A Bayesian Optimization Approach for Data-Efficient Flexibility Analysis of Expensive Black-Box Models

Akshay Kudva [a] and Joel A. Paulson [a,1]

[a] Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210

Abstract

The performance of a variety of emerging "next-generation" biochemical systems rests on their potential to quickly and accurately adapt to many different sources of uncertainties. Flexibility analysis represents a quantitative framework for determining if a system can maintain safe and feasible operation despite the presence of these uncertainties. The majority of work on flexibility analysis assume access to equation-oriented (white-box) models, which can be very difficult to obtain in many practical applications, especially those involving complex multi-scale models defined in terms of expensive computer simulations. In this paper, we propose a novel black-box flexibility analysis method that can overcome this challenge by simultaneously accounting for uncertain and recourse variables. The proposed method relies on a probabilistic surrogate to jointly model the effect of uncertain and recourse variables on a smooth approximation of the constraint aggregation function (which represents the objective function of interest in flexibility analysis problems). By utilizing confidence bounds predicted from the probabilistic surrogate, we can sequentially design our next sample points in a way that tradeoffs between exploration and exploitation of the unknown system model in the overall search space. Lastly, the advantages of the proposed method are demonstrated on two benchmark problems.

Keywords

Flexibility analysis; Bayesian optimization; Robust black-box optimization; Semi-infinite programming

Introduction

Many complex next-generation biochemical manufacturing processes are inevitably subjected to uncertainties from a variety of internal and external sources. Flexibility analysis provides a quantitative framework for identifying if a process design of interest is feasible over a range of uncertainty values, while taking into account feedback from control inputs (or recourse variables) available in the system. Since the notion of flexibility was introduced in the process systems engineering community in (Halemane and Grossmann, 1983), there have been a number of works on flexibility analysis over the past few decades including more efficient strategies for solving flexibility analysis problems (Grossmann and Floudas, 1987; Floudas et al., 2001) and extensions to stochastic (Pistikopoulos and Mazzuchi, 1990; Straub and Grossmann, 1990) and dynamic problems (Dimitriadis and Pistikopoulos, 1995).

An important (often implicit) assumption in the vast majority of current flexibility analysis methods, however, is that the structure of model is known and can be exploited by state-of-the-art optimization methods. There are many important engineering design problems in which this assumption is not valid since the model is defined in terms of some expensive computer simulation or experiment, whose structure is completely unknown to the modeler. Such cases are often referred to as "simulation-based" or "black-box" models and have been the focus of more recent flexibility analysis contributions such as (Boukouvala and Ierapetritou, 2012; Banerjee et al., 2010; Wang and Ierapetritou, 2017; Zhao et al., 2021). It is important to note that these works focus mostly on a special case of the flexibility analysis problem in which there are no recourse variables, which can be interpreted as a "feasibility test." This can yield a highly pessimistic view of flexibility in many relevant engineering problems in which there exists the ability to respond to external fluctuations. Therefore, the key goal of this paper is to develop a fully black-box flexibility analysis method that can simultaneously account for uncertain parameters and control inputs. By considering these competing sources at the same time, however, the flexibility problem becomes a tri-level "max-min-max" problem that is significantly harder to solve. In addition to being black-box in nature, we require a *data-efficient* algorithm that does not require the system model to be queried a large number of times. The rationale behind this restriction is that many system models are *expensive* in the sense that they take a long time and/or a lot of resources to be evaluated.

In this paper, we propose a novel black-box flexibility analysis method that is able to adress the challenging tri-level optimization structure without requiring a large number of system model evaluations. The proposed method has three main parts: (i) a smooth constraint aggregation approach is

[1] Corresponding author: J.A. Paulson (paulson.82@osu.edu).

used to reduce the flexibility problem to a two-level "max-min" problem without introducing non-differentiable components into the model; (ii) a probabilistic surrogate model is built for the constraint aggregation function since its structure is unknown; and (iii) upper and lower confidence bounds on the true (unknown) constraint aggregation function, constructed from the probabilistic surrogate model, are used to sequentially select the next best uncertainty and control input values to evaluate the system model. An important advantage of the proposed approach is its ability to systematically tradeoff between *exploration* (searching regions of the input where the surrogate model is most uncertain) and *exploitation* (searching near previous sampled input values where the surrogate model is confident that good solutions exist). Although the initial algorithm is specifically built to test for feasibility, we also discuss how this approach can be straightforwardly extended to determine a *flexibility index* for the system of interest, which is a quantitative measure of flexibility.

