TECHNOMETRICS
2025, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/00401706.2025.2561745

Solving Bayesian Inverse Problems Using Gaussian Process Regression with Goal-Oriented Active Learning

Paul Lartaud[a,b], Philippe Humbert[a], and Josselin Garnier[b]

[a]CEA DAM Île-de-France, Bruyères-le-Châtel, Arpajon, France; [b]Centre de Mathématiques Appliquées, École polytechnique, Institut Polytechnique de Paris, Palaiseau, France

ABSTRACT

Sequential design focuses on designing computer experiments with limited computational budgets. It aims to create efficient surrogate models to replace complex computer codes. Some sequential design strategies can be understood within the Stepwise Uncertainty Reduction (SUR) framework. In the SUR framework, each new design point is chosen by minimizing the expectation of a metric of uncertainty with respect to the yet unknown new data point. These methods offer an accessible framework for sequential experiment design, including almost sure convergence for common uncertainty functionals. This article introduces two goal-oriented strategies tailored for surrogate models used within a Bayesian inverse problem. The Constraint Set Query (CSQ) strategy is adapted from Maximum Mean Squared Error (MMSE) designs, where the search space is constrained in a ball for the Mahalanobis distance around the maximum a posteriori. The second, known as the IP-SUR (Inverse Problem SUR) strategy, uses a posterior-weighted integrated mean squared prediction error as the uncertainty metric and is derived from SUR methods. It is tractable for Gaussian process surrogates and comes with a theoretical guarantee for the almost sure convergence of the uncertainty functional. In various test cases, both strategies are shown to outperform standard goal-oriented designs for fine-tuning surrogate models.

ARTICLE HISTORY
Received April 2024
Accepted August 2025

KEYWORDS
Active learning; sequential design; Bayesian inverse problem; surrogate models

1. Introduction

In the field of uncertainty quantification, the resolution of ill-posed inverse problems with a Bayesian approach is a widely used method. It gives access to the posterior distribution of the uncertain inputs from which various quantities of interest can be calculated for decision-making (Kaipio and Somersalo 2006; Stuart 2010). However, sampling the posterior distribution often requires a large number of calls to the forward model. This limits the tractability of such methods when the forward model is a complex computer code. To overcome this obstacle, it is common practice to replace the forward model with a surrogate such as deep neural networks (Yan and Zhou 2019b), physics-informed neural networks (Li, Wang, and Yan 2023), polynomial chaos (Yan and Zhou 2019a), or Gaussian processes (Frangos et al. 2010; Drovandi, Moores, and Boys 2018; Chen, Kang, and Lin 2021; Sauer, Gramacy, and Higdon 2023). These surrogates help in reducing the computational burden of the inverse problem resolution. The quality of the uncertainty prediction is then directly linked to the quality of the surrogate model. In some applications (Mai, Konakli, and Sudret 2017), the computational budget implies a scarcity of the numerical data used to build the surrogate model, since they are given by calls to the forward model. To build the best possible surrogate model, adaptive strategies have been developed to make the most out of the given computational budget (Wan and Ren 2016; Wang and Li 2018).

This is the premise of sequential designs for computer experiments (Sacks, Schiller, and Welch 1989; Santner et al. 2003; Gramacy and Lee 2009). While space-filling designs based on Latin Hypercube Sampling (LHS) (Stein 1987) or minimax distance designs (Johnson, Moore, and Ylvisaker 1990) have been widely investigated, they are not suited for the specific problem at stake, where the region of interest only covers a small fraction of the input space. For such cases, criterion-based designs have been introduced in various domains, such as reliability analysis (Du and Chen 2002; Lee and Jung 2008; Dubourg, Sudret, and Deheeger 2013; Agrell and Dahl 2021; Azzimonti et al. 2021; Cole et al. 2023), Bayesian optimization (Shahriari et al. 2015; Imani and Ghoreishi 2020), contour identification (Ranjan, Bingham, and Michailidis 2008), or more recently in Bayesian calibration (Kennedy and O'Hagan 2001; Ezzat, Pourhabib, and Ding 2018; Sürer, Plumlee, and Wild 2024) or in multi-fidelity surrogate modeling (Stroh et al. 2022). Most common approaches include criteria based on the maximization of the information gain (Shewry and Wynn 1987), D-optimal criteria based on the minimization of the covariance matrix determinant of the estimated parameters (de Aguiar et al. 1995; Osio and Amon 1996; Mitchell 2000; Wang et al. 2006) or designs minimizing a functional of the mean squared prediction error (MSPE) (Kleijnen and Beers 2004). We also highlight Maximum Mean Squared Error (MMSE) designs, which select the training point that maximizes the mean squared predictive

CONTACT Paul Lartaud 1s2lartaudpaul@gmail.com CEA DAM Île-de-France, Bruyères-le-Châtel, 91297 Arpajon, France.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

error (Liu, Ong, and Cai 2018; Jin, Chen, and Sudjianto 2002). This design is also known as an Active Learning MacKay (ALM) design (MacKay 1992; Gramacy and Lee 2009). Weighted integrated MSPE optimal designs have been used to improve surrogate models and outperform traditional LHS sampling (Picheny et al. 2010). In Bayesian inverse problems, sequential design strategies have been investigated to produce estimators for the inverse problem likelihood (Sinsbeck, Cooke, and Nowak 2021) or the Bayesian model evidence (Sinsbeck and Nowak 2017) to provide insight on model selection. In Li and Marzouk (2014), the design points for the surrogate model are obtained iteratively from variational posterior distributions designed to minimize the Kullback-Leibler divergence with the true posterior.

In this article, two novel sequential design strategies named Inverse Problem SUR (IP-SUR) and Constraint Set Query (CSQ) are introduced for Gaussian process surrogate models in Bayesian inverse problems. The first one is an objective-oriented design based on the Stepwise Uncertainty Reduction (SUR) paradigm applied to the posterior-weighted integrated MSPE. It is shown to be tractable for GP surrogates. Its integration into the SUR paradigm makes it easily interpretable. Moreover, it is supported by a theoretical guarantee of almost sure convergence of the metric of uncertainty, which is rarely obtained for sequential design strategies. The CSQ design is a simpler strategy adapted from MMSE design strategies. It provides a more accessible design, though it lacks the convergence guarantee of IP-SUR. Overall, the two sequential design strategies presented in this work are complementary. CSQ appears more computationally efficient for higher dimensions because it circumvents the need for intermediate Monte Carlo Markov Chain (MCMC) samplings, whereas IP-SUR offers greater theoretical rigor. We insist that the proposed strategies be used to fine-tune a surrogate to a given inverse problem. They should not be used for the initial construction of a surrogate. For this task, space-filling or criterion-based designs, such as MMSE designs, are more relevant.

The following section presents the fundamentals of Gaussian process regression, Bayesian inverse problems, and SUR designs. Then, we present our two methods and develop the theoretical foundations of the IP-SUR strategy. Finally, we apply our strategies to several test cases and compare them to other designs (MMSE, Bayes risk minimization).

2. Bayesian Inverse Problems

2.1. Problem Statement

Consider the following inverse problem. We would like to identify some parameters $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ based on the observations of some vectorial quantity $\mathbf{y} \in \mathbb{R}^d$. The link between the inputs $\mathbf{x}$ and the outputs $\mathbf{y}$ is provided by a forward model $\mathbf{f}\colon \mathcal{X} \to \mathbb{R}^d$. Assume that we are given some noisy observations of the forward model in the form of an $N \times d$ matrix $\mathbf{Y} = (\mathbf{y}_k)_{1\le k\le N}$ for $N \ge 1$. Throughout this work, vectors are denoted by bold letters and matrices by bold capital letters. We would like to estimate the parameters $\mathbf{x}$ and quantify the uncertainty associated with this estimation. The traditional approach is to solve the inverse problem in a Bayesian framework (Scales and Tenorio 2001; Dashti and Stuart 2017).

In the Bayesian paradigm, the inputs $\mathbf{x}$ and the outputs $\mathbf{y}$ are considered random vectors, and the goal is to estimate the posterior distribution $p(\,\cdot\,|\mathbf{Y})$, which is the probability distribution of the inputs $\mathbf{x}$ conditioned by the noisy observations $\mathbf{Y}$ of the outputs. We assume the observations are independent and identically distributed with zero-mean Gaussian noise:

$$\mathbf{y}_k = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}_k \text{ with } \boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\text{obs}}) \tag{1}$$

for $1 \le k \le N$, and where $\mathbf{C}_{\text{obs}}$ is the $d \times d$ covariance matrix of the observations.

Let $\mathbf{x} \in \mathcal{X}$. The posterior distribution density $p(\mathbf{x}|\mathbf{Y})$ can be expressed with Bayes' theorem as the product of a prior distribution with density $p(\mathbf{x})$ and an analytically tractable likelihood $L(\mathbf{Y}|\mathbf{x})$:

$$p(\mathbf{x}|\mathbf{Y}) \propto p(\mathbf{x})L(\mathbf{Y}|\mathbf{x}) \propto p(\mathbf{x}) \exp\left(-\frac{1}{2}\sum_{k=1}^{N} \|\mathbf{y}_k - \mathbf{f}(\mathbf{x})\|^2_{\mathbf{C}_{\text{obs}}}\right)$$

where the notation $\|\mathbf{a}\|^2_{\mathbf{C}} = \mathbf{a}^T\mathbf{C}^{-1}\mathbf{a}$ refers to the Mahanalobis norm, for any vector $\mathbf{a} \in \mathbb{R}^d$ and positive definite matrix $\mathbf{C}$ of size $d \times d$. The prior $p(\,\cdot\,)$ is chosen by the user directly. It can incorporate expert knowledge or information from other observations. It can also be taken as mostly non-informative (Jeffreys 1946; Ghosh 2011; Consonni et al. 2018).

