# Beyond the Pareto Efficient Frontier: Constraint Active Search for Multiobjective Experimental Design

**Gustavo Malkomes** [1] **Bolong Cheng** [1] **Eric Hans Lee** [1] **Michael McCourt** [1]

## Abstract

Many problems in engineering design and simulation require balancing competing objectives under the presence of uncertainty. Sample-efficient multiobjective optimization methods focus on the objective function values in *metric space* and ignore the sampling behavior of the design configurations in *parameter space*. Consequently, they may provide little actionable insight on how to choose designs in the presence of metric uncertainty or limited precision when implementing a chosen design. We propose a new formulation that accounts for the importance of the parameter space and is thus more suitable for multiobjective design problems; instead of searching for the Pareto-efficient frontier, we solicit the desired minimum performance thresholds on all objectives to define regions of satisfaction. We introduce an active search algorithm called Expected Coverage Improvement (ECI) to efficiently discover the region of satisfaction and simultaneously sample diverse acceptable configurations. We demonstrate our algorithm on several design and simulation domains: mechanical design, additive manufacturing, medical monitoring, and plasma physics.

## 1. Introduction

Accelerated design and optimization via machine learning is receiving increasing interest in multiple science and engineering disciplines, such as materials design, drug discovery, and chemical engineering (Forrester et al., 2008; Negoescu et al., 2011; Molesky et al., 2018). Using numerical simulation to study the impact of design decisions prior to manufacturing has become a common strategy to reduce the number of fabrications, thus reducing the substantial cost of development cycles.

[1]SigOpt, an Intel company, San Francisco, CA, USA. Correspondence to: Gustavo Malkomes <gustavo.malkomes@intel.com>.

*Proceedings of the $38^{th}$ International Conference on Machine Learning*, PMLR 139, 2021. Copyright 2021 by the author(s).

In recent years, active learning has gained popularity in scientific communities. These adaptive methods more efficiently identify promising candidates compared to brute-force or evolutionary algorithms (Song et al., 2018; Attia et al., 2020; Haghanifar et al., 2020; Duris et al., 2020). Bayesian optimization (BO) is arguably the most prominent example of active learning, where the goal is to seek the optimizer $\arg\max_{\mathbf{x}} f(\mathbf{x})$ of an expensive-to-evaluate black-box objective $f$.

However, in most real world applications, multiple competing objectives need to be investigated (Koziel et al., 2014; Singh et al., 2016). Optimality of multiobjective problems is usually defined by the concept of *Pareto efficiency*, where no individual objective can be improved without loss in at least one other objective. Most work on optimizing multiple objectives focuses on searching for the Pareto efficient frontier (and the corresponding parameters). In general, for an $m$-objective problem, the (nearly) Pareto efficient solutions are assumed to be a dense set of points on an $(m-1)$-dimensional manifold (Konakovic-Lukovic et al., 2020). Unfortunately, using the Pareto frontier as the measurement of success is limiting in engineering and design applications; candidates near the Pareto frontier, unaccounted for in most optimization literature, are still of scientific interest in the real world (del Rosario et al., 2020).

In practice, design objectives can also be formulated as inequality constraints, where the decision makers seek to find design choices that satisfy known performance thresholds on each objective. This formulation is related to the topic of level set estimation (LSE) in the single objective setting. Its goal is to determine the decision boundary of parameter space that separates it above and below the threshold on the objective (Gotovos et al., 2013). However, the notion of level set does not naturally extend to the multiobjective settings; a naive approach of identifying the level set for each objective and taking the intersection is not computationally efficient. In addition, most LSE algorithms focus on searching for the decision boundary; in the multiobjective setting, this behavior cannot reveal the tradeoff among objectives, since there are not enough samples within the region of interest.

Figure 1 illustrates the limitations of multiobjective BO and

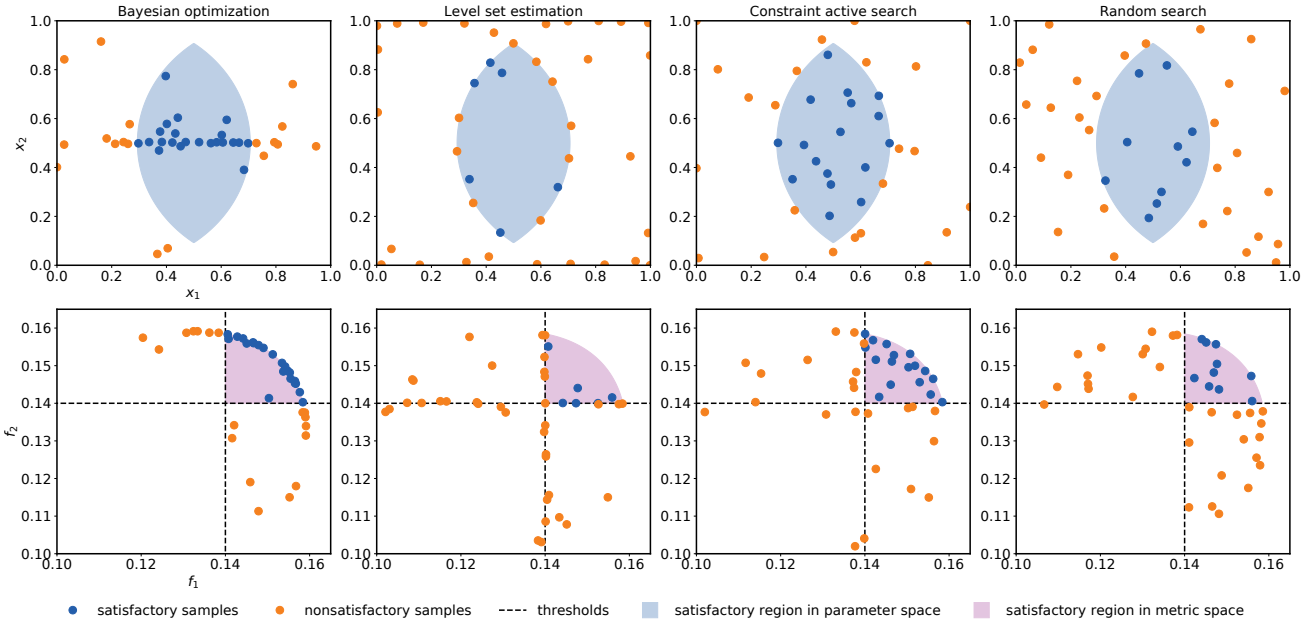


*Figure 1.* Comparison of BO, LSE, constraint active search, and random search behaviors on the same multiobjective problem. The top row shows the observed samples in the parameter space; the bottom row, in the metric space. The shaded region represents the satisfactory regions of interest, where both objectives exceed the known thresholds (labeled with dashed black lines in the metric space).

