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.

ACHIEVING DIVERSITY IN OBJECTIVE SPACE FOR SAMPLE-EFFICIENT SEARCH OF MULTIOBJECTIVE OPTIMIZATION PROBLEMS

Eric Hans Lee
Bolong Cheng
Michael McCourt

SigOpt, an Intel Company
San Francisco, CA, USA

ABSTRACT

Efficiently solving multi-objective optimization problems for simulation optimization of important scientific and engineering applications such as materials design is becoming an increasingly important research topic. This is due largely to the expensive costs associated with said applications, and the resulting need for sample-efficient, multiobjective optimization methods that efficiently explore the Pareto frontier to expose a promising set of design solutions. We propose moving away from using explicit optimization to identify the Pareto frontier and instead suggest searching for a diverse set of outcomes that satisfy user-specified performance criteria. This method presents decision makers with a robust pool of promising design decisions and helps them better understand the space of good solutions. To achieve this outcome, we introduce the Likelihood of Metric Satisfaction (LMS) acquisition function, analyze its behavior and properties, and demonstrate its viability on various problems.

1 INTRODUCTION

Simulation is a fundamental element to many product and system development processes. As mathematical, statistical, and machine learning algorithms leverage increasingly powerful computational hardware to perform elaborate tasks, simulation has grown to play a key role in fields such as materials science, operations research, industrial engineering, aerodynamics, pharmaceuticals, image processing, and many others. In particular, a key use of these simulations is to serve as a surrogate for the eventual implementation and/or manufacturing during the design optimization; running a computational simulation is likely much cheaper than actually conducting a physical experiment or fabrication (Forrester et al. 2008; Negoescu et al. 2011; Molesky et al. 2018; Haghanifar et al. 2020).

Computational simulations can, however, easily run for hours or days, making simulation itself an often costly proposition. The high cost of a single simulation is compounded by the frequent need to simulate many different systems to search for a set of desirable outcomes. This is the motivating force behind simulation optimization, which seeks to identify suitable system parameters to achieve a satisfactory system or effective simulation in a *sample-efficient* fashion, i.e., with as few simulations conducted as possible.

In practical situations, simulations almost always have multiple competing objectives which define success, and thus it is important for users to understand trade-offs between these competing objectives in order to make an informed design decision. Multiobjective optimization tackles this problem by identifying the *Pareto frontier*, which is the manifold in objective space such that improving one objective cannot occur without harming another. Unfortunately, using the Pareto frontier as the measurement of success may be limiting in engineering and design applications. Simulations always yield some gap from reality (both due to imprecision in the simulation and errors in the modeling itself) meaning that even if we could perfectly optimize the simulation, the eventual performance of the real, physical system would still suffer from a loss in optimality. This ubiquitous problem in single-objective optimization is more pronounced in the multiobjective setting. The Pareto frontier is sensitive to these ever-present inaccuracies, and as a result, candidates near the Pareto frontier, unaccounted for in most optimization literature, are still of scientific interest in the real world. For example, (del Rosario et al. 2020) introduces the notion of *Pareto shells*, the set of near Pareto optimal solutions, as the desired outcome of multiobjective optimization.

This sensitivity to both simulation noise and model error make the Pareto frontier inadequate for many practitioners, who seek a more robust method for efficient design-of-experiments. The purpose of this paper is to help bridge this discrepancy between theory and practice by offering an alternative formulation to multiobjective optimization, known as *constraint active search* (CAS) (Malkomes et al. 2021). The purpose of CAS is to account for the three factors described above: sample-efficiency, multiple objectives, and simulation inaccuracy. However, instead of explicit optimization, CAS searches for a diverse set of satisfactory points which can be considered for eventual manufacturing. In (Malkomes et al. 2021), the authors introduced CAS and tackled diversity in parameter space with the expected coverage improvement (ECI) acquisition function. CAS has been successfully used to help design nanostructured glass with multiple desirable optical and physical characteristics, in which it sought a diverse set of candidates in parameter space that fulfilled user-specified performance thresholds.

In this article, we propose an alternate definition of diversity to guide constraint active search. Rather than using diversity in parameter space (the input space), we seek diversity in *objective space* (the output space). Doing so gives us the ability to have greater understanding as to how the competing objectives are likely to interact among satisfactory configurations; this, in turn, gives us more confidence about the eventual manufacturing or deployment process. Our paper makes the following concrete contributions:

- We propose an alternative approach to multiobjective optimization, in which we solicit minimum performance thresholds and search for a diverse set of objective values that meet these thresholds.
- We use multi-output Gaussian process modeling combined with a novel acquisition function called Likelihood of Metric Satisfaction (LMS) to search for a diverse set of feasible objective values in a sample-efficient manner.
- We demonstrate the effectiveness on a set of synthetic problems as well as a real-world nuclear fusion simulation optimization problem.
- We identify key practical steps to take when using CAS-LMS to achieve diversity in objective space in adverse and/or complicated circumstances.

2 BACKGROUND

2.1 Bayesian Optimization

One of the best-known stochastic surrogate optimization methods for expensive, black functions is Bayesian optimization (BO). BO consists of two core components: a probabilistic model (commonly referred to as a *surrogate model*) that models the objective function and an *acquisition function* that uses the model to determine where next to sample.

BO research can be traced back to the Efficient Global Optimization (EGO) algorithm (Jones et al. 1998), which combined a Gaussian Process model with the Expected Improvement acquisition function (Mockus et al. 1978). Recent research has proposed the use of other probabilistic models, such as kernel density estimators (Bergstra et al. 2011) or random forests (Hutter et al. 2011). In practice, Gaussian processes (Rasmussen and Williams 2005) have become the predominant surrogate model of choice in the BO community, bolstered by the proliferation of modern GP software such as GPy (GPy 2012) and GpyTorch (Gardner et al. 2018). A significant portion of the BO literature focuses on proposing new acquisition functions, including Upper Confidence Bound (Srinivas et al. 2010), Knowledge Gradient (Scott et al. 2011), and Entropy Search (Villemonteix et al. 2009). Each of these methods have different treatment of the exploration exploitation tradeoff. As a result, there is no one default acquisition function agreed upon in practice, and researchers continue to develop new variations of these aforementioned acquisition functions for different applications. We refer the readers to (Garnett 2022) for a modern comprehensive review of BO.

