

Adaptive Sampling with Automatic Stopping for Feasible Region Identification in Engineering Design

Jixiang Qing · Nicolas Knudde ·
Federico Garbuglia · Domenico Spina ·
Ivo Couckuyt · Tom Dhaene

Received: date / Accepted: date

Abstract Engineering design is a complex process to find a suitable trade-off among different, and sometimes conflicting, design specifications. In reality, these requirements can be often considered as constraints of the design problem, that can be defined in terms of performance measures or geometrical characteristics of the device under study. In this paper, a new design space exploration methodology is presented for discovering feasible regions in the design space, where the term feasible region indicates the set of all design configurations satisfying all constraints of the design problem. The proposed method is based on Gaussian Process metamodels to estimate the feasible region and leverages a information-based adaptive sampling technique to sequentially refine the prediction accuracy, which is applicable for multiple constraints problems. In order to efficiently stop the adaptive sampling process, a novel framework to estimate the metamodel's prediction accuracy is proposed. The efficiency, accuracy and robustness of the proposed approach are compared with state-of-art techniques on suitable benchmark problems and practical engineering examples.

Keywords Feasible region · Gaussian Process · Adaptive sampling · Stopping criterion

1 Introduction

Modern engineering design is a complex and systematic task. Different disciplines are involved and various goals are set depending on the design stages. In the early stage of the design pipeline, a trade-off between conflicting design requirements must be found to generate satisfying solutions for the problem

J. Qing
Ghent University – imec, IDLab, Department of Information Technology (INTEC), Technologiepark – Zwijnaarde 15, 9052 Ghent, Belgium
E-mail: Jixiang.Qing@UGent.be

at hand. Without loss of generality, these requirements are presented as constraints. The designs satisfying such constraints are considered *feasible* and can be good candidates for detailed investigation. In practice, Autodesk [1] pointed out that, in the preliminary design stage, designers would use design optimization (e.g., [2, 3]) as a heuristic design tool. These algorithms are originally proposed to push the performance of a design problem to its extreme as a finalizing step in the design process [4], but now optimization has been misused to explore the space of feasible designs at the beginning of the design cycle. This interesting phenomenon stems from the fact that there is a need for a generative design process where designers are presented with several interesting design candidates.

Nevertheless, the sole emphasis on optimality by optimization algorithms may not be the best choice at this stage and there is a need of techniques identifying all the suitable designs. Several arguments support this claim. First, the optimal design configuration may be just used as the start of the design process [4], being too myopic at limited number of optima may ignore some interesting design candidates. Secondly, especially at the early design stage, the objectives of the system that need to be optimized may not yet be clearly defined [5] and designers need a strategy to gain understanding about the design space. In this case, unconventional but feasible designs may also be valuable and worth further investigation. Additionally, the obtained result may not transfer well to reality [6] (a phenomenon called the reality gap [7]). In these cases, a large collection of suitable design configurations can quickly provide a working solution.

This alternative interpretation can be formulated as a *feasible region identification* problem, where one seeks to make accurate predictions if a design configuration is feasible or not:

$$\begin{aligned} \text{Find } A &:= \{\mathbf{x} \in \mathcal{X} : I_A(f(\mathbf{x})) = 1\} \\ \text{where } I_A(f(\mathbf{x})) &= \begin{cases} 1 & \alpha \leq f(\mathbf{x}) \leq \beta \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (1)$$

$f : \mathcal{X} \rightarrow \mathbb{R}$ is a noise-free, continuous, black box constraint function defined on a compact set $\mathcal{X} \subset \mathbb{R}^d$. $I_A(\cdot)$ is an indicator function defined on an interval of the output space explicitly specified by the designer, which denotes if design requirements are satisfied. In the bounded space \mathcal{X} , a design \mathbf{x} is called *feasible* if its corresponding constraint function value $f(\mathbf{x})$ is in the interval $[\alpha, \beta]$. The region is called *feasible region* if all designs within are feasible. Specifically, we denote $\alpha, \beta \in \{\mathbb{R}, -\infty, \infty\}$ as *feasible boundary*, while $\{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}) \in \{\alpha, \beta\}\}$ is denoted as *feasible region boundary*. In general, more than one constraint f can be applied in a design problem, hence the feasible region will be the intersection of the ones corresponding to each different f .

In this framework, Computer Aided Design (CAD) simulations are often involved to evaluate if the constraints are satisfied, leading to feasible design configurations. However, given the complexity of modern engineering systems, CAD simulations can be both time- and resource-intensive. For instance, an

electromagnetic simulator or a computational fluid dynamic solver can take hours [8] or even days [9] to perform a single evaluation of f . Hence, it is of paramount importance to limit the number of (expensive) evaluations during the design phase, while also being able to provide a large number of feasible designs with high accuracy. In this scenario, *active learning* [10,11] (or adaptive sampling¹ [12–14]) is an elegant tool to solve this problem. Active learning (AL) is also known as the sequential experimental design method in statistics [15]: it sequentially extends the data set to explore the design space in a sample-efficient way, in order to reduce uncertainty with respect to certain properties. Such desired properties can either be the maximum location of a function, which some examples are shown in Bayesian Optimization (or surrogate based optimization) [16,17], the overall accuracy of regression model [18], or a feasible region satisfying all constraints. The last case is considered in our study.

In this paper, we introduce an information theoretic-based active learning framework to make accurate feasibility predictions within a bounded design space. In order to reach an efficient termination of the active learning process, a new framework of termination criterion, which we alternatively refer to as stopping criterion, is proposed to provide a probabilistic belief of the metamodel’s prediction accuracy. We theoretically investigate the relations between the proposed stopping criterion and the test accuracy metrics commonly used for binary classification problem. The performance of our novel method are validated through comparison with respect to state-of-the-art methods on synthetic experiments. Additionally, two real world design problems, including one involving multiple design constraints, are also presented.

This paper is structured as follows. First, the background and related techniques are described in Section 2. The novel active learning framework, including the active learning method and the framework of the stopping criteria, is introduced in Section 3. The numerical experiments and real world applications are presented in Section 4. Conclusions are summed up in Section 5.

2 Background and Related Work

The generic active learning framework for feasible region identification problems is shown in Fig. 1. Given the initial samples, the corresponding function values are evaluated by expensive black box functions. Subsequently, a metamodel is fitted to the data. Next, a new sample of the design parameters is selected according to some specified criterion built upon the model (*infill criterion* or *acquisition function*). The key characteristics of the active learning are two fold: the model provides a description of the feasible region, and one can utilize the model prediction to formulate an acquisition function, which provides on a measure how informative an unevaluated sample can be. Then, the acquisition function is used to choose the next sample to evaluate. The

¹ Note that in this paper the terms “active learning” and “adaptive sampling” are used interchangeably.

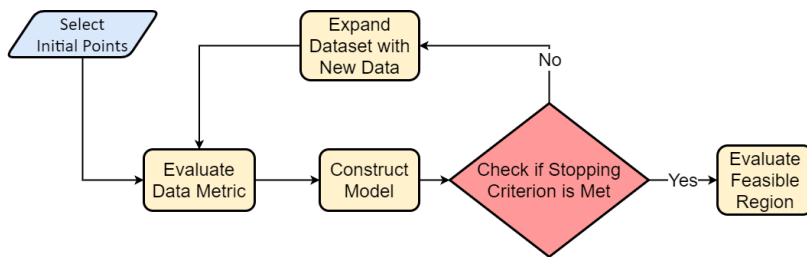


Fig. 1: Generic flowchart of active learning for feasible region identification

model is updated accordingly to provide a new (and more accurate) description of the feasible region. This process is repeated until a stopping criterion is met. Finally, the feasible region is accurately described by the metamodel.

2.1 Metamodels for characterizing feasibility

A metamodel, or surrogate model, is a numerical approximation of the black box constraint function. Making feasibility predictions with metamodels is a common strategy. This problem has a close relation with reliability analysis [19], where the computation complexity is mainly from the statistical estimation of f in Eq. 1. In other contexts, this problem is also widely studied [20] and is referred to as *excursion set* [21,22] and *level set* [23,24] estimation, where one wants to predict if an unevaluated observation's outcome has surpassed a specified level. From Eq. 1, a unified view can be given by regarding *level set estimation*, *excursion search* and *feasible region identification* as the same problem².

In order to make accurate predictions, several approaches have been proposed in literature [5, 13, 25, 26]. These methodologies are regression-based, rather than classification-based, since f is a real-valued continuous function in many relevant engineering applications. We refer [11, 27, 28] for classification based methods in case of f is discrete-valued,

2.2 Active feasible region identification

Once the metamodel has been constructed, active learning will improve the confidence in predicting the feasible region. Several techniques have been presented in literature to actively identify the feasible region. An early one actively selects queries to learn the boundaries that divide a function domain

² The main research of level set estimation is built on discrete data sets, where one tries to efficiently identify all the data of which the outcomes are above a given level. On the other hand, recent research, for instance [23, 24], has mentioned that the problem can be generalized for a continuous domain, which becomes very relevant for feasible region identification.

into regions where the function is above or below a given threshold [26]. A heuristic acquisition function has also been proposed to encourage sampling near the feasible region boundary, while having large uncertainty at the same time. A weighted integration mean square error (IMSE_T) strategy is proposed by Picheny [25] to measure the belief of the metamodel about the accuracy of the feasible region. A new sampling technique is then developed, based on the fact that model variance is not affected by the function response of new data. Several stepwise uncertainty reduction (SUR) strategies for the excursion set estimation problem have also been developed in [20], which employ sequential sampling to minimize the estimation error or uncertainty, measured by Bayesian risk. Some SUR methods' consistency property have also been established recently [29]. A variant of Probability of Feasibility (PoF) [30] has been proposed in [31], where a predicting variance term has been heuristically added to encourage exploring within the design space. Since most of the works assume the presence of the feasible region within a bounded design space, an active expansion sampling technique is introduced in [11] for unbounded design space problem.

2.3 Stopping criterion

Considering the active learning framework in Fig. 1, it is fundamental to define a metric to stop the iterative refinement of the metamodel.

In order to define such a stopping criterion, it is necessary to assess when the feasible region is accurately characterized by the metamodel, i.e., if the metamodel can accurately predict if a design configuration is feasible or not. As different performance metrics have been used for regression and classification problems, first the proper problem type must be determined. The goal is to predict whether a data is within the feasible region: $I_A(f(\cdot)) \rightarrow \{0, 1\}$. This can be formulated as a binary classification problem, thus evaluated by measurements for classification metrics such as the F1 score:

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (2)$$

where precision and recall are defined as:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (3)$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (4)$$

Either precision, recall or F1 score can be used as a performance measurement, with each one having its own merits. Precision (Eq. 3) describes to what extent the predicted feasible region actually corresponds to a region satisfying the feasibility criteria, nevertheless, it lacks information of indicating whether all the feasible regions in the design space have been captured by the model. On the other hand, the recall metric (Eq. 4) provides more emphasis on to what

extent all the feasible regions have been captured by the model, though it provides no indication when the models incorrectly predicts infeasible regions as feasible ones. The F1 score, defined as the harmonic mean of precision and recall, provides a good balance for considering all these circumstances. This metric has been widely used in several applications [5,11,32] and is particularly suitable in our scenario.