The structure of this paper is as follows. Section 2 provides a description of the flexibility analysis problem of interest. Section 3 describes our proposed data-efficient, black-box flexibility analysis algorithm, while Section 4 describes an extension of this algorithm to determine flexibility index values. We showcase the results of our proposed method on two illustrative example problems in Section 5 and conclude the paper and discuss future work in Section 6.

Problem Statement

The flexibility test problem, originally presented in Halemane and Grossmann (1983), can be formulated as follows

$$\chi = \max_{\theta \in \Theta} \min_{z \in \mathcal{Z}} \max_{j \in \mathcal{J}} f_j(\theta, z), \tag{1}$$

where $\theta \in \Theta \subset \mathbb{R}^{d_1}$ denotes the set of uncertain parameters (e.g., rate constants, initial conditions), $z \in \mathcal{Z} \subset \mathbb{R}^{d_2}$ denotes the set of control variables that can be adjusted during operation (e.g., flows, utility loads), and $\{f_j(\theta, z)\}_{j \in \mathcal{J}}$ denotes the set of inequality constraints that must satisfy $f_j(\theta, z) \leq 0$ for all $j \in \mathcal{J}$ for the system to achieve feasible operation. Therefore, if the value $\chi \leq 0$, it means that feasible operation can be attained over the full parameter range Θ (test is passed), whereas, if $\chi > 0$, feasible operation cannot be achieved for at least part of the range of Θ (test is failed).

In the absence of control variables ($d_2 = 0$), (1) reduces to a standard feasibility problem; however, the presence of z fundamentally changes its behavior. In addition to searching over all possible uncertain parameters θ, we must also search over all possible control inputs z that can compensate to the specific realization of θ. Therefore, (1) is a "max-min-max" optimization problem that can be thought of as a sequential three-player game with θ representing the first player, z representing the second player that can adapt to the decisions of the first player, and the index j representing the third player that can adapt to the decisions of the first and second players.

Previous methods developed for solving the flexibility test problem assume the functional form of the inequalities $\{f_j(\theta, z)\}_{j \in \mathcal{J}}$ are known. Typically, they are represented implicitly by the following set of inequalities

$$h(x, \theta, z) = 0, \;\; g(x, \theta, z) \leq 0, \tag{2}$$

where x denotes the internal state variables of the system (e.g., temperatures, concentrations), $h(x, \theta, z)$ denotes a set of equality constraints that uniquely define the state in terms of the uncertainty and control inputs (e.g., material and energy balances), and $g(x, \theta, z)$ denote the inequality constraints that often depend on the state directly (e.g., physical constraints, product specifications). Letting $x(\theta, z)$ denote the state variable values that satisfy $h(x, \theta, z) = 0$, we can recover the definition of the inequality constraints in (1) as follows

$$g_j(x(\theta, z), \theta, z) = f_j(\theta, z) \leq 0, \;\; \forall j \in \mathcal{J}. \tag{3}$$

However, in many practical applications, it is difficult or impossible to obtain accurate equation-oriented models whose structure can be exploited by the previously developed flexibility test methods. For example, many engineering design problems are defined in terms of complex (potentially integrated, multi-scale) computer simulations such that the functions $\{f_j(\theta, z)\}_{j \in \mathcal{J}}$ are expensive "black-box" models.

In this paper, we are interested in developing a new strategy for flexibility analysis that is generally applicable to black-box models involving expensive-to-evaluate functions. Instead of making structural assumptions about $\{f_j\}_{j \in \mathcal{J}}$, we consider the *bandit feedback* setting in which we can query $\{f_j\}_{j \in \mathcal{J}}$ at specific $(\theta, z) \in \Theta \times \mathcal{Z}$ values in sequential fashion. Since this querying process is assumed to be expensive, we would like to design an algorithm that can solve (1) for χ in as few iterations as possible. Our proposed solution is discussed in the next section, which involves a combination of constraint aggregation (to simplify the innermost maximization problem in (1)) and a robust surrogate-based optimization method that explicitly accounts for model uncertainty.

Proposed Approach for Black-Box Flexibility Analysis

In this section, we first present the constraint aggregation approach that is needed to reduce (1) to a two-level "max-min" robust optimization problem. We then review Gaussian process (GP) models that are used as the probabilistic surrogate for the unknown aggregated constraint function. Lastly, we summarize an efficient robust black-box optimization algorithm that uses GPs to accurately estimate the flexibility test value χ in a limited number of iterations.