Given a normalized or non-normalized expression of the posterior density, the posterior distribution can be sampled with MCMC methods. MCMC algorithms produce ergodic Markov chains whose invariant distribution is the target posterior distribution. MCMC methods eliminate the need to evaluate the multiplicative constant $\int p(\mathbf{x}')L(\mathbf{Y}|\mathbf{x}')d\mathbf{x}'$, which is often intractable. However, the Bayesian approach requires many calls to the forward model $\mathbf{f}$. For real-world applications, the forward model is often given by experiments or by complex computer codes. Thus, obtaining the posterior distribution is often too computationally expensive. In some cases, an analytical forward model can be found given strong assumptions, but it comes with a systematic bias that needs to be accounted for, lest the posterior distribution be inaccurate. A way to circumvent these obstacles is to rely on surrogate models.

2.2. Gaussian Process Surrogate Models

In this work, we consider Gaussian Process (GP) surrogate models because GPs provide a simple framework to quantify model uncertainty while having a simple and explicit predictive distribution. We briefly recall the basic concepts of GP surrogates. For a comprehensive review on the subject, we refer to Williams and Rasmussen (2006) and Gramacy (2020).

Let $\mathcal{X} \subset \mathbb{R}^p$ be an index set. A Gaussian process z is a stochastic process such that for any finite collection $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{X}^n$ with $n \ge 1$, the random vector $z(\mathbf{X}) = (z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n))$ of size n follows a multivariate normal distribution. The distribution of a scalar-valued Gaussian process is entirely determined by its mean function $m\colon \mathcal{X} \longrightarrow \mathbb{R}$ and its positive definite covariance function $k\colon \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$. In this work, we are considering multi-output GP models. These models are built with the Linear Model of Coregionalization (LMC) introduced in Bonilla, Chai, and Williams (2007). LMC

models allow correlations across outputs to be modeled with a linear combination of independent latent GPs. They produce a vectorial mean function $\mathbf{m}\colon \mathcal{X} \to \mathbb{R}^d$ and a multi-output covariance function $\mathbf{K}\colon \mathcal{X}\times\mathcal{X} \to \mathcal{M}_d(\mathbb{R})$ where $\mathcal{M}_d(\mathbb{R})$ is the set of real-valued square matrices of size $d \times d$. Throughout this work, a multi-output GP with mean function $\mathbf{m}$ and covariance function $\mathbf{K}$ is denoted by $\mathbf{z} \sim \mathcal{GP}\left(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}')\right)$. It refers to a d-dimensional vector-valued random function here.

Consider a collection of inputs $\mathbf{X} \in \mathcal{X}^n$ and a multi-output GP $\mathbf{z} \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{K}(\mathbf{x}, \mathbf{x}')\right)$. The distribution of the GP for the inputs $\mathbf{X}$ is written as

$$\mathbf{z}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}(\mathbf{X}, \mathbf{X})\right)$$

where $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is a $nd \times nd$ block matrix whose constitutive blocks are given by $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ for $1 \le i, j \le n$. Here $\mathbf{z}(\mathbf{X})$ is defined as a random vector of size nd where the ith block of length d contains the vector-valued output for input location $\mathbf{x}_i$ for $1 \le i \le n$.

Gaussian processes can be used to introduce a prior on a functional space. The covariance function defines the class of random functions represented by the GP prior. Standard covariance kernels are the stationary RBF and Matérn kernels. Depending on the expected properties of the target function (periodicity, regularity,…), various classes of covariance kernels can be used. The mean function is more straightforward and is often set to a constant, though more advanced approaches exist to improve the performance of the GP surrogate (Schobi, Sudret, and Wiart 2015). GP-based surrogate models are widely used as predictor models in uncertainty quantification because they provide predictive means and covariances that are analytically tractable and rely only on matrix products and inversions. Suppose we are given a training set of $n \ge 1$ input–output pairs $(\mathbf{X}, \mathbf{Z})$ with $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ a matrix of size $n \times p$ and $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ a matrix of size $n \times d$. The $\mathbf{z}_i$ are given by calls to the forward model. For any $\mathbf{x} \in \mathcal{X}$, we write the posterior predictive distribution of the GP model as a d-dimensional random vector whose distribution is

$$\mathbf{z}^{(n)}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{m}_n(\mathbf{x}), \mathbf{K}_n(\mathbf{x})\right) \tag{2}$$

where

$$\begin{aligned}\mathbf{m}_n(\mathbf{x}) &= \mathbf{K}(\mathbf{x}, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{Z}\\ \mathbf{K}_n(\mathbf{x}) &= \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{x}, \mathbf{X})^T.\end{aligned}$$

Here $\mathbf{K}(\mathbf{x}, \mathbf{X})$ is a block matrix of size $d \times nd$ with blocks $\mathbf{K}(\mathbf{x}, \mathbf{x}_i)$ for $1 \le i \le n$ and we recall that $\mathbf{K}(\mathbf{X})$ is a block matrix of size $nd \times nd$ with blocks $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ for $1 \le i, j \le n$. We also introduce the matrix $\mathbf{K}_n(\mathbf{x}, \mathbf{x}')$ of size $d \times d$. It encodes the covariances between outputs at $\mathbf{x}$ and $\mathbf{x}'$:

$$\mathbf{K}_n(\mathbf{x}, \mathbf{x}') = \mathbf{K}(\mathbf{x}, \mathbf{x}') - \mathbf{K}(\mathbf{x}, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{x}', \mathbf{X})^T. \tag{3}$$

The hyperparameters of the GP model are the latent covariance kernels' hyperparameters and the linear combination coefficients in the LMC model. For stochastic computer codes, we may add one hyperparameter per output dimension, which is known as the nugget noise level. The hyperparameters are tuned by maximization of the log-marginal likelihood $\log p(\mathbf{Z}|\mathbf{X})$ with gradient-descent type methods (Byrd et al. 1995). The log-marginal likelihood is given by:

$$\begin{aligned}\log p(\mathbf{Z}|\mathbf{X}) = &-\frac{nd}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}(\mathbf{X}, \mathbf{X})|\\ &- \frac{1}{2}\mathbf{Z}^T\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{Z}\end{aligned} \tag{4}$$

where $|\mathbf{K}(\mathbf{X}, \mathbf{X})|$ is the determinant of the matrix $\mathbf{K}(\mathbf{X}, \mathbf{X})$.

2.3. Surrogate Models in Bayesian Inverse Problems

How are GP surrogates used in practice in Bayesian inverse problems? Starting from the consideration that the forward model $\mathbf{f}$ is too costly, one can build a GP surrogate with some training data $(\mathbf{X}, \mathbf{Z})$ obtained by calls to the function $\mathbf{f}$. We wish to obtain a posterior distribution $p(\cdot\,|\mathbf{Y})$ given some noisy observations $\mathbf{Y}$ defined in (1). The objective is then to include the predictive distribution of the GP model in the likelihood of the inverse problem.

The uncertainties in the inverse problem resolution can be split between the aleatoric uncertainty, linked to the noise of the observations $\mathbf{Y}$, and the epistemic uncertainty, which is the uncertainty of the surrogate model itself (Lartaud, Humbert, and Garnier 2023). The two sources of uncertainty are independent from one another, thus, one can express the observations as noisy predictions of the GP predictive mean, with two independent sources of errors:

$$\mathbf{y}_k = \mathbf{m}_n(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{x}) + \boldsymbol{\varepsilon}_k$$

where $\boldsymbol{\eta}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}_n(\mathbf{x})\right)$ and $\boldsymbol{\varepsilon}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}_{\text{obs}}\right)$ are d-dimensional random vectors representing respectively the epistemic and aleatoric uncertainty terms.

This formulation expresses the idea that the observations are independent with respect to the observation noise, but are all linked together by the surrogate model error. The inverse problem can thus be solved by computing a posterior distribution $p_n(\cdot\,|\mathbf{Y})$ with a new likelihood that accounts for both sources of uncertainties:

$$\begin{aligned}p_n(\mathbf{x}|\mathbf{Y}) \propto L_n(\mathbf{Y}|\mathbf{x})p(\mathbf{x}) &\propto |\mathbf{C}_n(\mathbf{x})|^{-1/2}\\ &\times \exp\left(-\frac{1}{2}\|\overline{\mathbf{y}} - \mathbf{m}_n(\mathbf{x})\|^2_{\mathbf{C}_n(\mathbf{x})}\right)p(\mathbf{x})\end{aligned} \tag{5}$$

with $\mathbf{C}_n(\mathbf{x}) = \mathbf{K}_n(\mathbf{x}) + \frac{1}{N}\mathbf{C}_{\text{obs}}$ an effective covariance matrix of size $d \times d$ and $\overline{\mathbf{y}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{y}_k \in \mathbb{R}^d$. This posterior distribution quantifies both the uncertainties from the observations and the uncertainties of the surrogate model. It can be sampled by standard MCMC techniques.