Unfortunately, the direct application of the F1 score (as well as the precision and recall) is costly because of the requirement of suitable test data. Thus, common AL methods [5,11,19] use a maximum number of function evaluations as the stopping criterion. However, it is not possible to estimate upfront how many function evaluation are necessary to obtain an accurate estimate of the feasible region: adopting a predefined maximum number of function evaluations can lead to an inaccurate estimation of the feasible region or to perform too many (unnecessary) expensive evaluations of the function f . To circumvent this problem, a common approach is to determine the stopping criteria purely based on the metamodel. Perhaps the most relevant criterion is the $IMSE_T$ proposed in [25], in which the Kriging predicted variance is weighted by a probability specifying how likely a point is in the feasible region, and then averaged within the design space. The $IMSE_T$ metric presents as elegant way to measure model's belief about its accuracy within the feasible region. However, the use of this unscaled metric in reality still depends on expert knowledge to judge if the averaged variance is tolerable or not.

3 Active feasible region identification with stopping criterion

The proposed novel AL framework is presented in this section, where we first assume that the constraint functions are continuous-valued like in [5,33] and can be described via Gaussian Process regression. Then, the acquisition function is presented in (Section.3.2) and the new stopping criterion is proposed in (Section.3.3).

The main workflow of our active learning consists of the following steps:

1. Use design of experiments (DOE) techniques to select an initial set of design samples to evaluate.
2. In each subsequent iteration:
 - (a) Build a GP regression model on the evaluated samples (Section 3.1)
 - (b) Check if the stopping criterion has been satisfied (Section 3.3), if yes then proceed to step 3, otherwise continue.
 - (c) Select a new sample to be evaluated by maximizing the acquisition function (Section 3.2)
 - (d) Evaluate the new sample, augment the existing samples and go back to step 2(a)
3. Exit the AL process and return the final metamodel for prediction.

The proposed framework is illustrated in Fig. 2.

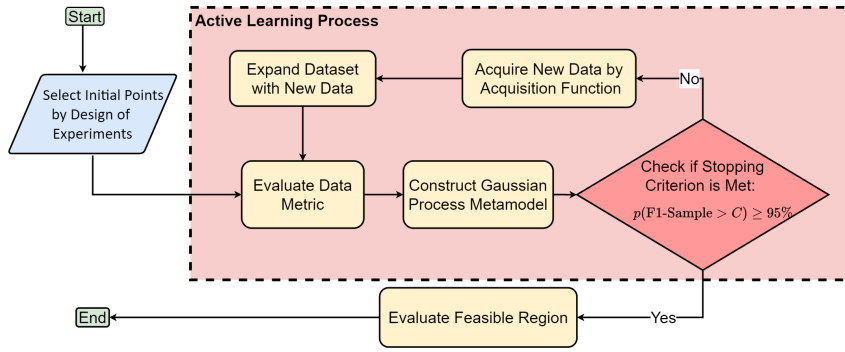


Fig. 2: Flowchart of the active learning framework: the Gaussian Process is utilized to provide a probabilistic description of the feasible region. The data is sequentially acquired by the information-theoretic based acquisition function. The whole active learning process will terminate until the metamodel is 95% confident that its prediction accuracy is larger than a predefined acceptable F1 score C .

3.1 Gaussian Process regression

There are many regression models that can be used to cheaply approximate the simulator, such as Bayesian neural networks, support vector regression, and random forest regression. Nevertheless, under the category of data-efficient machine learning, Gaussian Process (GP), also known as Kriging [34, 35], is arguably the most prevalent probabilistic technique. GPs have been widely used as a regression method in many different areas including electronics engineering [36], computational fluid dynamics [37] and chemistry [38].

A GP can be considered as a distribution over functions $f : \mathcal{X} \rightarrow \mathbb{R}$ that is completely defined by a suitable mean function $m : \mathcal{X} \rightarrow \mathbb{R}$ and covariance function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. Starting from the GP prior and the training data D_n , it is possible to derive a posterior belief of the function distribution aggregated at (possibly multiple) test data \mathbf{X}_* . This belief is Gaussian distributed:

$$\mathbf{Y}_* | D_n, \mathbf{X}_* \sim \mathcal{N}(K_{*n}^T (K_{nn} + \sigma^2 I)^{-1} \mathbf{y}, K_{**} - K_{*n}^T (K_{nn} + \sigma^2 I)^{-1} K_{n*}) \quad (5)$$

where \mathbf{X}_* denotes the (possible multiple) test point, K_{nn} is the kernel matrix between training samples D_n , K_{n*} is the kernel matrix between the training samples and the test point and σ^2 is the variance of the Gaussian likelihood, which usually represents the noise level of training samples. Note that, Eq. 5 also reveals that sampling a path of a Gaussian Process based on discrete observations is equivalent with sampling a known multivariate Gaussian distribution, this property will be heavily utilized for the proposed stopping criterion. The interested reader is referred to [39] for a detailed description of GPs.

Once the GP metamodel has been built, it can be utilized to make feasibility predictions. Since the GP posterior prediction is a Gaussian distribution,

we use its mean to make the feasibility predictions, i.e., the feasible region is modeled by:

$$A_\mu := \{\mathbf{x} \in \mathcal{X} : I_A(\mu(\mathbf{x})) = 1\} \quad (6)$$

Note that the use of the GP posterior mean as the feasibility predictor also embeds our assumption that, in case there is a misclassification of a design configuration's feasibility, the cost of either wrongly predict it as feasible (false positive) or wrongly predict it as infeasible (false negative) is equivalent from a decision theoretic perspective [39]. We refer to [21] in case these two costs are not equivalent.

3.2 Acquisition function based on uncertainty reduction

The information-theoretic based acquisition function presented in [5, 40, 41] is adopted. The main intuition behind it is to sample the point that maximizes the expected information gain of an interval g , given the new data \mathbf{x} . This is quantified by the reduction of differential entropy:

$$\alpha_g(\mathbf{x}) = \mathbb{H}(p(g|D_n)) - \mathbb{E}_{p(f_{\mathbf{x}}|D_n, \mathbf{x})} [\mathbb{H}(p(g|D_n \cup \{\mathbf{x}, f_{\mathbf{x}}\}))] \quad (7)$$

where $p(f_{\mathbf{x}}|D_n, \mathbf{x})$ denotes the predictive response distribution at \mathbf{x} , which in case of a GP is a Gaussian distribution, $\mathbb{E}(\cdot)$ denotes the operation of expectation through possible outcomes of $f_{\mathbf{x}}$, while $\mathbb{H}(\cdot)$ denotes the differential entropy, which can be used as an aggregated representation of a probability distribution's uncertainty.

However, the calculation of Eq. 7 is intractable, since the GP posterior must be updated to calculate $p(g|D_n \cup \{\mathbf{x}, f_{\mathbf{x}}\})$ for a certain location \mathbf{x} . In order to alleviate the computational burden, a common alternative representation [42] arises by noting that Eq. 7 actually represents the conditional mutual information: $I\{g, f_{\mathbf{x}}|D_n, \mathbf{x}\}$. Hence, Eq. 7 can be reformulated as its symmetric form:

$$\alpha_g(\mathbf{x}) = \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x})) - \mathbb{E}_{p(g|D_n)} [\mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x}, g))] \quad (8)$$

Let us suppose that the feasible boundary is $[\alpha, \beta]$; the acquisition function is then formulated to maximize the expected information gain about the region of interest and its complement³

$$\begin{aligned} \alpha(\mathbf{x}) &= \sum_{i=1}^3 \alpha_{g_i}(\mathbf{x}) \\ &= 3\mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x})) - \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x}, f_{\mathbf{x}} < \alpha)) \\ &\quad - \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x}, \alpha < f_{\mathbf{x}} < \beta)) \\ &\quad - \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x}, \beta < f_{\mathbf{x}})) \end{aligned} \quad (9)$$

³ For notation simplicity, we follow the original formulation in [5] and reuse symbol α and β in Eq. 3 with a different domain: $\alpha, \beta \in \mathbb{R}$. In practice, it can be trivially applicable in case g is defined on an infinite interval, e.g., $\alpha < f_{\mathbf{x}} < \infty$.

with intervals defined as:

$$g_1 = \{-\infty < f_{\mathbf{x}} < \alpha\}, \quad g_2 = \{\alpha < f_{\mathbf{x}} < \beta\}, \quad g_3 = \{\beta < f_{\mathbf{x}} < \infty\} \quad (10)$$

Now, computing Eq. 9 only involves calculating the differential entropy of a truncated Gaussian distribution. In particular, the following closed form expression holds [5]:

$$\begin{aligned} \alpha_{g_2}(\mathbf{x}) &= \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x})) - \mathbb{H}(p(f_{\mathbf{x}}|D_n, \mathbf{x}, \alpha < f_{\mathbf{x}} < \beta)) \\ &= -\log(Z) \\ &\quad - \frac{1}{2Z} [(\alpha - \mu(\mathbf{x}))\mathcal{N}(\mu(\mathbf{x})|\alpha, \sigma^2) - (\beta - \mu(\mathbf{x}))\mathcal{N}(\mu(\mathbf{x})|\beta, \sigma^2)] \end{aligned} \quad (11)$$

where $Z = \Phi\left[\frac{\beta - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right] - \Phi\left[\frac{\alpha - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right]$, and Φ denotes the cumulative distribution function of the standard normal distribution.

One advantage of this information-theoretic based acquisition function is that it can be conveniently extended to handle multiple constraints problem. Since the derivation of this condition is omitted in [5], we briefly demonstrate it here. Consider the following multi-constraint problem: $\mathbf{x} \rightarrow \mathbf{f}$, where the desired feasible region must satisfy all k black box constraints:

$$I_A(\mathbf{f}(\mathbf{x})) = \begin{cases} 1 & \alpha_1 \leq f_1(\mathbf{x}) \leq \beta_1, \alpha_2 \leq f_2(\mathbf{x}) \leq \beta_2, \dots, \alpha_k \leq f_k(\mathbf{x}) \leq \beta_k \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

by making use of the mutual information $I(\mathbf{g}; \mathbf{f}|\mathbf{x}, D_n)$, the acquisition function can be formulated as:

$$\begin{aligned} \alpha_{\mathbf{g}}(\mathbf{x}) &= \mathbb{H}(p(\mathbf{f}_{\mathbf{x}}|D_n, \mathbf{x})) - \mathbb{E}_{p(\mathbf{g}|D_n, \mathbf{x})} [\mathbb{H}(p(\mathbf{f}_{\mathbf{x}}|D_n \cup \{\mathbf{x}, \mathbf{g}\}))] \\ &= \frac{1}{2} \ln \det(2\pi e \Sigma(\mathbf{x})) - \mathbb{H}(p(\mathbf{f}_{\mathbf{x}}|D_n \cup \{\mathbf{x}, \mathbf{g}\})) \end{aligned} \quad (13)$$

$\Sigma(\mathbf{x})$ is the covariance matrix between each one dimensional f . Assuming the independence between the different components of \mathbf{f} , Eq. 11 can be written as:

$$\begin{aligned} \alpha_{\mathbf{g}_2}(\mathbf{x}) &= \frac{1}{2} \ln (2\pi e)^k \det(\Sigma(\mathbf{x})) - \sum_{j=1}^k \mathbb{H}(p(f_j|D_n \cup \{\mathbf{x}, g_{j_2}\})) \\ &= \sum_{j=1}^k \frac{1}{2} \ln (2\pi e) \sigma_j^2 - \sum_{j=1}^k \left[\frac{1}{2} \ln(2\pi e \sigma_j^2 Z_j^2) + \right. \\ &\quad \left. \frac{1}{2Z_j} [(\alpha_j - \mu_j)\mathcal{N}(\mu_j|\alpha_j, \sigma_j^2) - (\beta_j - \mu_j)\mathcal{N}(\mu_j|\beta_j, \sigma_j^2)] \right] \\ &= \sum_{j=1}^k \alpha_{g_{j_2}}(\mathbf{x}) \end{aligned} \quad (14)$$

Note that, for simplicity, the symbol \mathbf{x} is sometimes omitted in Eq. 14.