LSE respectively, and highlights constraint active search as a new perspective on multiobjective problems. BO focuses on searching for the Pareto efficient parameters. LSE focuses on searching for the boundary of the region satisfying the thresholds. Constraint active search attempts to sample from that region of interest.

Our paper makes the following contributions:

- We propose an alternative formulation for the multiobjective design problem as an active search problem. In this setting, the objectives are treated as constraints with known threshold values; the goal is to find diverse samples that satisfy all the thresholds. We call this formulation *constraint active search* (CAS).
- We develop an algorithm called *expected coverage improvement* (ECI) for constraint active search that preserves the sample efficiency of active search whilst also diversely sampling from the regions of interest. We provide theoretical analysis on the sample diversity of ECI; specifically, it guarantees a constant approximation ratio to the optimal sample diversity.
- We compare ECI to various baselines on a suite of synthetic multiobjective design benchmarks as well as real-world multiobjective design and simulation applications in materials science, medical monitoring, and plasma physics.

## 2. Related Work

### 2.1. Multiobjective Bayesian Optimization

Bayesian optimization (BO) is a popular method for optimizing expensive, black-box objective functions. BO consists of two components: a *probabilistic model*, to model the objective function $f$, and an *acquisition function*, to determine which parameters $\mathbf{x}$ to sample next. In this section, we focus on reviewing the literature of BO in the multiobjective setting. We refer the readers to Frazier (2018) for a comprehensive review of BO in the single objective setting.

There are several strategies at adapting BO to multiobjective optimization. The simplest approach is linear scalarization, where single objective BO strategies optimize a linear combination of the objectives. Knowles (2006) first proposes using a random scalarization of the objective functions to search for the entire frontier. This method is further improved in Paria et al. (2020), which models each individual objective with a separate Gaussian process (GP).

A related approach involves reformulating the problem as a constrained optimization problem, commonly known as the $\varepsilon$-constraint method (Haghanifar et al., 2020). In particular, this formulation leverages prior works from constrained Bayesian optimization literature (Gelbart, 2015; Gardner et al., 2014). In constrained BO, a probabilistic model is formed for each constraint and used as a multiplicative penalty term in the acquisition function (Letham et al., 2019). Recent works on constrained BO borrow ideas

from the mathematical optimization community such as the trust region (Eriksson & Poloczek, 2021) and Alternating Directions Methods of Multiplier (Ariafar et al., 2019).

A separate line of research in multiobjective BO improves the hypervolume of the Pareto efficient frontier directly (Emmerich et al., 2011; Konakovic-Lukovic et al., 2020; Daulton et al., 2020). Belakaria et al. (2019) adapts the entropy search acquisition function for the multiobjective setting, by maximizing the entropy of the Pareto frontier. Belakaria et al. (2020) and Fernández-Sánchez et al. (2021) build upon this work to incorporate additional black-box inequality constraints. Almost all multiobjective BO literature uses the hypervolume of the Pareto frontier as their performance criterion. While hypervolume provides a single numerical value for the ease of comparison, using it as the sole criterion of success misses the goal of multiobjective optimization — understanding tradeoffs between parameters and objectives.

### 2.2. Level Set Estimation

The goal of level set estimation is to determine the regions where the objective function lies above or below a known threshold level. Bryan et al. (2005) first uses a GP-based active learning approach to tackle this problem through the straddle heuristic, defined as

$$\textsc{straddle}(\mathbf{x}) = 1.96\sqrt{v(\mathbf{x})} - |f(\mathbf{x}) - \tau|,$$

where $\tau$ is the threshold of the level set and $v$ is the model uncertainty. Gotovos et al. (2013) generalizes the straddle heuristic and provides theoretical guarantees on the classification quality. Zanette et al. (2019) maximizes the expected increase in the volume of the predicted level set. Iwazaki et al. (2020) considers the LSE problem with input uncertainty, where the input parameter values are subject to perturbation from a known Normal distribution. This work is related to research in safe BO (Sui et al., 2015; 2018), where only configurations exceeding the threshold with high probability are sampled.

Almost all LSE methods focus on identifying the decision boundary. To our best knowledge, Wang et al. (2018) is the closest example of learning and diversely sampling from a region of interest (in the single objective setting). It divides the sampling procedure into two disjoint algorithms: one determines the level set and the other samples diversely from the interior of the learned level set.

The notion of a level set does not naturally extend beyond more than one objective. Bryan & Schneider (2008) only considers the multiobjective problem in a composite setting, where a single threshold is used for a linear combination of the objective functions. One naive and inefficient approach to extend existing LSE algorithms to multiobjective problems is to estimate the level set of each objective separately and then intersect each region of interest.

### 2.3. Active Search

Our proposed work is also related to the topic of active search (Garnett et al., 2012). 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.

Garnett et al. (2012) explores this problem using Bayesian decision theory, introduces a natural utility function for active search, and derives the optimal policy for this framework. In general, only myopic (one-step) approximations to the optimal policy are computationally tractable. More recently, Jiang et al. (2017) proposes an nonmyopic and efficient approximation to the optimal policy, which is later extended to the batch setting in Jiang et al. (2018). Active search, however, is not designed to solve multiobjective continuous problems — a problem that we address next.

## 3. Constraint Active Search

Scientists, engineers, and decision makers seldom rely on single-point observations to make final design decisions. Instead, they seek a diverse set of design configurations to enrich their understanding of each objective function's impact and its respective tradeoffs. As Figure 1 illustrates, neither BO nor LSE is well-suited to address these experimental design concerns. The former clusters around optima, and the latter clusters around the level set boundary.

We propose an alternative formulation to the multiobjective design problem better suited to gather a diverse spectrum of design configurations. The goal of this formulation is to efficiently cover an unknown, possibly disconnected, set $\mathcal{S}$ defined by the objective thresholds $\boldsymbol{\tau}$ in a sequential fashion. We call this problem *constraint active search*. In Section 5, we illustrate how CAS may enable a practitioner to make better decisions than those made by exclusively identifying the Pareto frontier on an additive manufacturing problem.

This concept is related to del Rosario et al. (2020) which advocates that the Pareto frontier itself is insufficient for scientific and engineering settings. That work introduces the notion of *Pareto shells*, the set of near Pareto optimal solutions, as the desired outcome of multiobjective optimization. We propose that, when thresholds $\boldsymbol{\tau}$ are available, studying the full region of constraint satisfaction[1] is preferable.

[1] We use the term satisfaction instead of feasibility merely to stress that the constraints are over the objectives and not explicitly defined over the parameter space.