2.2 Multiobjective Bayesian Optimization

The earliest attempt at adapting BO to multiobjective optimization is the Pareto Efficient Global Optimization (ParEGO) algorithm (Knowles 2006), which uses the aforementioned EGO algorithm to optimize a linear combination of the multiple objectives. This so-called linear scalarization does not work well when the Pareto frontier is non-convex in the objective space. Second, this method lacks interpretability in practice; finding the appropriate weighting of multiple objectives is a nontrivial task.

A separate approach leverages work from constrained Bayesian optimization literature (Gardner et al. 2014), by reformulating the problem as a constrained optimization problem, commonly known as the ε-constraint method. In this setting, the objectives are treated as inequality constraints of the form $f(\mathbf{x}) \leq \varepsilon$, for some known threshold ε. These constraints often appear naturally in real world applications, such as baseline performance metrics (Haghanifar et al. 2020).

Recent effort in multiobjective BO focuses on directly improving the hypervolume of the Pareto frontier directly (Emmerich et al. 2011; Daulton et al. 2020; Daulton et al. 2021). In particular, (Daulton et al. 2020) exploits modern parallel hardware (such as GPUs) to efficiently compute the expected hypervolume. Similar to ε-constraint methods, these hypervolume improvement methods also require knowing *a priori* thresholds on the objective values.

2.3 Active Search and Feasibility Determination

Our proposed work is also related to the topics of *active search* (Garnett et al. 2012; Jiang et al. 2017) and *feasibility determination* (Szechtman and Yücesan 2008; Szechtman and Yücesan 2016). Active search can be seen as a special case of Bayesian optimization, where one has binary observations and cumulative reward. The goal of active search is to sequentially discover members of a rare, desired class. Inspecting any element is assumed to be expensive, representing, for instance, the cost of performing a real world laboratory experiment.

2.4 Constraint Active Search with Parameter Diversity

Constraint active search is an alternate formulation of the multiobjective design problem first proposed in (Malkomes et al. 2021). In lieu of identifying the highest performing design (in a single objective setting) or the Pareto frontier (in a multiobjective setting), success is defined as the identification of multiple outcomes in a “satisfactory region”. User-defined upper or lower constraints on each objective implicitly define the satisfactory region (the region of satisfactory performance); we adopt the same structure. Designs which satisfy these constraints are termed satisfactory.

CAS proceeds in a sequential fashion, similarly to Bayesian optimization: first, each objective is modeled using the available data (we use a Gaussian process, though other models are viable); second, a next design at which to sample is suggested through optimization of an acquisition function. (Malkomes et al. 2021) introduced the expected coverage improvement (ECI) acquisition function – ECI is maximized for designs which have high probability of satisfying the user-defined constraints but also are sufficiently distinct from previously observed satisfactory designs in parameter space.

3 CONSTRAINT ACTIVE SEARCH WITH OBJECTIVE VALUE DIVERSITY

Given a target range of objective values, we seek to generate a diverse set of outcomes in the objective space that presents decision makers with a robust set of promising designs, so that they may better understand the space of good solutions for the problem at hand. We want to do this in a sample-efficient manner, which

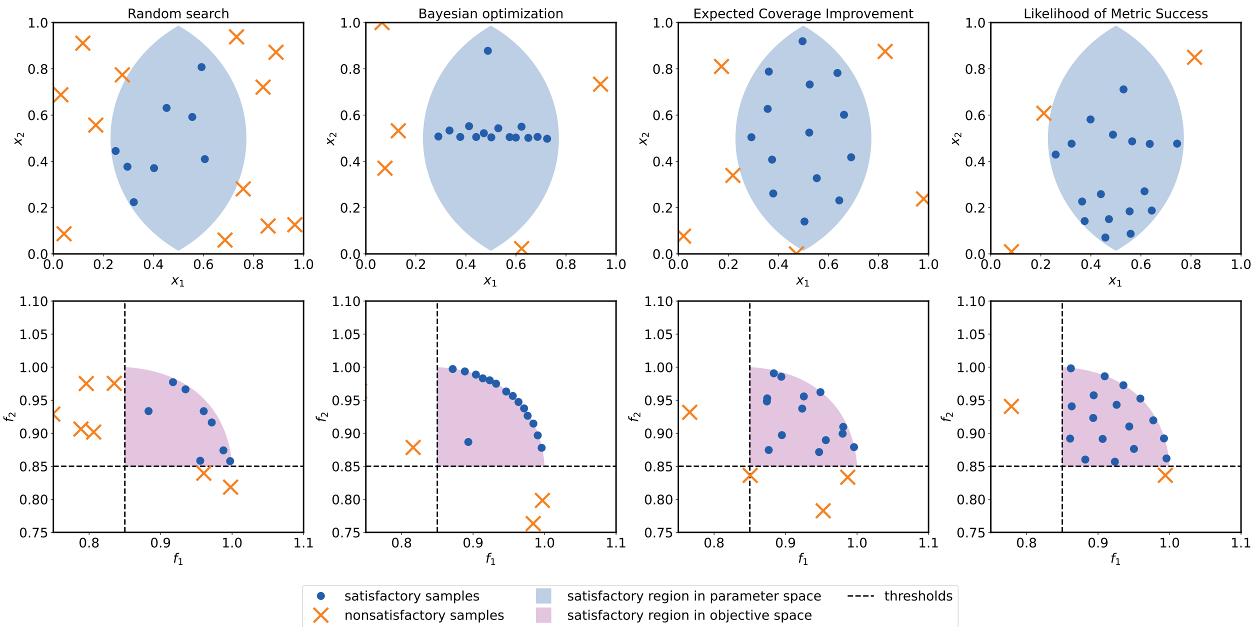

Figure 1: We compare random search, BO, CAS in parameter space, and CAS in objective space on the same multiobjective problem. The top row shows the observed samples in the parameter space (blue); the bottom row, in the objective space (red). BO seeks to isolate the Pareto frontier, leading to a tight concentration of points in both parameter and objective space. CAS in parameter space unsurprisingly achieves diversity in parameter space, but at the cost of poor spread in objective space. Spreading points out in objective space —CAS with objective value diversity— is the topic of this paper.

presents two additional challenges that are not present in the parameter space formulation of constraint active search.