Once the acquisition function is defined, it can be used in the learning loop (see Fig. 2). Indeed, at each iteration, the GP metamodel is refined with a new expensive simulation, performed for the maximizer $\bar{\mathbf{x}}$ of the acquisition function. In order to estimate $\bar{\mathbf{x}}$, the following procedure is adopted: the acquisition function is evaluated at 1000 candidate points chosen via Monte Carlo sampling, then the best candidate point is chosen as the starting point of an L-BFGS-B gradient based optimizer [43]. This optimization process can be performed very efficiently, since the acquisition function $\alpha(\mathbf{x})$ depends only on the GP model, which is very cheap to evaluate.

Let us remark that this information theoretic-based acquisition function has demonstrated competitive performance compared with existing active learning techniques for feasible region identification. The interested reader is referred to [5] for a detailed discussion on this topic.

3.3 Framework of metamodel-based stopping criterion

An accurate estimation of the precision, recall and F1 score requires to evaluate f for a large set of samples in the design space. Since evaluating f via simulations is expensive, it is not possible to directly use these metrics as stopping criterion of the proposed active learning strategy. In the following, it is first described how these metrics can be efficiently estimated from a Bayesian perspective, i.e., from the metamodel's belief. Then, these estimations can be adopted as stopping criteria of the AL process (see Fig. 2). For notation and introduction simplicity, we start with the description of the stopping criterion in case of single constraint function ($k=1$), and then we generalize it to multiple constraints.

For a problem defined by a single constraint $f : \mathcal{R}^{m \times d} \rightarrow \mathcal{R}^{m \times 1}$, the corresponding feasible region is $A_f := \{\mathbf{x} \in \mathcal{X} : \alpha \leq f(\mathbf{x}) \leq \beta\}$, and the feasible region predicted by the GP metamodel (with posterior mean $\mu : \mathcal{R}^{m \times d} \rightarrow \mathcal{R}^{m \times 1}$) is $A_\mu := \{\mathbf{x} \in \mathcal{X} : \alpha \leq \mu(\mathbf{x}) \leq \beta\}$. Furthermore, let us indicate with $\text{precision}_\mu | f, \alpha, \beta, X^*$ the precision of the metamodel identifying feasible region A_μ , compared with the actual feasible region A_f based on test data X^* . Let us use $\text{recall}_\mu | f, \alpha, \beta, X^*$ and $\text{F1}_\mu | f, \alpha, \beta, X^*$ to indicate recall and F1 score estimated by the metamodel, respectively. These three performance measurements depend on the given feasible boundary $[\alpha, \beta]$ and test data $X^* : \{X_F^*, X_{IF}^*\} \in \mathcal{R}^{n, d}$, where X_F^* and X_{IF}^* denotes predicted feasible and not feasible data by posterior mean μ , respectively, m can be the size of the test data, while d represents the problem dimensionality.

Let us further define a set operator $I_F[C] = \{c \in C | \alpha < c < \beta\}$, where $C \subset \mathcal{R}^{m \times 1}$, which returns the subset of elements satisfying the constraints. Let us define a similar operator for the metamodel $I_{A_\mu}(D) = D \cap A_\mu$, which returns the subset of elements of the input data $D \in \mathcal{R}^{m \times d}$ that satisfy the feasible region.

With slight abuse of the notations given above, the precision, recall and F1 score can then be represented as:

$$\text{precision}_\mu | f, \alpha, \beta, X^* = \frac{|I_F[f[I_{A_\mu}(X^*)]]|}{|I_F[\mu[X^*]]|} = \frac{|I_F[f[X_F^*]]|}{|X_F^*|} \quad (15)$$

$$\begin{aligned} \text{recall}_\mu | f, \alpha, \beta, X^* &= \\ &= \frac{|I_F[f[I_{A_\mu}(X^*)]]|}{|I_F[f[I_{A_\mu}(X^*)]]| + |I_F[f[I_{X \setminus A_\mu}(X^*)]]|} \quad (16) \\ &= \frac{|I_F[f[X_F^*]]|}{|I_F[f[X^*]]|} \end{aligned}$$

$$\begin{aligned} \text{F1}_\mu | f, \alpha, \beta, X^* &= 2 \frac{\text{precision}_\mu \cdot \text{recall}_\mu}{\text{precision}_\mu + \text{recall}_\mu} \quad (17) \\ &= 2 \frac{|I_F[f[X_F^*]]|}{|I_F[f[X^*]]| + |X_F^*|} \end{aligned}$$

where $|\cdot|$ denotes the number of rows, for instance, $|X^*| = n$.

These three performance metrics are difficult to compute, as a large number of samples needs to be evaluated by the expensive simulator as test data. In order to efficiently make use of these metrics, a Bayesian approach can be utilized to describe the expensive constraint f in Eq. 15-17 with the belief of the GP metamodel: f follows a generative approach from the GP posterior. In this framework, Eq. 15-17 become random variables.

First, we will present an analytical analysis of the precision metric under our setting. Specifically, we provide an unbiased analytical estimation of its mean and an analytical approximation of its variance.

Proposition 1. *Given the assumption that the black box function is a random sample path of a known GP posterior: $f \sim \mathcal{GP}$. We have:*

$$\begin{aligned} &\mathbb{E}_f(\text{precision}_\mu | f, \alpha, \beta, X^*) \\ &= \frac{1}{N_F} \sum_{i=1}^{N_F} \left[\Phi \left(\frac{\beta - \mu(\mathbf{x}_{F_i}^*)}{\sigma(\mathbf{x}_{F_i}^*)} \right) - \Phi \left(\frac{\alpha - \mu(\mathbf{x}_{F_i}^*)}{\sigma(\mathbf{x}_{F_i}^*)} \right) \right] \quad (18) \\ &= \mathbb{E}_f \left[\frac{1}{N_F} \sum_{i=1}^{N_F} I_F[f(\mathbf{x}_{F_i})] \right] \approx \mathbb{E}_f [V(A_{f \subseteq \mu}) / V(A_\mu)] \end{aligned}$$

where N_F is the size of the predicted feasible data set X_F^* , i.e., $|X_F^*|$. $\mathbf{x}_{F_i}^*$ denotes a specific point i in X_F^* , $V(\cdot)$ denotes the volume of a region, $A_{f \subseteq \mu}$ denotes the feasible region within bounded space A_μ

A generalized proof of proposition 1 (which is applicable in case of multiple constraints) is given in appendix (Section A.1). Proposition 1 explains three properties of the Bayesian treatment of the precision metric. First and foremost, the precision metric formulated by comparing the posterior mean and GP posterior that contains the black box function f is a random variable,

and its expectation equals the second line of Eq. 18, which provides an unbiased analytical expression of this expectation. Second, the last line of Eq. 18 reveals the connection of the precision metric with the excursion search method [20], which tries to provide an accurate estimation of the probability of failure within a given design space \mathcal{X} . This brings an intuitive geometric interpretation of this equation: with the increase of the test data, the analytical expression will eventually converge to the expected feasible volume ratio within the model predict feasible space A_μ . Third, based on the second property, the last line of Eq. 18 also provide a justification that one can use any test data when applying the metamodel, with Eq. 18 still as a meaningful reference even though it is originally conditioned on probably different test data \mathbf{X}^* . Since as long as these two datasets are condense distributed in the design space, applying Eq. 18 on different test data sets approximate the same geometric interpretation.

We now demonstrate an approximation of the variance of the precision metric, a generalized proof of which is given in appendix (Section A.2).

Proposition 2. *Based on the same assumptions and notations of proposition 1. The variance of the precision metric can be approximated as:*

$$\mathbb{V}_f(\text{precision}_\mu | f, \alpha, \beta, \mathbf{X}^*) \approx \frac{1}{N_F^2} \sum_{i=1}^{N_F} [W(\mathbf{x}_{F_i}^*) - W(\mathbf{x}_{F_i}^*)^2] \quad (19)$$

where $W(\mathbf{x}_{F_i}^*) := \left[\Phi\left(\frac{\beta - \mu(\mathbf{x}_{F_i}^*)}{\sigma(\mathbf{x}_{F_i}^*)}\right) - \Phi\left(\frac{\alpha - \mu(\mathbf{x}_{F_i}^*)}{\sigma(\mathbf{x}_{F_i}^*)}\right) \right]$, the equality can be taken when the conditional independence of GP posterior prediction holds, i.e., $\text{cov}(\mathbf{Y}_* | D_n, \mathbf{X}_*)$ is diagonal.

The benefits of using Bayesian perspective to describe the precision metric (as well as the recall and F1 score) is twofold. First, the Bayesian description of f by the GP posterior has circumvented the use of precalculated expensive test data. Second, the reformulation of the performance metric as random variables has also allowed the calculation of its variance, which is of practical interest since a confidence bound of it can then be extracted.

Unfortunately, similar analytical forms do not exist for the recall and F1 score metric. Hence, in order to make use of these metrics as stopping criteria, following the same Bayesian treatment, one can sample the possible f from the GP posterior to calculate the possible F1 scores, as shown in Algorithm 1. This new algorithm, referred to as F1-Sample, is used in this paper to describe to what extent the feasible region has been modeled.

In case the feasible region is defined by multiple (k in total) black box constraints: $A := \{\mathbf{x} \in \mathcal{X} : \bigcap_{j=1}^k \alpha_j < f_j(\mathbf{x}) < \beta_j\}$, a histogram of sampled F1 score could be evaluated for each one of the constraints. However, this does not provide an overall accuracy of the prediction model.

Hence, an aggregated F1 score is proposed in the following. We first introduce some additional definitions in order to properly describe the criterion for

Algorithm 1: Sampling the F1 score

Input: GP, sample number: N , feasible boundary: α, β , test data: \mathbf{X}^* ;
 Extract: $\mathbf{X}_F^*, \mathbf{X}_{IF}^*, |\mathbf{X}_F^*|$ from \mathbf{X}^* ;
 $F1 = \{\}$;
for $i := 1$ *to* N **do**
 | sample $f \sim \mathcal{GP}$ based on \mathbf{X}^* through Eq. 5;
 | calculate $F1_\mu | f, \alpha, \beta, \mathbf{X}^*$ through Eq. 17
 | $F1 = F1 \cup \{F1_\mu | f, \alpha, \beta, \mathbf{X}^*\}$;
end
Output F1

multi-constraint problems from a practical implementation perspective. Let us define a set operator $I_{\mathbf{A}_\mu}(D) = D \cap_{i=1}^k A_{\mu_j}$, which returns the subset of elements of input data $D \in \mathcal{R}^{m \times d}$ that are in the joint feasible region. Let $I'_{F_j}[\cdot] : \mathbb{R}^m \rightarrow \{0, 1\}^m$ denote whether function values are within a feasible region j , while the operator $|\cdot|_1$ returns the element number of 1. Assuming the same bounded design space \mathcal{X} for all functions and using the same test data \mathbf{X}^* for each of the constraints f , then the predictions of these models can be aggregated into a single final model by reformulating the F1 score as a random variable:

$$F1_\mu(\mathbf{f}) | \alpha, \beta, \mathbf{X}^* = 2 \frac{|\prod_{j=1}^k I'_{F_j}[f_j(I_{\mathbf{A}_\mu}[\mathbf{X}^*])]|_1}{|\prod_{j=1}^k I'_{F_j}[f_j[\mathbf{X}^*]|_1 + |\mathbf{X}_F^*|} \quad (20)$$

As this expression cannot be trivially decomposed, we can also sample its distribution as in Algorithm 2.