2.4. Stepwise Uncertainty Reduction

The Stepwise Uncertainty Reduction (SUR) paradigm was introduced in Bect et al. (2012) and Villemonteix, Vazquez, and Walter (2009). It serves as a general-purpose theoretical framework for the design of computer experiments. SUR strategies seek to iteratively select new design points that minimize a given metric of uncertainty. The main advantage of SUR is to

provide a theoretical foundation for many sequential design strategies. For instance, the SUR paradigm is applicable in Bayesian optimization with the efficient global optimization algorithm (EGO) (Jones, Schonlau, and Welch 1998; Snoek, Larochelle, and Adams 2012) or in reliability analysis (Chevalier et al. 2014; Bect, Li, and Vazquez 2017). Before diving into our contribution, we provide a brief introduction to the SUR framework in this section. In everything that follows, we will consider Gaussian processes with continuous sample paths. A comprehensive description of the SUR paradigm is provided in Section 1 of the supplementary material. It highlights the formal definition of SUR designs with random Gaussian measures and recalls the main convergence theorem stated in Bect, Bachoc, and Ginsbourger (2019), which is based on a supermartingale property for the uncertainty function. For the sake of clarity, this section describes only the essential information. The interested reader may refer to this supplementary material for further details.

SUR strategies are sequential design strategies constructed to minimize the expectation of a quantity of interest, known as an uncertainty functional H_n, to select a new design point $\mathbf{x}_{n+1} \in \mathcal{X}$. The quantity H_n depends on the training inputs $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and outputs $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ up to iteration n. At iteration $n+1$, $H_{n+1}(\mathbf{x}_{n+1}, \mathbf{z}_{n+1})$ depends on the new training location $\mathbf{x}_{n+1}$ and the yet unknown output $\mathbf{z}_{n+1}$.

SUR strategies are obtained by minimizing the expectation of the quantity $H_{n+1}(\mathbf{x}_{n+1}, \mathbf{z}_{n+1})$ where $\mathbf{z}_{n+1} \sim \mathbf{z}^{(n)}(\mathbf{x}_{n+1})$. In other words, the unknown output is assumed to follow the predictive distribution of the GP model at iteration n.

Definition 2.1. A SUR sequential design for the uncertainty functional H_n is defined as a sequential design such that for all $n \geq n_0$ with n_0 a given integer:

$$\mathbf{x}_{n+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\left[H_{n+1}(\mathbf{x}, \mathbf{z}^{(n)}(\mathbf{x}))\right]$$

where the expectation is taken with respect to the distribution of $\mathbf{z}^{(n)}(\mathbf{x})$, that is the conditional distribution of $\mathbf{z}(\mathbf{x})$ given $\{\mathbf{z}(\mathbf{x}_j) = \mathbf{z}_j, 1 \leq j \leq n\}$, which is given in (2).

The definition of H_n depends on the task. For example, in the SUR framework, the integrated mean squared prediction error takes the form $H_n = \int_{\mathcal{X}} |\mathbf{K}_n(\mathbf{x}')| p(\mathbf{x}') d\mathbf{x}'$. More generally, Bect, Bachoc, and Ginsbourger (2019) showed the almost-sure convergence of the quantity H_n to zero, given some assumptions on H_n presented in Section 1 of the supplementary material.

3. Sequential Design for Bayesian Inverse Problems

Let us get back to the main problem at stake here, which is the Bayesian resolution of an ill-posed inverse problem. In this work, the following question is asked. Given a limited computational budget, how can new design points be added to train the surrogate model? In this section, a first strategy based on an MMSE design criterion is presented, and then a second approach based on the SUR paradigm for a well-chosen metric of uncertainty is derived. This new strategy, named IP-SUR (Inverse Problem SUR), is shown to be tractable for GP surrogates and is supported by a theoretical guarantee of almost sure convergence to zero for the metric of uncertainty.

3.1. An Introductory Example

In this section, we consider a scalar inverse problem to highlight some limitations of standard design strategies. As usual, consider a GP surrogate we wish to improve with a few model calls. Several design strategies can be carried out, such as MMSE designs or I-optimal designs (Antognini and Zagoraiou 2010). MMSE designs select a new training point where the predictive variance of the GP surrogate is maximal. On the other hand, the goal of I-optimal designs is to minimize the integrated mean squared prediction error (IMPSE) at the next iteration. However, these methods do not account for the posterior distribution and may choose points that lie far from the posterior distribution. Although they may improve the surrogate, such design points are not valuable in this context. If they are in regions of low posterior densities, the improvements to the surrogate model will not be reflected in the MCMC sampling.

To illustrate the previous remark, we consider a toy example in which the forward model is a parabolic function $f(x) = \frac{x^2}{2}$ for $x \in \mathbb{R}$. We act as if this forward model were a complex computer code and replace it with a surrogate model. Given one observation of the model $y = f(x_{\text{th}}) + \varepsilon$, with $x_{\text{th}} = 3$ and $\varepsilon \sim \mathcal{N}(0, 0.5^2)$ we evaluate the posterior distribution $p(x|y)$ using the surrogate model. Then, we add two new design points with various design strategies. The newly added design points are plotted in Figure 1, along with the posterior density obtained by MCMC and the GP prediction. Between the first and second iterations, the training point selected by the IP-SUR strategy is added to the model.

The strategies investigated in this illustrative example are MMSE, I-optimal, and the two strategies proposed in this article, named CSQ and IP-SUR. They are presented respectively in Sections 3.2 and 3.3.

In this illustrative example, the two proposed strategies select points for which both the posterior density and the model uncertainty are large. On the other hand, MMSE and I-optimal strategies do not account for the posterior density and thus select points that may not improve the inverse problem resolution, because these points are selected in regions of low posterior density.

3.2. Constraint Set Query

Our first idea is to adapt MMSE design strategies to Bayesian inverse problems. To improve the surrogate in regions of high posterior density, the next training point $\mathbf{x}_{n+1} \in \mathcal{X}$ can be chosen to maximize the determinant of the predictive covariance matrix on a well-chosen subset $\mathcal{B} \subset \mathcal{X}$:

$$\mathbf{x}_{n+1} \in \operatorname{argmax}_{\mathbf{x}\in\mathcal{B}} |\mathbf{K}_n(\mathbf{x})|. \tag{6}$$

How can we choose $\mathcal{B}$? One could try to force the newly chosen design points to have a posterior density close to the density of the maximum a posteriori (MAP) $\mathbf{x}_*^{(n)}$ defined by:

$$\mathbf{x}_*^{(n)} \in \operatorname{argmax}_{\mathbf{x}\in\mathcal{X}} p_n(\mathbf{x}|\mathbf{Y}).$$

Then for any $h \in \mathbb{R}_+$ one can define the subset $\mathcal{B}_n(h)$ such as

$$\mathcal{B}_n(h) = \left\{\mathbf{x} \in \mathcal{X} | \log p_n\left(\mathbf{x}_*^{(n)}|\mathbf{Y}\right) - \log p_n(\mathbf{x}|\mathbf{Y}) \leq h\right\}. \tag{7}$$

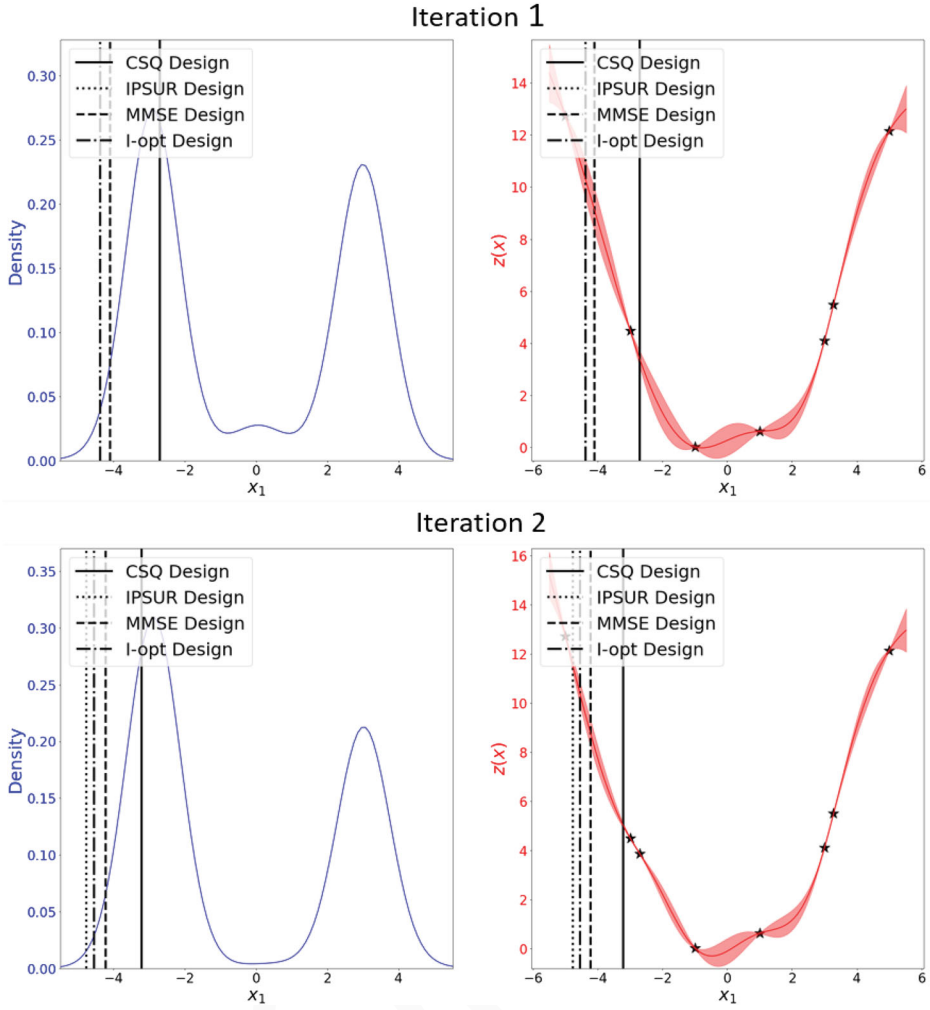

Figure 1. Example of two iterations of various design strategies. The GP mean and 95 % credible interval are plotted in red. The posterior density obtained by MCMC is plotted in blue. The vertical black lines represent the four points determined by the four different strategies. For the first iteration, the training points selected by IP-SUR and CSQ are very close, thus, the vertical lines overlap.