- **Objective Uncertainty:** Achieving diversity in objective space is subject to uncertainty because we do not know prior to evaluating the objective what those values will be. This stands in contrast to achieving parameter diversity, in which no uncertainty exists in the parameter values.
- **Objective Heteroskedasticity:** Sampling the objective uniformly across parameter space will very likely yield a non-uniform distribution of objective values. For example, a substantial portion of the domain in parameter space may map to a very concentrated cluster in objective space, making it difficult for an algorithm to spread objective values out efficiently.

To address these challenges, we propose a novel acquisition function criterion —which we call Likelihood of Metric Satisfaction (LMS)— to select a suitable next evaluation during constraint active search. LMS is defined over our optimization domain $\mathcal{X}$ and the value $\mathrm{LMS}(\mathbf{x})$ quantifies how much $\mathbf{x}$ might improve the diversity in objective values we seek.

We outline our constraint active search procedure with the following general steps:

1. Build a multi-output Gaussian process surrogate to model the observation data.
2. Choose the next evaluation to be $\mathbf{x}^* = \arg\max_{\mathcal{X}} \mathrm{LMS}(\mathbf{x})$.
3. Evaluate $\mathbf{y} = f(\mathbf{x})$, update the observation data, and repeat until budget exhausted.

In the following subsections, we lay out a more precise problem statement, describe the Gaussian process (GP) model, and then explain LMS at length.

3.1 Notation and Problem Statement

Suppose we want to search for design configurations in a d-dimensional, compact search space $\mathcal{X} \subset \mathbb{R}^d$. We may judge the quality of a design $\mathbf{x} \in \mathcal{X}$ by evaluating m expensive black-box objective functions $f_1, f_2, \ldots f_m$, each mapping $\mathcal{X}$ to $\mathbb{R}$. We seek designs $\mathbf{x}$ that yield satisfactory performance, defined by finite threshold values $\tau = [\tau_1, \tau_2, \ldots, \tau_m]^\top$. Specifically, we wish to sequentially select configurations from:

$$\mathcal{S} = \{\mathbf{x} \mid f(\mathbf{x}) \succeq \tau\},$$

where $f(\mathbf{x}) \succeq \tau := f_i(\mathbf{x}) \geq \tau_i$, $i = 1, \ldots, m$. We refer to $\mathcal{S}$ as the *satisfactory region*. We make the assumption that $f : \mathcal{X} \to \mathcal{Y}$ is continuous and that the set $\{\mathbf{y} \mid \mathbf{y} \succeq \tau\}$ is also compact.

The concrete goal of CAS using the LMS acquisition function is identify candidates in $\mathcal{S}$ and disperse them throughout the subregion of objective space that exceeds a certain user-defined performance threshold. The LMS acquisition function does this by attempting to guarantee that each point is at least distance r from any other point in objective space, where r is a user-defined resolution parameter.

3.2 Gaussian Process Models

A Gaussian process (GP) model is a specific type of stochastic surrogate model —to be more precise, a random field— that uses radial basis functions to model the value of an unseen point as a normal distribution (Fasshauer and McCourt 2015). In recent years, GPs have become a popular method for modeling simulation experiments (Binois et al. 2015; Binois et al. 2018), due to not only their ability to accurately approximate a wide range of continuous functions, but also due to their built-in uncertainty estimates.

We start by first describing the GP model of a scalar function. Note that for clarity, the notation used in this subsection will be slightly different from that of the rest of the paper. We assume that we are trying to model a function $y = f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{y} \in \mathbb{R}$, and $f(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$. We collect n observations of $f(\mathbf{x})$ in the pair of matrices $\{\mathrm{X}, \mathbf{y}\}$, where $\mathrm{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $\mathbf{y} = [y_1, \ldots y_n]^T$.

On this set of n observations we place a GP prior. Given this GP prior, the posterior distribution of any set k of function values is modeled by a random variable $y_{\mathbf{x}} \in \mathbb{R}^k$ with the following normal distribution:

$$y_{\mathbf{x}} \sim \mathcal{N}(\tilde{\mu}_{\mathbf{x}}, \tilde{K}_{\mathrm{XX}}) \mid \tilde{\mu}_{\mathbf{x}} = K_{\mathrm{X}\mathbf{x}}^T (K_{\mathrm{XX}} + \sigma^2 \mathsf{I}_d)^{-1} (\mathbf{y}_{\mathbf{x}} - \mu_{\mathbf{x}}) + \mu_{\mathbf{x}} \ , \ \tilde{K}_{\mathrm{XX}} = K_{\mathrm{X}\mathbf{x}}^T (K_{\mathrm{XX}} + \sigma^2 \mathsf{I}_d)^{-1} K_{\mathrm{X}\mathbf{x}}.$$

The entries of the vector $\mu_{\mathbf{x}} \in \mathbb{R}^n$ and the matrices $K_{\mathrm{X}x} \in \mathbb{R}^{n \times k}$, $K_{\mathrm{XX}} \in \mathbb{R}^{n \times n}$ are determined by a mean function $\mu(\mathbf{x})$ and a covariance kernel $k(\mathbf{x}, \mathbf{x}')$ respectively, which largely control the fit of the GP. The σ^2 term is a regularization parameter and I is the $n \times n$ identity matrix. The GP we have described models a scalar function. However, in this paper we deal with multi-objective functions. To model these, we simply use a collection of m independent GP models.

3.3 Constraint Active Search with Parameter Diversity

Constraint active search (CAS) was first introduced in (Malkomes et al. 2021), in which the authors searched for designs in $\mathcal{S}$ which were as dispersed as possible within $\mathcal{X}$. The ECI acquisition function was created in pursuit of this goal. ECI balances the exploitation and exploration tradeoffs by simultaneously preferring candidates that are a) likely to satisfy the performance thresholds and b) located in unexplored regions. This paper can be viewed as an extension of CAS to objective space instead of the parameter space considered in the original work. The challenges of achieving diversity in parameter and objective space are somewhat different given the heteroskedasticity and uncertainty in the latter, and therefore our approach in this paper differs from that of ECI. However, we maintain a resolution parameter r which defines a sense of locality through Euclidean distance surrounding satisfactory designs.