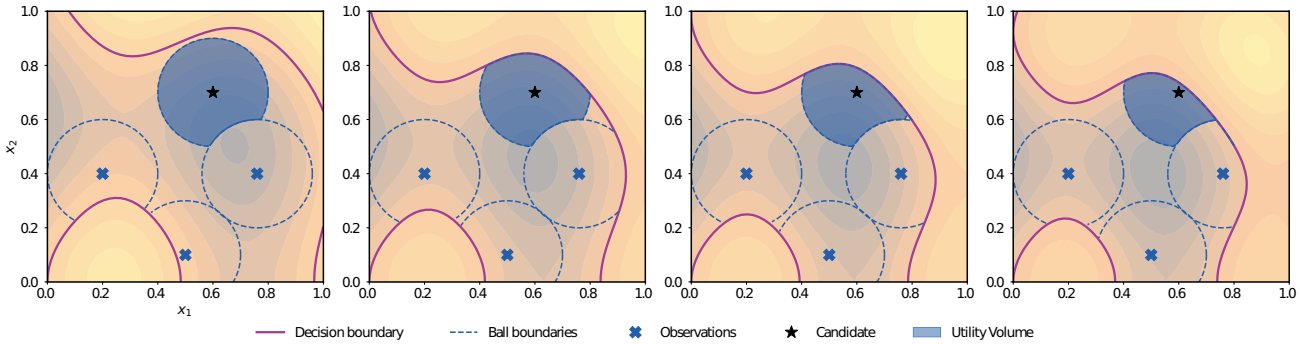


*Figure 2.* An illustration of our utility function when constraint active search is run on the 2D sinusoids problem in (Bryan et al., 2005). Each of the four subfigures above plots a single realization of the underlying stochastic model and outlines the decision boundary of the satisfactory region in magenta. The observations within the boundary are marked with blue crosses, and a dotted blue ball of radius $r = 0.2$ is drawn around each to indicate coverage. A ball is also drawn around the candidate point, itself marked with a black star. The improvement of the utility is the volume of the blue region, and expected coverage improvement is the average value of all realizations.

### 3.1. Problem Statement

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

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

where $f(\mathbf{x}) \succeq \boldsymbol{\tau} := f_i(\mathbf{x}) \geq \tau_i,\ i = 1, \dots, m$. At iteration $t$, we assume we have evaluated $t$ design configurations $\mathbf{X}_t = \{\mathbf{x}_j\}_{j=1}^t$ and their respective outputs $\mathbf{Y}_t = \{\mathbf{y}_j\}_{j=1}^t$, where $\mathbf{y}_j = [y_{j_1}, y_{j_2}, \dots, y_{j_m}]$ is the potentially noisy observation of the objective functions $f_1, f_2, \dots f_m$ evaluated at $\mathbf{x}_j$. These observations form the dataset $\mathcal{D}_t = (\mathbf{X}_t, \mathbf{Y}_t)$.

We assume we have $m$ independent probabilistic models that capture our prior beliefs about observations $y_i = f_i(\mathbf{x}_i) + \varepsilon_i$ for $i = 1, 2, \dots, m$, as a probability distribution over $p(\mathbf{y})$, where $\varepsilon_i$ is additive Gaussian noise. We also assume $p(\mathbf{y})$ allows posterior updates when $\mathcal{D}$ is observed, i.e., $p(\mathbf{y} \mid \mathcal{D})$. Let $Z$ be an indicator variable that tells us if a point $\mathbf{x}$ will satisfy the thresholds, $Z(\mathbf{x}) = \mathbb{1}\left[\mathbf{y}(\mathbf{x}) \succeq \boldsymbol{\tau}\right]$. Our model of $\mathbf{y}$ allows us to compute the probability of any point $\mathbf{x}$ belonging to the satisfactory region $\mathcal{S}$ after observing $\mathcal{D}_t$, which we denote by $p(Z(\mathbf{x}) = 1 \mid \mathcal{D}_t)$. We delay discussing the statistical model's details until Section 4, as they are immaterial to decision-making.

### 3.2. Expected Coverage Improvement

Our approach to constraint active search considers an additional *resolution* parameter $r$; in real-world problems one can often define $r$ in terms of the robustness of the design to perturbations. Typically, most experimental design problems possess a sense of known resolution, such as simulation accuracy or manufacturing precision/tolerance. This presumes that any design configuration within distance $r$ of another does not convey extra information about $\mathcal{S}$.

**Definition 3.1** (Coverage neighborhood)**.** The *coverage neighborhood* of any $\mathbf{x}$ is defined as

$$\mathbb{N}_r(\mathbf{x}) = \{\mathbf{x}' : d(\mathbf{x}, \mathbf{x}') < r\},$$

for an *a priori* fixed $r \in \mathbb{R}^+$ and an appropriate distance function $d : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$. The coverage neighborhood of a set of points $\mathbf{X}$ is defined as

$$\mathbb{N}_r(\mathbf{X}) = \bigcup_{\mathbf{x} \in \mathbf{X}} \mathbb{N}_r(\mathbf{x}).$$

For Euclidean distance, $\mathbb{N}_r(\mathbf{x})$ is simply a ball of radius $r$.

Constraint active search prioritizes datasets $\mathcal{D}$ that cover as much volume of the satisfactory region $\mathcal{S}$ as possible. We express this desire with the following utility function:

$$u(\mathcal{D}) = \mathrm{Vol}\left(\mathbb{N}_r(\mathbf{X}) \cap \mathcal{S}\right).$$

Intuitively, $u(\mathcal{D})$ measures the total volume of $\mathcal{S}$ that is covered by neighborhood $\mathbb{N}_r(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$. At iteration $t$, the expected *improvement* of this utility function for sampling a candidate configuration[2] $\mathbf{x}$ is:

$$\begin{aligned} \alpha(\mathbf{x} \mid \mathcal{D}_t) &= \mathbb{E}_{\mathbf{y}}\left[u\left(\mathcal{D}_t \cup (\mathbf{x}, \mathbf{y})\right) - u(\mathcal{D}_t)\right] \\ &= \mathbb{E}_Z\left[\mathrm{Vol}\left(\left\{\mathbb{N}_r(\mathbf{x}) \cap \mathcal{S}_Z\right\} \setminus \mathbb{N}_r(\mathbf{X_t})\right)\right]. \end{aligned} \tag{1}$$

We can also refer to (1) as an *acquisition function*. At iteration $t$, our policy selects the location $\mathbf{x}^*$ that maximizes the acquisition function, i.e., $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{X}} \alpha(\mathbf{x} \mid \mathcal{D}_t)$. We visualize the intuition behind our policy for different

[2]For clarity, we have overloaded the notation for $\mathcal{D}_t \cup (\mathbf{x}, \mathbf{y})$ to indicate a pairwise union, which we hope is clearly understood.

realizations of $Z$ in Figure 2. Our acquisition function considers the volume of $\mathcal{S}_Z$ that a new candidate location is *expected* to cover under our current belief about $Z$.

