qEI**, **sPI** and **qPI**, **sUCB** and **qUCB**, **sMES** and **qMES**, **qKG** denoting the sequential and batch versions of the discussed acquisition functions, respectively. We mainly focus on the batch version for real-world applications and hyperparameter tuning test problems due to their practicality. Other considered baselines, **SOP**, **EEPA**$^+$ and **DYCORS** are batch sampling techniques. In particular, for **DYCORS** we utilize the synchronous batch sampling proposed in Eriksson et al. (2019). For the surrogate model, we use GP with a constant mean and a Matérn Kernel.

5.1 Experiment Setup

We first consider three commonly used synthetic optimization test problems *Rosenbrock*, *Rastrigin*, and *Levy* (Surjanovic, S. and Bingham, D. 2013) for two low-dimensional ($d = 6$) and high-dimensional ($d = 30$) cases. Our purpose is to cover different topological characteristics such as shape and the number of local optimums. *Rosenbrock* function is a unimodal valley-shaped function, *Rastrigin* is a multimodal and complex function encompassing extremely distinct local minima, and the *Levy* function is challenging due to having several local minima despite its fast convergence to the near-optimal area. For each synthetic test problem, we use Latin Hypercube Design (LHD) (Joseph and Hung 2008) with maximin criterion to generate 30 different initial sets of labeled points with the size $2 * (d + 1)$ and replicate our evaluation 30 times. We consider a batch size of $|S| = 4$ across all experiments for this group of test problems. When $d = 6$, the budget and discretized solution space size are 400 and 1000, respectively. When $d = 30$, we use a budget of 500, and discretized space size is 5000.

We investigate two well-known benchmark problems in science and engineering domains, **DNA binding** and **Airfoil design** optimization, as BBO baselines for real-world applications. For **DNA binding**, Barrera et al. (2016) utilizes protein-binding microarrays to assess the DNA binding activity of all feasible length-8 DNA sequences on 201 protein (human transcription factors) targets. We consider CRX and VSX1 protein binding problems similar to Hashimoto et al. (2018). The evaluation budget and size of the initial set are 1000 and 18, respectively. We consider a batch size of 100, similar to Hashimoto et al. (2018), and replicate the results 10 times for each different initialization obtained from the LHD technique. For **Airfoil design**, we use the UIUC database (Selig 1996), which provides the geometries of approximately 1600 airfoil designs. The designs are then evaluated by the XFOIL 2-D aerodynamics simulator (Drela 1989). The objective is to maximize the lift divided by the drag fraction. We consider a 10-dimensional representation of airfoil

for perturbation and candidate generation based on NURBS parameterization proposed in Viswanath et al. (2011). The evaluation budget and size of the initial set are 300 and 1, respectively. In each iteration, we evaluate 30 airfoils ($|S| = 30$), and replicate the results 10 times under the same initial point.

Finally, we consider the MNIST image classification with Logistic Regression (**MNIST-Logit**), Deep Neural Network (**MNIST-DNN**), and Convolutional Neural Network (**MNIST-CNN**) for hyperparameter optimization. The evaluation budget and batch size equal 100 and 4, respectively, and we replicate 10 times for 10 different initial sets of size $d+1$ points where d is the number of hyperparameters. For **MNIST-Logit**, we consider *batch size*, *learning rate*, *L2 regularization*, and *number of epochs* to be optimized. For **MNIST-DNN**, we additionally consider *number of Dense units in the 1st layer and the 2nd layer, and dropout rate*. Furthermore, for **MNIST-CNN**, we consider *number of units in the convolution layer, and kernel size* besides the above-mentioned hyperparameters. The batch size range of $(32,512)$, learning rate $(0,0.1)$, L2 regularization $(0,1)$, number of dense/convolution units $(32,512)$, kernel size of $(2,9)$, dropout rate $(0,0.9)$, and number of epochs $(0,10)$ have been considered in all problems.