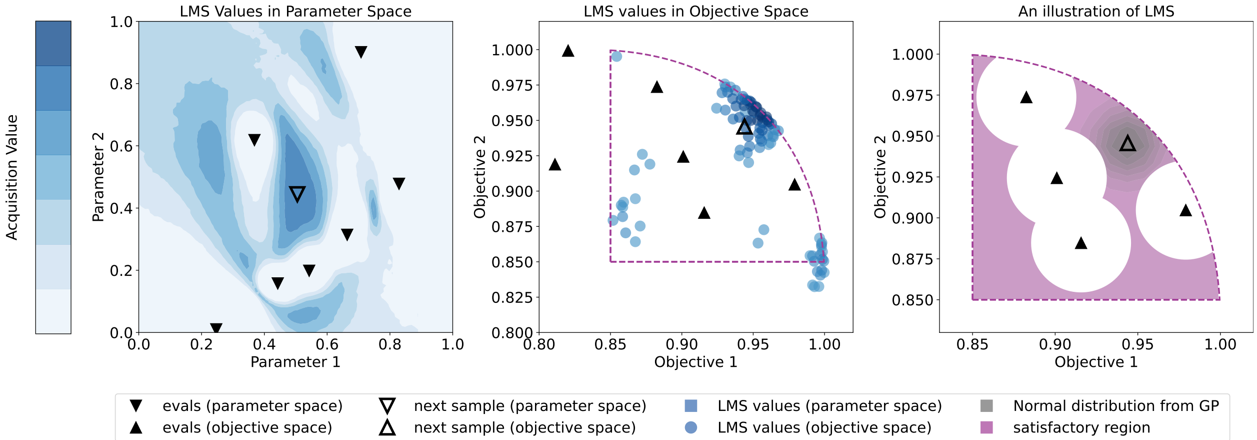

Figure 2: On the left, we plot the LMS values in parameter space. In the middle, we plot high values of LMS in objective space. On the right, we visualize computing LMS($\mathbf{x}^*$), where $\mathbf{x}^*$ is. LMS over our domain, in objective space. The LMS value is the volume of the Gaussian probability distribution that models $f(\mathbf{x}^*)$ centered around $\mu(\mathbf{x}^*)$ that is within the satisfactory region (magenta) and sufficiently far from existing observations in objective space (white circles).

3.4 Likelihood of Metric Satisfaction

LMS quantifies how likely a configuration $\mathbf{x}$ will be satisfactory (exceeding all thresholds) and promote diversity in objective space. This requires us to first define what diversity means. For LMS, we specify diversity using an additional parameter r, which informs the minimal distance between objective values it seeks to achieve. Given r, we define the following.

Definition 1 (Number of neighbors within r) *The number of neighbors within r of a point $\mathbf{y}$ and a finite set $\mathrm{Y} \subset \mathcal{Y}$ is defined as*

$$\mathbb{N}_r(\mathbf{y}, \mathrm{Y}) = |\{\mathbf{y}' \mid \mathbf{y}' \in \mathrm{Y}, d(\mathbf{y}, \mathbf{y}') < r\}|,$$

for an a priori fixed $r \in \mathbb{R}^+$ and an appropriate distance function $d : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}^+$.

Definition 2 (Average number of neighbors within r) *The average number of neighbors within r of a finite set $\mathrm{Y} \subset \mathcal{Y}$ is defined as*

$$\mathbb{N}_r(\mathrm{Y}) = \frac{1}{n} \sum_{i=1, \mathbf{y}_i \in \mathrm{Y}}^{n} \mathbb{N}_r(\mathbf{y}_i, \mathrm{Y}),$$

for an a priori fixed $r \in \mathbb{R}^+$ and an appropriate distance function $d : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}^+$.

Definition 3 (Comparing the spatial diversity of two sets) *Given a radius r, we say that Y_1 exhibits more spatial diversity than Y_2 if $\mathbb{N}_r(\mathrm{Y}_1) < \mathbb{N}_r(\mathrm{Y}_2)$.*

In other words, Y_1 exhibits greater spatial diversity than Y_2 if it has fewer neighbors within r on average. More generally, a set of points that are well-spaced apart will be more spatially diverse than a set of points that are very tightly clustered. Note that this definition is somewhat counter-intuitive because a lower $\mathbb{N}_r$ implies higher diversity.

We need one last definition and that is the half space whose points are greater than the threshold.

Definition 4 (Half space exceeding thresholds) *$H_\tau \subset \mathbb{R}^m$ is the half-space greater than thresholds τ:*

$$H_\tau = \{\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^m, \mathbf{y}_i \geq \tau_i, i = 1, \dots\ m\}.$$

Having now defined all these things, we can now precisely define LMS. Assume we already have a set of in observations in objective space Y. We want the objective values of the next observation $\mathbf{y}$ to decrease

$\mathbb{N}_r(\mathrm{Y} \cup \{\mathbf{y}\})$, which is equivalent to increasing diversity in objective space. This is guaranteed to occur if $\mathbb{N}_r(\mathbf{y}, \mathrm{Y}) = 0$; $\mathbf{y}$ is not within r of any other observation. We also want $\mathbf{y}$ to satisfy our thresholds i.e., $\mathbf{y} \in H_\tau$. LMS is the probability of these two events occurring with respect to a GP distribution on $\mathbf{y}$.

Definition 5 (Likelihood of Metric Satisfaction) *The LMS value of a point* $\mathbf{x}$ *and it's associated objective values* $\mathbf{y} = f(\mathbf{x})$ *is the probability that* $\mathbf{y}$ *has no neighbors within r and lies in* H_τ*:*

$$\mathrm{LMS}(\mathbf{x}) = Pr\Bigg(\mathbb{N}_r(\mathbf{y}, \mathrm{Y}) = 0\ ,\ \mathbf{y} \in H_\tau\Bigg).$$

Thus, the higher the LMS value, the greater the probability that the observed next objective value $\mathbf{y}$ will decrease $\mathbb{N}_r$; thus, it may help to think of LMS as attempting to minimize $\mathbb{N}_r$ in a sample-efficient manner.