Smooth Constraint Aggregation using the KS Function

The innermost $\max_{j \in \mathcal{J}}$ operator in (1) introduces a non-smooth element in the flexibility test optimization problem. To avoid this complication, we replace $\max_{j \in \mathcal{J}}$ with a smooth conservative approximation. In particular, we rely on the Kreisselmeier–Steinhsauser (KS) function (Wrenn, 1989):

$$\text{KS}(\theta, z; \rho) = M + \frac{1}{\rho} \ln \left[\sum_{j \in \mathcal{J}} \exp\left(\rho(f_j(\theta, z) - M)\right) \right], \tag{4}$$

where $\rho > 0$ represents the "aggregation" parameter whose value determines the degree of conservatism of the approximation and $M \approx \max_{j \in \mathcal{J}} f_j(\theta, z)$ is a constant used to reduce overflow/underflow errors in the exponential function. Using the established properties of the KS function in (Raspanti et al., 2000), its value can be bounded above and below by the max operator as follows

$$\max_{j\in\mathcal{J}} f_j(\theta,z) \leq \text{KS}(\theta,z;\rho) \leq \max_{j\in\mathcal{J}} f_j(\theta,z) + \frac{\ln|\mathcal{J}|}{\rho}. \tag{5}$$

From this, we see $KS(\theta,z;\rho) \leq 0$ underapproximates the feasible region $f(\theta,z) \leq 0$ and the KS function becomes equivalent to $\max_{j\in\mathcal{J}}$ as $\rho \to \infty$. In theory, we could select ρ to be large enough to yield nearly 0 error; however, this significantly increases the curvature of the function that can lead to numerical difficulties. Therefore, $\rho = 50$ is a typical value selected that leads to relatively small errors.

We can now replace the original flexibility test problem (1) with the following approximation in terms of the KS function

$$\hat{\chi} = \max_{\theta\in\Theta} \min_{z\in\mathcal{Z}} \text{KS}(\theta,z;\rho). \tag{6}$$

Due to the properties in (5), we know $\chi \leq \hat{\chi}$ such that we can guarantee the flexibility test is passed whenever $\hat{\chi} \leq 0$. Although we cannot definitively say the flexibility test is failed when $\hat{\chi} > 0$, values near 0 imply the system is at the boundary of the required degree of flexibility, suggesting additional investigation is warranted.

Gaussian Process Regression

Given that $\text{KS}(\mathbf{x})$, where $\mathbf{x} = (\theta,z)$ is the stacked vector of uncertainty and control input values and the parameter ρ is dropped for simplicity, will be a smooth (differentiable) function whenever the original constraint functions are smooth, we can now leverage Gaussian process (GP) surrogate models to approximate $\text{KS}(\mathbf{x})$. GP models are are an uncountable collection of random variables, any finite subset of which has a joint Gaussian distribution. Given a GP prior for $\text{KS}(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}),k(\mathbf{x},\mathbf{x}'))$, where $\mu(\mathbf{x})$ and $k(\mathbf{x},\mathbf{x}')$ denote the prior mean and covariance functions, respectively, the posterior distribution $\text{KS} \mid \mathbf{X}_t,\mathbf{y}_t$ will remain a GP where $\mathbf{X}_t = [\mathbf{x}_1,\dots,\mathbf{x}_t]$ is a set of t input values and $\mathbf{y}_t = [y_1,\dots,y_t]$ is a set of t output observations, i.e., $y_i = \text{KS}(\mathbf{x}_i)$. The posterior predictive distribution at any predicted test point $\mathbf{x}$ remains a Gaussian with the following mean and variance:

$$\mu_t(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{k}_t^\top(\mathbf{x})\mathbf{K}_t^{-1}(\mathbf{y}_t - \mu(\mathbf{x})), \tag{7a}$$

$$\sigma_t^2(\mathbf{x}) = k(\mathbf{x},\mathbf{x}) - \mathbf{k}_t^\top(\mathbf{x})\mathbf{K}_t^{-1}\mathbf{k}_t(\mathbf{x}), \tag{7b}$$

where $\mathbf{k}_t(\mathbf{x}) = [k(\mathbf{x}_1,\mathbf{x}),\dots k(\mathbf{x}_t,\mathbf{x})]^\top$ contains the covariance between the test input $\mathbf{x}$ and the observed data points $\mathbf{X}_t$ and $\mathbf{K}_t$ is the covariance matrix between the observation input data points with elements $[\mathbf{K}_t]_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$.

Note that the properties of the GP posterior (7) will be determined by the chosen class of covariance functions. In most situations, one selects the covariance function to be stationary from the Matern class (e.g., squared exponential function). In practice, the prior k will be defined in terms by several hyperparameters that must be trained. We follow the standard maximum likelihood estimation framework in this work (see, e.g., (Rasmussen and Williams, 2006)). Therefore, we can directly construct the GP posterior model for $\text{KS}(\theta,z)$ as long as we can query samples of it from a potentially expensive simulation or experiment.