6 RESULTS

Figure 2 demonstrates the performance of our considered baselines for the global optimization test problems plotted every 8 iterations. It is worth noting that the performance of acquisition functions diminishes as the dimension and complexity of problems increase. In general, distance-based acquisition functions (**EEPA**$^+$, **SOP**, **DYCORS**) outperform variance-based acquisition functions for these problems, as a stronger global exploration is necessary for such complex functions. Note that for higher-dimensional problems, there are observable performance gaps between distance-based and variance-based baselines, making the choice of distance-based approaches even more practical. In Figure 2(e), **qUCB** has a competitive performance with **EEPA**$^+$**, SOP, DYCORS**. Although **qUCB** is a variance-based acquisition function, it considers an explicit trade-off between exploration and exploitation criteria through the control parameter β_t. Furthermore, **qKG** is inapplicable for high-dimensional test cases due to the expensive computation.

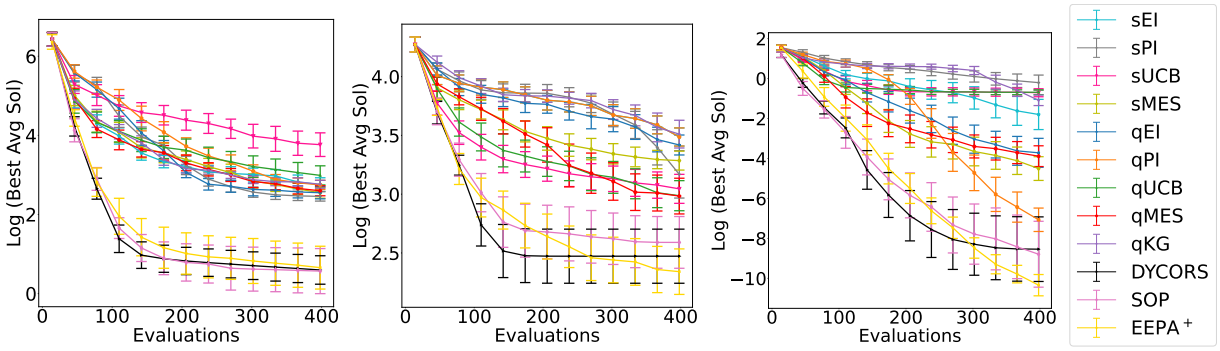
In Figure 3, we illustrate the performance of the considered acquisition functions for real-world optimization test problems **DNA binding** and **Airfoil design**. Overall, we can observe a similar pattern as in global optimization test problems comparing distance-based versus variance-based acquisitions except for the **SOP**. Although **SOP** has a promising performance for synthetic test problems, it lacks adequate flexibility to maintain its exploration power when batch size is large (e.g., 100). If the batch size is greater than the number of centers of **SOP**, it requires manipulation to adjust the size of the selected samples from each cluster, which adversely affects the performance of the algorithm. Moreover, **SOP** considers multiple heuristic steps to filter out non-promising centers that need adjustments to be suitable for real-world applications. We can observe that distance-based approaches such as **EEPA**$^+$ and **DYCORS** are still effective in this case. Note that **qUCB**, has relatively promising convergence performance due to its adaptive consideration of the exploration-exploitation trade-off parameter (β_t).

Figure 4 presents our empirical study on MNIST image classification task (per 8 iterations). The results indicate that the variance-based acquisition functions are more competitive for hyperparameter optimization. In particular, our proposed **EEPA**$^+$ has competitive performance in all three hyperparameter tuning problems. **qKG** also outperforms many other strategies for both **MNIST-DNN** and **MNIST-CNN**. Unlike synthetic test cases, **SOP** and **DYCORS** are not as competitive for hyperparameter tuning. Although **DYCORS** is in a reasonable range of performance (specifically in **MNIST-CNN**), its wide confidence intervals indicate high variability. The results indicate that our considered baselines are equally good for hyperparameter tuning problems since the response surface is easier to optimize as discussed in Section 3.