We assume we have m independent Gaussian process models that capture our prior beliefs about observations $y_i = f_i(\mathbf{x}_i) + \varepsilon_i$ for $i = 1, 2, \ldots, m$, as a probability distribution over $p(\mathbf{y})$, where ε_i is additive Gaussian noise. This set of GPs models each point $\mathbf{y}(\mathbf{x}) = [y_1(\mathbf{x}) \ldots y_m(\mathbf{x})]$ as the following distribution:

$$y_i(\mathbf{x}) \sim \mathcal{N}(\tilde{\mu}_{\mathbf{x}}^{(m)}, \tilde{\Sigma}_{\mathbf{x},\mathbf{x}}^{(m)}) \mid \tilde{\mu}_{\mathbf{x}}^{(m)} \in \mathbb{R}^m, \tilde{\Sigma}_{\mathbf{x},\mathbf{x}}^{(m)} \in \mathbb{R}^{m \times m},$$

where $\tilde{\mu}_{\mathbf{x}}^{(m)}$ is vector of GP means for each objective and $\tilde{\Sigma}^{(m)}$ is the matrix of GP variances for each objective. Then $\mathrm{LMS}(\mathbf{x})$ is the the volume of the PDF of $\mathcal{N}(\tilde{\mu}_{\mathbf{x}}^{(m)}, \tilde{\Sigma}_{\mathbf{x},\mathbf{x}}^{(m)})$ that is above the thresholds and not within radius r of any existing observation in objective space. This is visualized in Figure 2, in which we plot the LMS values over the parameter and objective spaces.

To compute LMS, we rewrite it as the following integral of an indicator function:

$$\mathrm{LMS}(\mathbf{x}) = \mathbb{E}_{\mathbf{y}}[\mathbb{1}_{\tau,\mathrm{Y}}(\mathbf{y}_{\mathbf{x}})] = \int_{\mathbb{R}^m} \mathbb{1}_{\tau,\mathrm{Y}}(\mathbf{y}_{\mathbf{x}}) p(\mathbf{y}_{\mathbf{x}}) d\mathbf{x},$$

where $\mathbb{1}_{\tau,\mathrm{Y}}(\mathbf{y}_{\mathbf{x}})$ is 1 when $\mathbf{y}$ is above the thresholds and outside r of any observations, and 0 otherwise. We compute LMS in a straightforward manner with Monte Carlo (MC) integration —sample $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ from $\mathcal{N}(\tilde{\mu}_{\mathbf{x}}^{(m)}, \tilde{\Sigma}_{\mathbf{x},\mathbf{x}}^{(m)})$ and sum the samples: $\mathrm{LMS}(\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\tau,\mathrm{Y}}(\mathbf{y}_i)$. We note that computing the acquisition function via MC is very common in Bayesian optimization (Emmerich et al. 2011; Daulton et al. 2020; Garnett 2022) and is indeed supported by popular BO packages such as BoTorch (Balandat et al. 2020).

3.5 Scaling in Objective Space

When ECI is used for diversity in parameter space, the user benefits from knowing *a priori* what range the parameter values may take —indeed, they are the ones who define the search domain $\mathcal{X}$. As a result, ECI is able to choose an appropriate measure of distance $d(\mathbf{x}, \mathbf{x}')$ such that scale of each axis is the same —for example, by normalizing $\mathcal{X}$ to be the unit hypercube $[0, 1]^d$.

No such luxury exists in the case of LMS, which works with no prior knowledge of the range of $\mathcal{Y}$. If objective f_1 represents simulation time and objective f_2 represents a residual, then f_1 may be on the order of 10^6 while f_2 may be on the order of 10^{-6}. In this particular case, if the user selects $d(\mathbf{y}, \mathbf{y}')$ to be standard Euclidean distance, any perturbation to f_1 will far exceed the largest of perturbations to f_2; this heavily biases diversity towards f_1.

There is no perfect solution to this problem, and we advocate for dynamic scaling of each objective's axis to address this problem. At each iteration, we consider the minimal bounding box of our observed objective values and scale each dimension to be unit length. We provide an experimental study in Section 4.1 to illustrate the practical improvements that dynamic scaling provides.

4 EXPERIMENTS

In this section, we present numerical experiments to analyze the efficacy of our method. We compare our LMS acquisition function against three baselines: Random search, expected hypervolume improvement (Daulton et al. 2020), and the expected coverage improvement (Malkomes et al. 2021).

Table 1: Select experimental results from three problems: HC22, described in Section 4.1, RE33 (Tanabe and Ishibuchi 2020), and a plasma physics simulation optimization problem, with an optimization budget of 30, 50, and 100, respectively. The median over ten trials is provded below. We see that LMS achieves the lowest fill distance in objective space. It also generally achieves the lowest number of neighbors within $r_{\mathbf{y}}$ (random search does this trivially by identifying very few feasible points, thus guaranteeing that they are spaced far apart). MOBO achieves the highest hypervolume.

Function	d	m	Methods	Fill distance ↓	# Satisfactory ↑	# Neighbor ↓	Hypervolume ↑
HC22	2	2	RND	4.31	2	**0.0**	0.16
			BO	11.4	**27**	9.20	**1.71**
			ECI($r_{\mathbf{x}} = 0.1$)	4.84	20	2.5	1.50
			LMS($r_{\mathbf{y}} = 0.1$)	**2.61**	21	1.0	1.50
RE33	3	3	RND	7.18	5	**0.0**	5.71
			BO	2.53	32	3.06	**15.02**
			ECI ($r_{\mathbf{x}} = 0.1$)	10.01	**45**	4.71	10.52
			LMS ($r_{\mathbf{y}} = 0.26$)	**1.50**	33	2.48	6.93
STELL	9	3	RND	37.62	7	**0.0**	17.04
			BO	31.31	30	15.46	**36.83**
			ECI ($r_{\mathbf{x}} = 0.1$)	32.66	**32**	14.88	25.99
			LMS ($r_{\mathbf{y}} = 0.5$)	**30.79**	20	**0.0**	29.18

We emphasize that no single criterion can sufficiently convey the full strength of any methodology; this is especially true in the multiobjective setting, in which different performance criteria already exist to quantify different algorithmic goals. We hope to impart a nuanced comparison of LMS to existing baselines. To that end, we consider the following four standard criteria.