Efficient Robust Black-Box Optimization Algorithm

We are interested in leveraging the (non-parametric) GP model, described in the previous section, to construct an efficient black-box flexibility test algorithm. First, we need to use the GP to define the following upper u_t and lower confidence bounds l_t on the KS function:

$$u_t(\theta,z) = \mu_t(\theta,z) + \kappa\sigma_t(\theta,z), \tag{8a}$$

$$l_t(\theta,z) = \mu_t(\theta,z) - \kappa\sigma_t(\theta,z), \tag{8b}$$

where $\kappa \geq 0$ is a so-called "exploration" parameter. The value of κ is important because it controls the degree to which the uncertainty in the prediction $\sigma_t(\theta,z)$ should offset the nominal prediction $\mu_t(\theta,z)$. For a sufficiently large κ value, we can ensure that the KS function must lie within these bounds with high probability (Srinivas et al., 2009). As commonly done in the literature, we select a value of $\kappa = 2$ throughout this work, though other values can be used.

We can straightforwardly construct u_t and l_t given past data $\{\theta_1,z_1,y_1,\dots,\theta_t,z_t,y_t\}$. The key question is then: How can we use these confidence bounds (8) to select the next best location to sample (θ_{t+1},z_{t+1})? To do this, we propose to build upon a recent algorithm developed by our group (Paulson et al., 2021) that involves sequentially solving the following two optimization problems:

$$\theta_{t+1} = \operatorname*{argmax}_{\theta\in\Theta} \min_{z\in\mathcal{Z}} u_t(\theta,z), \tag{9a}$$

$$z_{t+1} = \operatorname*{argmin}_{z\in\mathcal{Z}} l_t(\theta_{t+1},z). \tag{9b}$$

Once we have solved these problems, we evaluate the KS function at the selected sample point to obtain new data to add to our list, i.e., $y_{t+1} = \text{KS}(\theta_{t+1},z_{t+1})$. We can use this new data to update our GP model and the entire process can be repeated until we exhaust our budget of maximum number of function evaluations. It has been shown that this type of alternating confidence bound approach will converge to the true solution $\hat{\chi}$ under certain assumptions (Bogunovic et al., 2018; Paulson et al., 2021). The first step (9a) attempts to make "optimistic" selections for the uncertain parameters under uncertainty (due to lack of knowledge of the exact structure of the KS function). The second step (9b), on the other hand, makes "pessimistic" selections for the control inputs under uncertainty. In the context of the flexibility test problem, this second step actually corresponds to overestimating the potential for recourse, which is needed to ensure sufficient exploration is achieved in the $z \in \mathcal{Z}$ space.

Assuming the proposed algorithm is run for a total of T steps, we can use the final GP model to estimate $\hat{\chi}$:

$$\hat{\chi} \approx \max_{\theta\in\Theta} \min_{z\in\mathcal{Z}} \mu_T(\theta,z) + \kappa_{\text{rec}}\sigma_T(\theta,z), \tag{10}$$

where κ_{rec} is a constant that controls the conservatism of the final approximation, with larger values producing more conservative estimates. We fix $\kappa_{\text{rec}} = \kappa = 2$ for simplicity, though we plan to study the impact of κ_{rec} in future work.

Extension to Flexibility Index Problems

Definition of Flexibility Index

The main drawback of the flexibility test (1) is that only determines if a design can or cannot flexibly operate over a specified parameter range Θ. The flexibility index F is a quantity that provides a quantitative measure of the flexibility that can be achieved in a given system. The flexibility index is defined as follows

$$F = \max_{\delta} \ \delta \ \text{ subject to: } \ \chi(\delta) \le 0, \ \ \delta \ge 0, \tag{11}$$

where $\chi(\delta)$ is a slightly modified version of the flexibility test

$$\chi(\delta) = \max_{\theta \in \Theta(\delta)} \min_{z \in \mathcal{Z}} \max_{j \in \mathcal{J}} f_j(\theta, z), \tag{12}$$

defined in terms of a variable parameter set $\Theta(\delta)$ given by

$$\Theta(\delta) = \{\theta : \theta_N - \delta\Delta\theta \le \theta \le \theta_N + \delta\Delta\theta\}, \tag{13}$$

for some nominal parameter value θ_N and expected deviations $\Delta\theta > 0$. The geometric interpretation of $\Theta(F)$ is that it is the largest box of uncertainty values that can be dealt with by the system without leading to infeasible operation.