To get an intuition of the influence of h, let us look at the acceptance probability $\alpha_n(\mathbf{x}, \mathbf{x}')$ for a jump from point $\mathbf{x}$ to $\mathbf{x}'$ in the context of Metropolis-Hastings sampling with a symmetric proposal distribution:

$$\alpha_n(\mathbf{x}, \mathbf{x}') = \min\left\{1, \frac{p_n(\mathbf{x}'|\mathbf{Y})}{p_n(\mathbf{x}|\mathbf{Y})}\right\} = \min\left\{1, \frac{L_n(\mathbf{Y}|\mathbf{x}')p(\mathbf{x}')}{L_n(\mathbf{Y}|\mathbf{x})p(\mathbf{x})}\right\}.$$

Then $\mathcal{B}_n(h) = \left\{\mathbf{x} \in \mathcal{X} | \log \alpha_n(\mathbf{x}_*^{(n)}, \mathbf{x}) \geq -h\right\} = \left\{\mathbf{x} \in \mathcal{X} | \alpha_n(\mathbf{x}_*^{(n)}, \mathbf{x}) \geq e^{-h}\right\}$. Thus, e^{-h} for $h \in \mathbb{R}_+$ represents the lowest possible acceptance probability of a Metropolis-Hastings jump from the MAP to $\mathbf{x} \in \mathcal{B}_n(h)$. The CSQ method requires solving the optimization problem (6). This is done with the dual annealing algorithm (Xiang et al. 1997). The choice of a log-decreasing temperature scheme in the annealing algorithm guarantees that the Markov chain asymptotically reaches any neighborhood of the set of minima of the function to optimize (Bohachevsky, Johnson, and Stein 1986; Locatelli 2000).

This constraint set query (CSQ) method is a simple approach for active learning in the context of Bayesian inverse problems. It can be understood as an MMSE sequential design strategy restricted to a subset of the domain. We could also rely on maximizing the trace of the predictive covariance instead of the determinant. However, the strong correlations across outputs in our application point toward the use of the determinant. Using the trace would neglect those correlations. The CSQ method offers a simple framework for the refinement of a GP surrogate model. It can be easily implemented for a large variety of tasks, such as reliability analysis or model calibration. The simple criterion offers the same intuitive framework as the MMSE (or ALM) design. The MMSE design tends to select points on the boundary of the design space, which is not the case with the CSQ design. Additionally, the CSQ strategy accounts for correlations across output channels. The main downside of the CSQ method is the addition of the hyperparameter h. Despite being easily interpretable, no rigorous selection criterion for h is available. Finally, contrary to the majority of design strategies, our CSQ strategy is tailored to the given task, which is the resolution of an inverse problem in this article. Note that the same framework could be easily applied to a calibration task.

3.3. Inverse Problem SUR (IP-SUR) Strategy

Now, let us use the SUR framework to derive a second sequential design well-suited for Bayesian inverse problems. We introduce the quantity H_n defined by:

$$H_n = \int_{\mathcal{X}} |\mathbf{K}_n(\mathbf{x}')| p_n(\mathbf{x}'|\mathbf{Y}) d\mathbf{x}'. \tag{8}$$

This quantity is a weighted integrated Mean Squared Prediction Error (weighted IMSPE). The weight is the posterior density $p_n(\mathbf{x}'|\mathbf{Y})$, which focuses the attention on the region of interest for the inverse problem.

At iteration $n+1$, we add a new design point at $\mathbf{x}_{n+1}$. The quantity H_{n+1} then depends on $\mathbf{x}_{n+1}$ and the unknown output $\mathbf{z}_{n+1}$. We want to write a SUR design for this quantity. We would like to take a new design point $\mathbf{x}_{n+1}$ such that we minimize the expectation of H_{n+1}:

$$\mathbf{x}_{n+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\left[H_{n+1}(\mathbf{x}, \mathbf{z}^{(n)}(\mathbf{x}))\right]$$

However, building a sequential design with $(H_n)_{n\geq 0}$ leads to a technical difficulty. The SUR criterion is not tractable. The reason lies in the normalization constant κ_n of the posterior distribution defined by:

$$\kappa_n = \int_{\mathcal{X}} p(\mathbf{x}') L_n(\mathbf{Y}|\mathbf{x}') d\mathbf{x}'. \tag{9}$$

At iteration $n+1$, this normalization constant κ_{n+1} depends on $\mathbf{z}_{n+1}$ and the expectation of H_{n+1} cannot be easily evaluated. To overcome this difficulty, we introduce the sequence $(D_n)_{n\geq 0}$ defined by:

$$D_n = \kappa_n \times H_n. \tag{10}$$

We modify our objective, and now we seek to write a SUR design for this quantity. We will show that the expectation of $D_{n+1}(\mathbf{x}, \mathbf{z}^{(n)}(\mathbf{x}))$ is simpler to evaluate, and that we can obtain the convergence of $(H_n)_{n\geq 0}$ to zero with this design.

Consider a new training point $(\mathbf{x}, \mathbf{z})$. For $\mathbf{x}' \in \mathcal{X}$, the updated mean and covariance functions of the newly conditioned GP are:

$$\mathbf{m}_{n+1}(\mathbf{x}'|\mathbf{x}, \mathbf{z}) = \mathbf{m}_n(\mathbf{x}') + \mathbf{K}_n(\mathbf{x}', \mathbf{x})\mathbf{K}_n(\mathbf{x})^{-1}(\mathbf{z} - \mathbf{m}_n(\mathbf{x})) \tag{11}$$

$$\mathbf{K}_{n+1}(\mathbf{x}'|\mathbf{x}) = \mathbf{K}_n(\mathbf{x}') - \mathbf{K}_n(\mathbf{x}', \mathbf{x})\mathbf{K}_n(\mathbf{x})^{-1}\mathbf{K}_n(\mathbf{x}', \mathbf{x})^T \tag{12}$$

where $\mathbf{K}_n(\mathbf{x}, \mathbf{x}')$ is defined in (3). When adding a new training point $(\mathbf{x}, \mathbf{z})$, the quantity of interest becomes:

$$D_{n+1}(\mathbf{x}, \mathbf{z}) = \int_{\mathcal{X}} |\mathbf{K}_{n+1}(\mathbf{x}'|\mathbf{x})| L_{n+1}(\mathbf{Y}|\mathbf{x}', \mathbf{x}, \mathbf{z}) p(\mathbf{x}') d\mathbf{x}'$$

where $L_{n+1}(\mathbf{Y}|\mathbf{x}', \mathbf{x}, \mathbf{z})$ is the updated likelihood obtained from (5) by replacing $\mathbf{m}_n(\mathbf{x}')$ and $\mathbf{K}_n(\mathbf{x}')$ by $\mathbf{m}_{n+1}(\mathbf{x}'|\mathbf{x}, \mathbf{z})$ and $\mathbf{K}_{n+1}(\mathbf{x}'|\mathbf{x})$ using (11) and (12).

Following a SUR design, we seek to minimize the expectation of $D_{n+1}(\mathbf{x}, \mathbf{z}^{(n)}(\mathbf{x}))$. To that end, we introduce

$$F_n(\mathbf{x}) = \mathbb{E}\left[D_{n+1}(\mathbf{x}, \mathbf{z}^{(n)}(\mathbf{x}))\right].$$

We recall that $\mathbf{z}^{(n)}(\mathbf{x}) \sim \mathcal{N}(\mathbf{m}_n(\mathbf{x}), \mathbf{K}_n(\mathbf{x}))$. The SUR strategy associated to the sequence $(D_n)_{n\geq 0}$ is thus given by:

$$\mathbf{x}_{n+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} F_n(\mathbf{x}). \tag{13}$$

This criterion defines the IP-SUR strategy. We can claim the IP-SUR strategy is interesting if we can solve the following two problems. First, we need to make sure the quantity $F_n(\mathbf{x})$ can be evaluated (up to a multiplicative constant) to solve (13). Otherwise, a suitable approximation should be used as is done in Zhang, Lin, and Ranjan (2019) for an expected improvement criterion, for example. Second, we need to show the convergence of $(H_n)_{n\geq 0}$ for that particular design. This second goal is interesting as the quantity H_n is the posterior-weighted predictive covariance determinant. If this quantity goes to zero, the surrogate model uncertainty on the posterior distribution goes to zero. This means the predictive uncertainties tend to zero in the region of interest for our inverse problem resolution.