- **Number of neighbors within** r: Defined in Section 3.4, LMS attempts to keep this quantity small. Note that this is a particular metric for the more general notion of coverage, which is a commonly used criteria to judge the spread of points in a continuous space (Sayın 2000)
- **Fill distance**: Fill distance is a standard measure of spatial diversity in a simple domain $\mathcal{S}$. Given a set of sample points Y, the fill distance is formally defined as the following: $\text{FILL}(\mathrm{Y}, \mathcal{S}) = \sup_{\mathbf{y} \in \mathcal{S}} \min_{\mathbf{y}_j \in \mathrm{Y}} d(\mathbf{y}_j, \mathbf{y})$. In Euclidean space, $\text{FILL}(\mathrm{Y}, \mathcal{S})$ is the radius of the largest empty ball one can fit in $\mathcal{S}$, and measures the spacing of X in $\mathcal{S}$. The smaller a set's fill, the better distributed it is within $\mathcal{S}$. Special sets that achieve low fill in simple domains include low-discrepancy sequences and Latin hypercubes, used frequently in simulation optimization (Niederreiter 1992).
- **Positive samples**: The number of points whose objective values exceed τ.
- **Hypervolume**: We measure the hypervolume of region in objective space bounded by the Pareto frontier and the defined thresholds. In particular, we conjecture that algorithms which excel at maximizing the hypervolume may underperform on other criteria, and vice versa.

4.1 Demonstration of Likelihood of Metric Satisfaction

Here we compare the algorithms on a simple two objective problem, with each objective $f : R^2 \to R$,

$$f_1(\mathbf{x}) = \exp(((x_1 - 0.2)^2 + (x_2 - 0.5)^2)/2),$$
$$f_2(\mathbf{x}) = \exp(((x_1 - 0.8)^2 + (x_2 - 0.5)^2)/2).$$

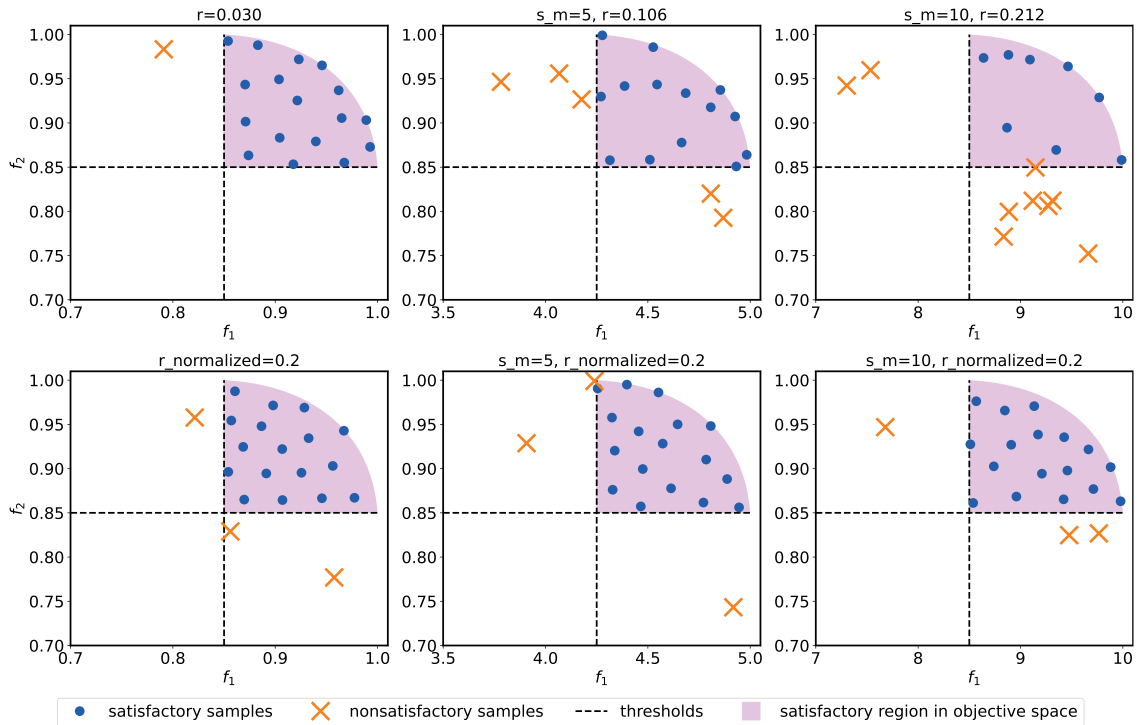

Figure 3: Comparison of non-normalized LMS (top row) and normalized (bottom row) LMS for different scaling of $f_1(\mathbf{x})$. We show the samples and satisfactory region in objective space for scaling $s_m = \{5, 10\}$.

We set the thresholds to be $f_1(\mathbf{x}) \geq 0.85$ and $f_2(\mathbf{x}) \geq 0.85$. We run each method for 20 iterations and plot the samples in both parameter space and objective space in Figure 1 of Section 2.4. Visually, we see that multiobjective Bayesian optimization seeks to isolate the Pareto frontier, leading to a tight concentration of points in both parameter and objective space. CAS using ECI does a good job of spreading points out in parameter space, but not objective space. Conversely, using LMS instead yields our desired result —a diverse spread of points in objective space.

4.2 Scaling in Objective Space

Next, we investigate how the difference in the range of f can impact the LMS algorithm. We repeat the test functions in Section 4.1, but scale $f_1(\mathbf{x})$ and its corresponding threshold by a factor of $s_m = \{5, 10\}$. For the non-normalized LMS, we multiply the resolution parameter r by a factor of $s_m/\sqrt{2}$. For the normalized LMS, we fix the $r_{normalized} = 0.2$. We demonstrate the different behaviors of the normalized and non-normalized LMS algorithms in Figure 3.

We can observe that the non-normalized LMS algorithm tends to search along f_1 when scaled, since it is more likely find new satisfactory samples that are at least r away from the observed samples because the axis along f_1 is longer. In contrast, the normalized LMS algorithm consistently produces more "evenly" distributed samples in the objective space despite the scaling of f_1.

4.3 Relationship between Parameter and Objective Satisfactory Region

We then investigate how LMS perform for different parameter satisfactory regions. More specifically, we want to see if LMS still works as the relative satisfactory region in the parameter space decreases. We modify the objective functions in Section 4.1 with a s_m scaling factor. As s_m increases, the satisfactory region in the parameter space decreases, but the satisfactory region in the objective space remains the same.