Combining Flexibility Test with Bisection Method

Since $\chi(\delta)$ is a monotonically increasing function of δ, we can equivalently search for F by finding $\chi(F) = 0$ (assuming a feasible solution for the nominal parameter values exists). This scalar root finding problem can be straightforwardly tackled with bisection, which involves the following steps given an interval $[\delta_L, \delta_U]$ that contains the root:

1. Calculate the midpoint $\delta_M = \frac{\delta_L + \delta_U}{2}$.
2. Solve the flexibility test problem at the midpoint $\chi(\delta_M)$.
3. Examine the sign of $\chi(\delta_M)$ to update bound:
   (a) If $\chi(\delta_M) \le 0$, then update lower bound $\delta_L \leftarrow \delta_M$.
   (b) If $\chi(\delta_M) > 0$, then update upper bound $\delta_U \leftarrow \delta_M$.
4. Stop if convergence is satisfactory ($\chi(\delta_M)$ is small) and return δ_L as the best feasible estimate for F.

Starting δ_L and δ_U values can be chosen by either intuition or random sampling. Since the key step of this bisection method is the evaluation of the flexibility test $\chi(\delta_M)$, we can directly use our proposed black-box algorithm in Section 3. It is worth noting that we can easily reuse $\text{KS}(\theta, z)$ values obtained at each step of the bisection method since each flexibility test is defined in terms of a common set of functions. For now, we assume that enough samples T are available at each bisection iteration to ensure an accurate estimate of $\chi(\delta)$ can be obtained. Future work will investigate more systematic stopping criteria using a combination of the upper and lower confidence bound functions.

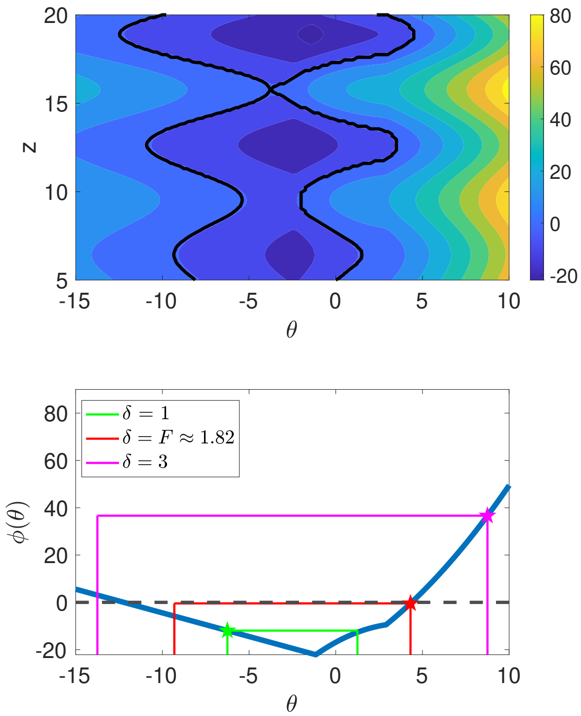


Figure 1: (Top) Contour plot for the KS function. The black line representing a value of 0. (Bottom) The minimum KS aggregated constraint violation value for every θ, which is given by (15). Three different boxes are shown for three different δ values, with the value of θ that leads to the worst-case violation shown with a star.

Case Studies

Illustrative Benchmark Problem

We consider the following system of the form (1):

$$f_1(\theta, z) = -2\theta - 15 - 0.5z\cos(z) \le 0, \tag{14a}$$

$$f_2(\theta, z) = \frac{\theta^2}{3} + 4\theta - 5 - z\cos(z) \le 0, \tag{14b}$$

$$f_3(\theta, z) = -\frac{(\theta - 4)^2}{2} + 10 - z\cos(z) \le 0. \tag{14c}$$

We assume a parameter set of the form (13) with $\theta_N = -2.5$ and $\Delta\theta = 3.75$. The control input (recourse variable) is assumed to satisfy the following bounds $5 \le z \le 20$. A contour plot of the KS function (4) is shown in Figure 1 (top). We see that the function is highly nonlinear but overall smooth. Since z represents our recourse action, we also plot the lowest possible constraint value for any $\theta \in \Theta(\delta)$ in Figure 1 (bottom). We define this function as $\phi(\theta)$:

$$\phi(\theta) = \min_{z \in \mathcal{Z}} \text{KS}(\theta, z). \tag{15}$$

The flexibility test is only passed if $\phi(\theta) \le 0, \forall \theta \in \Theta(\delta)$ such that it is passed for $\delta = 1$ but fails for $\delta = 3$. We can easily identify the δ for which $\phi(\theta) = 0$ from this plot as $\delta = F \approx 1.8$, which corresponds to the flexibility index value.