We start by computing $F_n(\mathbf{x})$. Since in our work, we have access to an ergodic Markov chain for the posterior distribution, we want to write $F_n(\mathbf{x})$ as an expectation with respect to this posterior and use the ergodicity of the chain to evaluate the expectation.

Proposition 3.1. The quantity $F_n(\mathbf{x})$ is given for $\mathbf{x} \in \mathcal{X}$ by:

$$\begin{aligned} F_n(\mathbf{x}) &= \kappa_n \times \int_{\mathcal{X}} |\mathbf{K}_{n+1}(\mathbf{x}'|\mathbf{x})| p_n(\mathbf{x}'|\mathbf{Y}) d\mathbf{x}' \\ &= \int_{\mathcal{X}} |\mathbf{K}_{n+1}(\mathbf{x}'|\mathbf{x})| L_n(\mathbf{Y}|\mathbf{x}') p(\mathbf{x}') d\mathbf{x}'. \end{aligned} \tag{14}$$

Proof. The proof is given in Section 2 of the supplementary material. □

The SUR strategy is unchanged when multiplying by a constant (i.e., a quantity independent of $\mathbf{x}$). The criterion (13) is thus equivalent to the criterion:

$$\mathbf{x}_{n+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} \frac{F_n(\mathbf{x})}{\kappa_n}.$$

With (14), we show that this new criterion boils down to an optimization problem where the target function $\frac{F_n(\mathbf{x})}{\kappa_n}$ can be evaluated by:

$$\frac{F_n(\mathbf{x})}{\kappa_n} \simeq \frac{1}{L}\sum_{l=1}^{L} |\mathbf{K}_{n+1}(\tilde{\mathbf{x}}_l^{(n)}|\mathbf{x})|$$

where $(\tilde{\mathbf{x}}_l^{(n)})_{1\leq l\leq L}$ is an ergodic Markov chain whose invariant distribution is the posterior $p_n(\cdot|\mathbf{Y})$. This Markov chain can be obtained by standard MCMC sampling methods. Once again, the dual annealing algorithm is used to solve the optimization problems.

Therefore, we showed that the SUR criterion is tractable. Then, let us investigate the convergence of $(H_n)_{n\geq 0}$. Theorem 3.12 in Bect, Bachoc, and Ginsbourger (2019) is our main tool to prove the convergence. This theorem is presented in the supplementary material.

Theorem 3.1. Consider the uncertainty functional H_n defined in (8) and a SUR sequential design $(\mathbf{x}_n)_{n\geq 1}$ given by the strategy (13), associated to the functional D_n. Then, the sequence $(H_n)_{n\geq 0}$ converges almost surely to 0:

$$H_n \xrightarrow[n\to+\infty]{a.s.} 0.$$

Proof. The proof is given in Section 2 of the supplementary material. □

Remark. Theorem 3.12 from Bect, Bachoc, and Ginsbourger (2019) can be extended to quasi-SUR sequential design, as stated by Bect, Bachoc, and Ginsbourger (2019). $(\mathbf{x}_n)_{n\geq 1}$ is a quasi-SUR sequential design if there exists a sequence $(\varepsilon_n)_{n\geq 1}$ of non-negative real numbers such that $\varepsilon_n \xrightarrow{n\to+\infty} 0$, and if there exists $n_0 \in \mathbb{N}$ such that $(\mathbf{x}_n)_{n\geq 1}$ verifies $F_n(\mathbf{x}_{n+1}) \leq \inf_{\mathbf{x}\in\mathcal{X}} F_n(\mathbf{x}) + \varepsilon_n$ for all $n \geq n_0$. Thus, the almost sure convergence proven in Theorem 3.1 is also valid for quasi-SUR designs.

This remark is crucial for numerical applications since the annealing process in finite time for the optimization step reaches only a neighborhood of the global optimum. The convergence theorem for the quasi-SUR designs is more flexible in that regard and provides more robustness for numerical applications.

Equation (14) and Theorem 3.1 are the foundation for the IP-SUR method. The IP-SUR design thus provides an interpretable design strategy to fine-tune surrogate models in the form of the likelihood-weighted criterion in (13). It is a goal-specific strategy since we provide design points in regions of high posterior density, within the resolution of an inverse problem. We make sure the new design points are relevant by merging the adaptive design with the inverse problem resolution, at the cost of intermediate MCMC sampling. The IP-SUR design offers an exploitation/exploration tradeoff that is not dependent on additional hyperparameters Singh, Deschrijver, and Dhaene (2013) or arbitrary rules (Lin et al. 2004; Crombecq et al. 2011). This design is valid for any GP surrogate models and does not need stationarity, contrary to Beck and Guillas (2016), and could be extended to nonstationary GP models (Gramacy and Lee 2008). The convergence guarantee is a supplementary result hinting at good asymptotic properties. This last part is important, considering we seek to fine-tune surrogate models. Even with a large density of training points, the inherent exploration/exploitation tradeoff of IP-SUR guarantees improvements to the surrogate model posterior-weighted MSE.

4. Application

4.1. Set-Up and Test Cases

The IP-SUR strategy presented in this article is applied to various test cases for different shapes of posterior distributions and is then compared to the CSQ approach described in Section 3.2, to MMSE and I-optimal designs, and to a state-of-the-art sequential design based on Bayes risk minimization. These applications rely on multi-output Gaussian process surrogate models. In the first two test cases, we are considering simple test cases in two dimensions. In the last test case, the strategies are applied to a more difficult example taken from neutron noise analysis, a technique whose goal is to identify fissile material from measurements of temporal correlations between neutrons.

In everything that follows, the forward model is always considered too costly to be called directly and is replaced by a GP surrogate. The observations of the forward model are noisy with a known noise covariance $\mathbf{C}_{\text{obs}}$. The multi-output Gaussian process surrogate model is based on the Linear Model of Coregionalization described in Bonilla, Chai, and Williams (2007). To build this surrogate model, a training dataset of $n_0 = 10$ input–output pairs $(\mathbf{X}, \mathbf{Z})$ is accessible. The surrogate model predictive distribution obtained is denoted $\mathbf{z}^{(0)}$. For the third case, which is in a higher dimension, we start with $n_0 = 20$. These initial training sizes are deliberately low compared to the standard rule of thumb of $n_0 \simeq 10\times d$ in order to better highlight the improvements of the surrogate models involved. This initial surrogate model is our starting point from which a sequential design is built. At each iteration, a new design point is acquired with the tested strategies, and the GP predictive distribution is updated to $\mathbf{z}^{(n)}$. The posterior distribution density obtained with the initial GP is denoted $p_0(\mathbf{x}'|\mathbf{Y})$, and after $n \geq 1$ iterations, the new posterior distribution is $p_n(\mathbf{x}'|\mathbf{Y})$. Similarly, $L_0(\mathbf{Y}|\mathbf{x}')$ and $L_n(\mathbf{Y}|\mathbf{x}')$ are the initial and updated likelihoods in the inverse problem.

At each iteration, a new Markov chain $(\tilde{\mathbf{x}}_l^{(n)})_{1\leq l\leq L}$ is generated with $L = 2\times 10^5$ with the Adaptive Metropolis algorithm of Haario, Saksman, and Tamminen (2001). The prior in the inverse problem is always uniform on the input domain $\mathcal{X}$. To compare the different approaches for sequential design, performance metrics are needed. The first obvious choice is the weighted IMPSE (WIMPSE) as introduced in (8). The iterative values H_n of the metric are evaluated for each iteration. The next metric considered is the differential entropy of the posterior distribution:

$$S_n = -\int_{\mathcal{X}} p_n(\mathbf{x}'|\mathbf{Y}) \log p_n(\mathbf{x}'|\mathbf{Y}) d\mathbf{x}'.$$

Because we have access to ergodic Markov chains $(\tilde{\mathbf{x}}_l^{(n)})_{1\leq l\leq L}$ and the prior is uniform, the differential entropy can be estimated by:

$$S_n \simeq \widehat{S_n} = -\frac{1}{L}\sum_{l=1}^{L} \log p_n(\tilde{\mathbf{x}}_l^{(n)}|\mathbf{Y})$$

where the posterior densities are obtained from kernel density estimation using the MCMC samples.

