Received: 20 June 2022 | Revised: 13 November 2022 | Accepted: 18 November 2022

DOI: 10.1002/qre.3245

SPECIAL ISSUE ARTICLE

WILEY

Dispersion-enhanced sequential batch sampling for adaptive contour estimation

Yiming Che[1] | **Juliane Müller**[2] | **Changqing Cheng**[1]

[1]Department of Systems Science and industrial Engineering, Binghamton University, Binghamton, New York, USA

[2]Computational Science Center, National Renewable Energy Laboratory, Golden, Colorado, USA

Correspondence
Changqing Cheng, Department of Systems Science and industrial Engineering, Binghamton University, Binghamton, NY 13902, USA.
Email: ccheng@binghamton.edu

Abstract

In computer simulation and optimal design, sequential batch sampling offers an appealing way to iteratively stipulate optimal sampling points based upon existing selections and efficiently construct surrogate modeling. Nonetheless, the issue of near duplicates poses tremendous quandary for sequential learning. It refers to the situation that selected critical points cluster together in each sampling batch, which are individually but not collectively informative towards the optimal design. Near duplicates severely diminish the computational efficiency as they barely contribute extra information towards update of the surrogate. To address this issue, we impose a dispersion criterion on concurrent selection of sampling points, which essentially forces a sparse distribution of critical points in each batch, and demonstrate the effectiveness of this approach in adaptive contour estimation. Specifically, we adopt Gaussian process surrogate to emulate the simulator, acquire variance reduction of the critical region from new sampling points as a dispersion criterion, and combine it with the modified expected improvement (EI) function for critical batch selection. The critical region here is the proximity of the contour of interest. This proposed approach is vindicated in numerical examples of a two-dimensional four-branch function, a four-dimensional function with a disjoint contour of interest and a time-delay dynamic system.

KEYWORDS

Batch sampling, contour estimation, Gaussian process surrogate, sequential optimal design

1 | INTRODUCTION

While computer simulations (e.g., finite element method) are widely used to aid the optimal design of real-world complex systems,[1,2] those high-fidelity simulators generally entail prohibitive computational cost. Alternatively, low-fidelity surrogate modeling has been brought to the limelight to emulate the expensive-to-evaluate simulators, which seeks to mimic the mathematical relation between response surface and design inputs with only a paltry of the computational expense.[3–7] Surrogate models can be derived from sequential or non-sequential sampling strategies. For non-sequential design, the sampling points are selected in a one-shot fashion, and a large number of them are generally needed for accurate fitting. By contrast, in sequential sampling, the design or sampling inputs are selected sequentially, and the surrogate is continuously updated with the addition of new sampling points. Central to sequential modeling is how to select the most informative

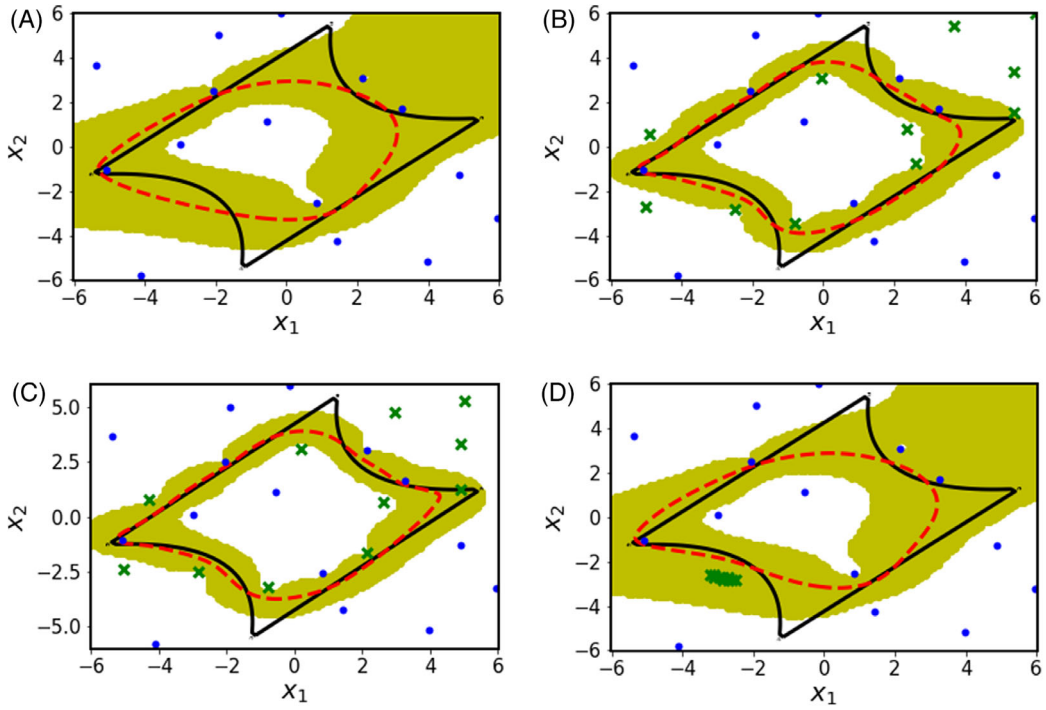


FIGURE 1 Comparison of the fitted iso-surface (the dashed red curve) against the underlying truth $y = 0$ (the solid black curve): (A) the prediction by GP from the initial design (the blue dots); (B) prediction by GP with addition of the critical batch $\boldsymbol{X}_b^*$ selected by the proposed method (the green crosses); (C) prediction by GP with addition of the critical batch $\boldsymbol{X}_b^*$ selected by WKMS (the green crosses); and (D) prediction by GP with addition of the critical batch $\boldsymbol{X}_b^*$ selected by the modified EI (the green crosses). The shaded area signifies the critical region $\boldsymbol{X}_s$

design points across the design space $\{\boldsymbol{x} \in \mathcal{X}\}$ in each sampling iteration. As such, only a fraction of the sampling points is required to achieve accuracy comparable to the one-shot design.

On the other hand, identification of a prescribed contour $\{\boldsymbol{x} \in \mathcal{X} : y(\boldsymbol{x}) = a \in \mathcal{Y}\}$ is a profound quest in a host of industrial applications. For instance, it represents the limit state in reliability engineering: $\{\boldsymbol{x} \in \mathcal{X} : y(\boldsymbol{x}) \geq a\}$ denotes the set of stable / functioning design points, and $\{\boldsymbol{x} \in \mathcal{X} : y(\boldsymbol{x}) < a\}$ indicates unstable / failure design points.[8–10] That said, only local accuracy around the limit state, instead of the global response surface, is required to build the surrogate. This is also referred to as contour or iso-surface estimation. The iso-surface can be approximated from either a regression or classification perspective, depending on the continuous or discrete property of the output of the design points. Accordingly, different criteria can be utilized to select the design points, including the modified EI[11] and the integrated mean square error (IMSE).[12] In classification setting, the iso-surface is considered as the decision boundary of a classifier, and closeness of design points to the boundary has been incorporated in active learning to locate the decision boundary in an iterative manner.[13] In regression setting, integration of modified EI function and the nonparametric Gaussian process (GP) has been investigated to select the most informative design points.[11,14]

Notably, GP is a widely used surrogate model, whose predictive uncertainty furnishes the quintessential information in sequential design.[15–18] Whereas EI function was originally developed for efficient global optimization,[5] the modified EI involves a utility function, which considers not only expected proximity to the iso-surface but also the associated uncertainty to stipulate the most critical design point in each iteration to update the GP surrogate. Nonetheless, such one-at-a-time sampling strategies engender a considerable number of iterations to fit the surrogate and approximate the iso-surface, particularly for complicated contours. It is noteworthy that fitting of GP surrogate will become cumbersome with the accumulation of sampling points. This bottleneck can be resolved by batch sampling, which is naturally compatible with parallel computing. One naïve solution is to select a batch of the most informative design points according to the modified EI function in each iteration, which, however, inevitably leads to the near-duplicate issue, in that the selected critical points are fairly close to each other (for illustration see Figure 1D). Intuitively, if one sampling point is selected via the modified EI, then points sufficiently close to this one will have similar utility values and tend to be selected in the same batch, provided that the underlying function is smooth. As a result, the near duplicates carry substantially redundant

information, and they are individually-but-not-collectively informative towards the update of the surrogate and estimate of the contour of interest. Of a similar ilk, other one-critical-point-per-selection strategies bear the same issue of near duplicates, including the local IMSE.[12] Indeed, most existing improvement criteria do not lend themselves to sequential batch sampling.[11,19]

In this present study, we aim to sequentially select a batch of the most informative points for contour estimation and avoid succumbing to the issue of near duplicates. Remarkably, a dispersion concept is included to sparsely distribute the selected points in each batch. A similar notion of diversity among design points has been explored in the literature, mostly under the framework of classification. For instance, Bayesian active learning by disagreement (BALD) selects the most critical design point according to the mutual information between the responses and model parameters.[20] Based upon this, Batch BALD is formulated to explicitly maximize the collective information carried by the selected batch to eschew near duplicates.[21] Unfortunately, the criteria used in BALD and Batch BALD are intractable in regression settings. Similarly, BADGE and Core-Set are specially designed for classification as well.[22,23] Zhdanov formulated weighted K-means (WKMS) clustering on the candidate points that are possibly in close proximity to the contour of interest, and then the batch was selected across the clusters; the batch size is equivalent to the number of clusters; the weight of WKMS can be any informativeness score in active learning, for example, entropy.[24] Whereas this method is also applicable in regression problems and represents one of state-of-the-art approaches, it does not explicitly gauge the uncertainty associated with the estimated iso-surface. Some other works related to batch sampling focus on the fitting of the global surface[19,25,26] and they may not be easily adapted to the fitting of the iso-surface.