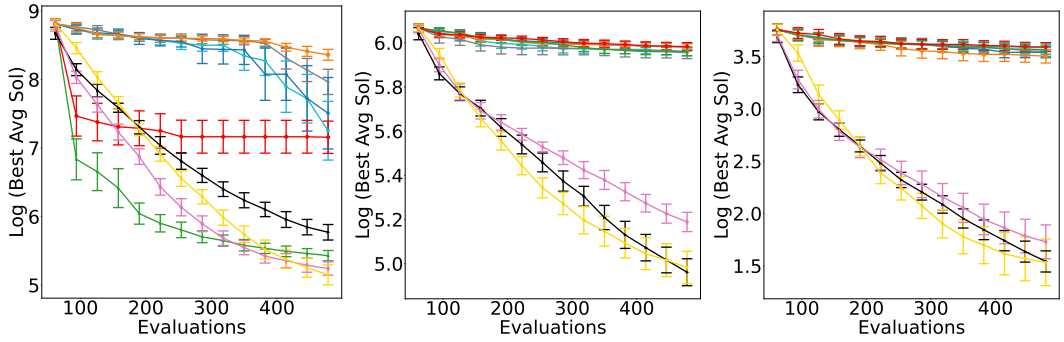
Finally, we assign a rank to each considered acquisition function in Figure 5 based on the mean and variance of the final best-known solution obtained across all replications to summarize the results. We can confirm that the variance-based acquisition functions are outperformed by the distance-based ones for complex optimization settings (both real applications and synthetic test problems), which makes distance-based approaches, specifically **EEPA**$^+$, a more viable choice. For hyperparameter tuning problems, we

can observe a more competitive performance of variance-based acquisition functions. In particular, **qKG** is competitive to **EEPA**$^+$. However, our proposed **EEPA**$^+$ outperforms most of the acquisition functions across all test problems. **EEPA**$^+$ provides greater flexibility for the exploration-exploitation trade-off through the non-dominated set construction and the dynamic discretization scheme.

The average running time of **EEPA**$^+$ for **DNA binding** problems are 1200 and 2400 sec, which is faster than the variance-based strategies with $2800-6000$ sec. The performance of **EEPA**$^+$ for **Airfoil design** is close to the performance of variance-based methods with an average time of 830 sec.

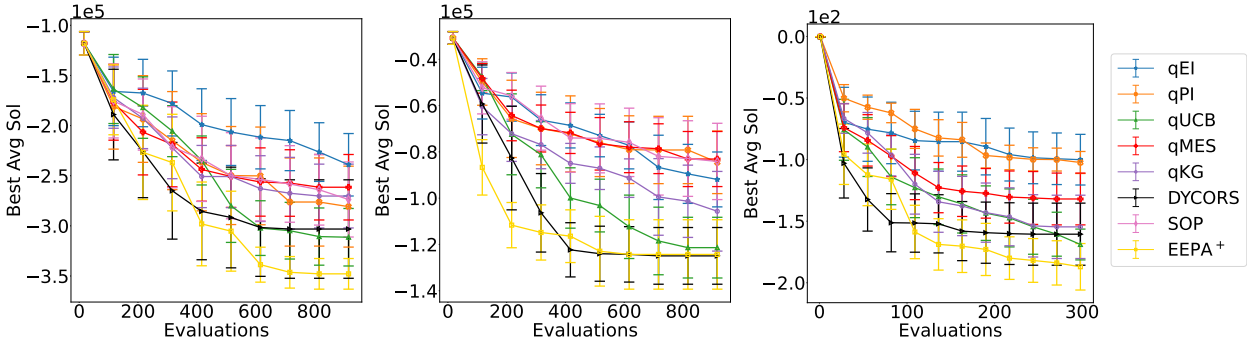


(a) RosenBrock $d=6$ (b) Rastrigin $d=6$. (c) Levy $d=6$.