The other performance metric considered is the Kullback-Leibler (Kullback and Leibler 1951) divergence $\mathrm{KL}\left[p_n(\cdot|\mathbf{Y})\|p_\infty(\cdot|\mathbf{Y})\right]$ between a posterior $p_n(\cdot|\mathbf{Y})$ and a reference posterior distribution $p_\infty(\cdot|\mathbf{Y})$ obtained from a GP surrogate $\mathbf{z}^{(\infty)}$ trained with infinitely many data points. For the numerical applications, we do not have access to p_∞. We compare $p_n(\cdot|\mathbf{Y})$ with the posterior obtained with a well-trained GP $p_{1000}(\cdot|\mathbf{Y})$ instead. We thus define:

$$\begin{aligned} d_n &= \mathrm{KL}\left[p_n(\cdot|\mathbf{Y})\|p_{1000}(\cdot|\mathbf{Y})\right] \\ &= \int_{\mathcal{X}} p_n(\mathbf{x}'|\mathbf{Y}) \log\left(\frac{p_n(\mathbf{x}'|\mathbf{Y})}{p_{1000}(\mathbf{x}'|\mathbf{Y})}\right) d\mathbf{x}'. \end{aligned}$$

To estimate the KL, the ergodicity of the Markov chain is used once again, in combination with the kernel density estimates for the posterior densities:

$$d_n \simeq \widehat{d_n} = \frac{1}{L}\sum_{l=1}^{L} \log\left(\frac{p_n(\tilde{\mathbf{x}}_l^{(n)}|\mathbf{Y})}{p_{1000}(\tilde{\mathbf{x}}_l^{(n)}|\mathbf{Y})}\right).$$

To avoid introducing bias, the samples $\tilde{\mathbf{x}}_l^{(n)}$ used to evaluate performance metrics should be independent of those used in the design strategies. However, using the same samples did not lead to any change in the final results.

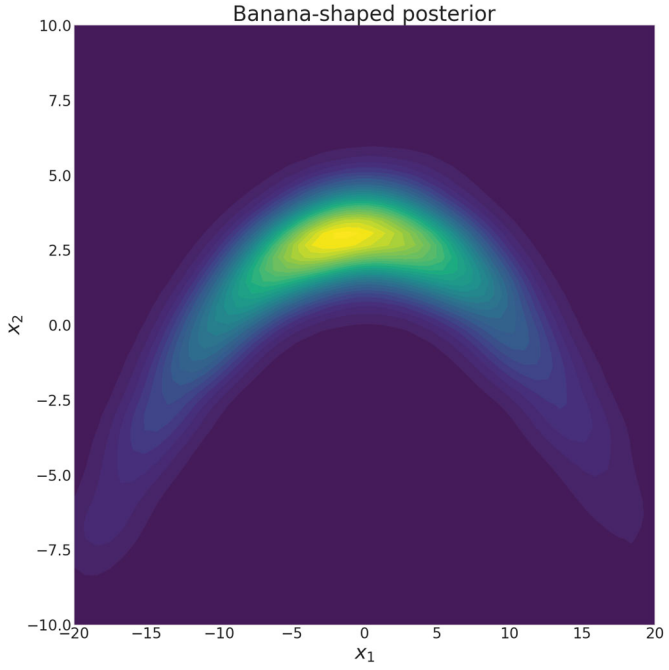

Figure 2. Banana-shaped target posterior distribution.

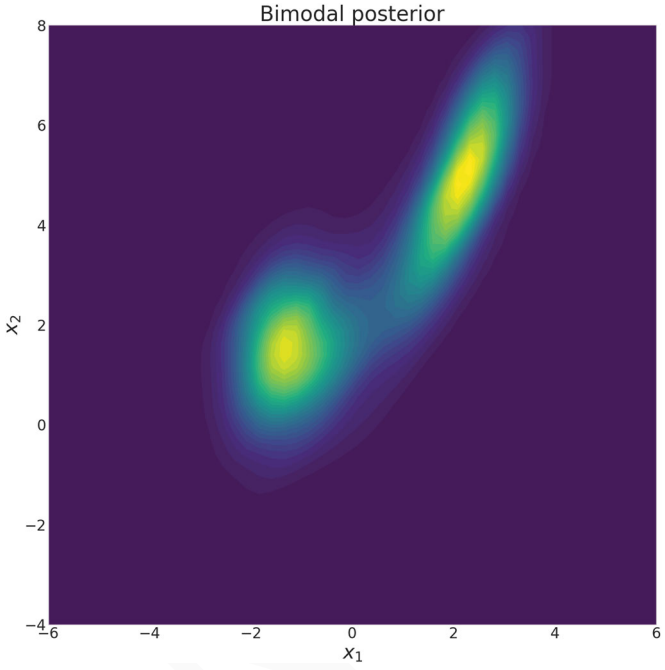

Figure 3. Bimodal target posterior distribution.

4.1.1. Banana Posterior Distribution

In this first test case, the target posterior distribution has a banana shape as displayed in Figure 2.

This posterior distribution is similar to the one introduced in Sürer, Plumlee, and Wild (2024) and is described by the following analytical forward model:

$$\mathbf{f}_{\mathrm{b}} \colon \mathcal{X}_{\mathrm{b}} \longrightarrow \mathbb{R}^2$$
$$\mathbf{x} = (x_1, x_2) \longmapsto (x_1, x_2 + 0.03x_1^2)$$

where $\mathcal{X}_{\mathrm{b}} = [-20, 20] \times [-10, 10] \subset \mathbb{R}^2$. For a single observation $\mathbf{y} = (y_1, y_2)$, the posterior has the density:

$$p_{\mathrm{b}}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\frac{(x_1 - y_1)^2}{100} - \frac{1}{2}\left(x_2 + 0.03x_1^2 - y_2\right)^2\right).$$

The observations $\mathbf{Y} = (\mathbf{y}^{(k)})_{1 \leq k \leq N}$ are generated with $N = 5$ and such that for $1 \leq k \leq N$ we have $\mathbf{y}^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_{\mathrm{obs}})$ with $\boldsymbol{\mu} = (0, 3)$ and:

$$\mathbf{C}_{\mathrm{obs}} = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}.$$

4.1.2. Bimodal Posterior Distribution

In this second test case, the target posterior is bimodal as plotted in Figure 3. The corresponding forward model is $\mathbf{f}_{\mathrm{m}}$ defined by:

$$\mathbf{f}_{\mathrm{m}} \colon \mathcal{X}_{\mathrm{m}} \longrightarrow \mathbb{R}^2$$
$$\mathbf{x} = (x_1, x_2) \longmapsto (x_2 - x_1^2, x_2 - x_1)$$

where $\mathcal{X}_{\mathrm{m}} = [-6, 6] \times [-4, 8] \subset \mathbb{R}^2$. For a single observation $\mathbf{y} = (y_1, y_2)$, the posterior has the density:

$$p_{\mathrm{m}}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{\sqrt{0.2}}{10}(x_2 - x_1^2 - y_1)^2 - \frac{\sqrt{0.75}}{10}\left(x_2 + 0.03x_1^2 - y_2\right)^2\right).$$

The observations $\mathbf{Y} = (\mathbf{y}^{(k)})_{1 \leq k \leq N}$ are generated with $N = 10$ and such that for $1 \leq k \leq N$ we have $\mathbf{y}^{(k)} \sim \mathcal{N}(\mu, \mathbf{C}_{\mathrm{obs}})$ with $\boldsymbol{\mu} = (0, 2)$ and:

$$\mathbf{C}_{\mathrm{obs}} = \begin{pmatrix} \frac{5}{\sqrt{0.2}} & 0 \\ 0 & \frac{5}{\sqrt{0.75}} \end{pmatrix}.$$

4.1.3. A Practical Application to Neutron Noise Analysis

Now let us consider an example taken from a practical problem in neutron noise analysis (Pázsit and Pál 2007). In this application, the forward model is the analytical approximation of a more complex forward model, which involves a full 3D neutron transport model.

Let us consider the following problem. Let $\mathbf{f}_{\mathrm{p}} : \mathcal{X}_{\mathrm{p}} \mapsto \mathbb{R}^3$ be the analytical forward model defined in Section 4 of the supplementary material, with $\mathcal{X}_{\mathrm{p}} \subset \mathbb{R}^4$. Let $\mathbf{Y} = (\mathbf{y}_k)_{1 \leq k \leq N}$ be the $N = 20$ noisy observations of the forward model:

$$\mathbf{y}_k = \mathbf{f}_{\mathrm{p}}(\mathbf{x}_{\mathrm{th}}) + \boldsymbol{\varepsilon}_k \text{ with } \boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathrm{obs}})$$

where $\mathbf{x}_{\mathrm{th}} \in \mathcal{X}_{\mathrm{p}}$ is the true value of the inputs. The domain of interest for the input parameters is

$$\mathcal{X}_{\mathrm{p}} = [0.7, 0.9] \times [0.01, 0.10] \times [1 \times 10^5, 2 \times 10^5] \times [0.1, 0.9].$$

Some two-dimensional marginals of the posterior distribution p_{1000} obtained with a well-trained GP are shown in Figure 4.

This posterior distribution is more difficult to sample as it is concentrated close to a one-dimensional manifold subspace in the four-dimensional parameter space. The decorrelation time τ is significantly larger, as can be seen in Section 3 of the supplementary material, which presents MCMC diagnostics. Thus, the number of MCMC samples is increased to 4×10^5 iterations per run for that specific test case.