This problem is straightforward if we exactly know the functions in (14); however, we assume that we do not have this information for the purposes of testing our proposed algorithm. We only assume that we can query these functions at specific (θ, z) values. For a fixed δ value, we run the black-box flexibility test algorithm in Section 3 for a total of $T = 50$ iterations. To ensure that we are able to train the GP model for the KS function initially, we draw 10 random samples uniformly from the space $\Theta(\delta) \times \mathcal{Z}$ and evaluate the KS function at these points for our starting dataset. The max-min optimization in (9a) is solved using the semi-infinite programming method from (Mitsos, 2011), while (9b) is solved using a standard multi-start heuristic.

Since the final estimated $\hat{\chi}$ value from (10) depends on the random initial samples, we repeat the entire algorithm 10 times to get an estimate of the average $\hat{\chi}$ versus number of iterations (with corresponding confidence bounds estimated using the standard error formula). The results are shown in Figure 2 for three different $\delta \in \{1, 3, 1.8125\}$ values for our proposed data-efficient search method. For comparison purposes, we also show the results when the samples $\{\theta_1, z_1 \ldots, \theta_T, z_T\}$ are selected randomly, while still using the same final recommendation procedure (10). We see that our algorithm reliably shows substantial improvements over random search in all three cases, i.e., significantly faster convergence to the true χ value with substantially less variance. Since the flexibility test could be accurately solved in less than 20 function evaluations for 3 very different δ values, the overall cost of the flexibility index method (Section 4) would also be substantially reduced compared to random search.

Bubble Column Fermentation Problem

To highlight potential value of our proposed method on real-world problems, we also apply it to a simulation-based bubble column reactor model developed in (Chen et al., 2018). A simple illustration of the bioreactor system is shown in Figure 3. We are interested in performing a flexibility test of the form (1) on this system, where the reactor temperature $\theta \in [307.25, 313.25]$ Kelvin represents a key source of uncertainty and the gas velocity $z \in [9.84, 14.76]$ m/hr is an adjustable control variable that can be used to compensate for temperature fluctuations. There are three main constraints that must be satisfied to ensure feasible operation:

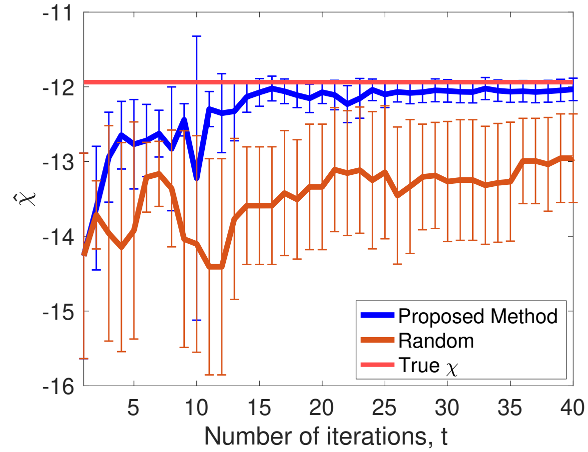
$$f_1(\theta, z) = 14 \text{ g/L} - C_E(\theta, z) \leq 0, \quad (16a)$$

$$f_2(\theta, z) = \text{TSS}(\theta, z) - 830 \text{ hours} \leq 0, \quad (16b)$$

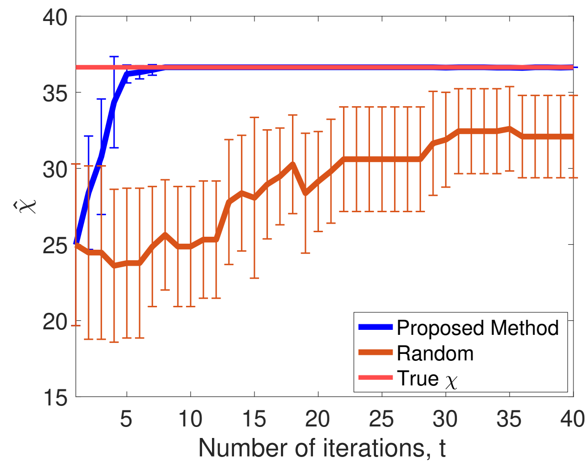
$$f_3(\theta, z) = 1.455 \text{ g Ethanol/g Acetate} - S_{E/A} \leq 0, \quad (16c)$$

where C_E, TSS, and $S_{E/A}$ denote the steady-state ethanol concentration, time-to-steady-state, and steady-state selectivity of ethanol to acetate, respectively. These quantities can be computed directly by solving the complex simulation-based model from (Chen et al., 2018) at any desired (θ, z) value.