(e) RosenBrock $d=30$. (f) Rastrigin $d=30$ (g) Levy $d=30$.

Figure 2: Performance of different sampling techniques for global optimization test problems.

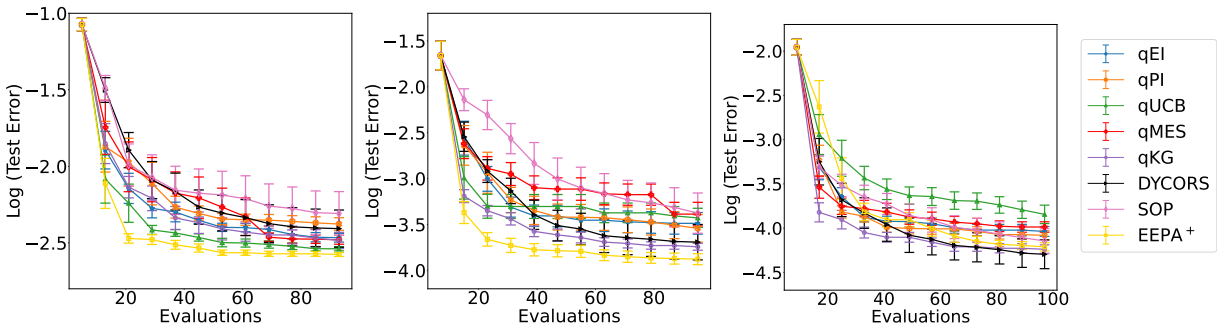


(a) Binding to CRX Protein. (b) Binding to VSX1 Protein. (c) Airfoil Design.

Figure 3: Performance of different sampling techniques for real-world optimization problems.

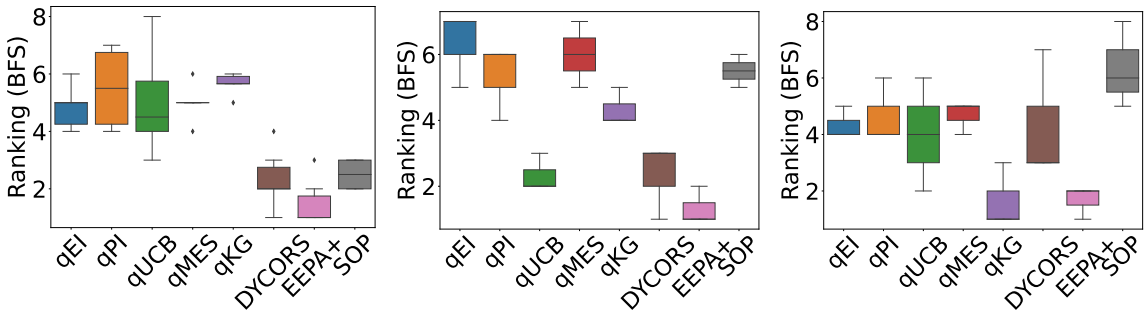
7 FINAL REMARKS

In this paper, we introduced new perspectives on evaluating the exploration component of acquisition functions and the discretization scheme in various settings. Our empirical and analytical analysis revealed that variance-based acquisitions (e.g., **qKG**) are more suitable for hyperparameter tuning settings while multi-criteria Pareto-based (e.g., **EEPA**$^+$), and weighted-score acquisition with dynamic discretization scheme (e.g., **DYCORS**), are more promising in finding the optimal regions of in high-dimensional complex optimization problems. Moreover, considering *distance-based* exploration metrics in acquisition functions yields promising results for engineering design problems. We noticed that some existing adaptive acquisition functions lack the adequate flexibility of being conveniently adopted for different problems. **SOP** with a large batch size (which is necessary for certain problems) is not applicable and requires further adjustments. The well-known **qKG** is computationally inefficient with a large finite discretization scheme. Furthermore, an explicit exploration-exploitation trade-off, similar to **qUCB**, can significantly improve the performance of variance-based approaches. However, tuning the critical control parameter for weighted score acquisition functions including **qUCB** is a challenging task and requires extensive investigation. In this paper, we also investigated the impact of the discretization scheme on the performance of different acquisition functions. We discovered that discretization has a greater influence on distance-based acquisition functions than variance-based acquisition functions. Moreover, we noted that a dynamic discretization scheme is an effective approach for improving the performance of distance-based acquisition functions. In light of this, we proposed an enhanced modification of an existing Pareto-based acquisition algorithm supported by empirical effectiveness in different settings.