A common practice in the LSE and active search literature is to work on discrete search spaces: $\mathcal{X}$ could be discretized by sampling a pool of $n$ points. In this discretized setting, (1) simplifies to

$$\begin{aligned}\alpha(\mathbf{x} \mid \mathcal{D}_t) &= \mathbb{E}_Z\Big[\sum_{\mathbf{x}' \in \mathbb{N}(\mathbf{x}) \setminus \mathbb{N}_r(\mathbf{X})} Z(\mathbf{x}')\Big] \\ &= \sum_{\mathbf{x}' \in \mathbb{N}(\mathbf{x}) \setminus \mathbb{N}_r(\mathbf{X})} \mathbb{E}_Z\big[Z(\mathbf{x}')\big] \\ &= \sum_{\mathbf{x}' \in \mathbb{N}(\mathbf{x}) \setminus \mathbb{N}_r(\mathbf{X})} p(Z(\mathbf{x}') = 1 \mid \mathcal{D}_t). \qquad (2)\end{aligned}$$

In practice, the argmax of $\alpha(\mathbf{x} \mid \mathcal{D}_t)$ is often nonunique, especially at the start of search. This reflects the intuition that when only a few observations sparsely cover $\mathcal{S}$, multiple —and possibly many— candidates will contribute equally to the expected coverage. Crucially, if this occurs, we break ties among the multiple optimum points by selecting the one furthest from all observations in $\mathbf{X}_t$, which encourages selecting a diverse set of points. We call this policy *expected coverage improvement* (ECI).

### 3.3. Theoretical Analysis of ECI

We present three theoretical results for ECI (proved in the supplement). All assume that $\mathcal{S}$ is fixed *a priori* — that $\mathcal{S}$ does not change as iteration proceeds.

**Theorem 1.** If we have a budget of one iteration left, then ECI is one-step Bayes-optimal among all feasible policies.

This result is not particularly surprising. The next pair of less trivial theorems concern the convergence of CAS using ECI. Because CAS seeks to sample a diverse set of points, we measure this diversity with *fill distance*. Fill distance is the standard measure of point diversity in the quasi-Monte Carlo (Joy et al., 1996), experimental design (Pronzato & Müller, 2012), and meshfree approximation (Fasshauer & McCourt, 2015) communities. The convergence of the fill distance of ECI thus describes the limiting behavior of its sample diversity.

Given a set of sample points $\mathbf{X}$, the fill distance is formally defined as the following:

$$\textsc{fill}(\mathbf{X}, \mathcal{S}) = \sup_{\mathbf{x} \in \mathcal{S}} \min_{\mathbf{x}_j \in \mathbf{X}} d(\mathbf{x}_j, \mathbf{x}). \qquad (3)$$

In Euclidean space, $\textsc{fill}(\mathbf{X}, \mathcal{S})$ is the radius of the largest empty ball one can fit in $\mathcal{S}$, and measures the spacing of $\mathbf{X}$ in $\mathcal{S}$. The smaller a set's fill, the better distributed it is within $\mathcal{S}$. Minimal fill is defined in the following sense:

$$\rho = \min_{\mathbf{X} \in \mathcal{X}^n} \textsc{fill}(\mathbf{X}, \mathcal{S}).$$

Low-discrepancy sequences and Latin hypercubes achieve low fill in simple domains. However, computing the minimal fill is generically NP-hard (Pronzato & Müller, 2012).

Having defined fill distance, we can now frame convergence in terms of the limiting behavior of $\textsc{fill}(\mathbf{X}, \mathcal{S})$ as $n$ approaches infinity. This is formalized as follows.

**Theorem 2.** If $\mathcal{S}$ is dense, ECI produces a dense sequence of evaluation points in $\mathcal{S}$. Equivalently, the limit as $n$ approaches infinity of $\textsc{fill}(\mathbf{X}, \mathcal{S})$ is zero.

A more interesting remark is that we are able to bound the *rate* at which ECI fills the set $\mathcal{S}$. We can guarantee that ECI produces a fill bounded above by $4\rho$, and that for a sufficiently large iteration budget, ECI produces a fill that is no more than $2\rho$.

**Theorem 3.** For a fixed and known $\mathcal{S}$, let $\rho = \min_{\mathbf{X}} \textsc{fill}(\mathbf{X}, \mathcal{S})$. For any $n$, ECI produces a fill bounded above by $4\rho$. Furthermore, there exists a finite $\mathbf{n}^*$ for which ECI produces a 2-approximation ratio of $\rho$ for all $n \geq \mathbf{n}^*$.

## 4. Experiments

In our experiments, we build an independent Gaussian process (GP) to model each objective (Rasmussen & Williams, 2005). In particular, for a GP model and a single objective,

$$p(Z(\mathbf{x}) = 1 \mid \mathcal{D}_t) = \int_{\tau}^{\infty} \phi(y \mid \mu(\mathbf{x}), v(\mathbf{x}))dy, \qquad (4)$$

where $\phi$ is the Normal PDF, and $\mu(\mathbf{x})$ and $v(\mathbf{x})$ are the GP's posterior mean and variance, respectively, evaluated at $\mathbf{x}$. For $m$ objectives modeled by $m$ independent GPs, $p(Z(\mathbf{x}) = 1 \mid \mathcal{D}_t)$ is the product of probabilities associated with each model. Our GPs use a zero mean function and a $C^4$ Matérn covariance kernel with length scales fixed *a priori* (more information is provided in the supplement).

### 4.1. Baseline Methodologies

In addition to random search (RND), we compare our algorithm to several (modified) baseline algorithms aforementioned in Section 2: $\varepsilon$-constraint BO (Haghanifar et al., 2020), STRADDLE (Bryan et al., 2005), and one-step active search (Garnett et al., 2012). We extend STRADDLE to the multiobjective setting by alternating the objective functions at each iteration. For one-step active search, we maximize the point with highest probability $p(Z(\mathbf{x}) = 1 \mid \mathcal{D}_t)$ at each iteration, which we call ONE-S. In addition, we develop two information theoretic baselines:

**Mutual information** (EZ)**:** We consider the mutual information between $\mathbf{y}$ and $Z$. Assuming noiseless observations $f_i(\mathbf{x}) = \mathbf{y}_i, \forall i$, this simplifies to the entropy of $Z$:

$$\textsc{mi}(\mathbf{y}; Z) = H(\mathbf{y}) - \mathbb{E}_Z\big[H(\mathbf{y} \mid Z)\big] = H(Z),$$