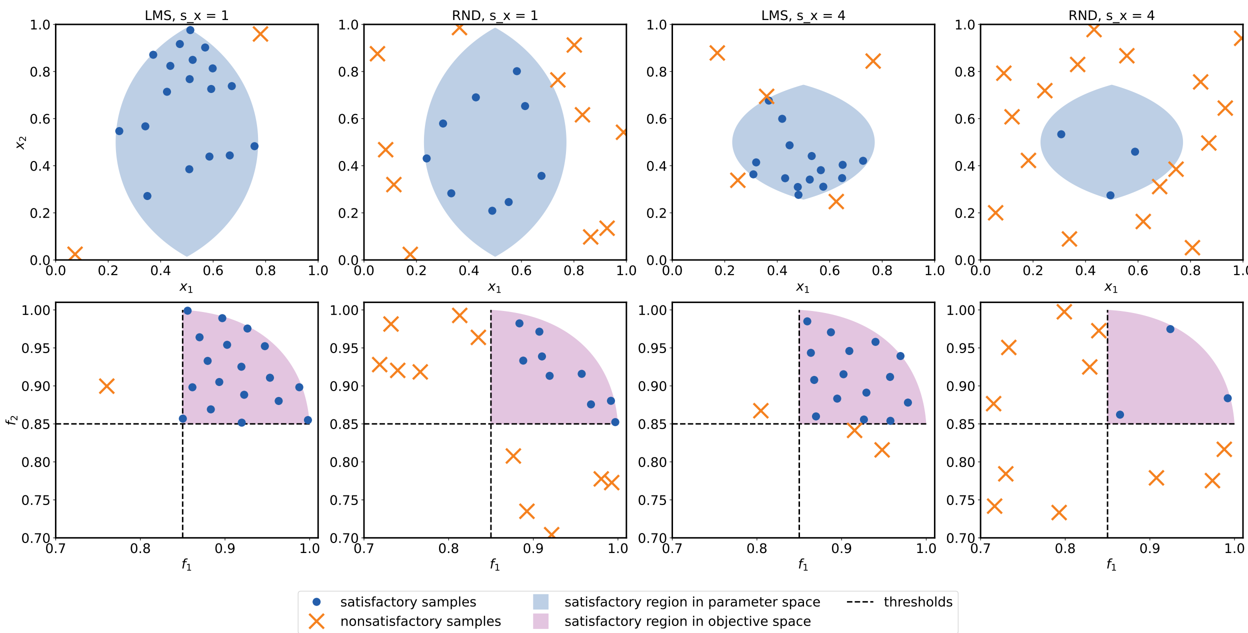

Figure 4: Comparison of LMS and RND for different parameter satisfactory volumes (same satisfactory region in objective space).

$$f_1(\mathbf{x}) = \exp(((x_1 - 0.2)^2 + s_m(x_2 - 0.5)^2)/2),$$
$$f_2(\mathbf{x}) = \exp(((x_1 - 0.8)^2 + s_m(x_2 - 0.5)^2)/2).$$

We demonstrate in Figure 4 that LMS can consistently sample diversely in the satisfactory region in the objective space, despite the satisfactory region in the parameter space shrinking. In contrast, random search is unable to handle the shrinking satisfactory region.

4.4 Plasma Physics

A stellarator is device that uses a set of magnetic coils to confine a plasma hot enough to sustain nuclear fusion (Spitzer Jr 1958). Stellarator coils lack rotational symmetry, and possess notably warped shapes due to the complex, quasi-symmetric magnetic field they must produce. Thus, determining suitable stellarator designs through simulation is crucial to produce a stellarator candidate for real-world production.

We use the PyPlasmaOpt (Giuliani et al. 2022) simulator to generate a diverse set of stellarator designs, where each stellarator is a set of coils. Each coil is represented by the curve in 3D Cartesian coordinates $\Gamma(\theta) = (x(\theta), y(\theta), z(\theta))$, where each coordinate admits the following Fourier expansion, e.g., for x:

$$x(\theta) = c_0 + \sum_{k=1}^{n_{order}} s_k \sin(k\theta) + c_k \cos(k\theta).$$

For each coil, the parameters c_k and s_k are the search parameters. Given fixed coils, the simulator solves a certain first order, nonlinear ordinary differential equation outputs summary information in three functions:

$$f(\mathbf{x}) = [F_{\text{magnetic}}(\mathbf{x}), F_{\text{transform}}(\mathbf{x}), F_{\text{shape}}(\mathbf{x})].$$

The first quantifies the quasi-symmetry of the magnetic field —the smaller it is, the more desirable the resulting field. The second locks the solution into a target rotational transform. The third penalizes overly complex coils too impractical to manufacture in real life. We test LMS on a small stellarator simulation of six repeated coils and present results in Table 1.

5 CONCLUSION

In this paper, we tackle an alternative formulation multi-objective optimization problems for simulation optimization by generating a diverse set of designs in objective space instead of explicitly searching for the Pareto frontier. This presents decision makers with a robust pool of promising design decisions and helps them better understand the space of good solutions.

We do so through an acquisition called Likelihood of Metric Satisfaction (LMS), which quantifies the utility of a candidate point as the probability that it is both above user-defined performance thresholds and sufficiently far from other observations in objective space. We then illustrated the strength of LMS on a few synthetic simulation optimization problems as well as one application in stellarator design for nuclear fusion. Finally, we examine the performance of LMS under adverse conditions. We believe LMS and more generally, the presentation of multiple, diverse solutions to a design problem, is a promising research direction. Future work includes simultaneously considering diversity in both parameter and objective space.

REFERENCES

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 *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Volume 33. December 6$^{\text{th}}$-12$^{\text{th}}$, virtual, 21524–21538.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl. 2011. "Algorithms for Hyper-Parameter Optimization". In *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Volume 24. Dec 12$^{\text{th}}$-17$^{\text{th}}$, Grenada, Spain.

Binois, M., D. Ginsbourger, and O. Roustant. 2015. "Quantifying Uncertainty on Pareto Fronts with Gaussian Process Conditional Simulations". *European Journal of Operational Research* 243(2):386–394.

Binois, M., R. B. Gramacy, and M. Ludkovski. 2018. "Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments". *Journal of Computational and Graphical Statistics* 27(4):808–821.

Daulton, S., M. Balandat, and E. Bakshy. 2020. "Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization". In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin. December 6$^{\text{th}}$-12$^{\text{th}}$, virtual, 9851–9864.

Daulton, S., D. Eriksson, M. Balandat, and E. Bakshy. 2021. "Multi-Objective Bayesian Optimization over High-Dimensional Search Spaces". *arXiv preprint arXiv:2109.10964*. Accessed: Jan 30, 2022.

del Rosario, Z., M. Rupp, Y. Kim, E. Antono, and J. Ling. 2020. "Assessing the frontier: Active learning, model accuracy, and multi-objective candidate discovery and optimization". *The Journal of Chemical Physics* 153(2):024112.