(a) MNIST-Logit. (b) MNIST-DNN. (c) MNIST-CNN.

Figure 4: Performance of different sampling techniques for hyperparameter optimization problems.



(a) Synthetic Optimization. (b) Real Optimization. (c) Hyperparameter Optimization.

Figure 5: Ranking of BBO strategies across different problems.

REFERENCES

Alarie, S., C. Audet, A. E. Gheribi, M. Kokkolaras, and S. Le Digabel. 2021. "Two Decades of Blackbox Optimization Applications". *EURO Journal on Computational Optimization* 9:100011.

Anahideh, H., J. Rosenberger, and V. Chen. 2022. "High-dimensional Black-box Optimization Under Uncertainty". *Computers & Operations Research* 137:105444.

Angermueller, C., D. Belanger, A. Gane, Z. Mariet, D. Dohan, K. Murphy, L. Colwell, and D. Sculley. 2020, 13–18 Jul. "Population-Based Black-Box Optimization for Biological Sequence Design". In *Proceedings of the 37th International Conference on Machine Learning*, edited by H. D. III and A. Singh, Volume 119, 324–334. Virtual Event: Journal of Machine Learning Research.

April, Glover, Kelly, and Laguna. 2003. "Practical Introduction to Simulation Optimization". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. Sanchez, D. Ferrin, and D. Morrice, Volume 1, 71–78. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers Inc.

Balandat, M., B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy. 2020. "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization". In *Proceedings of the 33th Conference on Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Volume 33, 21524–21538. New York: Curran Associates Inc.

Bardenet, R., and B. Kégl. 2010. "Surrogating the Surrogate: Accelerating Gaussian-process-based Global Optimization with a Mixture Cross-entropy Algorithm". In *27th International Conference on Machine Learning 2010*, edited by T. J. Johannes Fürnkranz, 55–62. Madison: Omnipress.

Barrera, L. A., A. Vedenko, J. V. Kurland, J. M. Rogers, S. S. Gisselbrecht, E. J. Rossin, J. Woodard, L. Mariani, K. H. Kock, S. Inukai et al. 2016. "Survey of Variation in Human Transcription Factors Reveals Prevalent DNA Binding Changes". *Science* 351(6280):1450–1454.

Chen, W., K. Chiu, and M. Fuge. 2019. "Aerodynamic Design Optimization and Shape Exploration Using Generative Adversarial Networks". In *American Institute of Aeronautics and Astronautics Scitech 2019 Forum*, 2351. San Diego, California: American Institute of Aeronautics and Astronautics.

Dickson, J. F. 2014. *An Exploration and Exploitation Pareto Approach to Surrogate Optimization*. Ph.D. thesis, Department of Industrial and Manufacturing Systems, University of Texas at Arlington, Arlington, Texas. https://rc.library.uta.edu/uta-ir/handle/10106/24947, accessed 09.01.2022.

Drela, M. 1989. "XFOIL: An Analysis and Design System for Low Reynolds Number Airfoils". In *Low Reynolds Number Aerodynamics*, edited by T. J. Mueller, Volume 54, 1–12. New York: Springer.

Eriksson, D., D. Bindel, and C. A. Shoemaker. 2019. "PySOT and POAP: An Event-driven Asynchronous Framework for Surrogate Optimization". https://arxiv.org/abs/1908.00420, accessed 09.01.2022.