Herein, we develop a dispersion-enhanced sequential batch sampling approach to adaptively estimate the iso-surface, by integrating the modified EI and a variance reduction technique for the critical region, which is defined as the proximity of the contour of interest. The modified EI plays the role of weight in variance reduction of the critical region. Thus, the variance reduction effectively exacts the dispersion "force" among all design points in each batch, and account for the predictive uncertainty in the estimated iso-surface simultaneously. We show that this novel approach is superior to WKMS, particularly for intricate contours.

The rest of this paper is organized as follows. In Section 2, we revisit GP and the modified EI function. Section 3 provides the methodology of the proposed method, and the comparison of WKMS, the one-shot design and our approach are given in Section 4. Section 5 concludes the paper.

2 | BACKGROUND

2.1 | Problem statement

Define a training set $\{(\boldsymbol{x_i}, y_i)\}_{i=1}^{n}$ with $\boldsymbol{x_i} \in \mathcal{X} \subset \mathbb{R}^d$ and $y_i \in \mathcal{Y} \subset \mathbb{R}$, annotated from the simulator, and an unlabeled dataset $\mathcal{D} \subset \mathcal{X}$, the iso-surface of interest is given as $\{\boldsymbol{x} \in \mathcal{X} : y(\boldsymbol{x}) = a \in \mathcal{Y} \subset \mathbb{R}\}$. A surrogate model is trained to learn $y(\boldsymbol{x})$ along with the uncertainty for any $\boldsymbol{x} \in \mathcal{D}$. In sequential optimal design, we seek to sketch the contour or iso-surface $\{\boldsymbol{x} \in \mathcal{X} : \hat{y}(\boldsymbol{x}) = a \in \mathcal{Y} \subset \mathbb{R}\}$ iteratively with the least number of sampling points. In each iteration, a batch of critical design points are selected from $\mathcal{D}$ and appraised by the simulator, which will then be annexed into the training set to update the surrogate model. Mathematically, a batch $\boldsymbol{X_b^*} = \{\boldsymbol{x}_1^*, \dots, \boldsymbol{x_b^*}\} \subseteq \mathcal{D}$ with cardinality $|\boldsymbol{X_b^*}| = n_b$ will be selected sequentially based on an acquisition function A,

$$\boldsymbol{X_b^*} = \underset{\boldsymbol{X_b} \in D}{\operatorname{argmax}}\, A\left(\boldsymbol{X_b}\right). \tag{1}$$

Here, GP is adopted as the surrogate model, as it provides a full predictive distribution as opposed to merely a point estimate. Subsequently, a modified EI function based upon the GP is utilized as the acquisition function to estimate the iso-surface.[11] Note that we use bold letters for vectors or sets and non-bold letters for scalars or functions.

2.2 | Gaussian process surrogate

GP is a collection of random variables $\{f(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{X}\}$, any finite number of which have a joint Gaussian distribution. It is specified by the mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$, that is, $f(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$. The noisy

observation $y(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is the noise variance. In Bayesian learning, a Gaussian prior $p(\boldsymbol{f}|\boldsymbol{X})$ is placed on $f(\boldsymbol{x})$ and the posterior $p(\boldsymbol{f}|\boldsymbol{y}, \boldsymbol{X})$ can be derived from assimilating observation data, where $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ and $\boldsymbol{y} \in \mathbb{R}^n$ are the set of design points $\boldsymbol{x}$ and the corresponding responses $y(\boldsymbol{x})$, respectively. For simplicity, we assume the prior $\boldsymbol{f}|\boldsymbol{X} \sim \mathcal{N}(0, k(\boldsymbol{X}, \boldsymbol{X}))$, and $k(\boldsymbol{X}, \boldsymbol{X})$ is $n \times n$ symmetric positive-definite covariance matrix. The likelihood is given as $\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{f}, \sigma^2\boldsymbol{I})$, which leads to the fact that the marginal likelihood is also Gaussian, that is, $p(\boldsymbol{y}|\boldsymbol{X}) = \int p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{X})p(\boldsymbol{f}|\boldsymbol{X})d\boldsymbol{f}$, and

$$\boldsymbol{y}|\boldsymbol{X} \sim \mathcal{N}\left(0, \boldsymbol{K} + \sigma^2\boldsymbol{I}\right), \tag{2}$$

where $\boldsymbol{K}\coloneqq k(\boldsymbol{X}, \boldsymbol{X})$ is the $n \times n$ covariance matrix for design points $\boldsymbol{X} \in \mathbb{R}^{n\times d}$. For the training of GP, we maximize the marginal likelihood in Equation ((2) in terms of the hyperparameters in the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$. The posterior distribution $p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{y})$ is the prior joint distribution $p(\boldsymbol{y}, \boldsymbol{f}|\boldsymbol{X}) = p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{X})p(\boldsymbol{f}|\boldsymbol{X})$ conditioning on observations. Consequently, posterior GP is specified by the posterior mean and covariance of nominal function $f(\boldsymbol{x})$,

$$\hat{y}(\boldsymbol{x}) = \bar{f}(\boldsymbol{x}) = \boldsymbol{K}_{\boldsymbol{x}}\left(\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{y} \tag{3}$$

$$Cov(\hat{y}(\boldsymbol{x})) = Cov\left(\bar{f}(\boldsymbol{x})\right) = \boldsymbol{K}_{\boldsymbol{xx}} - \boldsymbol{K}_{\boldsymbol{x}}\left(\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{K}_{\boldsymbol{x}}^T, \tag{4}$$

where $\boldsymbol{K}_{\boldsymbol{x}}\coloneqq k(\boldsymbol{x}, \boldsymbol{X})$ is the $1 \times n$ covariance vector between the response $\boldsymbol{y}(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{d}}$ and $\boldsymbol{y}(\boldsymbol{X}), \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{d}}$. $\boldsymbol{K}_{\boldsymbol{xx}}\coloneqq k(\boldsymbol{x}, \boldsymbol{x})$is the covariance at design point $\boldsymbol{x}$. $\boldsymbol{I}$ is the $n \times n$ identify matrix. Hence, the nominal function $f(\boldsymbol{x}) \sim \mathcal{N}(\bar{f}(\boldsymbol{x}), Cov(f(\boldsymbol{x})))$.[18]

2.3 | Modified expected improvement function for contour estimation

To quantify the importance or informativeness of a design point $\boldsymbol{x}$ towards estimation of the iso-surface, a utility function derived from the GP framework is employed[11]:

$$I(\boldsymbol{x}) = S^2(\boldsymbol{x}) - \min\left((y(\boldsymbol{x}) - a)^2,\ S^2(\boldsymbol{x})\right), \tag{5}$$

where a is the response of the iso-surface of interest.