Algorithm 2: Sampling the aggregated F1 score

Input: GP_1, GP_2, \dots, GP_k , sample number: N , feasible boundaries: α, β , test data: \mathbf{X}^*
 Extract: $\mathbf{X}_F^*, \mathbf{X}_{IF}^*, |\mathbf{X}_F^*|$ from \mathbf{X}^*
 $F1 = \{\}$
for $i := 1$ *to* N **do**
 | **for** $j := 1$ *to* k **do**
 | | sample $f_j \sim \mathcal{GP}_j$ based on \mathbf{X}^* through Eq. 5;
 | **end**
 | calculate $F1_\mu | f, \alpha, \beta, \mathbf{X}^*$ through Eq. 20
 | $F1 = F1 \cup \{F1_\mu | f, \alpha, \beta, \mathbf{X}^*\}$
end
Output F1

The recall metric can also be sampled by following a similar procedure. Note that the proposed strategy is also applicable to any other suitable probabilistic models, while in this research, we restrict our discussion within the framework of Gaussian Process.

Since the proposed F1-Sample measure heavily depends on sampling from the Gaussian Process defined in Eq. 5, it is also worthwhile to describe the

computational complexity of calculating the proposed stopping criterion. For the general multi-constraint case, the computational complexity⁴ in case of exact sampling (A. 2 of [39]) is $\mathcal{O}(n^3 Nk)$, where $n = |X^*|$ represents test sample numbers, N is the path sample number of GP posterior and k denotes the number of constraints. More efficient F1 score estimation is a topic of future research, we refer to [45–47] for further discussion.

In order to utilize the samples from Algorithm 1 and 2 as an indicator of the metamodels’ prediction accuracy, the sampled F1 score (or recall) can be averaged as the unbiased estimator of Eq. 20’s mean, which is referred to as $\overline{\text{F1-Sample}}$. Meanwhile, as the F1-sample measure relies on the assumption that the real function f follows the generation of a GP (Proposition 1), in reality this assumption can (possibly) result in either overestimation or underestimation of the real F1 score. Thus, in order to provide a more robust estimation, several different strategies can be adapted. For instance, instead of using a fixed test data set, it can be randomly generated for each F1 sample. On the other hand, one can employ a fully Bayesian approach by placing a prior distribution on the kernel hyperparameters and sample the posterior GP to estimate the F1 score. Moreover, one can take a step further by sampling from a mixture of possible kernels (e.g., [48]). Nevertheless, such approaches will result in much more extra computational effort within the current AL framework. Hence, in order to efficiently make use of the proposed F1-sample metric as a stopping criterion, as shown in Fig. 2, a conservative approach can be adopted by investigating when the metamodel is 95% confident that its predicted F1 score is larger than an acceptable F1 score, which can be obtained from the 5th percentile of the F1-Sample as a point estimation. Once this condition is continuously reached for several iterations (e.g., 5 iterations are used in this paper), it is suitable to stop the AL iterations.

4 Numerical examples and application examples

First synthetic experiments are conducted to validate the $\overline{\text{F1-Sample}}$ metric, i.e., to investigate to what extent it is able to capture the real F1 score. Afterwards, applications to Electronic Design Automation (EDA) examples are presented, where the feasible region identification is of practical interest during design space exploration. In particular, an antenna and a distributed filter are studied, both operating in the microwave spectrum. Different design constraints are applied to their frequency response, simulated using *ADS Momentum* [49]. In the first EDA example, the feasible region is defined by a single constraint condition on the device Return Loss (RL). In the second EDA example, a feasible region determined by multiple constraints must be identified,

⁴ The computational complexity involving sampling from multivariate normal distribution depends mainly on the (e.g. Cholesky) decomposition of the $n \times n$ covariance matrix (a detailed derivation can be found for instance at A. 4 of [44]). Nevertheless, given a fixed test data location X^* (i.e., a fixed covariance matrix), this decomposition only need to be calculate once upfront and can be reused to generate different samples with $\mathcal{O}(n^2)$.

based on the filter operating bandwidth. Both geometrical parameters and material properties are used as design parameters. Microwave designers can find feasible solutions for different manufacturing processes or layout techniques.

All the experiments are conducted on a computer with *Intel(R) Core(TM) i7-8650u @ 1.9GHz* processor and 16 GB RAM. The AL framework is implemented using GPFlowOpt [50]. Regarding the GP hyperparameters, the mean function m is chosen equal to zero, while the Automatic Relevance Detection (ARD) [39] version of the Squared Exponential (SE) kernel is employed. The kernel hyperparameters are determined using maximum likelihood estimation with an L-BFGS-B optimizer. For the test data X^* of F1-Sample metric, we use 10000 random samples. Then, to calculate the F1-Sample score, $N = 500$ function samples are drawn from the posterior GP. Finally, 20 experiments are conducted independently to provide robust results.

4.1 Numerical Examples

The numerical experiments are conducted on seven different benchmark functions, described in Table 1 along with their feasible regions. In order to give an indication of the difficulty of the problems studied here, the feasible ratio and the feasible variance ratio are also reported in Table 1. The feasible region ratio indicates the ratio between the volumes of the feasible region and the total bounded design space, and it is defined as:

$$R_{fr} = V(A)/V(\mathcal{X}) \text{ (\%)} \quad (21)$$

where the operator $V(\cdot)$ calculates the volume of a specified region. The *feasible variance ratio* is the function variance within the feasible region over the variance in the design spaces, and is defined as:

$$\begin{aligned} R_{fv} &= \Delta\{f|A\}/\Delta\{f|\mathcal{X}\} \text{ (\%)} \\ \text{where } \Delta\{f|\mathcal{X}\} &= \max_{\mathbf{x} \in \mathcal{X}} f - \min_{\mathbf{x} \in \mathcal{X}} f \\ \Delta\{f|A\} &= \max_{\mathbf{x} \in A} f - \min_{\mathbf{x} \in A} f \end{aligned} \quad (22)$$

4.1.1 Assessment of stopping criterion on random distributed data

The accuracy of the estimated F1 score by $\overline{\text{F1-Sample}}$ is assessed by training a one-shot GP model on random data sets with an increasing number of samples T . Adopting random samples to build the GP, makes it difficult to accurately model the feasible region, especially in cases when the feasible ratio is small: it might happen that no point is located in the feasible region. This allows a fair evaluation of the F1-Sample score even when the metamodel is less confident. The F1-Sample score is compared against the deterministic F1 score estimated on a random test data set of 10000 samples.

Table 1: Benchmark functions settings. As the R_{fv} metric is not well defined for multiple constraints problems, it is reported as *Not Applicable (NA)*

Function	Design Space	$[\alpha, \beta]$	$R_{fr}(\%)$	$R_{fv}(\%)$
Rosenbrock [51]	$[-2, 2]^2$	$[-\infty, 100]$	34.89	2.79
Hosaki [11]	$[0, 1]^2$	$[-\infty, -1]$	7.71	0.32
Curved Function [52]	$[0, 1]^3$	$[1, 2]$	4.74	2.21
Branin-Hoo [51]	$[0, 1]^2$	$[20, 40]$	24.68	6.61
Goldstein-Price [53]	$[-2, 2]^2$	$[-\infty, 500]$	15.21	0.05
Six-hump camel back [53]	$[-3, 3] \times [-2, 2]$	$[1, 10]$	35.04	5.63
Multi-Constraint Functions				
f_1 : Branin-Hoo &	$[0, 1]^2$	$\{[-\infty, 30],$	5.45	NA
f_2 : Quadratic: $y = x_1^2 + x_2^2$	$[0, 1]^2$	$[0.9, 1.1]\}$		

As a comparison, we also investigate the performance of the IMSE_T , which has widely been applied as an accuracy measurement of the feasible region within the design space:

$$\text{IMSE}_T = \int_{\mathcal{X}} s_k^2(\mathbf{x}) W(\mathbf{x}) d\mathbf{x} \quad (23)$$

where the so-called *interval* weight: $W(\mathbf{x}) = \Phi(\frac{\beta - \mu(\mathbf{x})}{\sigma(\mathbf{x})}) - \Phi(\frac{\alpha - \mu(\mathbf{x})}{\sigma(\mathbf{x})})$ (Eq.21 of [25]) is the probability that a response \mathbf{x} is inside the interval.

In order to evaluate the performance of the metric estimation, two criteria will be considered. First, a good prediction accuracy measurement should follow the same trend as the test data, based on a suitable evaluation metric, which is the F1 score in our setting. This can be retrieved from the linear correlation between the stopping criterion and the F1 score. Moreover, the deviation of the estimation should be sufficiently small with respect to the F1 score, since either overestimating or underestimating the F1 score will provide misleading results to designers.

Results in the form of linear correlation are provided in Table 2. The deviation with respect to the F1 score is also provided in Fig. 3. This progress plot is constructed by linearly interpolating the F1 score versus the F1 score predicted by the two metrics. Since the metamodel is built by random sampling, all the experiments are independent, hence the F1 score prediction is only reported with its mean. One immediate advantage which can be seen here is that the IMSE_T metric, while also measures the metamodel's accuracy, cannot be used to describe the F1 score directly, as it is unscaled and negatively correlated with the F1 score. Thus a scaled IMSE_T is used here for fair comparison and denoted as $\text{IMSE}_T^+ := 1 - \text{IMSE}_T / \max(\text{IMSE}_T)$. Note that, this comparison is not performed for the Curved benchmark function due to the small volume of its feasible region, which requires an unaffordable number of samples to provide an accurate metamodel. For all the other benchmark functions, it can be shown that the proposed F1-Sample clearly provides a better correlation with