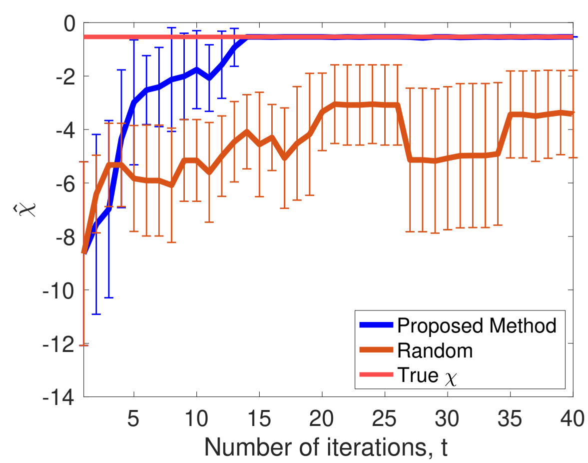
It is important to note that we have no prior knowledge about the structure of these functions and each simulation is fairly expensive, which makes it a great candidate for our proposed approach. We use the same settings as in the illustrative example for the black-box flexibility test algorithm in Section 3. We again evaluate performance by estimating the average $\hat{\chi}$ value over 10 randomly generated



(a) $\delta = 1$



(b) $\delta = 3$



(c) $\delta = 1.8125$

Figure 2: Predicted estimates of the worst-case aggregated constraint satisfaction $\hat{\chi}$ using (10) for (a) $\delta = 1$, (b) $\delta = 3$, and (c) $\delta = 1.8125$. The proposed method is shown in blue, a random search baseline is shown in orange, and the true χ value is shown in red. The reported values show estimates of the average $\hat{\chi}$ versus number of iterations with corresponding confidence bounds shown with error bars.

sets of initial points. Since the true $\hat{\chi}$ is unknown in this problem, we estimate it using a fine grid search. The re-

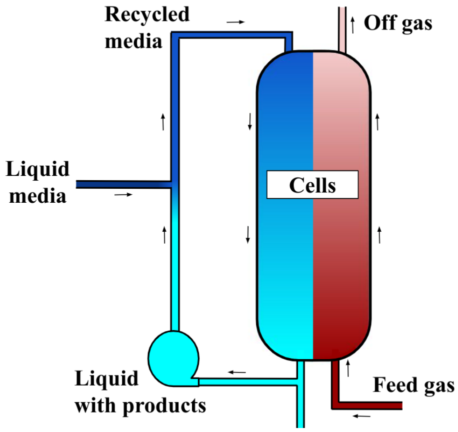


Figure 3: Schematic of the bubble column reactor system. Reproduced from (Kudva et al., 2022).

sults are shown in Figure 4. Similarly to the previous example, we see that our proposed method shows faster convergence and less variability when compared to random search. The average maximum possible improvement for both algorithms is around 0.036 (i.e., difference between the true optimal point and the starting estimate). Thus, by iteration 50, our method achieves 0.034 improvement, whereas random search only achieves 0.014 improvement, implying a substantial $\frac{0.034}{0.036} - \frac{0.014}{0.036} \approx 55\%$ performance increase by using a confidence bound-based search approach to select the query points. This reduction in the number of evaluations directly leads to computational savings in this problem since many fewer simulations would need to be run in practice to construct an accurate estimate for the flexibility test.

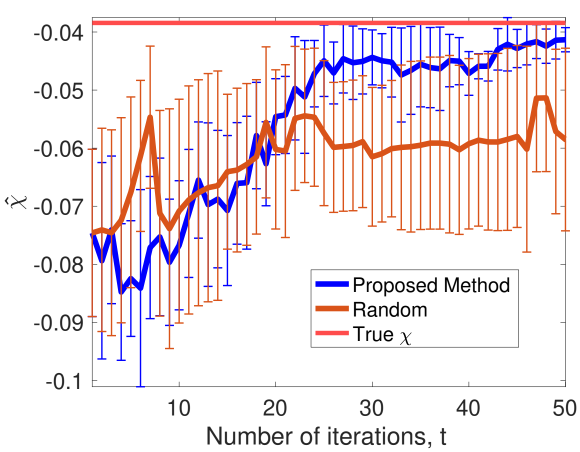


Figure 4: Predicted estimates of the worst-case aggregated constraint satisfaction $\hat{\chi}$ using (10) for our proposed method and random search.