Here, $y(\boldsymbol{x}) \sim \mathcal{N}(\hat{y}(\boldsymbol{x}), Var(\hat{y}(\boldsymbol{x})))$ is the predictive response for design point $\boldsymbol{x} \in \mathcal{D}$ derived from the GP surrogate. Define $s(\boldsymbol{x}) = \sqrt{Var(\hat{y}(\boldsymbol{x}))}$, then the uncertainty measure $S(\boldsymbol{x}) = \beta s(\boldsymbol{x})$ with β a positive constant weight which controls the scale of the uncertainty measure. That said, the range of selection of $\boldsymbol{x}$ can be wider if a larger β is adopted. Maximization of the utility function tends to select $\boldsymbol{x}^*$ that finds the balance between the proximity to the iso-surface $(y(\boldsymbol{x}) - a)^2$ and the uncertainty measure $S^2(\boldsymbol{x})$. Considering the predictive uncertainty for $y(\boldsymbol{x})$, we define an acquisition function $A(\boldsymbol{x}) = E_y[I(\boldsymbol{x})]$. Therefrom, selection of $\boldsymbol{x}^*$ can be represented as

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in D}{\operatorname{argmax}}\, E_y\left[I(\boldsymbol{x})\right]. \tag{6}$$

The closed-form expression of the acquisition function $E_y[I(\boldsymbol{x})]$ is given as[11]

$$\begin{aligned} E_y\left[I(\boldsymbol{x})\right] = &\left(S^2(\boldsymbol{x}) - (\hat{y}(\boldsymbol{x}) - a)^2\right)(\Phi(u_b) - \Phi(l_b)) - s^2(\boldsymbol{x})\int_{l_b}^{u_b} v^2\phi(v)\,dv \\ &+ 2(\hat{y}(\boldsymbol{x}) - a)\,s(\boldsymbol{x})(\phi(u_b) - \phi(l_b)), \end{aligned} \tag{7}$$

where $v = \frac{y(\boldsymbol{x})-\hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}$, $u_b = \frac{a-\hat{y}(\boldsymbol{x})+S(\boldsymbol{x})}{s(\boldsymbol{x})}$ and $l_b = \frac{a-\hat{y}(\boldsymbol{x})-S(\boldsymbol{x})}{s(\boldsymbol{x})}$. Equation ((7) is obtained when $(y(\boldsymbol{x}) - a)^2 \le S^2(\boldsymbol{x})$, wherein the u_b and l_b are derived from $(y(\boldsymbol{x}) - a)^2 \le S^2(\boldsymbol{x})$ and then normalized. $\Phi(\cdot)$ and $\phi(\cdot)$ are standard Gaussian density function and cumulative density function, respectively. Equation ((7) can be further simplified as $E_y[I(\boldsymbol{x})] = (S^2(\boldsymbol{x}) - (\hat{y}(\boldsymbol{x}) - a)^2 - s^2(\boldsymbol{x}))(\Phi(u_b) - \Phi(l_b)) + s^2(\boldsymbol{x})(u_b\phi(u_b) - l_b\phi(l_b)) + 2(\hat{y}(\boldsymbol{x}) - a)s(\boldsymbol{x})(\phi(u_b) - \phi(l_b))$: the first term

dominates the value of $E_y[I(\boldsymbol{x})]$ if the predicted response $\hat{y} \approx a$, i.e., the sampling point is close to the iso-surface; the second term gains sway for the design point with large predictive uncertainty while it is in the proximity $[a - \mathcal{S}, a + \mathcal{S}]$ to the iso-surface; and the third term promotes global exploration of regions that are not in the proximity of the iso-surface but with high uncertainty.[11]

In batch sampling, a set of n_b critical points $\boldsymbol{X}_b^* = \{\boldsymbol{x}_1^*, \dots, \boldsymbol{x}_b^*\}$ can be selected via

$$\boldsymbol{X}_b^* = \underset{\boldsymbol{X}_b \in D}{\operatorname{argmax}} \sum_{\boldsymbol{x} \in \boldsymbol{X}_b, |\boldsymbol{X}_b| = n_b} E_y\left[I\left(\boldsymbol{x}\right)\right]. \tag{8}$$

In contrast to Equation ((6) for the selection of one single best sampling point, the top n_b points with maximal sum of expected utility is sought after. Nonetheless, as alluded earlier, this selection criterion inevitably leads to near duplicates, that are informative individually but not collectively. This considerably suppresses the efficiency of surrogate learning, particularly for the GP, and barely contributes to update of the learned iso-surface.

3 | METHODOLOGY

It is noted that only the selection of individual critical design points via Equation ((6) is considered for sequential design in.[11] We instead investigate sequential batch sampling in this present study. The innovation of our work lies in the first-of-a-kind dispersion concept to distribute sampling points in each batch for variance reduction in the critical region, and the modified EI prescribes the informativeness for those design points during the optimization process.

3.1 | Nystrom approximation of covariance for the critical region

To eschew the near duplicates, the concept of dispersion for the critical batch $\boldsymbol{X}_b^*$ is investigated by Nystrom approximation. As the modified EI tends to select the critical points with response $y(\boldsymbol{x}^*) \in [a - \mathcal{S}(\boldsymbol{x}^*), a + \mathcal{S}(\boldsymbol{x}^*)]$, we define a critical region $\{\boldsymbol{x} \in \boldsymbol{X}_s \subset \mathbf{X} : \hat{y}(\boldsymbol{x}) \in [a - \mathcal{S}(\boldsymbol{x}), a + \mathcal{S}(\boldsymbol{x})]\}$, which consists of a set of n_s design points. Ideally, addition of a new batch $\boldsymbol{X}_b$ should reduce uncertainty of the critical region $\boldsymbol{X}_s$ to a maximal extent. Given that $y(\boldsymbol{X}_s) \equiv \boldsymbol{Y}_s$ and $y(\boldsymbol{X}_b) \equiv \boldsymbol{Y}_b$ have a joint Gaussian distribution $\begin{bmatrix} \boldsymbol{Y}_s \\ \boldsymbol{Y}_b \end{bmatrix} \sim \mathcal{N}(0, \begin{bmatrix} \boldsymbol{K}_{ss} & \boldsymbol{K}_{sb} \\ \boldsymbol{K}_{sb}^T & \boldsymbol{K}_{bb} \end{bmatrix})$, the uncertainty of critical region $\boldsymbol{X}_s$ with addition of a batch $\boldsymbol{X}_b$ can be expressed as

$$U = \frac{Tr\left(Cov(\boldsymbol{X}_s|\boldsymbol{X}_b)\right)}{n_s} = \frac{Tr\left(\boldsymbol{K}_{ss} - \boldsymbol{K}_{sb}\boldsymbol{K}_{bb}^{-1}\boldsymbol{K}_{sb}^T\right)}{n_s}, \tag{9}$$

where Tr is the trace of a matrix, and $\boldsymbol{K}_{ss} \in \mathbb{R}^{n_s \times n_s}$ and $\boldsymbol{K}_{bb} \in \mathbb{R}^{n_b \times n_b}$ are the covariance matrixes of $\boldsymbol{X}_s$ and $\boldsymbol{X}_b$, respectively. $\boldsymbol{K}_{sb} \in \mathbb{R}^{n_s \times n_b}$ is the covariance matrix between $\boldsymbol{X}_s$ and $\boldsymbol{X}_b$. The entities of the covariance matrices are derived from the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$. $\boldsymbol{K}_{sb}\boldsymbol{K}_{bb}^{-1}\boldsymbol{K}_{sb}^T$ is the Nystrom approximation of covariance matrix $\boldsymbol{K}_{ss}$.[27] U can be interpreted as the variance of $\boldsymbol{X}_s$ when $\boldsymbol{X}_b$ is included. It is noteworthy that Equation ((9) bears the resemblance to selection of inducing points in sparse GP, which comprises of a subset of the most informative training data.[28] The optimal batch is selected to minimize U for better Nystrom approximation. To minimize U, the selected critical points in $\boldsymbol{X}_b$ should not be close to each other, resulting in large entities of $\boldsymbol{K}_{sb}\boldsymbol{K}_{bb}^{-1}\boldsymbol{K}_{sb}^T$. We use exponential kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2l^2})$ as an example: if $\boldsymbol{x}$ and $\boldsymbol{x}'$ are close, $k(\boldsymbol{x}, \boldsymbol{x}')$ will be close to 0 for a fixed length scale l. In this regard, min U can be deemed as a dispersion force among the sampling points in $\boldsymbol{X}_b$, namely, $\boldsymbol{X}_b^* = \underset{\boldsymbol{X}_b \subseteq \boldsymbol{X}_s}{\operatorname{argmin}}\, U$. Furthermore, Equation ((9) can be rewritten as $U = \frac{Tr(Cov(\boldsymbol{X}_{s\setminus b}|\boldsymbol{X}_b))}{n_s}$, where $\boldsymbol{X}_{s\setminus b}$ is the difference between set $\boldsymbol{X}_s$ and $\boldsymbol{X}_b$, namely $\boldsymbol{X}_s = \boldsymbol{X}_{s\setminus b} \cup \boldsymbol{X}_b$. To start with, we set $\boldsymbol{X}_s$ as the unlabeled set $\mathcal{D}$, and it will be refined as more sampling points are included to update the contour and the surrogate.