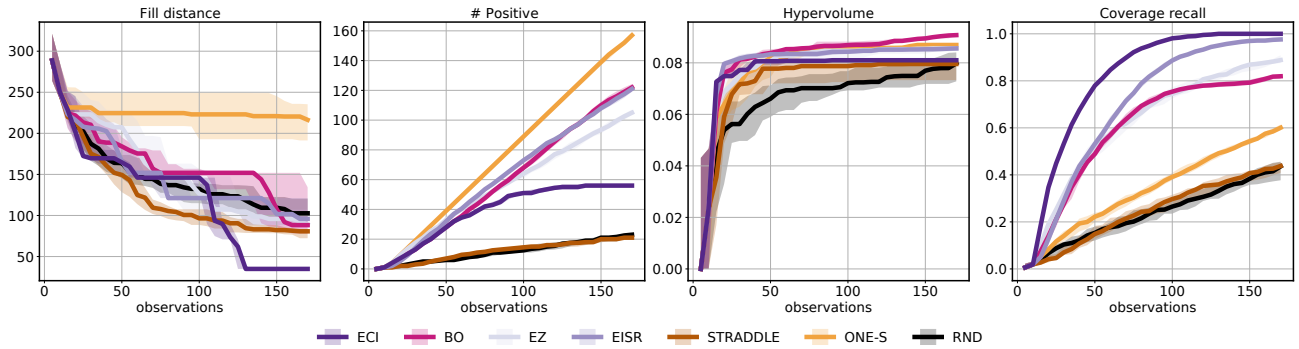


*Figure 3.* Median metric values across 20 repetitions for the additive manufacturing application. We plot the fill distance, number of positive samples, hypervolume, and coverage recall respectively. Shaded region corresponds to values between lower and upper quartiles.

where $H(\cdot)$ is the entropy of a random variable. This would be the one-step optimal strategy that recovers the most information about the boundaries of $\mathcal{S}$.

**Entropy inside $\mathcal{S}$ (EISR):** We consider the utility:

$$u_{\text{EISR}}(\mathbf{x}) = \begin{cases} 0 & Z = 0, \\ H(f \mid \mathbf{x}) & Z = 1. \end{cases}$$

$u_{\text{EISR}}$ rewards sampling configurations that are inside $\mathcal{S}$ *and* have high entropy. The corresponding acquisition function can be written as

$$\alpha_{\text{EISR}} = \mathbb{E}_Z\left[u_{\text{EISR}}(\mathbf{x})\right] = p(Z(\mathbf{x}) = 1)H(\mathbf{y}).$$

### 4.2. Performance Criteria

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 ECI to existing baselines. To that end, we consider the following four criteria.

**Fill distance**: We want an algorithm that can effectively explore the entire satisfactory region. This is quantified by fill distance, which is defined in Equation 3.

**Positive samples**: The number of satisfactory sampled points is commonly used in active search literature to judge the quality of the algorithm. In our experiments, these are the observed samples inside of $\mathcal{S}$.

**Hypervolume**: We measure the hypervolume of region in metric 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.

**Coverage recall**: Finally, we want to combine the idea of maximizing positive samples and minimizing fill distance to measure sample diversity. We propose a recall metric that considers coverage. Given a known $\mathcal{S}$ and samples $\mathbf{X}$, we define the coverage recall as

$$\text{Coverage recall} = |\mathbb{N}_r(\mathbf{X}) \cap \mathcal{S}| \,/ |\mathcal{S}|.$$

Note that the numerator is exactly the utility function that we have proposed in Section 3.1. We will simply refer to this as recall in the following text.

### 4.3. Multiobjective Design Problem Suite

We adopt eight multiobjective engineering design problems presented in the REPROBLEM suite (Tanabe & Ishibuchi, 2020). These synthetic functions are specifically designed to mimic real-world applications and avoid having unrealistic properties of traditional benchmark functions. We follow the naming convention from the original text and detail the problems we selected in the supplement.

### 4.4. Additive Manufacturing

Haghanifar et al. (2020) studies a numerical simulation to help design an additive manufacturing strategy for minimizing the reflection of light at multiple angles of incidence. We conduct the same search for different nanostructures which balance a desire for minimizing normal reflection with reflection at an oblique angle.

The simulation, conducted using the Lumerical FDTD simulator (Lumerical Inc., 2020), has a level of inaccuracy associated with the numerical methods. In addition, there is a level of imprecision associated with the actual manufacturing process which means that the desired design parameters are only realized to limited precision during fabrication. Our goal for this reflection minimization problem is to find nanostructured glass that satisfying $R_{\text{normal}}(\mathbf{x}) \leq \tau_{\text{normal}}$ and $R_{\text{oblique}}(\mathbf{x}) \leq \tau_{\text{oblique}}$. The design parameters are the dimension of the nanostructures (a cone): maximum cone width, bottom diameter, top diameter, and height of the structure.

*Table 1.* Selected experimental results, consisting of the mean over 20 independent trials. Hypervolume has been rescaled for ease of viewing: RE33 $\times 10^{-5}$, Additive $\times 10^{-2}$, EEG $\times 10^{2}$, Plasma $\times 10^{0}$. We denote with bold text the best mean for each problem and each metric. Please refer to the supplementary material for the complete set of experimental results and information regarding the variation around the mean (standard error). For each problem, ECI has the best recall performance.

| Function | $m$ | $d$ | $r$ | Methods | Fill distance ↓ | # Positive ↑ | Coverage Recall ↑ | Hypervolume ↑ |
|---|---|---|---|---|---|---|---|---|
| RE33 | 3 | 4 | 0.08 | RND | 0.34 | 23.80 | 0.14 | 5.53 |
| | | | | ONE-S | 0.60 | **101.55** | 0.18 | 7.03 |
| | | | | EZ | 0.26 | 34.05 | 0.25 | 2.77 |
| | | | | STRADDLE | 0.35 | 13.70 | 0.10 | 2.92 |
| | | | | BO | 0.42 | 36.55 | 0.11 | **8.65** |
| | | | | EISR | **0.24** | 100.05 | 0.23 | 7.46 |
| | | | | ECI | 0.27 | 100.95 | **0.73** | 7.91 |
| Additive | 2 | 4 | 35.0 | RND | 108.22 | 22.40 | 0.42 | 7.94 |
| | | | | ONE-S | 211.86 | **156.75** | 0.60 | 8.77 |
| | | | | EZ | 92.37 | 106.05 | 0.89 | 8.52 |
| | | | | STRADDLE | 79.30 | 21.10 | 0.43 | 7.78 |
| | | | | BO | 107.84 | 121.95 | 0.82 | **9.07** |
| | | | | EISR | 102.01 | 121.40 | 0.98 | 8.56 |
| | | | | ECI | **34.92** | 56.00 | **1.00** | 8.07 |
| EEG | 2 | 6 | 0.25 | RND | **0.38** | 4.40 | 0.58 | 3.17 |
| | | | | ONE-S | 0.48 | **31.20** | 0.50 | 3.45 |
| | | | | EZ | 0.44 | 17.95 | 0.58 | 3.45 |
| | | | | STRADDLE | 0.41 | 2.00 | 0.38 | 2.28 |
| | | | | BO | 0.52 | 4.30 | 0.35 | 2.40 |
| | | | | EISR | 0.45 | 22.85 | 0.52 | 3.55 |
| | | | | ECI | **0.38** | 8.30 | **0.79** | **3.65** |
| Plasma[3] | 1 | 9 | 0.50 | RND | 1.02 | 10.85 | 0.19 | 5.63 |
| | | | | ONE-S | 1.01 | 12.80 | 0.20 | 5.88 |
| | | | | EZ | 1.07 | 10.70 | 0.17 | 5.48 |
| | | | | STRADDLE | 1.07 | 10.70 | 0.17 | 5.48 |
| | | | | BO | 1.04 | 12.60 | 0.15 | 5.74 |
| | | | | EISR | 1.03 | 10.55 | 0.14 | 6.08 |
| | | | | ECI | **0.90** | **31.25** | **0.65** | **6.64** |