Emmerich, M. T. M., A. H. Deutz, and J. W. Klinkenberg. 2011. "Hypervolume-Based Expected Improvement: Monotonicity Properties and Exact Computation". In *2011 IEEE Congress of Evolutionary Computation (CEC)*. June 5$^{\text{th}}$-8$^{\text{th}}$, New Orleans, LA, CA, 2147-2154.

Fasshauer, G. E., and M. J. McCourt. 2015. *Kernel-based Approximation Methods Using MATLAB*. Singapore: World Scientific.

Forrester, A., A. Sobester, and A. Keane. 2008. *Engineering Design via Surrogate Modelling: A Practical Guide*. New York: John Wiley & Sons.

Gardner, J. R., M. J. Kusner, Z. E. Xu, K. Q. Weinberger, and J. P. Cunningham. 2014. "Bayesian Optimization with Inequality Constraints". In *Proceedings of the 31th International Conference on Machine Learning (ICML 2014)*, Volume 32. June 21$^{\text{st}}$-26$^{\text{th}}$, Beijing, China, 937–945.

Gardner, J. R., G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. 2018. "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration". In *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Dec 3$^{\text{rd}}$8$^{\text{th}}$, Montreal, Canada.

Garnett, R. 2022. *Bayesian Optimization*. Cambridge, UK: Cambridge University Press. in preparation.

Garnett, R., Y. Krishnamurthy, X. Xiong, J. Schneider, and R. Mann. 2012. "Bayesian Optimal Active Search and Surveying". In *Proceedings of the 29th International Conference on International Conference on Machine Learning (ICML 2012)*. June 26$^{\text{th}}$July 1$^{\text{st}}$, Edinburgh, Scotland, 843850.

Giuliani, A., F. Wechsung, A. Cerfon, G. Stadler, and M. Landreman. 2022. "Single-Stage Gradient-Based Stellarator Coil Design: Optimization for Near-Axis Quasi-Symmetry". *Journal of Computational Physics* 459:111147.

GPy since 2012. "GPy: A Gaussian process framework in python". http://github.com/SheffieldML/GPy. accessed Jan 12$^{\text{th}}$, 2022.

Haghanifar, S., M. McCourt, B. Cheng, J. Wuenschell, P. Ohodnicki, and P. W. Leu. 2020. "Discovering High-Performance Broadband and Broad Angle Antireflection Surfaces by Machine Learning". *Optica* 7(7):784–789.

Hutter, F., H. H. Hoos, and K. Leyton-Brown. 2011. "Sequential Model-Based Optimization for General Algorithm Configuration". In *International Conference on Learning and Intelligent Optimization*, edited by C. A. C. Coello. Jan 17th-21st, Rome, Italy, 507–523.

Jiang, S., G. Malkomes, G. Converse, A. Shofner, B. Moseley, and R. Garnett. 2017. "Efficient Nonmyopic Active Search". In *Proceedings of the 34th International Conference on Machine Learning, (ICML 2017)*, edited by D. Precup and Y. W. Teh, Volume 70. August 6th-11th, Sydney, Australia, 1714–1723.

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

Knowles, J. 2006. "ParEGO: A Hybrid Algorithm with On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems". *IEEE Transactions on Evolutionary Computation* 10(1):50–66.

Malkomes, G., B. Cheng, E. H. Lee, and M. McCourt. 2021. "Beyond the Pareto Efficient Frontier: Constraint Active Search for Multiobjective Experimental Design". In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, edited by M. Meila and T. Zhang, Volume 139. July 18th-24th, virtual, 7423–7434.

Mockus, J., V. Tiesis, and A. Zilinskas. 1978. "The Application of Bayesian Methods for Seeking the Extremum". *Towards Global Optimization* 2(117-129):2.

Molesky, S., Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez. 2018. "Inverse Design in Nanophotonics". *Nature Photonics* 12(11):659–670.

Negoescu, D. M., P. I. Frazier, and W. B. Powell. 2011. "The Knowledge-Gradient Algorithm for Sequencing Experiments in Drug Discovery". *INFORMS Journal on Computing* 23(3):346–363.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Rasmussen, C. E., and C. K. I. Williams. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: The MIT Press.

Sayın, S. 2000. "Measuring the Quality of Discrete Representations of Efficient Sets in Multiple Objective Mathematical Programming". *Mathematical Programming* 87(3):543–560.

Scott, W., P. Frazier, and W. Powell. 2011. "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters using Gaussian Process Regression". *SIAM Journal on Optimization* 21(3):996–1026.

Spitzer Jr, L. 1958. "The Stellarator Concept". *The Physics of Fluids* 1(4):253–264.

Srinivas, N., A. Krause, S. Kakade, and M. Seeger. 2010. "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design". In *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML 2010)*, edited by J. Fürnkranz and T. Joachims. June 21st-24th, Haifa, Israel, 1015–1022.

Szechtman, R., and E. Yücesan. 2008. "A New Perspective on Feasibility Determination". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. Mason, R. Hill, L. Mönch, and O. Rose, 273–280. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Szechtman, R., and E. Yücesan. 2016. "A Bayesian Approach to Feasibility Determination". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 782–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tanabe, R., and H. Ishibuchi. 2020. "An Easy-to-Use Real-World Multi-Objective Optimization Problem Suite". *Applied Soft Computing* 89:106078.

Villemonteix, J., E. Vazquez, and E. Walter. 2009, aug. "An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions". *Journal of Global Optimization* 44(4):509534.

AUTHOR BIOGRAPHIES

ERIC HANS LEE is a research scientist at SigOpt. His interests largely lie in Gaussian process modeling and global optimization. His email address is eric.lee@intel.com.

BOLONG CHENG is the technical lead of the research team at SigOpt. He works on productionizing Bayesian optimization, and more broadly, sequential decision making problems. Currently, he is interested in applying sequential optimization techniques in scientific and engineering domains such as materials simulation and design. His email address is harvey.cheng@intel.com.

MICHAEL MCCOURT is a senior principal engineer at Intel and the general manager of SigOpt. His recent work has focused on sample-efficient methods for design, involving both optimization and search. Additionally, he has worked on matrix computations and kernel methods in math and statistics. His email address is mccourt@sigopt.com.