3.2 | Weighted variance reduction of critical region

Next, we integrate the idea of greedy selection of inducing variables in sparse GP[29] with the modified EI function for critical batch selection. Given a batch $\boldsymbol{X_{b\setminus j}} \subset \boldsymbol{X_s}$, the variance reduction of the critical region $\boldsymbol{X_s}$ with addition of a new design point $\boldsymbol{x_j} \in \boldsymbol{X_s}$ is given as

$$\Delta_j = \frac{Tr\left(Cov(\boldsymbol{X_s}|\boldsymbol{X_{b\setminus j}}) - Cov(\boldsymbol{X_s}|\boldsymbol{X_b})\right)}{n_s}, \tag{10}$$

where $\boldsymbol{X_b} = \boldsymbol{X_{b\setminus j}} \cup \{\boldsymbol{x_j}\}, j = 1, \dots, n_s - n_{b\setminus j}$ and $|\boldsymbol{X_{b\setminus j}}| = n_{b\setminus j}$. The rescaled EI function $\tilde{E}_y[I(\boldsymbol{x_j})] = \frac{E_y[I(\boldsymbol{x_j})] - \min(\boldsymbol{E_y})}{\max(\boldsymbol{E_y}) - \min(\boldsymbol{E_y})} \in [0, 1]$, where $\boldsymbol{E_y} = [E_y[I(\boldsymbol{x}_1)], E_y[I(\boldsymbol{x}_2)], \dots, E_y[I(\boldsymbol{x_{n_s}})]]$, serves as an informativeness score. Then, the acquisition function is given as

$$A\left(\boldsymbol{X_b}\right) = \sum_{\boldsymbol{x_j} \in \boldsymbol{X_b}} \tilde{E}_y\left[I\left(\boldsymbol{x_j}\right)\right] \Delta_j. \tag{11}$$

That said, the rescaled EI function plays a role of weight in variance reduction in the critical region with the batch $\boldsymbol{X_b}$ in the GP surrogate framework, and the weight $\tilde{E}_y[I(\boldsymbol{x_j})]$ provide the importance of $\boldsymbol{x_j}$ in estimation of the iso-surface. Hence, the design points, that maximally reduce the uncertainty of the critical region and are more likely to update the predicted iso-surface, are selected. Note that Equation ((11) is a non-decreasing linear modular function, and we propose a greedy algorithm for iterative selection of sampling points for each batch in Algorithm 1. Here, $\tilde{E}_y[I(\boldsymbol{x_j})] \in [0, 1]$ is the rescaled modified EI and $\Delta_j > 0$ is the variance reduction after addition of $\boldsymbol{x_j}$. Therefore, $A(\boldsymbol{X_b})$ is positive. To start with, we set the critical region as $\boldsymbol{X_s} = \mathcal{D}$, and $\boldsymbol{X_b^*}$ is initialized as $\emptyset$, thus $Cov(\boldsymbol{X_s}|\emptyset) = \boldsymbol{K_{ss}}$. With selection of sampling points in each batch, the GP surrogate model and the critical region $\boldsymbol{X_s}$, hence the estimated iso-surface, are updated consequently.

Algorithm 1:

Input: acquisition size n_b, $\boldsymbol{X_s} = \mathcal{D}$
Initialize $\boldsymbol{X_b^*} = \emptyset$, $j = 1$
While $j \le n_b$:
$\boldsymbol{X_{b\setminus j}} := \boldsymbol{X_b^*}$
For $\boldsymbol{x_j} \in \mathcal{D}\setminus\boldsymbol{X_{b\setminus j}}$:
$\tilde{\boldsymbol{X}}_{\boldsymbol{b}} = \boldsymbol{X_{b\setminus j}} \cup \{\boldsymbol{x_j}\}$
$\Delta_j = \frac{Tr(Cov(\boldsymbol{X_s}|\boldsymbol{X_{b\setminus j}}) - Cov(\boldsymbol{X_s}|\tilde{\boldsymbol{X}}_{\boldsymbol{b}}))}{n_s}$
End for
$\boldsymbol{x}^* = \underset{\boldsymbol{x_j} \in \mathcal{D}\setminus\boldsymbol{X_{b\setminus j}}}{\operatorname{argmax}} \sum_{\boldsymbol{x_j} \in \boldsymbol{X_b}} \tilde{E}_y[I(\boldsymbol{x_j})]\, \Delta_j$
$\boldsymbol{X_b^*} = \boldsymbol{X_{b\setminus j}} \cup \{\boldsymbol{x}^*\}$
$j{+}{+}$
End While
Output: $\boldsymbol{X_b^*}$

Algorithm 1 requires calculation of $Cov(\boldsymbol{X_s}|\tilde{\boldsymbol{X}}_{\boldsymbol{b}})$ that involves inversion operation $\boldsymbol{K_{bb}^{-1}}$ repeatedly. This could be cumbersome when n_b is large. We suggest a lower-diagonal-upper (LDU) decomposition, that is, $\boldsymbol{K_{bb}^{-1}} = \begin{bmatrix} k(\boldsymbol{x_b}, \boldsymbol{x_b}) & k(\boldsymbol{x_b}, \boldsymbol{X_s}) \\ k(\boldsymbol{X_s}, \boldsymbol{x_b}) & \boldsymbol{K_{b\setminus j}} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -\boldsymbol{K_{b\setminus j}^{-1}}\boldsymbol{k_b} & \boldsymbol{I_{b\setminus j}} \end{bmatrix} \begin{bmatrix} \frac{1}{k_b - \boldsymbol{k_b^T}\boldsymbol{K_{b\setminus j}^{-1}}\boldsymbol{k_b}} & 0 \\ 0 & \boldsymbol{K_{b\setminus j}^{-1}} \end{bmatrix} \begin{bmatrix} 1 & -\boldsymbol{k_b^T}\boldsymbol{K_{b\setminus j}^{-1}} \\ 0 & \boldsymbol{I_{b\setminus j}} \end{bmatrix}$. Here, $\boldsymbol{K_{b\setminus j}}$ is the $n_{b\setminus j} \times n_{b\setminus j}$ covariance matrix of $\boldsymbol{X_{b\setminus j}}$, $k_b := k(\boldsymbol{x_b}, \boldsymbol{x_b})$ and $\boldsymbol{k_b} = k(\boldsymbol{X_s}, \boldsymbol{x_b})$ is $n_s \times 1$ vector of covariance between $\boldsymbol{X_s}$ and $\boldsymbol{x_b}$. The LDU decomposition only involves the inversion of $\boldsymbol{K_{b\setminus j}}$ for each single selection in a batch.

In our study, uncertainty reduction of the critical region $\boldsymbol{X_s}$ effectively plays the role of the dispersion force. The critical region $\boldsymbol{X_s} \coloneqq \{\boldsymbol{x} \in \boldsymbol{X_s} : \hat{y}(\boldsymbol{x}) \in [a - \mathcal{S}(\boldsymbol{x}), a + \mathcal{S}(\boldsymbol{x})]\}$ is a potential region in which the contour of interest lies. Here, a is the target and $\mathcal{S}(\boldsymbol{x}) = \beta s(\boldsymbol{x})$, with β a positive constant weight that controls the scale of the uncertainty measure, i.e., width of the region $\boldsymbol{X_s}$. Suppose the set $\boldsymbol{X_b}$ is an optimal representation of the critical region $\boldsymbol{X_s}$, then the uncertainty or variance in the critical region should be reduced maximally by $\boldsymbol{X_b}$. Mathematically, the variance that $\boldsymbol{X_b}$ explains is $Tr(\boldsymbol{K_{sb}}\boldsymbol{K_{bb}^{-1}}\boldsymbol{K_{sb}^T})$, which is the Nystrom approximation. Then, $Tr(\boldsymbol{K_{ss}})$, the variance in critical region $\boldsymbol{X_s}$, is reduced to $Tr(\boldsymbol{K_{ss}} - \boldsymbol{K_{sb}}\boldsymbol{K_{bb}^{-1}}\boldsymbol{K_{sb}^T})$ with addition of $\boldsymbol{X_b}$. Minimization of $Tr(\boldsymbol{K_{ss}} - \boldsymbol{K_{sb}}\boldsymbol{K_{bb}^{-1}}\boldsymbol{K_{sb}^T})$ is equivalent to minimization of the average $U = \frac{Tr(\boldsymbol{K_{ss}} - \boldsymbol{K_{sb}}\boldsymbol{K_{bb}^{-1}}\boldsymbol{K_{sb}^T})}{n_s}$. In the Nystrom approximation $Tr(\boldsymbol{K_{sb}}\boldsymbol{K_{bb}^{-1}}\boldsymbol{K_{sb}^T})$, points in the set $\boldsymbol{X_b}$ should not be close to each other, because $\boldsymbol{K_{bb}} \coloneqq k(\boldsymbol{X_b}, \boldsymbol{X_b})$ considers the distance among those design points. In implementation, the greedy algorithm is used to approximate the NP-hard problem, i.e., the selection of $\boldsymbol{X_b}$. We use the modified expected improvement $E_y[I(\boldsymbol{x})]$ as weight for each candidate $\boldsymbol{x}$ to select $\boldsymbol{X_b}$ in the vicinity of the contour of interest to obtain a weighted representation of $\boldsymbol{X_s}$. For unweighted representation, that is, Nystrom approximation, the weight is 1 for each candidate.