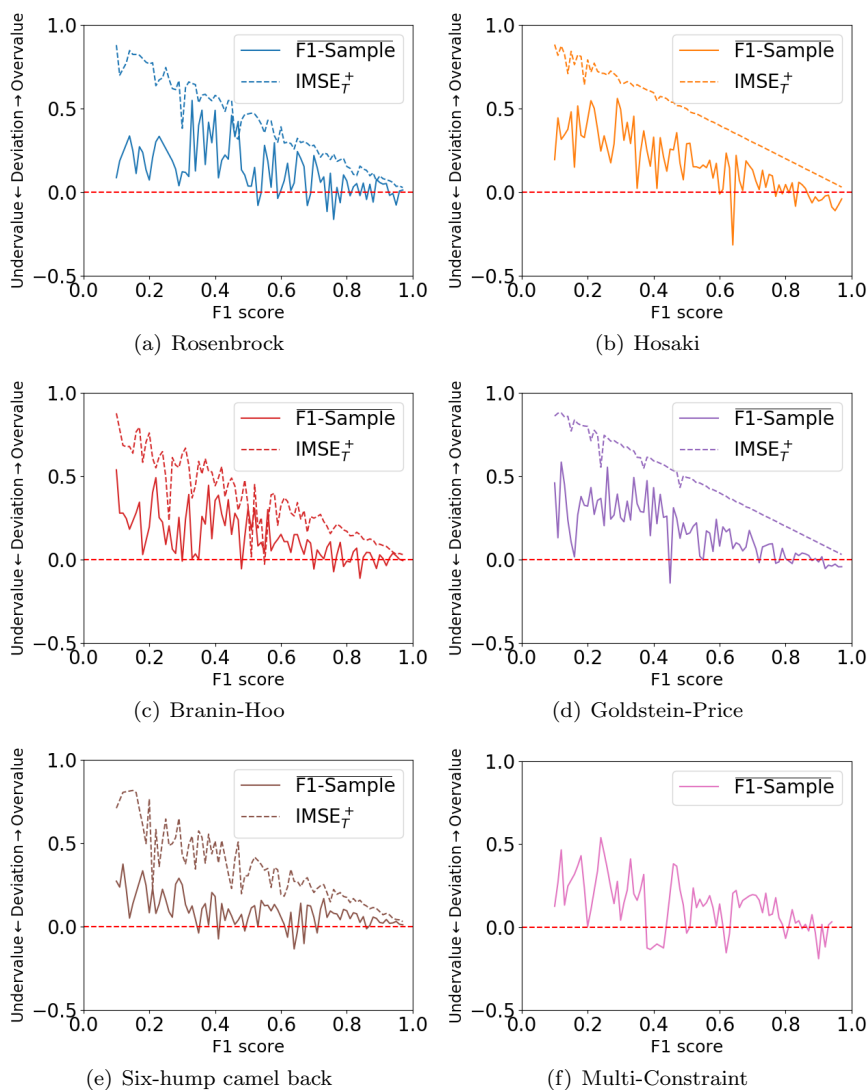


Fig. 3: Deviation progress plot of random sampling: The deviation of averaged stopping criteria predicted versus F1 score is shown with the increase of F1 score, a positive deviation means an overestimation of the F1 score; a negative deviation means an underestimation.

Table 2: Benchmark results of random sampling. Averaged results from 20 experiments are shown as the Pearson correlation coefficient using the true F1 score, where the bold denotes a higher correlation.

Function	IMSE $_T^+$	F1-Sample	T range
Rosenbrock	0.4832	0.8517	5-60
Hosaki	0.3139	0.5196	5-95
Curved Function	-	-	-
Branin-Hoo	0.6410	0.8495	5-60
Goldstein-Price	0.2816	0.7345	5-95
Six-hump camel back	0.6832	0.8539	5-60
Multi-Constraint Functions			
f_1 : Branin-Hoo &	NA	0.9043	5-35
f_2 : Quadratic: $y = x_1^2 + x_2^2$			

the F1 score than the IMSE $_T^+$. Taking a look at the deviation progress plot in Fig. 3, both stopping criteria converge when the model accuracy improves, i.e. when the F1 score approaches to 1. Nevertheless, $\overline{\text{F1-Sample}}$ has a smaller deviation for almost all the processes.

4.1.2 Assessment of stopping criterion in active learning

To further demonstrate the advantage of F1-Sample in the AL schemes, experiments for active feasible region identification are conducted, now based on the proposed entropy-based acquisition function. The initial data are generated by max-min Latin Hypercube Design (LHD). The experiments starting and stopping sample numbers, as well as the corresponding results are reported in Table 3. For some benchmark functions like the Goldenstein-Price function, low correlation is shown by both metrics. The main reason is that the feasible region is especially difficult to model in this case, as the function variance is considerably larger than the specified feasible region variance. Additionally, in earlier stages the metamodel accuracy is not high enough for the criterion to closely follow the F1 score trend, which changes along with the learning steps. This issue will gradually disappears as the model accuracy grows.

As a complementary benchmark, a multi-constraint experiment is executed. Since there is no trivial extension of IMSE $_T$ to this setting it is not compared here and reported as *Not Applicable* (NA). Results with random sampling and active learning are also provided in Table 3 and Fig 4. A high correlation can be observed consistently.

4.2 Design of a microstrip patch antenna

We now demonstrate the usage of the proposed AL framework in the electronic design of a microstrip antenna, which consists of a metal patch on a dielectric

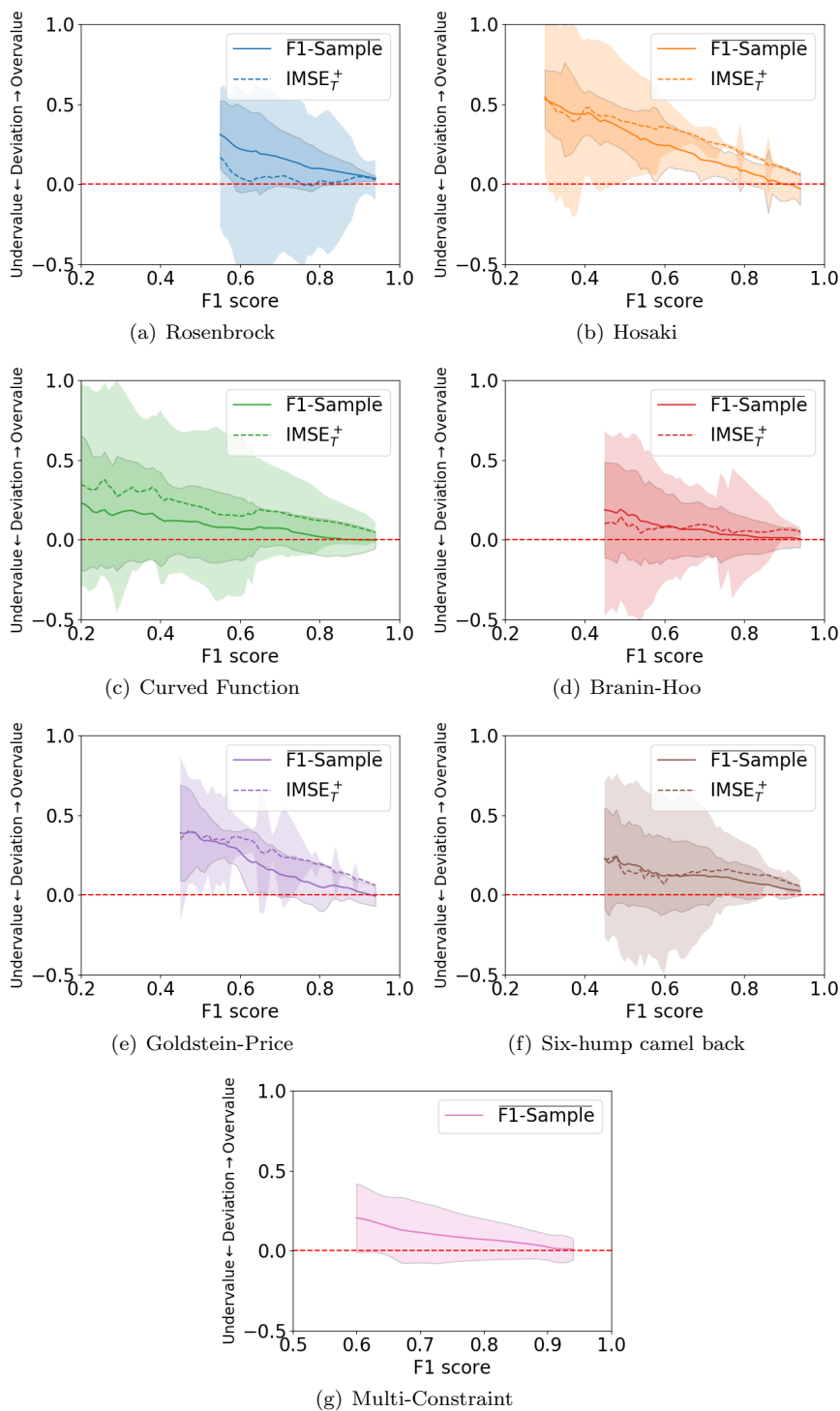


Fig. 4: Progress plot of the active learning process: The deviation of the stopping criteria versus the F1 score is shown. A positive deviation means an overestimation of the F1 score; a negative deviation means an underestimation. The shaded areas with black edge represent 2σ of F1-Sample metrics.

Table 3: Benchmark results of active learning. Averaged results from 20 AL experiments are shown as the Pearson correlation coefficient using the true F1 score, where the bold denotes a higher correlation (compared by mean).

Function	IMSE $_T^+$	F1-Sample	T range
Rosenbrock	0.4571 \pm 0.2350	0.6027 \pm 0.2328	5-20
Hosaki	0.3601 \pm 0.1502	0.4357 \pm 0.2690	5-80
Curved Function	0.6828 \pm 0.1449	0.9321 \pm 0.0450	5-65
Branin-Hoo	0.6479 \pm 0.2028	0.8098 \pm 0.0801	5-35
Goldstein-Price	0.4025 \pm 0.1167	0.4320 \pm 0.3453	5-85
Six-hump camel back	0.4643 \pm 0.1661	0.7133 \pm 0.1434	5-45
Multi-Constraint Functions			
f_1 : Branin-Hoo &	NA	0.8175 \pm 0.1654	5-25
f_2 : Quadratic: $y = x_1^2 + x_2^2$			

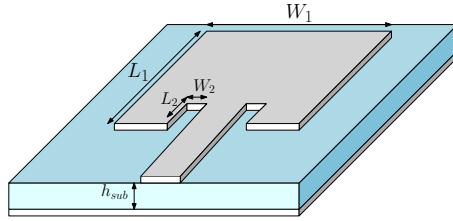


Fig. 5: Microstrip patch antenna geometry and design parameters

substrate, with a ground metal layer on the backside side. The patch geometry is shown in Fig. 5, while the design parameters can be found in Table 4. The feed line for the signal is matched with a 50Ω impedance at 2.4 GHz. The antenna is simulated in the range $[2 - 3]$ GHz with ADS Momentum, and a typical simulation requires about 45 seconds. Thus, the objective function is the frequency for which the Return Loss (RL) is minimal. If this value is included in $[2.35, 2.45]$ GHz, the design is considered feasible. The mathematical notation is shown as follows:

$$\begin{aligned}
 &\text{Find } A = \{\mathbf{x} \in \mathcal{X} | I_A(f(\mathbf{x})) = 1\} \\
 &\text{where } I_A(f(\mathbf{x})) = \begin{cases} 1 & 2.35 \leq f(\mathbf{x}) \leq 2.45 \\ 0 & \text{otherwise} \end{cases} \\
 &f(\mathbf{x}) := \underset{\zeta \in [2,3]}{\operatorname{argmin}}(RL_{\mathbf{x}}(\zeta))
 \end{aligned} \tag{24}$$

where $RL_{\mathbf{x}}(\cdot)$ denotes the RL curve for a given design \mathbf{x} , $f(\mathbf{x})$ denotes the RL curves minimum between frequency ζ range: $[2, 3]$ GHz.