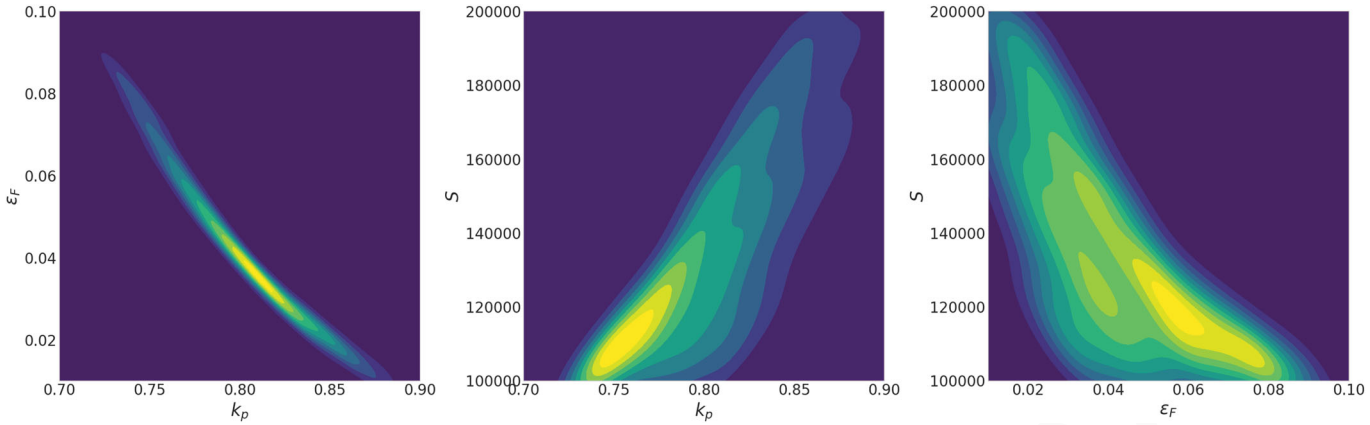

Figure 4. Two-dimensional marginal distributions from the target posterior distribution in the neutron noise analysis case study with $\mathbf{x} = (k_p, \varepsilon_F, S, x_s)$.

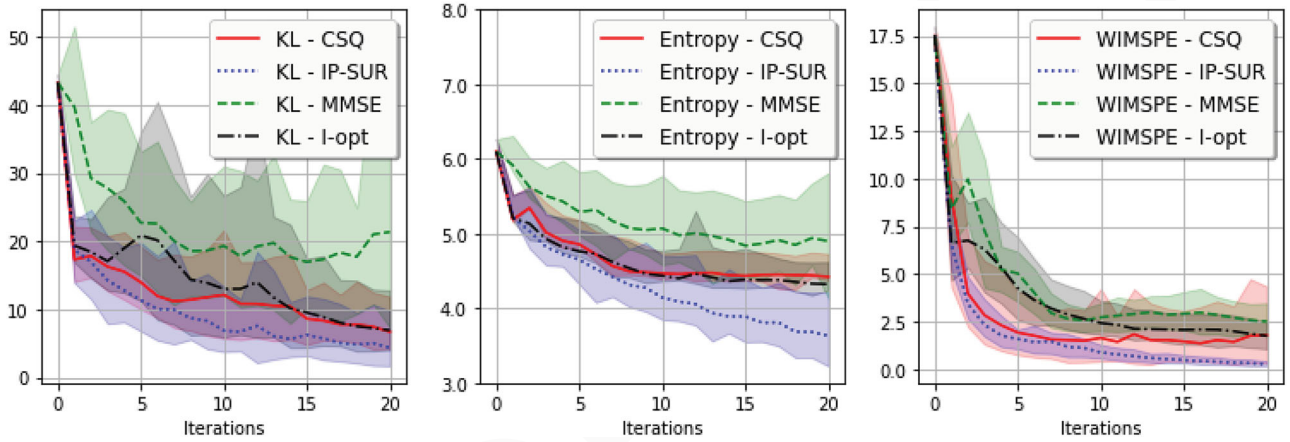

Figure 5. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR, MMSE, and I-optimal design strategies, replicated for 100 experiments—Banana test case.

Table 1. Average running times (in seconds) for one iteration of the various strategies, on each of the test cases.

	CSQ	IP-SUR	MMSE	I-optimal	Sinsbeck et al.
Banana case	16	235	3	136	180
Bimodal case	14	265	2	92	156
Neutronic case	18	1676	16	440	1235

4.2. Results

The IP-SUR sampling strategy and the CSQ method both require solving an optimization problem with a non-convex function. We recall that the optimization problems are solved with the dual annealing algorithm as implemented in *scipy*. The same is true for the other strategies. The IP-SUR strategy and the CSQ method are applied iteratively 20 times to produce new design points. For the CSQ strategy, the hyperparameter h introduced in (7) is set to $h = 3$. The influence of h is discussed afterward.

The evolutions of the performance metrics are plotted in Figures 5–7 for all tested strategies. The empirical 95% confidence interval for each metric and strategy is also displayed. The confidence intervals were obtained by running the numerical experiment 100 times in parallel.

Both the CSQ and SUR strategies perform better than the MMSE and I-optimal designs, which tend to target away from the posterior distribution. This increased performance is especially noticeable in the bimodal case and in the higher-dimensional situation of the neutronic test case.

Based on these results, one could argue that the CSQ strategy can be situationally better as it is easier to set up while providing similar performance in the end. However, two counter-arguments can be pointed out. First of all, the IP-SUR strategy does exhibit a guarantee for the convergence of the integrated variance, which offers a strong theoretical foundation. Besides, though the acquisition function in the IP-SUR strategy is more computationally intensive, the method does not rely on the prior setting of an arbitrary hyperparameter, while the CSQ design is based on the hyperparameter $h \in \mathbb{R}_+$ introduced in the definition of the bounding set $\mathcal{B}_n(h)$ in (7). This hyperparameter quantifies how far away from the MAP the optimization problem can search. The choice of this hyperparameter can impact quite drastically the performance of the sequential design. A lower value leads to a more constrained set and thus an easier optimization problem. However, the new design points are confined to a smaller region of the parameter space.

To investigate the influence of h, the CSQ strategy is used once again to provide 20 new design points for varying values of $h \in \{1, 2, 3\}$. The performance metrics obtained for the banana-shaped and bimodal posterior distributions

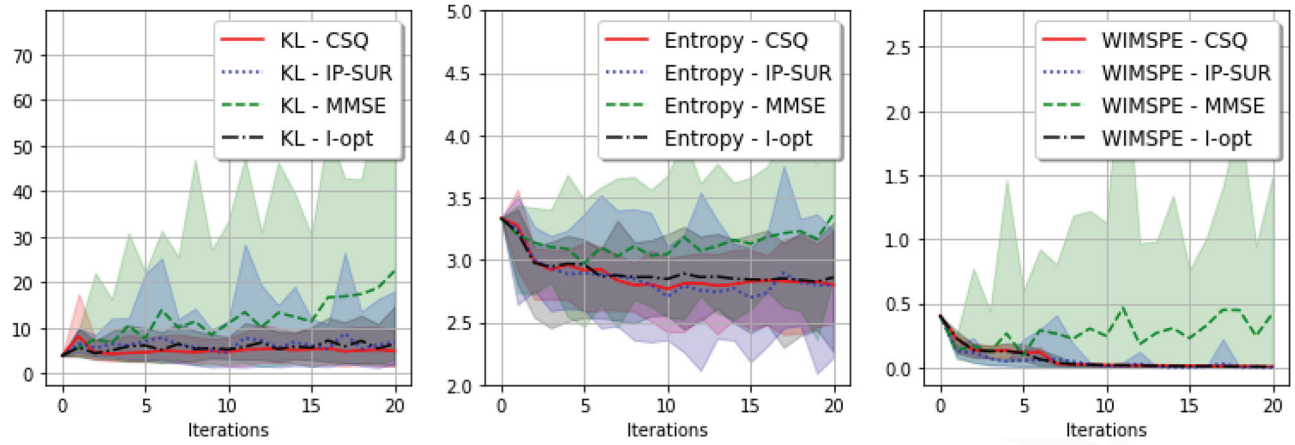

Figure 6. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR, MMSE, and I-optimal design strategies, replicated for 100 experiments—Bimodal test case.

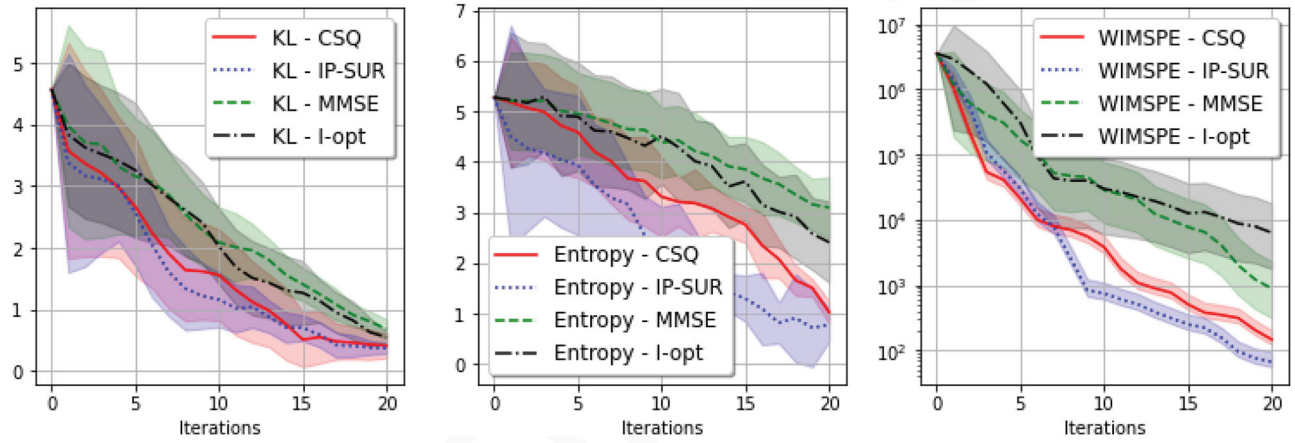

Figure 7. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR, MMSE, and I-optimal design strategies, replicated for 100 experiments—Neutronic test case.