4 | NUMERICAL STUDIES

In this section, we demonstrate the effectiveness of the proposed approach for adaptive contour estimation in three numerical studies, that is, the four-branch function, a disjoint iso-surface, as well as the stability analysis for time-delay dynamics. It is noted that given the contour of interest $\{\boldsymbol{x} : y(\boldsymbol{x}) = a\}$, it is no-trivial to solve the inverse problem to glean the design points that are on the iso-surface. To this end, we include a new set $\boldsymbol{X}_{test}$ of n_{test} test points to assess the quality of contour estimation:

$$e = 1 - \frac{\sum_x 1(\boldsymbol{x})}{n_{test}} \tag{12}$$

where $\boldsymbol{x} \in \boldsymbol{X}_{test}$ and $1(\boldsymbol{x}) = \begin{cases} 1, (\hat{y}(\boldsymbol{x}) - a)(y(\boldsymbol{x}) - a) \geq 0 \\ 0, \ otherwise \end{cases}$. $\hat{y}(\boldsymbol{x})$ is the predictive response from GP. Ideally, the overlap of the target contour $\{\boldsymbol{x} : y(\boldsymbol{x}) = a\}$ and the surrogate estimation $\{\boldsymbol{x} : \hat{y}(\boldsymbol{x}) = a\}$ induces $e = 0$. $\boldsymbol{X_{test}}$ is constructed from a large number of mesh grid points in the two-dimensional examples in Case 1 and Case 3, and from Latin hypercube design (LHD)[30] for the four-dimensional example in Case 2. $\boldsymbol{X_{test}}$ is also used to establish the ground truth of the iso-surface. We also note that β controls the width of the critical region, and large β implies huge computational cost, especially for high-dimensional cases. It also depends on the volume of initial design, or the cold-start problem. Hence, for high-dimensional problems (e.g., even the 4D Case 2 here), we suggest a relatively large number of initial design points, which only entails a small β and further reduces the computational cost.

We compare the proposed method with the state-of-the-art method WKMS and the one-shot design. The result shows that the proposed method can achieve even better result especially when the iso-surface is complicated and achieves similar result when the iso-surface is simple compared to WKMS selection. Also, the proposed method can obtain significantly better result compared to the one-shot design.

The batch size $|\boldsymbol{X_b}|$ and the number of initial training points may depend on the dimension of problems. There are no specific rules to determine the number of batch size $|\boldsymbol{X_b}|$ and the number of initial training points. The fundamental is that we need sufficient number of initial training points to cover the whole sampling space, and in each batch, the new addon points should cover the critical region as much as possible. Hence, we may select a larger number of points on problems with higher dimension.

4.1 | The four-branch function

The four-branch function represented in Equation ((13) is widely used in structural reliability analysis. It captures the limit state of a series system with four distinct components.[31]

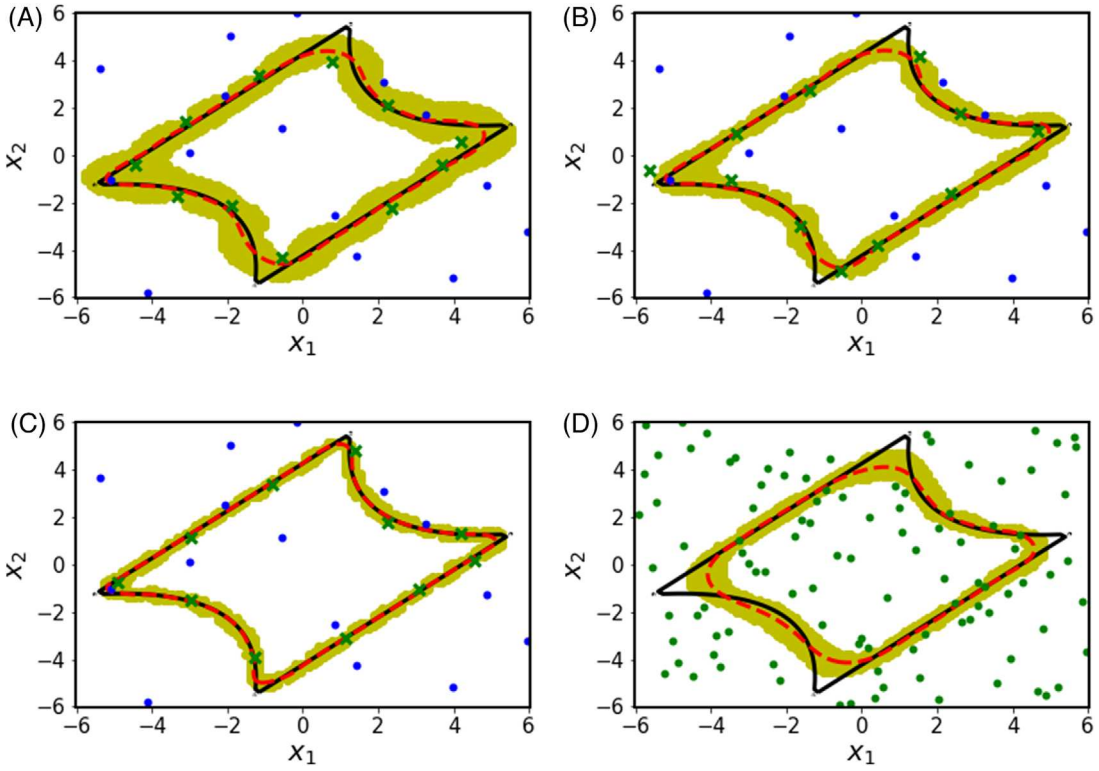


FIGURE 2 Evolution of the predicted iso-surface using the proposed method at (A) 2nd, (B) 4th and (C) 8th iteration, and (D) the predicted iso-surface of the one-shot design. The black solid curve is the true iso-surface and the red dashed curve is the predicted iso-surface. The green crosses are critical batch $\boldsymbol{X}_b^*$ and yellow area is $\boldsymbol{X}_s$. Blue dots are the initial design points in (A), (B), (C), and green dots are the one-shot design points in (D). To avoid the clutter, we only show the selected critical batch $\boldsymbol{X}_b^*$ for each batch and hide those selected in previous batches

$$(\boldsymbol{x}) = \min \begin{bmatrix} 3 + 0.1(x_1 - x_2)^2 - \dfrac{x_1 + x_2}{\sqrt{2}} \\ 3 + 0.1(x_1 - x_2)^2 + \dfrac{x_1 + x_2}{\sqrt{2}} \\ x_1 - x_2 + \dfrac{6}{\sqrt{2}} \\ x_2 - x_1 + \dfrac{6}{\sqrt{2}} \end{bmatrix}, \tag{13}$$

Here, $\boldsymbol{x} = [x_1, x_2]$ are the design variables, and the system fails when $y(\boldsymbol{x}) \leq 0$. Hence, the iso-surface of interest is defined as $\{\boldsymbol{x} : y(\boldsymbol{x}) = 0\}$. The true iso-surface is derived from evaluation of Equation ((13) on a 100×100 mesh grid of $\boldsymbol{X}_{test} \in [-6, 6]^2$, amounting to 10^4 evaluations in total. For the sequential design, $n = 15$ initial design points (the blue dots in Figures 1 and 2) are selected by LHD with min-max distance criterion. Next, $n_b = 11$ critical design points (green crosses in Figures 1 and 2) are selected via the proposed method to form the critical batch $\boldsymbol{X}_b^*$ in each iteration, which will be evaluated according to Equation ((13). Following this, the GP surrogate is updated at each iteration until the predicted iso-surface converges to the true iso-surface. To obtain the critical region $\boldsymbol{X}_s \subset \boldsymbol{X}_{test}$ (the shaded yellow area), we set $\beta = 0.95$, that is, $\hat{y}(\boldsymbol{x}_s) \in [0 - 0.95 \times s(\boldsymbol{x}_s), 0 + 0.95 \times s(\boldsymbol{x}_s)], \boldsymbol{x}_s \in \boldsymbol{X}_s$.