Conclusions

This paper presents a novel algorithm for fully black-box (derivative-free) flexibility analysis of expensive-to-evaluate systems that can simultaneously capture the impact of uncertainty and recourse variables. The algorithm consists of three main concepts. First, we take advantage of a smooth constraint aggregation function to reduce the flexibility problem from a tri-level to a bi-level optimization problem. Second, we use the notion Gaussian process (GP) surrogate models to construct a probabilistic representation of the smooth constraint aggregation function. Third, we take advantage of the uncertainty quantification performed by the GP model to construct upper and lower confidence bounds for the true (unknown) constraint aggregation function, which can be used to efficiently sample the uncertainty and recourse variable space simultaneously. We further extend this method to the so-called flexibility index problem using a simple bisection procedure. Lastly, we demonstrate the advantages of our proposed approach on two illustrative benchmark problems and show that it is able to substantially outperform existing alternatives. There are several interesting directions for future work including more systematic convergence analysis, the development of robust stopping criteria, and applications to more challenging simulation models.

References

Banerjee, I., S. Pal, and S. Maiti (2010). Computationally efficient black-box modeling for feasibility analysis. *Computers Chemical Engineering 34*(9), 1515–1521. Selected papers from the 7th International Conference on the Foundations of Computer-Aided Process Design (FOCAPD, 2009, Breckenridge, Colorado, USA.

Bogunovic, I., J. Scarlett, S. Jegelka, and V. Cevher (2018). Adversarially robust optimization with gaussian processes. *Advances in Neural Information Processing Systems 31*.

Boukouvala, F. and M. G. Ierapetritou (2012, January). Feasibility analysis of black-box processes using an adaptive sampling kriging-based method. *Computers &amp Chemical Engineering 36*, 358–368.

Chen, J., J. Daniell, D. Griffin, X. Li, and M. A. Henson (2018, January). Experimental testing of a spatiotemporal metabolic model for carbon monoxide fermentation with clostridium autoethanogenum. *Biochemical Engineering Journal 129*, 64–73.

Dimitriadis, V. D. and E. N. Pistikopoulos (1995, December). Flexibility analysis of dynamic systems. *Industrial &amp Engineering Chemistry Research 34*(12), 4451–4462.

Floudas, C. A., Z. H. Gümüş, and M. G. Ierapetritou (2001, May). Global optimization in design under uncertainty: feasibility test and flexibility index problems. *Industrial &amp Engineering Chemistry Research 40*(20), 4267–4282.

Grossmann, I. E. and C. A. Floudas (1987). Active constraint strategy for flexibility analysis in chemical processes. *Computers and Chemical Engineering*, 675–693.

Halemane, K. and I. Grossmann (1983, May). Optimal process design under uncertainty. *AIChE Journal*, 425–433.

Kudva, A., F. Sorourifar, and J. A. Paulson (2022, August). Constrained robust bayesian optimization of expensive noisy black-box functions with guaranteed regret bounds. *AIChE Journal*.

Mitsos, A. (2011, October). Global optimization of semi-infinite programs via restriction of the right-hand side. *Optimization 60*(10-11), 1291–1308.

Paulson, J. A., G. Makrygiorgos, and A. Mesbah (2021). Adversarially robust bayesian optimization for efficient auto-tuning of generic control structures under uncertainty. *AIChE Journal*, e17591.

Pistikopoulos, E. and T. Mazzuchi (1990, September). A novel flexibility analysis approach for processes with stochastic parameters. *Computers &amp Chemical Engineering 14*(9), 991–1000.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning.* Cambridge, MA: MIT Press.

Raspanti, C., J. Bandoni, and L. Biegler (2000, October). New strategies for flexibility analysis and design under uncertainty. *Computers &amp Chemical Engineering 24*(9-10), 2193–2209.

Srinivas, N., A. Krause, S. M. Kakade, and M. Seeger (2009). Gaussian process optimization in the bandit setting: No regret and experimental design.

Straub, D. and I. Grossmann (1990, September). Integrated stochastic metric of flexibility for systems with discrete state and continuous parameter uncertainties. *Computers &amp Chemical Engineering 14*(9), 967–985.

Wang, Z. and M. Ierapetritou (2017). A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach. *AIChE Journal 63*(2), 532–550.

Wrenn, G. A. (1989). An indirect method for numerical optimization using the kreisselmeir-steinhauser function.

Zhao, F., I. E. Grossmann, S. García-Muñoz, and S. D. Stamatis (2021, February). Flexibility index of scpblack-box/scp models with parameter uncertainty through scpderivative-free/scp optimization. *AIChE Journal 67*(5).