Feurer, M., and F. Hutter. 2019. "Hyperparameter Optimization". In *Automated Machine Learning*, edited by L. K. Frank Hutter and J. Vanschoren, 3–33. Cham: Springer, Cham.

Forrester, A. I., N. W. Bressloff, and A. J. Keane. 2006. "Optimization Using Surrogate Models and Partially Converged Computational Fluid Dynamics Simulations". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 462(2071):2177–2204.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-gradient Policy for Sequential Information Collection". *Society for Industrial and Applied Mathematics Journal on Control and Optimization* 47(5):2410–2439.

Golzari, A., M. H. Sefat, and S. Jamshidi. 2015. "Development of An Adaptive Surrogate Model for Production Optimization". *Journal of Petroleum Science and Engineering* 133:677–688.

Griffiths, R.-R., and J. M. Hernández-Lobato. 2020. "Constrained Bayesian Optimization for Automatic Chemical Design Using Variational Autoencoders". *Chemical Science* 11(2):577–586.

Hashimoto, T., S. Yadlowsky, and J. Duchi. 2018. "Derivative free Optimization via Repeated Classification". In *International Conference on Artificial Intelligence and Statistics*, edited by A. Storkey and F. Perez-Cruz, Volume 84, 2027–2036. Playa Blanca, Lanzarote: Proceedings of Machine Learning Research.

Hennig, P., and C. J. Schuler. 2012. "Entropy Search for Information-Efficient Global Optimization". *Journal of Machine Learning Research* 13(6):1809–1837.

Hernández-Lobato, J. M., M. W. Hoffman, and Z. Ghahramani. 2014. "Predictive Entropy Search for Efficient Global Optimization of Black-box Functions". In *Proceedings of the 27th Conference on Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Volume 1, 918–926. Cambridge: MIT Press.

Jiang, P., Q. Zhou, and X. Shao. 2020. "Surrogate-model-based Design and Optimization". In *Surrogate Model-Based Engineering Design and Optimization*, edited by S.-B. C. . H. D. . Y. F. . C. G. . V. . J.-Q. S. . Y. W. Kwon, 135–236. Gateway East, Singapore: Springer Singapore.

Jones, D. R. 2001. "A Taxonomy of Global Optimization Methods Based on Response Surfaces". *Journal of Global Optimization* 21(4):345–383.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455–492.

Joseph, V. R., and Y. Hung. 2008. "Orthogonal-maximin Latin Hypercube Designs". *Statistica Sinica* 18(1):171–186.

Krityakierne, T., T. Akhtar, and C. A. Shoemaker. 2016. "SOP: Parallel Surrogate Global Optimization with Pareto Center Selection for Computationally Expensive Single Objective Problems". *Journal of Global Optimization* 66(3):417–437.

Kushner, H. J. 1964. "A New Method of Locating the Maximum Point of An Arbitrary Multipeak Curve in the Presence of Noise". 86(1):97–106.

Liu, J., N. Ploskas, and N. V. Sahinidis. 2019. "Tuning BARON Using Derivative-free Optimization Algorithms". *Journal of Global Optimization* 74(4):611–637.

Lockwood, O. 2022. "An Empirical Review of Optimization Techniques for Quantum Variational Circuits". https://arxiv.org/abs/2202.01389, accessed 09.01.2022.

Luong, P., S. Gupta, D. Nguyen, S. Rana, and S. Venkatesh. 2019. "Bayesian Optimization with Discrete Variables". In *Australasian Joint Conference on Artificial Intelligence*, edited by J. Liu and J. Bailey, 473–484. Cham: Springer, Cham.

Müller, J., and M. Day. 2019. "Surrogate Optimization of Computationally Expensive Black-box Problems with Hidden Constraints". *INFORMS Journal on Computing* 31(4):689–702.