In both Figures 1 and 2, the solid / black curve represents the true iso-surface $\{\boldsymbol{x} : y(\boldsymbol{x}) = 0\}$ and red / dashed curve is the predicted iso-surface $\{\boldsymbol{x} : \hat{y}(\boldsymbol{x}) = 0\}$ from GP surrogate after addition of each critical batch. In Figure 1(A), the predicted iso-surface by GP after the initial design is shown for comparison. Figure 1(B) shows the design point selection in the first iteration using the proposed method, which exhibits the dispersion effect, leading to significant update of the surrogate model. In Figure 1(C), the WKMS[24] is included for comparison and the scaled modified EI function value serves as the weight. The WKMS achieves similar predicted iso-surface with the proposed method after the ini-

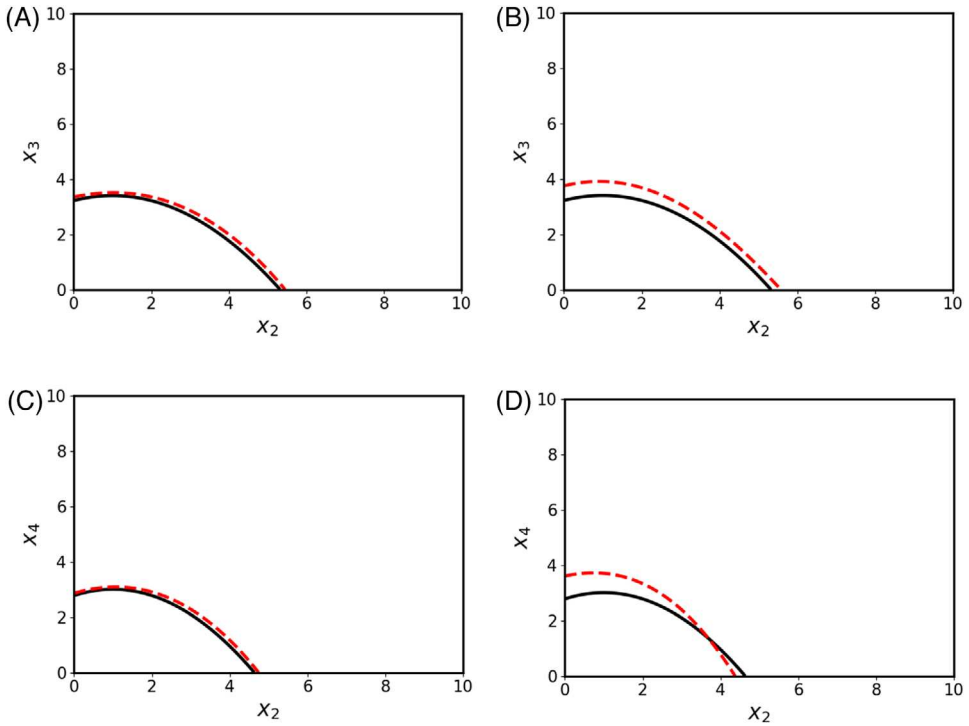


FIGURE 3 Comparison of the predicted iso-surface between the proposed method and WKMS on 2 dimensions. (A) and (B) are the predicted iso-surface obtained by the proposed method and WKMS on (x_2, x_3) while (x_1, x_4) fixed at 0.5. (C) and (D) are the predicted iso-surface obtained by the proposed method and WKMS on (x_2, x_4) while (x_1, x_3) fixed at 0.5. The black solid curve is the true iso-surface and the red dashed curve is the predicted iso-surface

tial design in this simple example. The issue of near duplicates from batch selection via modified EI is exemplified in Figure 1(D), and those near duplicates barely contribute to the update of the surrogate. The evolution of the predicted iso-surface using the proposed method at different iterative steps are shown in Figure (2 A-D) shows the predicted iso-surface using one-shot design with the same total number of design points. Here, the one-shot design is conducted by LHD with min-max distance criterion. With 8 iterations, the estimated iso-surface almost converges to the true iso-surface, and totally, $15 + 8 \times 11 = 103$ sampling points are included. The proposed method achieves $e = 0.006$ and $e = 0.012$ is obtained for the WKMS. The one-shot design obtains $e = 0.034$ with 103 sampling points. In this simple case, the proposed method obtains a slightly better result compared to the WKMS and a significantly better result compared to the one-shot design.

4.2 | A disjoint iso-surface

In this case, we adopt the four-dimensional function with a disjoint iso-surface:

$$y(\boldsymbol{x}) = \frac{1}{4}\left(\sin(x_1 - 3)(x_2 - 1)^2 + (x_3 - 1)x_4\right) - 3, \tag{14}$$

where $\boldsymbol{x} = [x_1, x_2, x_3, x_4] \in [0, 10]^4$. The iso-surface of interest is defined as $\{\boldsymbol{x} : y(\boldsymbol{x}) = 0\}$. To start with, we select $n = 40$ design points from LHD with min-max distance criterion, define $\beta = 0.5$ to reduce the computational cost and set the size of each batch as $n_b = 5$ in the sequential batch sampling. To quantify the performance of the proposed method and WKMS, we select $\boldsymbol{X_{test}}$ with size 10^5 from LHD. To further visualize the part of iso-surface, the four-dimensional iso-surface is projected onto a two-dimensional space. Here, we select 2 angles. For the first one, we set $[x_2, x_3] \in [0, 10]^2$ and fix $[x_1, x_4] = 0.5$. For the other, we set $[x_2, x_4] \in [0, 10]^2$ and fix $[x_1, x_3] = 0.5$. After 6 iterations, $40 + 5 \times 6 = 70$ design points are used, and the proposed method obtains the error measure $e = 0.01$ while $e = 0.02$ for WKMS. The predicted iso-surface are showcased in the Figure 3.

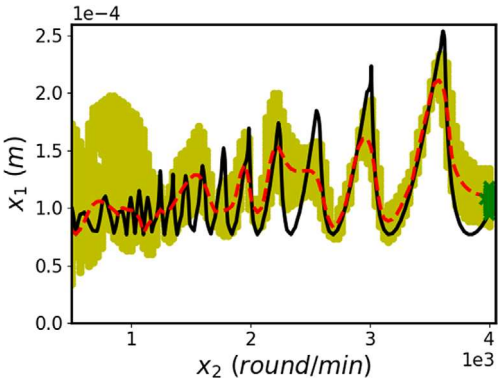


FIGURE 4 The near duplicates (green clustered crosses) caused by selecting top 11 best critical design points according to the modified EI after the initial design

The proposed method is slightly better than WKMS in terms of fitting the iso-surface and achieves smaller error with the same experiment setting. In this case, the iso-surface is relatively simple because it requires less than 100 design points to obtain a small error e for both methods. Hence, both methods do not show a large difference in this case. The proposed method obtains significantly better result compared to the one-shot design.

4.3 | Stability identification of time-delay dynamic systems

In this example, we present the proposed method on a complicated and bumpy iso-surface: the stability boundary in a machining or material removal process.[9] The cutting tool vibration is modeled by a second-order delay differential equation

$$\ddot{z}(t) + 2\zeta\omega_n \dot{z}(t) + \omega_n^2 z(t) = -\mathcal{K}w\left(x_1 - z(t) + z(t-\tau)\right). \tag{15}$$

Here, $z(t) \in \mathbb{R}$ is the tool displacement relative to the nominal position in feed direction during the machining process. $\omega_n = 600\pi$ Hz is the vibration natural frequency, $\zeta = 0.02$ is the damping ratio, $\mathcal{K} = 2 \times 10^{11}\ N/m^2$ represents the force coefficient, and the constant cutting width is $w = 5 \times 10^{-2}\ \mu m$. Thus, the righthand side of Equation ((15) is the instantaneous cutting force, which is proportional to $\mathcal{K}$ and instantaneous cutting area $w(x_1 - z(t) + z(t-\tau))$. The design variables are $\boldsymbol{x} = [x_1, x_2]$: x_1 represents the nominal feed (μm) and x_2 is the spindle speed (round / minute), which further determines the time delay τ, that is, $\tau = \frac{2\pi}{x_2}$. In this study, only x_1 and x_2 are tunable and other parameters are fixed once the machine is set up. The stability problem, whether the dynamics of $z(t)$ explodes or dies off, can be solved via temporal finite element method (TFEM) as demonstrated in our recent studies,[8,14] in which Equation ((15) is reduced to a compact matrix form as

$$\boldsymbol{N}\boldsymbol{a}^{\boldsymbol{n}} = \boldsymbol{P}\boldsymbol{a}^{\boldsymbol{n-1}} + \boldsymbol{Q} \tag{16}$$

where $\boldsymbol{a}^{\boldsymbol{n}}$ and $\boldsymbol{a}^{\boldsymbol{n-1}}$ are the coefficient vectors for the polynomial bases in TFEM in the n^{th} and $(n-1)^{th}$ revolution, respectively. $\boldsymbol{N}$, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the matrices for integration terms derived from Galerkin projection. Define λ_G as the maximum absolute eigenvalue of the transition matrix $\boldsymbol{G} = \boldsymbol{N}^{-1}\boldsymbol{P}$, then stable cutting requires $\lambda_G \leq 1$. Conversely, $\lambda_G > 1$ implies unstable cutting, manifesting in fierce tool vibration, which could considerably diminish quality of the machined products. Therefore, we seek the iso-surface $\{\boldsymbol{x} : \lambda_G(\boldsymbol{x}) = 1\}$ to separate the stable and unstable cutting.