Table 4: Microstrip patch antenna design parameters range

Parameter	Description	Range
W1	main patch width mm	[32.0, 38.0] mm
L1	main patch length	[27.5, 33.5] mm
W2	feed width	[2.2, 3.6]mm
L2	feed length	[8.2, 9.2]mm
h_{sub}	substrate height	[1, 2] mm
ϵ_{sub}	substrate relative permittivity	[2, 5]

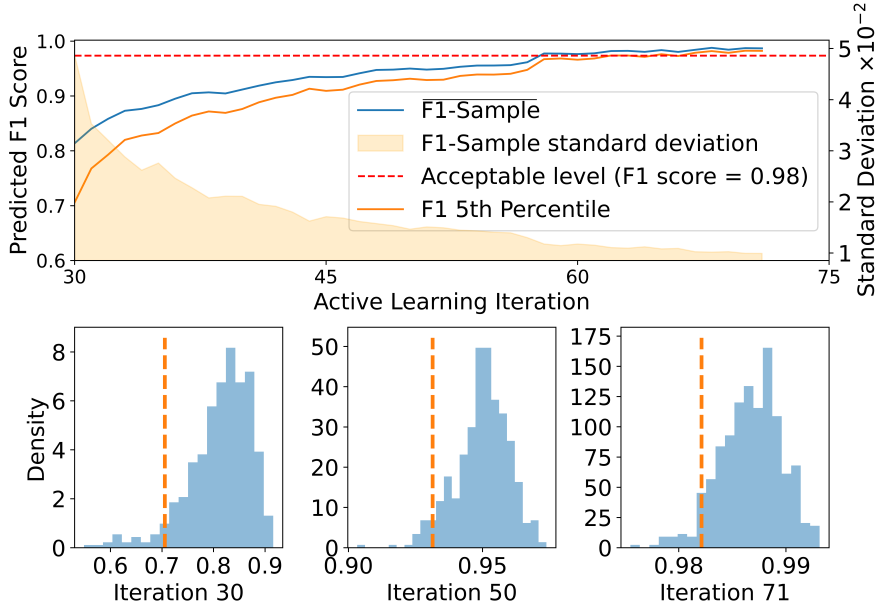


Fig. 6: Progress plot of active feasible region identification for the microstrip patch antenna. Sampled F1 histogram are shown at different iterations, where the vertical dashed line denotes the 5th percentile of the histogram.

An initial GP model is built using 30 initial samples, which are generated by LHD. The progress plot of the predicted F1 score is shown in Fig. 6. It can be seen that the F1-Sample keeps increasing with the number of iterations. The sampled F1 score histogram with its 5th percentile are also shown at three different stage of the AL process. By applying the proposed stopping criterion, the six dimensional AL process has stopped in 71 iterations, and the whole AL process lasts about 62 minutes.

To evaluate the final learned GP's performance, 3000 antenna configurations are randomly generated in the design space and the final GP metamodel is adopted to predict their feasibility. Only 372 have been predicted as fea-

sible. We evaluated these predicted feasible designs performance by the ADS Momentum simulator, the results of RL curves are shown in Fig. 7. The final learned GP metamodel has a precision score of 98.12%, a F1 score of 96.69%.

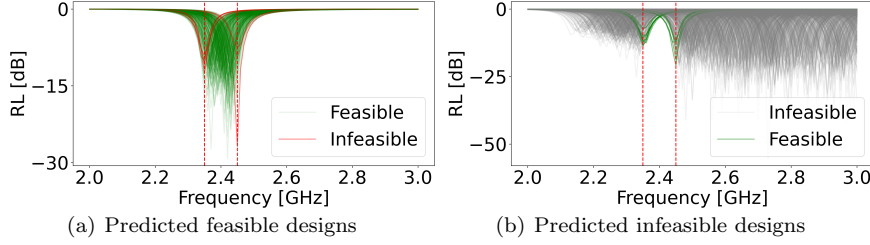


Fig. 7: Microstrip patch antenna: RL of the 372 designs predicted as feasible by the metamodel. Among these, 365 are true positive (in green), while 7 are false positive (in red). Besides, 18 of 2628 predicted infeasible designs are false negative. Note that both the false positive and false negative are near the edge of the feasible region.

4.3 Design of a double-folded microstrip band-stop filter

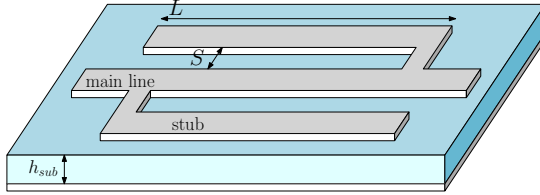


Fig. 8: Double-folded Microstrip band-stop filter geometry and parameter design parameters

The second application example is a double-folded microstrip band-stop filter [54], formed by a dielectric substrate between a top metallization and a bottom ground plane. As illustrated in Fig. 8, the geometry of the upper layer consists of two stubs, with identical length and spacing, folded onto the two sides of a transmission line. Here, the proposed AL framework can be applied to identify all possible parametric configurations for which the band-stop presents the desired width and location. Hence, the feasible region is defined by the following conditions:

$$BW \leq 7 \text{ GHz}; \quad 13.5 \leq f_0 \leq 14.5 \text{ GHz} \quad (25)$$

$$BW = f_H - f_L; \quad f_0 = \frac{f_H + f_L}{2} \quad (26)$$

where BW is the bandwidth and f_0 is the central frequency. The filter response is computed at 101 frequency points over the frequency range [5 – 25] GHz. Then, BW and f_0 can be defined w.r.t. the higher and lower band-stop frequencies (f_H, f_L), which correspond to -3 dB of magnitude response.

Table 5: Double-folded Microstrip band-stop filter design parameters range

Parameter	Description	Range
L	stub length	[1.98, 2.38] mm
S	line-stub spacing mm	[0.081, 0.161] mm
h_{sub}	substrate height mm	[0.110, 0.150] mm
e_{sub}	substrate relative permittivity	[9.6, 11]

The design space has four parameters as shown in Table 5, and the evaluation of a single design configuration takes about 42 seconds. The initial GP model is built upon 10 initial samples from a LHD and updated by the proposed entropy search-based method. The corresponding progress plot are depicted in Fig. 9, showing a rapid convergence of the F1-Sample score for this multi-constraint problem. The AL process is halted when F1-Sample is continuously larger than 0.98 for 5 consecutive iterations. The overall AL process takes about 14 minutes.

The accuracy of the computed metamodel is verified by comparison with 1000 ADS simulations. The results are shown in Fig. 10, indicating a precision score of 98.22%, a F1 score of 98.08%.

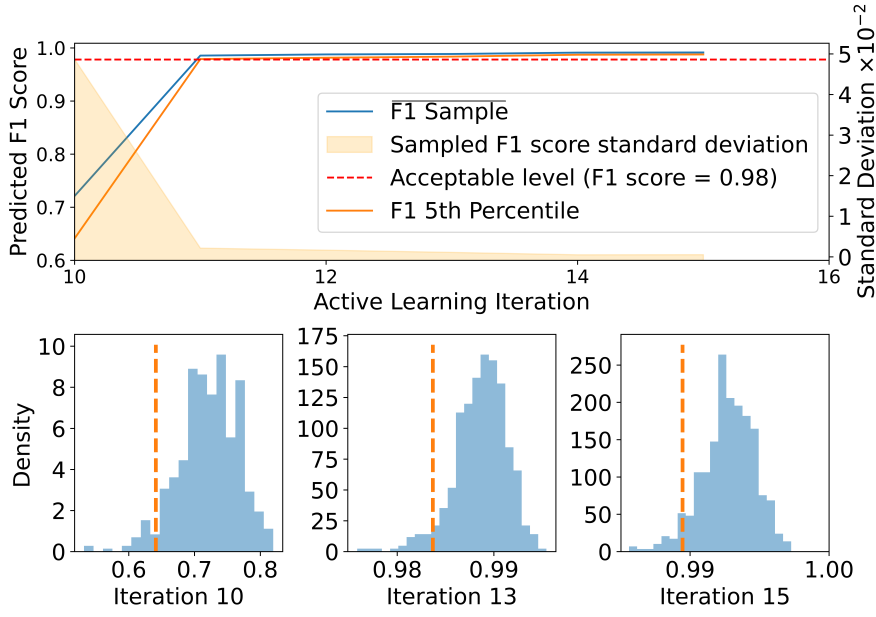


Fig. 9: Progress plot of active feasible region identification for the double-folded microstrip band-stop filter. Sampled F1 histogram are shown at different iterations, where the vertical dashed line denotes the 5th percentile of the histogram.

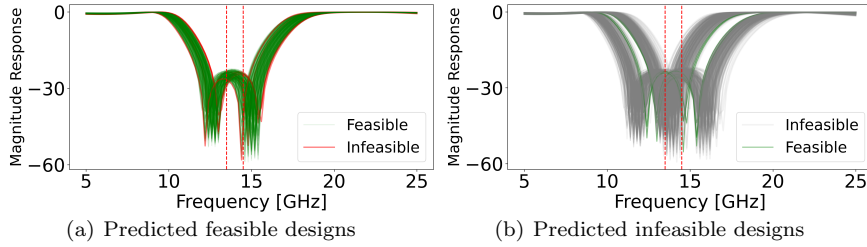


Fig. 10: Double-folded microstrip band-stop filter: the magnitude of the element S_{21} of the scattering matrix (representing the frequency response of the filter) for the 338 designs predicted as feasible by the metamodel from 1000 randomly distributed designs. Among these, 332 are true positive (in green), while 6 are false positive (in red). Besides, 7 of 668 predicted infeasible designs are false negative.

5 Conclusion

An information-theoretic based active learning method has been introduced to sequentially learn an accurate Gaussian Process metamodel for generating feasible design candidates. In particular, feasible regions can be identified with an efficient, robust and automated procedure, where multiple constraints can be specified at the same time. Furthermore, a novel AL stopping criterion has been proposed, based on the probabilistic description of the prediction accuracy. The proposed framework has been applied to two real-world electromagnetic design problems, a six-dimensional design space with a single constraint, and a four-dimensional problem with two constraints. In both applications, the method is able to reach a high-precision representation. Moreover, the number of samples needed to compute the model is minimized thanks to a novel sampling strategy, as well as the stopping criterion which allows for online supervising the metamodel’s belief about the feasible region, thus stopping the algorithm efficiently. This indicates that the model is suitable to simulate the behaviour of the real system on a confined region, and it can be successfully employed to generate a large number of good design candidates in complex engineering problems.

There are still several limitations on proposed algorithm. For the F1-Sample itself, the stopping criterion depends on numerical integration in the input domain, which makes it difficult to apply it in high dimensional problems. Moreover, the sampling process used to approximate the distribution requires additional computation. Future research will be focused on overcoming these limitations.

Acknowledgements This work is partially supported by the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” programme, and Chinese Scholarship Council under the grant number 201906290032.