### 4.5. EEG for Brain Activity Reconstruction

Electroencephalography (EEG) is a non-invasive strategy for reconstructing cerebral electric currents from their resulting electrostatic potential measured on a human's scalp (Ala et al., 2017). This process consists of solving an inverse problem, requiring localization of the electric dipoles' locations and moments, subject to the associated model of dipole/potential interaction (Sorrentino & Piana, 2017). The limited spatial resolution of EEG is a well-known problem with its practical viability (Samuelsson et al., 2021), making constraint active search a potentially interesting tool for identifying regions of brain activity.

We use a meshfree solver to generate a scalp potential from a proposed dipole location through a quasi-stationary approximation to Maxwell's equations (Ala et al., 2015). We then apply our strategy to identify the proposed dipole which most closely matches the true potential created by the true dipole, which is fixed at the start but hidden. We measure scalp potential at 28 locations and try to minimize the 25th and 75th percentiles of absolute deviation between the proposed and true potential; this has the effect of pushing both ends of the interquartile range as low as possible.

### 4.6. 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,

[3] Since the plasma problem is single objective, the hypervolume is computed as the difference between the threshold and the minimum value, $|\tau - \min \mathbf{y}|$.

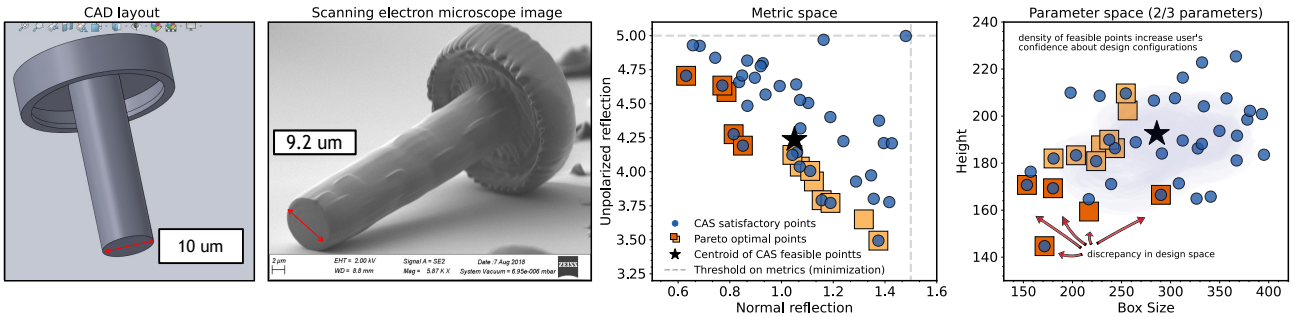


*Figure 4.* Left two panels: difference from the expected design (in CAD) and the fabricated sample of a nanostructure. Right two panels: comparison of acceptable configurations in metric space and in the parameter space. Small variations in the Pareto optimal points can lead to significant discrepancies in the design space. CAS points provide more information about satisfactory design configuration.

and possess notably warped shapes due to the complex, quasi-symmetric magnetic field they must produce.

Recently, (Giuliani et al., 2020) developed a stellarator simulator, PyPlasmaOpt, which significantly streamlines the simulation procedure of previous packages such as StellOpt (Strickler et al., 2003). PyPlasmaOpt parameterizes each coil as a curve in 3D Cartesian coordinates $\Gamma(\theta) = (x(\theta), y(\theta), z(\theta))$, where each coordinate admits the following Fourier expansion, e.g., in the case of $x$:

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

For each coil, the parameters $c_k$ and $s_k$ represent the search space that we must search over. We consider a relatively simple, order one coil which has nine parameters total.

We compute the objective function to be minimized by simulating the stellarator using the given coil shapes (which solves a certain first order, nonlinear ordinary differential equation). The resulting summary information from the simulation is our objective $f$ and contains three terms:

$$f(\mathbf{x}) = R_{\text{magnetic}}(\mathbf{x}) + R_{\text{transform}}(\mathbf{x}) + R_{\text{shape}}(\mathbf{x}).$$

The first term quantifies the quasi-symmetry of the magnetic field —the smaller the first term, the more desirable the resulting field. The second term locks the solution into a target rotational transform. The third term penalizes overly complex coils too impractical to manufacture in real life.

### 4.7. Discussion

We exhaustively tested ECI and other baselines on eleven total problems, but due to limited space, we present a representative subset of our experimental results in this section. Our analysis holds true for the seven additional problems, whose results are in the supplement along with accompanying experimental setups for reproducibility purposes.

Specifically, Table 1 shows the mean value over 20 trials for each performance metric at termination for the RE33 function from the design problem suite and all our real-world applications. These results demonstrate that no algorithm can be the best in all performance metrics. In fact, it highlights each method's strengths and weaknesses, confirming our thesis that evaluating the performance of multiobjective problems is itself multiobjective.

As expected, BO typically outperforms all methods on the hypervolume criterion, which only judges the Pareto frontier. However, it is consistently worse than its peers when judged by diversity (fill distance) and coverage (recall). Our proposed method ECI, on the other hand, is effective at diverse sampling inside the satisfactory region $\mathcal{S}$; it is almost always the best performing method for coverage recall and fill distance. Unsurprisingly, ONE-S excels at accumulating satisfactory samples since it greedily maximizes the number of positive samples by design. These samples are highly clustered, as revealed by the high fill distance and low coverage recall values.

Each policy's objective is evident when we consider the sample progression. Figure 3 shows the variation on performance as a function of the number of observed samples for the additive manufacturing problem. For instance, ECI maximizes the coverage recall and also minimizes the fill distance but stagnates at maximizing hypervolume and number of positives.

Our investigation reveals that different multiobjective algorithms prioritize different performance criteria. Therefore, one can tailor a strategy (perhaps an ensemble of methods) according to each application's need. If metric optimization is of ultimate relevance, we recommend multiobjective Bayesian optimization. If the decision maker's goal is to identify rare elements regardless of diversity, then active search is the best choice. However, we argue that selecting a diverse set of satisfactory configurations is the most suited objective for design problems. We provide an example to support this claim in the following section.