The initial design contains $n = 100$ design points selected by LHD with min-max distance criterion and we set the batch size $n_b = 11$. For demonstration, we showcase the near duplicates in Figure 4, which are the batch sampling points generated by the conventional modified EI. Evolution of the predicted boundary using the proposed method and the WKMS are showcased in Figures 5 and 6. The black solid curve is the true boundary or the iso-surface $\{\boldsymbol{x} : \lambda_G(\boldsymbol{x}) = 1\}$, obtained on a 100×100 mesh grid which is also $\boldsymbol{X_{test}}$. The red dashed curve is the predicted boundary $\{\boldsymbol{x} : \hat{\lambda}_G(\boldsymbol{x}) = 1\}$

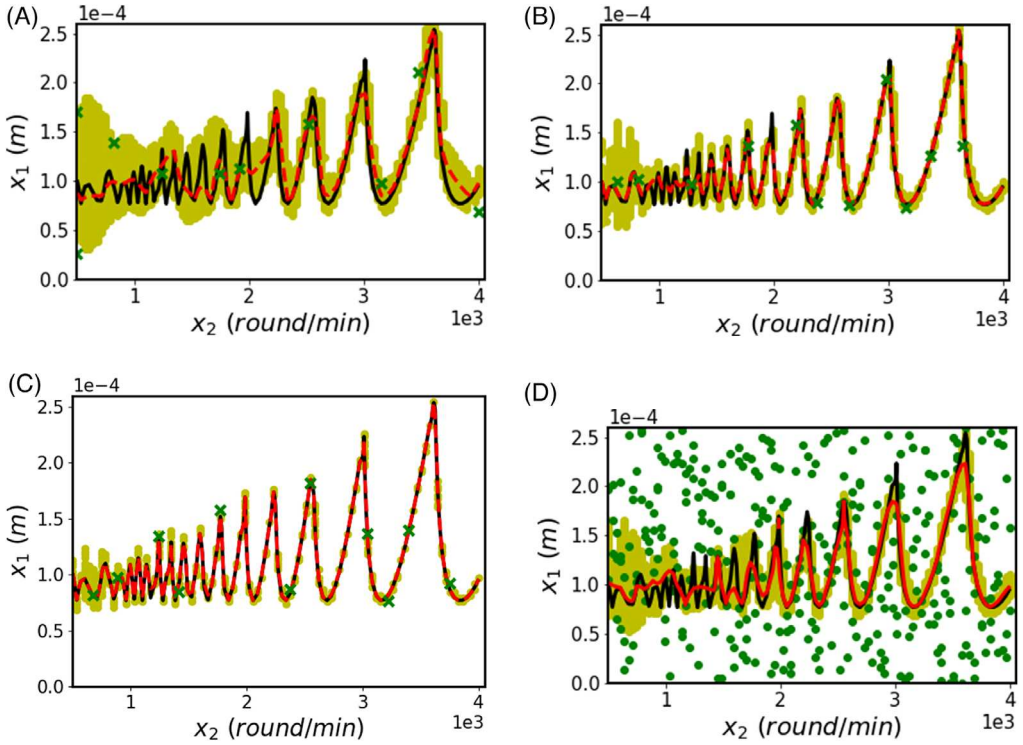


FIGURE 5 Comparison between one-shot design and the proposed method. The evolution of predicted boundary using the proposed method at the iteration (A) 1, (B) 11, (C) 21, and (D) the predicted iso-surface of the one-shot design. Solid black and dashed red curves are true and predicted boundary respectively. Green crosses are the selected design points in the critical batch and shaded area is $\boldsymbol{X}_s$. Green dots are the one-shot design points in (D). To avoid the clutter, we show the selected critical batch $\boldsymbol{X}_b^*$ for each batch only and hide those selected in previous batches

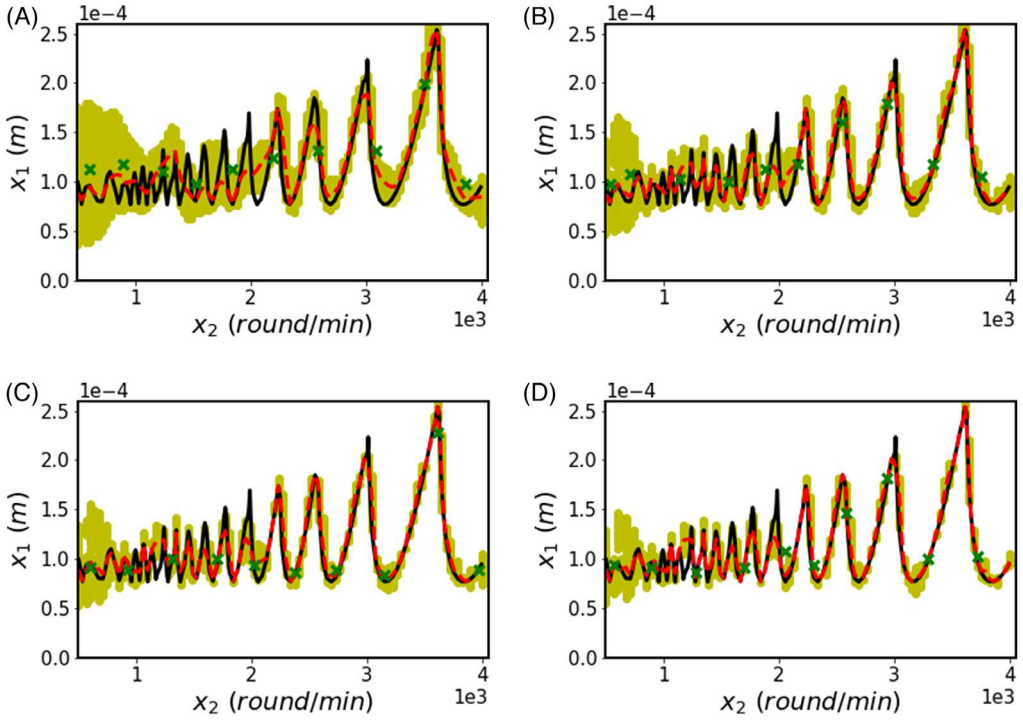


FIGURE 6 The evolution of predicted boundary using the WKMS at the iteration (A) 1, (B) 8, (C) 15, and (D) 21. Black and dashed red curves are true and predicted boundary respectively. Green crosses are the selected design points in the critical batch and yellow area is $\boldsymbol{X}_s$. To avoid the clutter, we show the selected critical batch $\boldsymbol{X}_b^*$ for each batch only and hide those selected in previous batches

via GP surrogate. Green crosses are the selected batch at each iteration while the shaded area is $\boldsymbol{X}_{\boldsymbol{s}}$. In the experiment, we set weight of predicted uncertainty $\beta = 0.95$ to obtain $\boldsymbol{X}_{\boldsymbol{s}}$.

As portrayed in Figure 5, the surrogate successfully finds the true boundary via the proposed method in the 21st iteration with $100 + 11 \times 21 = 331$ and the critical batch exhibits the dispersion. Even though the critical batch in Figure 6 via WKMS shows the diversity, the predicted boundary does not overlap with the true boundary with the same experiment setting, i.e., with the same weight, batch size, total number of sampling design points. The result shows that the proposed method significantly outperforms the WKMS in the same experiment setting in terms of the accuracy of predicted iso-surface. Also, the proposed method has smaller error measure $e = 0.0017$, while $e = 0.0177$ for WKMS at the 21st iteration. As aforementioned, the proposed method can achieve better result when the iso-surface is complicated. This is because WKMS only aims to spread out the sampling points in each batch, without considering the uncertainty of the predicted iso-surface.