A Appendix

A.1 Proof of proposition 1

Here, we directly prove a generalized proposition, which is applicable to multiple constraint problems; then, we show that proposition 1 is a special case of this general proposition. We first introduce the notation of precision metric in multiple constraints problems, by following the same notation of Eq. 20, it can be defined as:

$$\text{precision}_{\mu}|\mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, X^* = \frac{|\prod_{j=1}^k I'_{F_j}[f_j(I_{\mathbf{A}, \mu}[X^*])]|_1}{|X_F^*|} \quad (27)$$

Proposition 3. *Given the assumption that each black box constraint function is a random sample path of a known GP posterior: $\forall f_j \in \mathbf{f} : f_j \sim \mathcal{GP}_j$, if $\forall f_i, f_j \in \mathbf{f}, \forall \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$: $\text{cov}(f_i(\mathbf{x}), f_j(\mathbf{x})) = 0$, then the following analytical expectation exists:*

$$\begin{aligned}
& \mathbb{E}_{\mathbf{f}}(\text{precision}_{\mu} | \mathbf{f}, \boldsymbol{\alpha}, \beta, X^*) \\
&= \frac{1}{N_F} \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left[\Phi \left(\frac{\beta_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) - \Phi \left(\frac{\alpha_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) \right] \right] \\
&= \mathbb{E}_{\mathbf{f}} \left[\frac{1}{N_F} \sum_{i=1}^{N_F} \left[\prod_{j=1}^k I_F[f(\mathbf{x}_{F_i})] \right] \right] \approx \mathbb{E}_{\mathbf{f}} [V(A_{\mathbf{f} \subseteq \mu}) / V(A_{\mu})]
\end{aligned} \tag{28}$$

where N_F is the size of the predicted feasible data set X_F^* , i.e., $|X_F^*|$. $\mathbf{x}_{F_i}^*$ denotes a specific point i in X_F^* , $V(\cdot)$ denotes the volume of a region, $A_{\mathbf{f} \subseteq \mu}$ denotes the feasible region within bounded space A_{μ} .

Proof Start with the precision shown in Eq. 27.

While the denominator is a constant, for the numerator, we have:

$$\left| \prod_{j=1}^k I'_{F_j}[f_j(I_{A_{\mu}}[X^*])] \right|_1 = \sum_{i=1}^{N_F} \left[\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_i}^*)] \right] \tag{29}$$

By taking the expectation of this term w.r.t. \mathbf{f} and by making use of the linearity of expectation and the independent assumption, it becomes:

$$\begin{aligned}
\mathbb{E}_{\mathbf{f}} &= \sum_{i=1}^{N_F} \mathbb{E}_{\mathbf{f}} \left[\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_i}^*)] \right] \\
&= \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left[\int_{\alpha_j}^{\beta_j} f(\mathbf{x}_{F_i}^*) df_j \right] \right] \\
&= \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left[\Phi \left(\frac{\beta_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) - \Phi \left(\frac{\alpha_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) \right] \right]
\end{aligned} \tag{30}$$

which has obtained the second line of the Eq. 28. The last line of Eq. 18 and Eq. 28 can be conveniently obtained by substituting Eq. 29 into Eq. 28 and by taking the expectation, and can be regarded as the Monte Carlo approximation of the feasible ratio defined in the last term. With this proof, it can be seen that proposition 1 generally holds when $k = 1$.

A.2 Proof of proposition 2

Here, we follow the same strategy adopted in (Appendix A.1) by introducing a generalized proposition, which is applicable to multiple constraint problems, and then we show that proposition 2 is a special case of this general proposition. We begin with some necessary lemmas.

Lemma 1 *Slepian inequality: let X_1, X_2, \dots, X_n be normal with zero means and covariance matrix $\{\rho_{i,j}\}$, and let Y_1, Y_2, \dots, Y_n be normal with zero means and covariance matrix $\{l_{i,j}\}$. Let $\rho_{i,i} = l_{i,i}$, $i = 1, 2, \dots, n$, If $\rho_{i,i} \geq l_{i,j}$, $i, j = 1, 2, \dots, n$, then*

$$\Phi(\alpha_1, \alpha_2, \dots, \alpha_k; \{\rho_{i,j}\}) \geq \Phi(\alpha_1, \alpha_2, \dots, \alpha_k; \{l_{i,j}\})$$

Proof The proof of Slepian inequality can be found at [55].

Lemma 2 *Given the random variables $\{X_1, X_2\}$, which follow a joint multivariate normal distribution, each of them has the following marginal distribution:*

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1), X_2 \sim \mathcal{N}(\mu_2, \sigma_2). \quad (31)$$

If $\text{Cov}(X_1, X_2) \geq 0$, then the following inequality holds:

$$\Pr(\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2) \geq \Pr(\alpha_1 \leq X_1 \leq \beta_1) \cdot \Pr(\alpha_2 \leq X_2 \leq \beta_2),$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are constants.

Proof

$$\begin{aligned} & \Pr(\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2) \\ &= F_{X_1 X_2}(\beta_1, \beta_2) - F_{X_1 X_2}(\alpha_2, \beta_1) - F_{X_1 X_2}(\alpha_1, \beta_2) + F_{X_1 X_2}(\alpha_1, \alpha_2) \end{aligned} \quad (32)$$

Where $F_X(\cdot)$ denotes the cumulative density function of X . Substituting $\{\rho_{i,j}\}, \{\tau_{i,i}\}$ in Lemma 1 with the covariance matrix $\text{cov}(X_1, X_2)$, and the diagonal matrix of $\text{cov}(X_1, X_2)$ respectively. Since $\text{cov}(X_1 - \mu_1, X_2 - \mu_2) = \text{cov}(X_1, X_2)$, one can utilize Lemma 1 to have:

$$\begin{aligned} \Pr(X_1 \leq \beta_1, X_2 \leq \beta_2) &= \Pr(X_1 - \mu_1 \leq \beta_1 - \mu_1, X_2 - \mu_2 \leq \beta_2 - \mu_2) \\ &\geq \Pr(X_1 - \mu_1 \leq \beta_1 - \mu_1) \Pr(X_2 - \mu_2 \leq \beta_2 - \mu_2) \\ &= \Pr(X_1 \leq \beta_1) \Pr(X_2 \leq \beta_2) \end{aligned} \quad (33)$$

Hence, the following holds:

$$\begin{aligned} & \Pr(\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2) \\ &\geq F_{X_1}(\beta_1) F_{X_2}(\beta_2) - F_{X_1}(\alpha_2) F_{X_2}(\beta_1) - F_{X_1}(\alpha_1) F_{X_2}(\beta_2) + F_{X_1}(\alpha_1) F_{X_2}(\alpha_2) \\ &= (F_{X_1}(\beta_1) - F_{X_1}(\alpha_1))(F_{X_2}(\beta_2) - F_{X_2}(\alpha_2)) \\ &= \Pr(\alpha_1 \leq X_1 \leq \beta_1) \Pr(\alpha_2 \leq X_2 \leq \beta_2) \end{aligned} \quad (34)$$

Lemma 3 *With the same notation of Lemma 2, if $\text{Cov}(X_1, X_2) < 0$, then the following inequality holds:*

$$\Pr(\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2) < \Pr(\alpha_1 \leq X_1 \leq \beta_1) \Pr(\alpha_2 \leq X_2 \leq \beta_2),$$

Proof The Proof follows the same steps as for Lemma 2.

Finally, we can show our main proposition.

Proposition 4. *With the same assumptions and notation of proposition 3. the variance of precision defined by multiple constraints can be approximated as:*

$$\begin{aligned} & \mathbb{V}_{\mathbf{f}}(\text{precision}_{\mu} | \mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, X^*) \approx \\ & \frac{1}{N_F^2} \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left(W_j(\mathbf{x}_{F_i}^*) (1 - W_j(\mathbf{x}_{F_i}^*)) + W_j(\mathbf{x}_{F_i}^*)^2 \right) - \prod_{j=1}^k W_j(\mathbf{x}_{F_i}^*)^2 \right] \end{aligned} \quad (35)$$

where $W_j(\mathbf{x}_{F_i}^*) := \left[\Phi \left(\frac{\beta_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) - \Phi \left(\frac{\alpha_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) \right]$, the equality can be taken when the conditional independence of GP posterior prediction holds, i.e., $\forall GP_j : \text{cov}(\mathbf{Y}_{j^*} | D_n, \mathbf{X}_*)$ is diagonal.

Proof

$$\begin{aligned}
& \mathbb{V} \left[\frac{|\prod_{j=1}^k I'_{F_j}[f_j(I_{\mathbf{A}\mu}[X^*])]|_1}{|X_F|} \right] \\
&= \frac{1}{N_F^2} \mathbb{V} \left[\sum_{i=1}^{N_F} \left[\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_i}^*)] \right] \right] \\
&= \frac{1}{N_F^2} \sum_{i_1=1}^{N_F} \sum_{i_2=1}^{N_F} \text{cov} \left[\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)], \prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)] \right] \\
&= \frac{1}{N_F^2} \sum_{i=1}^{N_F} \mathbb{V} \left[\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_i}^*)] \right] + \frac{1}{N_F^2} \sum_{i_1=1}^{N_F} \sum_{i_2 \neq i_1}^{N_F} \left[\mathbb{E} \left(\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)] \cdot I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)] \right) \right. \\
&\quad \left. - \mathbb{E} \left(\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)] \right) \mathbb{E} \left(\prod_{j=1}^k I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)] \right) \right] \\
&= \frac{1}{N_F^2} \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left(\mathbb{V}(I_{F_j}[f_j(\mathbf{x}_{F_i}^*)]) + \mathbb{E}((I_{F_j}[f_j(\mathbf{x}_{F_i}^*)])^2) \right) - \prod_{j=1}^k \mathbb{E}((I_{F_j}[f_j(\mathbf{x}_{F_i}^*)])^2) \right] \\
&\quad + \frac{1}{N_F^2} \sum_{i_1=1}^{N_F} \sum_{i_2 \neq i_1}^{N_F} \left[\prod_{j=1}^k \mathbb{E} \left(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)] \cdot I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)] \right) \right. \\
&\quad \left. - \prod_{j=1}^k \mathbb{E}(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)]) \mathbb{E}(I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)]) \right] \tag{36}
\end{aligned}$$

By substituting with $W_j(\mathbf{x}_{F_i}^*) := \left[\Phi \left(\frac{\beta_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) - \Phi \left(\frac{\alpha_j - \mu_j(\mathbf{x}_{F_i}^*)}{\sigma_j(\mathbf{x}_{F_i}^*)} \right) \right]$, we have

$$\begin{aligned}
& \mathbb{V} \left[\frac{|\prod_{j=1}^k I'_{F_j}[f_j(I_{\mathbf{A}\mu}[X^*])]|_1}{|X_F|} \right] \\
&= \frac{1}{N_F^2} \sum_{i=1}^{N_F} \left[\prod_{j=1}^k \left(W_j(\mathbf{x}_{F_i}^*) (1 - W_j(\mathbf{x}_{F_i}^*)) + W_j(\mathbf{x}_{F_i}^*)^2 \right) - \prod_{j=1}^k W_j(\mathbf{x}_{F_i}^*)^2 \right] \\
&\quad + \frac{1}{N_F^2} \sum_{i_1=1}^{N_F} \sum_{i_2 \neq i_1}^{N_F} \left[\prod_{j=1}^k \left(W_j(\mathbf{x}_{F_{i_1}}^*) \cdot W_j(\mathbf{x}_{F_{i_2}}^*) + \text{cov}(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)], I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)]) \right) \right. \\
&\quad \left. - \prod_{j=1}^k W_j(\mathbf{x}_{F_{i_1}}^*) W_j(\mathbf{x}_{F_{i_2}}^*) \right] \tag{37}
\end{aligned}$$

For the covariance term:

$$\begin{aligned}
\text{cov}(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)], I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)]) &= \mathbb{E}(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)], I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)]) \\
&\quad - \mathbb{E}(I_{F_j}[f_j(\mathbf{x}_{F_{i_1}}^*)]) \cdot \mathbb{E}(I_{F_j}[f_j(\mathbf{x}_{F_{i_2}}^*)]) \\
&= \Pr(\alpha_j \leq f_j(\mathbf{x}_{F_{i_1}}^*) \leq \beta_j, \alpha_j \leq f_j(\mathbf{x}_{F_{i_2}}^*) \leq \beta_j) \\
&\quad - \Pr(\alpha_j \leq f_j(\mathbf{x}_{F_{i_1}}^*) \leq \beta_j) \Pr(\alpha_j \leq f_j(\mathbf{x}_{F_{i_2}}^*) \leq \beta_j) \tag{38}
\end{aligned}$$

With Lemma 2 and 3, since the posterior covariance between different points can either be positive or negative, a diagonal assumption of $\text{cov}(\mathbf{Y}_{j^*} | D_n, \mathbf{X}_*)$ is hence made to elimi-

nate Eq. 38 and provide an analytical approximation of the variance. One can also observe that proposition 2 can be generally obtained with $k = 1$.