## 5. Constraint Active Search in Practice

This research was motivated, in part, by a collaboration with materials scientists using additive manufacturing to fabricate nanostructured glass with multiple desirable optical and physical characteristics. Multiobjective problems are intrinsically subjective — if we could quantify the tradeoffs between each metric precisely *a priori*, we would solve a single-objective problem instead. Therefore, multiobjective methodologies provide the decision makers with a *set* of points at termination. Traditional multiobjective optimization offers a Pareto optimal set of points; our alternative gives a set of diverse points in parameter space that meet the user constraints. The decision maker, a domain expert, must then decide which one or more of these points to promote to physical manufacturing, often considering two key complications:

- Computational and manufacturing imprecision prevent an exact physical realization of the results. See Figure 4 (left two panels) for an example of the CAD model and the associated fabricated component, which suffered from limited manufacturing precision.
- Some metrics are unavailable during optimization – maybe they can only be computed after manufacturing (such as anti-fogging or durability) or require costly expert analysis (such as the supervisor stating expert opinions on a manufacturing scale).

Both of these complications are better addressed through sample diversity in parameter space (not in metric space) — the primary motivation of CAS.

Inaccuracies arising from the transfer of computational results to manufactured outcomes are best dealt with by choosing points most likely to achieve acceptable performance after fabrication. Because CAS explores the parameter space, it has a greater ability to select results far (e.g., in Euclidean sense) from unacceptable outcomes than multiobjective optimization. When only studying the Pareto frontier, no energy is spent exploring the safety of those points from undesirable results. Figure 4 (right) shows the numerical simulation results for the additive manufacturing example; interior points in parameter space are more likely to be selected for physical manufacturing.

For metrics only computable after fabrication, users benefit from a diverse set of designs in parameter space so they may manufacture a select few and still hope to achieve good results on these previously unavailable metrics.

## 6. Conclusion and Future Work

We introduce constraint active search as a novel methodology to explore multiobjective experimental design options while satisfying minimum performance criteria. One key differentiation between constraint active search and existing multiobjective optimization literature is the emphasis on the design parameters: constraint active search values sample diversity in the satisfactory region.

We propose a new search policy, expected coverage improvement, designed to efficiently improve the coverage of sampled designs over the satisfactory region; theoretical properties of this strategy include a constant approximation ratio to the optimal sample diversity. We define and study several performance criteria when judging the performance of proposed algorithms in multiobjective settings, and we explain the implications of prioritizing each. Using these criteria, we compare our new policy against existing strategies such as Bayesian optimization, level set estimation, and active search on synthetic and real world problems.

There are many open problems in the topic of constraint active search for design and simulation; here we prioritize a few topics for future investigation.

**Distinctness/Precision:** A key element of ECI (as well as the active search baseline) is the choice of the resolution $r$, through which the design precision is defined. This is acceptable, but limited: categorical parameters may exist; different parts of the domain may have different manufacturing precision; some parameters may have different levels of precision than others. Eventually, it would be beneficial to allow for a more abstract definition of design distinctness to support more complicated design circumstances.

**Unknown precision:** In most design/manufacturing settings, the resolution $r$ is known *a priori*. It is possible, however, that this will need to be discovered over the course of the search process (or, maybe, after the search process). For this situation, our proposed policy would benefit from an adaptive $r$ definition, which might start at a very large value e.g., 10% of the maximum distance between points in the domain, and then scale down as the search progresses.

**Parallelism:** ECI is limited to the sequential setting; extending the search policy to the batch setting and finding a computationally tractable policy is non-trivial.

**Judging performance:** Some of our proposed performance criteria are only computable for benchmark problems where the true satisfactory region can be estimated. In practice, judging the results of a search algorithm is less straightforward. We believe creating more practical performance criteria can accelerate our understanding of this problem.

## Acknowledgements

We want to thank Sajad Haghanifar and Paul Leu for providing the scanning electron microscope images and motivating the research presented in this paper.

## References


Ala, G., Fasshauer, G., Francomano, E., Ganci, S., and McCourt, M. J. The method of fundamental solutions in solving coupled boundary value problems for M/EEG. *SIAM Journal on Scientific Computing*, 37(4): B570–B590, 2015.

Ala, G., Fasshauer, G. E., Francomano, E., Ganci, S., and McCourt, M. J. An augmented MFS approach for brain activity reconstruction. *Mathematics and Computers in Simulation*, 141:3–15, 2017.

Ariafar, S., Coll-Font, J., Brooks, D., and Dy, J. ADMMBO: Bayesian optimization with unknown constraints using ADMM. *Journal of Machine Learning Research*, 20(123): 1–26, 2019.

Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., Chen, M. H., Cheong, B., Perkins, N., Yang, Z., Herring, P. K., Aykol, M., Harris, S. J., Braatz, R. D., Ermon, S., and Chueh, W. C. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795):397–402, 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-1994-5.

Belakaria, S., Deshwal, A., and Doppa, J. R. Max-value entropy search for multi-objective Bayesian optimization. In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 7823–7833, 2019.

Belakaria, S., Deshwal, A., and Doppa, J. R. Max-value entropy search for multi-objective Bayesian optimization with constraints. Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), 2020.

Bryan, B. and Schneider, J. G. Actively learning level-sets of composite functions. In Cohen, W. W., McCallum, A., and Roweis, S. T. (eds.), *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, volume 307 of *ACM International Conference Proceeding Series*, pp. 80–87. ACM, 2008. doi: 10.1145/1390156.1390167.

Bryan, B., Schneider, J. G., Nichol, R., Miller, C. J., Genovese, C. R., and Wasserman, L. A. Active learning for identifying function threshold boundaries. In *Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]*, pp. 163–170, 2005.

Daulton, S., Balandat, M., and Bakshy, E. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

del Rosario, Z., Rupp, M., Kim, Y., Antono, E., and Ling, J. Assessing the frontier: Active learning, model accuracy, and multi-objective candidate discovery and optimization. *The Journal of Chemical Physics*, 153(2):024112, 2020. doi: 10.1063/5.0006124.

Duris, J., Kennedy, D., Hanuka, A., Shtalenkova, J., Edelen, A., Baxevanis, P., Egger, A., Cope, T., McIntire, M., Ermon, S., and Ratner, D. Bayesian optimization of a free-electron laser. *Phys. Rev. Lett.*, 124:124801, 2020. doi: 10.1103/PhysRevLett.124.124801.

Emmerich, M. T. M., Deutz, A. H., and Klinkenberg, J. W. Hypervolume-based expected improvement: Monotonicity properties and exact computation. In *2011 IEEE Congress of Evolutionary Computation (CEC)*, pp. 2147–2154, 2011. doi: 10.1109/CEC.2011.5949880.