5 | CONCLUSION

Whereas sequential sampling plays a paramount role in surrogate modeling and optimal design for computer simulations, the sequential batch design has not been extensively carried out, particularly in the framework of GP surrogates. It is noted that fitting the GP incurs a huge computational cost with accumulation of the sampling points, thus stymieing its applicability. This can be readily solved by batch sampling, which is amenable to parallel computing. We have shown in this study that the naive extension of the conventional top-one sequential sampling strategies to batch selection inevitably leads to near duplicates, which considerably compromises the efficiency of surrogate learning and contour estimation. Whereas the lion's share of existing research in batch sampling is focused on classification problems, we develop a first-of-a-kind algorithm for sequential batch sampling in regression settings. The novel dispersion concept is adopted to sparsely spread the sampling points in each batch via integration of the modified EI function and variance reduction in the critical region for iso-surface estimation in the framework of GP surrogate. The proposed method is corroborated in three numerical case studies, and we have consistently shown that it outperforms the WKMS method in approximating the iso-surface of interest, especially when the iso-surface is complicated. This is attributed to the fact that WKMS achieves the sparse distribution of design points for each batch by maximizing the distance between them and does not account for the uncertainty of the predicted iso-surface. By contrast, the proposed dispersion-enhanced approach concurrently spreads the design points in each batch and reduces the uncertainty of the critical region in estimation of the iso-surface.

ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation (Award Number 2119334) of the United States, the Smart Energy Transdisciplinary Area of Excellence at Binghamton University, and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article, as no new data were created or analyzed in this study.

ORCID

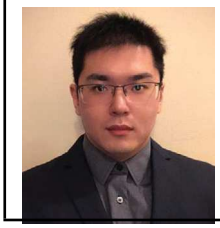
Changqing Cheng https://orcid.org/0000-0001-7810-7279

REFERENCES

1. Cheng C, Bukkapatnam ST, Raff LM, Hagan M, Komanduri R. Monte Carlo simulation of carbon nanotube nucleation and growth using nonlinear dynamic predictions. *Chem Phys Lett*. 2012;530:81-85.
2. Mak S, Sung C-L, Wang X, et al. An efficient surrogate model for emulation and physics extraction of large eddy simulations. *J Am Statist Assoc*. 2018;113(524):1443-1456.
3. Gauthier B, Pronzato L. Spectral approximation of the imse criterion for optimal designs in Kernel-Based interpolation models. *SIAM/ASA J Uncertain Quantif*. 2014;2(1):805-825.
4. Gauthier B, Pronzato L. Approximation of IMSE-Optimal designs via quadrature rules and spectral decomposition. *Commun Stat- Simul Comput*. 2016;45(5):1600-1612.

5. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J Global Optim*. 1998;13(4):455-492.
6. Müller J, Day M. Surrogate optimization of computationally expensive black-box problems with hidden constraints. *INFORMS J Comput*. 2019;31(4):689-702.
7. Park J-S. Optimal latin-hypercube designs for computer experiments. *J Stat Plan Inference*. 1994;39(1):95-111.
8. Che Y, Cheng C. Uncertainty quantification in stability analysis of chaotic systems with discrete delays. *Chaos, Solitons Fractals*. 2018;116:208-214.
9. Cheng C, Wang Z, Hung W, Bukkapatnam ST, Komanduri R. Ultra-Precision machining process dynamics and surface quality monitoring. *Procedia Manuf*. 2015;1:607-618.
10. Schöbi R, Sudret B, Marelli S. Rare event estimation using polynomial-chaos kriging. *ASCE-ASME J Risk Uncertain Eng Syst A: Civ Eng*. 2017;3(2):D4016002.
11. Ranjan P, Bingham D, Michailidis G. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*. 2008;50(4):527-541.
12. Picheny V, Ginsbourger D, Roustant O, Haftka RT, Kim N-H. Adaptive designs of experiments for accurate approximation of a target region. *J Mech Des*. 2010;132(7).
13. Settles B. *Active learning literature survey*. Computer Science Technical Report. University of Wisconsin-Madison; 2009.
14. Che Y, Liu J, Cheng C. multi-fidelity modeling in sequential design for stability identification in dynamic time-delay systems. *Chaos*. 2019;29(9):093105.
15. Jones B, Johnson RT. Design and analysis for the gaussian process model. *Qual Reliab Eng Int*. 2009;25(5):515-524.
16. Imani F, Cheng C, Chen R, Yang H. Nested gaussian process modeling and imputation of high-dimensional incomplete data under uncertainty. *IISE Trans Healthc Syst Eng*. 2019;9(4):315-326.
17. Gramacy RB, Lee HKH. Bayesian treed gaussian process models with an application to computer modeling. *J Am Statist Assoc*. 2008;103(483):1119-1130.
18. Rasmussen C, Williams C. *Gaussian Processes for Machine Learning*. MIT Press; 2006.
19. Loeppky JL, Moore LM, Williams BJ. Batch sequential designs for computer experiments. *J Stat Plan Inference*. 2010;140(6):1452-1464.
20. Houlsby N, Huszár F, Ghahramani Z, Lengyel M. Bayesian Active learning for classification and preference learning. *arXiv preprint*. ar2011;Xiv:1112.5745.
21. Kirsch A, Van Amersfoort J, Gal Y. Batchbald: efficient and diverse batch acquisition for deep bayesian active learning. *Adv Neural Inf Process Syst* . 2019;32:7026-7037.
22. Ash JT, Zhang C, Krishnamurthy A, Langford J, Agarwal A, Deep batch active learning by diverse, uncertain gradient lower bounds. *International Conference on Learning Representations*. 2019
23. Sener O, Savarese S, Active learning for convolutional neural networks: a Core-Set Approach. *International Conference on Learning Representations*. 2018
24. Zhdanov F. Diverse mini-batch active learning. *arXiv preprint*. 2019. arXiv:1901.05954.
25. Duan W, Ankenman BE, Sanchez SM, Sanchez PJ. Sliced Full Factorial-Based Latin Hypercube Designs as a Framework for a Batch Sequential Design Algorithm. *Technometrics*. 2017;59(1):11-22.
26. Crombecq K, De Tommasi L, Gorissen D, Dhaene T, A novel sequential design strategy for global surrogate modeling. *Proceedings of the 2009 Winter Simulation Conference (WSC)*, IEEE, 2009;pp. 731-742.
27. Williams C, Seeger M, Using the nyström method to speed up kernel machines. Proceedings of the 14th Annual Conference on Neural Information Processing Systems, 2001. pp. 682-688.
28. Smola AJ, Bartlett PL. Sparse greedy gaussian process regression. *Adv Neural Inf Process Syst*. 2001:619-625.
29. Titsias M. Variational learning of inducing variables in sparse gaussian processes. *Artif Intell Stat*. 2009;PMLR:567-574.
30. McKay MD, Beckman RJ, Conover WJ. A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 2000;42(1):55-61.
31. Schueremans L, Van Gemert D. Benefit of splines and neural networks in simulation based structural reliability analysis. *Struct Saf*. 2005;27(3):246-261.

AUTHOR BIOGRAPHIES



Yiming Che received the B.S. degree from the Capital University of Economics and Business, Beijing, China, in 2017, and the Master of Engineering degree in Industrial Engineering from the Binghamton University, New York, USA, in 2018. He is currently a PhD Candidate in the Department of Systems Science and Industrial Engineering at the Binghamton University, New York, USA. His research interests include statistical modeling, machine learning/deep learning, uncertainty quantification, and dynamical systems.

Juliane Mueller is the manager of the Artificial Intelligence, Learning, and Intelligent Systems (ALIS) group within the Computational Science Center at NREL. Her background is in the development of numerical optimization algorithms for black-box and compute-intensive problems where analytic descriptions of objective and constraint functions are not available. Dr. Mueller's algorithm developments include surrogate modeling and active learning. In the past, she has applied these optimization algorithms to a variety of U.S. Department of Energy-relevant problems, including environmental applications, fuel search, quantum computing, and high-energy physics. Most recently, Dr. Mueller's work has focused on tuning deep learning model architectures with the goal to find models that make robust and reliable predictions. As group leader of ALIS, it is her goal to develop optimization and machine learning capabilities that enable researchers across all NREL applications to accelerate their science.

Dr. Changqing Cheng is currently an assistant professor in the Department of Systems Science and Industrial Engineering at State University of New York, Binghamton. His research focus is on simulation and modeling for optimal design of complex networked systems, particularly the integration of physical and machine learning models, uncertainty quantification, and active learning.

How to cite this article: Che Y, Müller J, Cheng C. Dispersion-enhanced sequential batch sampling for adaptive contour estimation. *Qual Reliab Eng Int*. 2024;40:131–144. https://doi.org/10.1002/qre.3245