References

1. E. Bradner, F. Iorio, and M. Davis, "Parameters tell the design story: ideation and abstraction in design optimization," in *Proceedings of the symposium on simulation for architecture & urban design*, vol. 26, Society for Computer Simulation International, 2014.
2. Q. Zhou, J. Wu, T. Xue, and P. Jin, "A two-stage adaptive multi-fidelity surrogate model-assisted multi-objective genetic algorithm for computationally expensive problems," *Engineering with Computers*, pp. 1–17, 2019.
3. A. Ceruti, "Meta-heuristic multidisciplinary design optimization of wind turbine blades obtained from circular pipes," *Engineering with Computers*, vol. 35, no. 2, pp. 363–379, 2019.
4. A. Gaier, A. Asteroth, and J.-B. Mouret, "Data-efficient design exploration through surrogate-assisted illumination," *Evolutionary computation*, vol. 26, no. 3, pp. 381–410, 2018.
5. K. Nicolas, C. Ivo, D. Tom, and S. Kohei, "Active learning for feasible region discovery," in *2019 International Conference on Machine Learning and Applications (ICMLA)*, Ieee, 2019.
6. A. Cully and Y. Demiris, "Quality and diversity optimization: A unifying modular framework," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 2, pp. 245–259, 2017.
7. S. Koos, J.-B. Mouret, and S. Doncieux, "The transferability approach: Crossing the reality gap in evolutionary robotics," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 122–145, 2012.
8. B. Yuepeng, S. Wenping, H. Zhonghua, Y. Zhang, and L. Zhang, "Aerodynamic/aeroacoustic variable-fidelity optimization of helicopter rotor based on hierarchical kriging model," *Chinese Journal of Aeronautics*, 2019.
9. P. Song, J. Sun, and K. Wang, "Axial flow compressor blade optimization through flexible shape tuning by means of cooperative co-evolution algorithm and adaptive surrogate model," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 228, no. 7, pp. 782–798, 2014.
10. D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of artificial intelligence research*, vol. 4, pp. 129–145, 1996.
11. W. Chen and M. Fuge, "Active expansion sampling for learning feasible domains in an unbounded input space," *Structural and Multidisciplinary Optimization*, vol. 57, no. 3, pp. 925–945, 2018.
12. B. J. Larson and C. A. Mattson, "Design space exploration for quantifying a system model's feasible domain," *Journal of Mechanical Design*, vol. 134, no. 4, 2012.
13. X. Zhuang and R. Pan, "A sequential sampling strategy to improve reliability-based design optimization with implicit constraint functions," *Journal of Mechanical Design*, vol. 134, no. 2, 2012.
14. T. H. Lee and J. J. Jung, "A sampling technique enhancing accuracy and efficiency of metamodel-based rbd: Constraint boundary sampling," *Computers & Structures*, vol. 86, no. 13-14, pp. 1463–1476, 2008.
15. J. Chen, L. Kang, and G. Lin, "Gaussian process assisted active learning of physical laws," *arXiv preprint arXiv:1910.03120*, 2019.
16. J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani, "Predictive entropy search for efficient global optimization of black-box functions," in *Advances in neural information processing systems*, pp. 918–926, 2014.
17. Z. Wang and S. Jegelka, "Max-value entropy search for efficient bayesian optimization," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 3627–3635, JMLR. org, 2017.

18. Q. Zhou, Y. Wang, S.-K. Choi, P. Jiang, X. Shao, and J. Hu, "A sequential multi-fidelity metamodeling approach for data regression," *Knowledge-Based Systems*, vol. 134, pp. 199–212, 2017.
19. K. Song, Y. Zhang, X. Zhuang, X. Yu, and B. Song, "An adaptive failure boundary approximation method for reliability analysis and its applications," *Engineering with Computers*, pp. 1–16, 2020.
20. J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez, "Sequential design of computer experiments for the estimation of a probability of failure," *Statistics and Computing*, vol. 22, no. 3, pp. 773–793, 2012.
21. D. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, and Y. Richet, "Adaptive design of experiments for conservative estimation of excursion sets," *Technometrics*, pp. 1–14, 2020.
22. A. Marco, A. von Rohr, D. Baumann, J. M. Hernández-Lobato, and S. Trimpe, "Excursion search for constrained bayesian optimization under a limited budget of failures," *arXiv preprint arXiv:2005.07443*, 2020.
23. I. Bogunovic, J. Scarlett, A. Krause, and V. Cevher, "Truncated variance reduction: A unified approach to bayesian optimization and level-set estimation," in *Advances in neural information processing systems*, pp. 1507–1515, 2016.
24. M. Senadeera, S. Rana, S. Gupta, and S. Venkatesh, "Level set estimation with search space warping," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 827–839, Springer, 2020.
25. V. Picheny, D. Ginsbourger, O. Roustant, R. Haftka, and N.-H. Kim, "Adaptive designs of experiments for accurate approximation of a target region," *Journal of Mechanical Design*, vol. 132, no. 7, 2010.
26. B. Bryan, R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman, "Active learning for identifying function threshold boundaries," in *Advances in neural information processing systems*, pp. 163–170, 2006.
27. P. Singh, J. Van Der Herten, D. Deschrijver, I. Couckuyt, and T. Dhaene, "A sequential sampling strategy for adaptive classification of computationally expensive data," *Structural and Multidisciplinary Optimization*, vol. 55, no. 4, pp. 1425–1438, 2017.
28. J. van der Herten, I. Couckuyt, D. Deschrijver, and T. Dhaene, "Adaptive classification under computational budget constraints using sequential data gathering," *Advances in Engineering Software*, vol. 99, pp. 137–146, 2016.
29. J. Bect, F. Bachoc, D. Ginsbourger, *et al.*, "A supermartingale approach to gaussian process based sequential design of experiments," *Bernoulli*, vol. 25, no. 4A, pp. 2883–2919, 2019.
30. A. Forrester, A. Sobester, and A. Keane, *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.
31. A. Kaintura, K. Foss, I. Couckuyt, T. Dhaene, O. Zografos, A. Vaysset, and B. Sorée, "Machine learning for fast characterization of magnetic logic devices," in *2018 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS)*, pp. 1–3, Dec 2018.
32. D.-K. Thai, T. M. Tu, T. Q. Bui, and T.-T. Bui, "Gradient tree boosting machine learning on predicting the failure modes of the rc panels under impact loads," *Engineering with Computers*, pp. 1–12, 2019.
33. A. Rahat and M. Wood, "On bayesian search for the feasible space under computationally expensive constraints," *arXiv preprint arXiv:2004.11055*, 2020.
34. J. Qian, J. Yi, Y. Cheng, J. Liu, and Q. Zhou, "A sequential constraints updating approach for kriging surrogate model-assisted engineering optimization design problem," *Engineering with Computers*, vol. 36, pp. 993–1009, 2020.
35. A. Kaintura, D. Spina, I. Couckuyt, L. Knockaert, W. Bogaerts, and T. Dhaene, "A kriging and stochastic collocation ensemble for uncertainty quantification in engineering applications," *Engineering with Computers*, vol. 33, no. 4, pp. 935–949, 2017.
36. A. Kaintura, K. Foss, I. Couckuyt, T. Dhaene, O. Zografos, A. Vaysset, and B. Sorée, "Machine learning for fast characterization of magnetic logic devices," in *2018 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS)*, pp. 1–3, Ieee, 2018.

37. J. Qing, Y. Hu, Y. Wang, Z. Liu, X. Fu, and W. Liu, "Kriging assisted integrated rotor-duct optimization for ducted fan in hover," in *AIAA Scitech 2019 Forum*, p. 0007, 2019.
38. R. Meyer and A. W. Hauser, "Geometry optimization using gaussian process regression in internal coordinate systems," *The Journal of Chemical Physics*, vol. 152, no. 8, p. 084112, 2020.
39. C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.
40. J. Qing, N. Knudde, I. Couckuyt, D. Spina, and T. Dhaene, "Bayesian active learning for electromagnetic structure design," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, IEEE, 2020.
41. J. Qing, N. Knudde, I. Couckuyt, T. Dhaene, and K. Shintani, "Batch bayesian active learning for feasible region identification by local penalization," accepted at *2020 Winter Simulation Conference (WSC)*, IEEE, 2020.
42. N. Houlsby, F. Huszar, Z. Ghahramani, and J. M. Hernández-Lobato, "Collaborative gaussian processes for preference learning," in *Advances in neural information processing systems*, pp. 2096–2104, 2012.
43. C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-bfgs-b: Fortran sub-routines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 4, pp. 550–560, 1997.
44. V. Perrone, I. Shcherbatyi, R. Jenatton, C. Archambeau, and M. Seeger, "Constrained bayesian optimization with max-value entropy search," *arXiv preprint arXiv:1910.07003*, 2019.
45. A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems*, pp. 1177–1184, 2008.
46. J. T. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth, "Efficiently sampling functions from gaussian process posteriors," *arXiv preprint arXiv:2002.09309*, 2020.
47. K. Wang, G. Pleiss, J. Gardner, S. Tyree, K. Q. Weinberger, and A. G. Wilson, "Exact gaussian processes on a million data points," in *Advances in Neural Information Processing Systems*, pp. 14648–14659, 2019.
48. A. Wilson and R. Adams, "Gaussian process kernels for pattern discovery and extrapolation," in *International conference on machine learning*, pp. 1067–1075, 2013.
49. E. Agilent EEsof, "Advanced design system (ads)," *Agilent Technologies Inc, Santa Rosa, CA (USA)*.
50. N. Knudde, J. van der Herten, T. Dhaene, and I. Couckuyt, "Gpflowopt: A bayesian optimization library using tensorflow," *arXiv preprint arXiv:1711.03845*, 2017.
51. V. Picheny, T. Wagner, and D. Ginsbourger, "A benchmark of kriging-based infill criteria for noisy optimization," *Structural and Multidisciplinary Optimization*, vol. 48, no. 3, pp. 607–626, 2013.
52. H. Dette and A. Pepelyshev, "Generalized latin hypercube design for computer experiments," *Technometrics*, vol. 52, no. 4, pp. 421–429, 2010.
53. M. Molga and C. Smutnicki, "Test functions for optimization needs," *Test functions for optimization needs*, vol. 101, 2005.
54. D. Spina, F. Ferranti, T. Dhaene, L. Knockaert, G. Antonini, and D. Vande Ginste, "Variability analysis of multiport systems via polynomial-chaos expansion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 2329–2338, Aug. 2012.
55. S. S. Gupta, "Probability integrals of multivariate normal and multivariate t1," *The Annals of mathematical statistics*, pp. 792–828, 1963.