Müller, J., J. Park, R. Sahu, C. Varadharajan, B. Arora, B. Faybishenko, and D. Agarwal. 2021. "Surrogate Optimization of Deep Neural Networks for Groundwater Predictions". *Journal of Global Optimization* 81(1):203–231.

Owen, A. B. 1998. "Scrambling Sobol'and Niederreiter–Xing Points". *Journal of Complexity* 14(4):466–489.

Pronzato, L., and W. G. Müller. 2012. "Design of Computer Experiments: Space Filling and Beyond". *Statistics and Computing* 22(3):681–701.

Rasmussen, C. E. 2004. *Gaussian Processes In Machine Learning*, 63–71. Berlin, Heidelberg: Springer Berlin Heidelberg.

Regis, R. G., and C. A. Shoemaker. 2007. "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions". *INFORMS Journal on Computing* 19(4):497–509.

Regis, R. G., and C. A. Shoemaker. 2013. "Combining Radial Basis Function Surrogates and Dynamic Coordinate Search in High-dimensional Expensive Black-box Optimization". *Engineering Optimization* 45(5):529–555.

Rios, L. M., and N. V. Sahinidis. 2013. "Derivative-free Optimization: A Review of Algorithms and Comparison of Software Implementations". *Journal of Global Optimization* 56(3):1247–1293.

Selig, M. 1996. *UIUC Airfoil Data Site*. Urbana, Illinois: Department of Aeronautical and Astronautical Engineering University of Illinois at Urbana-Champaign. https://m-selig.ae.illinois.edu/ads/coord_database.html, accessed 09.01.2022.

Snoek, J., H. Larochelle, and R. P. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms". Volume 2, 2951–2959. New York: Curran Associates Inc.

Srinivas, N., A. Krause, S. M. Kakade, and M. Seeger. 2009. "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design". https://arxiv.org/abs/0912.3995, accessed 09.01.2022.

Surjanovic, S. and Bingham, D. 2013. "Virtual Library of Simulation Experiments: Test Functions and Datasets". Retrieved September 30, 2021, from http://www.sfu.ca/~ssurjano.

Tolson, B. A., and C. A. Shoemaker. 2007. "Dynamically Dimensioned Search Algorithm for Computationally Efficient Watershed Model Calibration". *Water Resources Research* 43(1).

Viswanath, A., A. I. Forrester, and A. Keane. 2011. "Dimension Reduction for Aerodynamic Design Optimization". *American Institute of Aeronautics and Astronautics Journal* 49(6):1256–1266.

Wang, Z., and S. Jegelka. 2017. "Max-value Entropy Search for Efficient Bayesian Optimization". In *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh, Volume 70, 3627–3635. Sydney, Australia: Journal of Machine Learning Research.

Wistuba, M., N. Schilling, and L. Schmidt-Thieme. 2016. "Hyperparameter Optimization Machines". In *2016 Institute of Electrical and Electronics Engineers International Conference on Data Science and Advanced Analytics*, edited by O. R. Zaiane and S. Matwin, 41–50. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers Inc.

AUTHOR BIOGRAPHIES

Nazanin Nezami is a third-year PhD Student in Mechanical and Industrial Engineering Department at the University of Illinois at Chicago (UIC). She obtained her M.S. degree in Industrial Engineering from University of Minnesota-Twin Cities. Her main research interests are Black-Box Optimization, Machine Learning (ML), and Fairness in ML. Her email address is nnezam2@uic.edu.

Hadis Anahideh is a Research Assistant Professor of the Mechanical and Industrial Engineering Department at the University of Illinois at Chicago. She received her Ph.D. degree in Industrial Engineering from the University of Texas at Arlington. Her research objectives center around Black-box Optimization, Sequential Optimization, Active Learning, Statistical Learning, Explainable AI, and Algorithmic Fairness. Her email address is hadis@uic.edu and her homepage is https://mie.uic.edu/profiles/anahideh-hadis/.