Eriksson, D. and Poloczek, M. Scalable constrained Bayesian optimization. In Banerjee, A. and Fukumizu, K. (eds.), *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*, volume 130 of *Proceedings of Machine Learning Research*, pp. 730–738. PMLR, 2021.

Fasshauer, G. E. and McCourt, M. J. *Kernel-based Approximation Methods Using Matlab*. World Scientific, 2015. doi: 10.1142/9335.

Fernández-Sánchez, D., Garrido-Merchán, E. C., and Hernández-Lobato, D. Improved max-value entropy search for multi-objective Bayesian optimization with constraints, 2021. URL https://arxiv.org/abs/2011.01150.

Forrester, A., Sobester, A., and Keane, A. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

Frazier, P. I. Bayesian optimization. In Gel, E. and Ntaimo, L. (eds.), *Recent Advances in Optimization and Modeling of Contemporary Problems*, pp. 255–278. INFORMS, 2018.

Gardner, J. R., Kusner, M. J., Xu, Z. E., Weinberger, K. Q., and Cunningham, J. P. Bayesian optimization with inequality constraints. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014,*

*Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pp. 937–945. JMLR.org, 2014.

Garnett, R., Krishnamurthy, Y., Xiong, X., Schneider, J., and Mann, R. Bayesian optimal active search and surveying. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pp. 843–850, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

Gelbart, M. A. *Constrained Bayesian optimization and applications*. PhD thesis, Harvard, 2015.

Giuliani, A., Wechsung, F., Cerfon, A., Stadler, G., and Landreman, M. Single-stage gradient-based stellarator coil design: Optimization for near-axis quasi-symmetry, 2020. URL https://arxiv.org/abs/2010.02033.

Gotovos, A., Casati, N., Hitz, G., and Krause, A. Active learning for level set estimation. In Rossi, F. (ed.), *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pp. 1344–1350. IJCAI/AAAI, 2013.

Haghanifar, S., McCourt, M., Cheng, B., Wuenschell, J., Ohodnicki, P., and Leu, P. W. Discovering high-performance broadband and broad angle antireflection surfaces by machine learning. *Optica*, 7(7):784–789, 2020. doi: 10.1364/OPTICA.387938.

Iwazaki, S., Inatsu, Y., and Takeuchi, I. Bayesian experimental design for finding reliable level set under input uncertainty. *IEEE Access*, 8:203982–203993, 2020. doi: 10.1109/ACCESS.2020.3036863.

Jiang, S., Malkomes, G., Converse, G., Shofner, A., Moseley, B., and Garnett, R. Efficient nonmyopic active search. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1714–1723. PMLR, 2017.

Jiang, S., Malkomes, G., Abbott, M., Moseley, B., and Garnett, R. Efficient nonmyopic batch active search. In Bengio, S., Wallach, H. M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 1107–1117, 2018.

Joy, C., Boyle, P. P., and Tan, K. S. Quasi-Monte Carlo methods in numerical finance. *Management Science*, 42 (6):926–938, 1996.

Knowles, J. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2006. doi: 10.1109/TEVC.2005.851274.

Konakovic-Lukovic, M., Tian, Y., and Matusik, W. Diversity-guided multi-objective Bayesian optimization with batch evaluations. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Koziel, S., Bekasiewicz, A., Couckuyt, I., and Dhaene, T. Efficient multi-objective simulation-driven antenna design using co-Kriging. *IEEE Transactions on Antennas and Propagation*, 62(11):5900–5905, 2014.

Letham, B., Karrer, B., Ottoni, G., and Bakshy, E. Constrained Bayesian optimization with noisy experiments. *Bayesian Anal.*, 14(2):495–519, 2019. doi: 10.1214/18-BA1110.

Lumerical Inc. Lumerical device multiphysics simulation suite, FDTD solver, 2020. URL https://www.lumerical.com/products/.

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

Negoescu, D. M., Frazier, P. I., and Powell, W. B. The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS Journal on Computing*, 23(3):346–363, 2011.

Paria, B., Kandasamy, K., and Póczos, B. A flexible framework for multi-objective Bayesian optimization using random scalarizations. In Adams, R. P. and Gogate, V. (eds.), *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pp. 766–776, Tel Aviv, Israel, 2020. PMLR.

Pronzato, L. and Müller, W. G. Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701, 2012.

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

Samuelsson, J. G., Peled, N., Mamashli, F., Ahveninen, J., and Hämäläinen, M. S. Spatial fidelity of MEG/EEG source estimates: A general evaluation approach. *NeuroImage*, 224:117430, 2021.

Singh, P., Couckuyt, I., Elsayed, K., Deschrijver, D., and Dhaene, T. Shape optimization of a cyclone separator using multi-objective surrogate-based optimization. *Applied Mathematical Modelling*, 40(5-6):4248–4259, 2016.

Song, J., Tokpanov, Y. S., Chen, Y., Fleischman, D., Fountaine, K. T., Atwater, H. A., and Yue, Y. Optimizing photonic nanostructures via multi-fidelity Gaussian processes. NeurIPS Workshop on Machine Learning for Molecules and Materials, 2018.

Sorrentino, A. and Piana, M. Inverse modeling for MEG/EEG data. In *Mathematical and Theoretical Neuroscience*, pp. 239–253. Springer, 2017.

Spitzer Jr, L. The stellarator concept. *The Physics of Fluids*, 1(4):253–264, 1958.

Strickler, D., Berry, L., and Hirshman, S. Integrated plasma and coil optimization for compact stellarators. Technical report, International Atomic Energy Agency (IAEA), 2003.

Sui, Y., Gotovos, A., Burdick, J. W., and Krause, A. Safe exploration for optimization with Gaussian processes. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 997–1005. JMLR.org, 2015.

Sui, Y., Zhuang, V., Burdick, J. W., and Yue, Y. Stagewise safe bayesian optimization with Gaussian processes. In Dy, J. G. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4788–4796. PMLR, 2018.

Tanabe, R. and Ishibuchi, H. An easy-to-use real-world multi-objective optimization problem suite. *Applied Soft Computing*, 89:106078, 2020. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2020.106078.

Wang, Z., Garrett, C. R., Kaelbling, L. P., and Lozano-Perez, T. Active model learning and diverse action sampling for task and motion planning. In *International Conference on Intelligent Robots and Systems (IROS)*, 2018.

Zanette, A., Zhang, J., and Kochenderfer, M. J. Robust super-level set estimation using Gaussian processes. In Berlingerio, M., Bonchi, Francescoand Gärtner, T., Hurley, N., and Ifrim, G. (eds.), *Machine Learning and Knowledge Discovery in Databases*, pp. 276–291, Cham, 2019. Springer International Publishing. ISBN 978-3-030-10928-8.