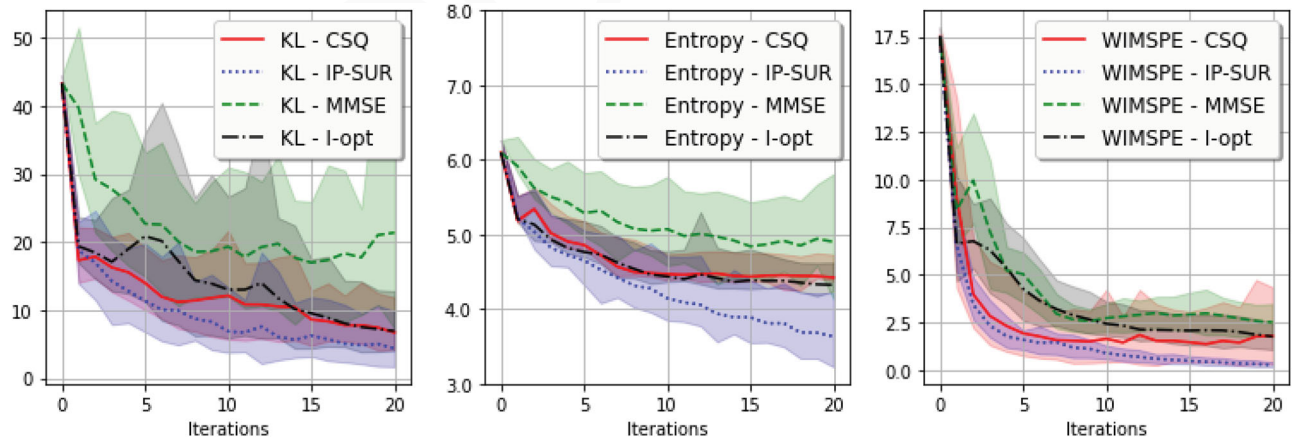

Figure 8. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ strategy for $h \in \{1, 2, 3\}$, replicated for 100 experiments—Banana test case.

are shown respectively in Figures 8 and 9. One can see that for $h = 1$, the design is significantly worse. We recall that all the previous test cases were conducted with $h = 3$, which provides the best results among the selected values. However, the optimal choice of h is likely dependent on the application and cannot be found easily. For this reason, the IP-SUR strategy presented in this work seems superior as it does not carry the burden of the selection of a hyperparameter.

Finally, let us compare the performance of our IP-SUR method with the design strategy introduced in Sinsbeck and Nowak (2017). This sequential design strategy is based on the minimization of the Bayes risk with respect to a loss function measuring the variance of the likelihood estimate with the

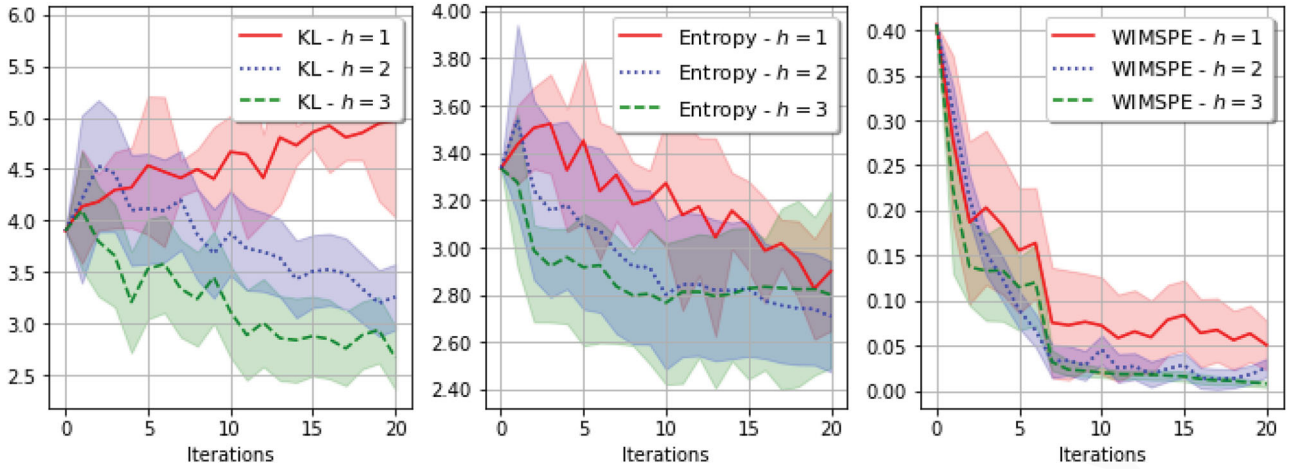

Figure 9. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ strategy for $h \in \{1, 2, 3\}$, replicated for 100 experiments—Bimodal test case.

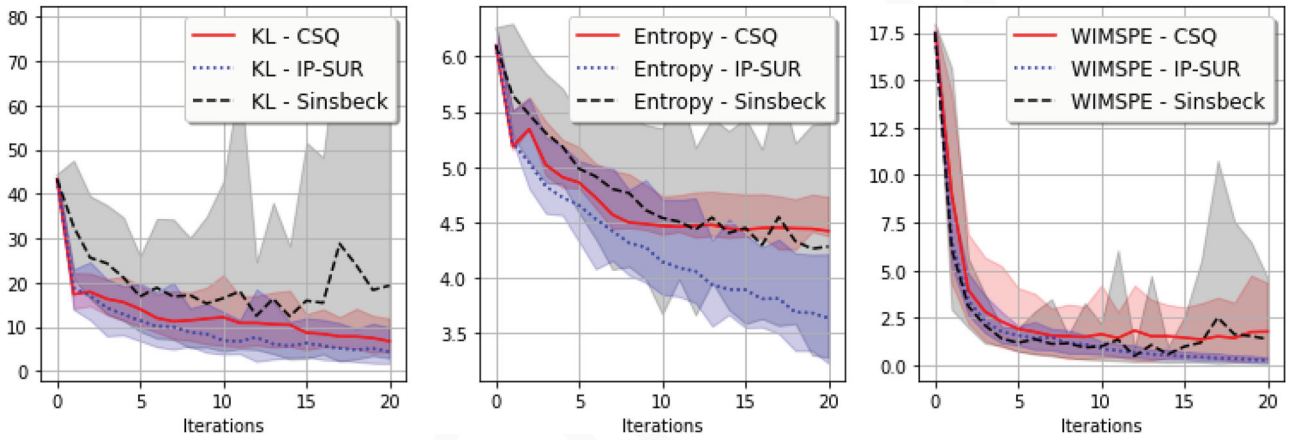

Figure 10. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR and Sinsbeck et al. strategy—Banana test case.

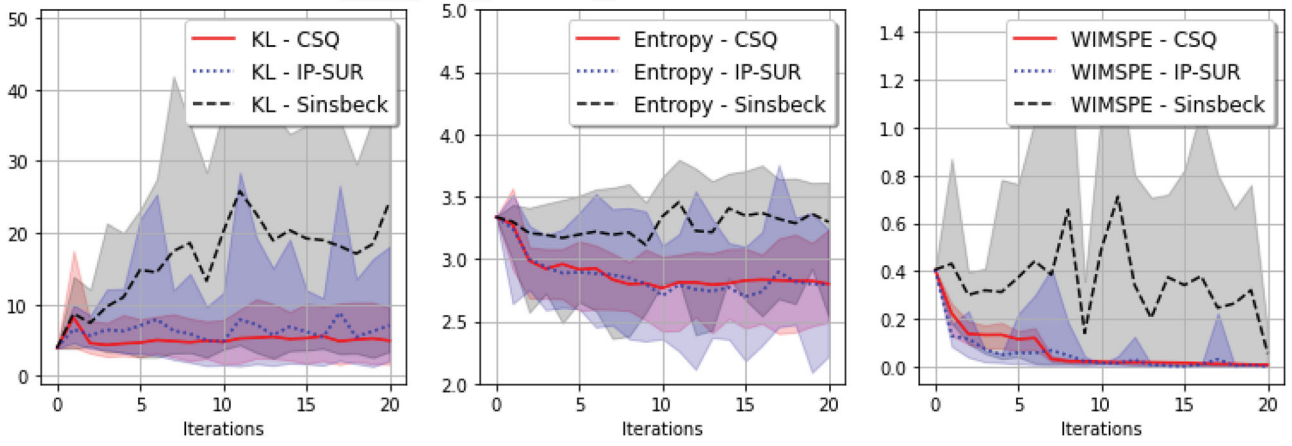

Figure 11. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR and Sinsbeck et al. strategy—Bimodal test case.

surrogate model. The same performance metrics are used for comparison.

The evolutions of the performance metrics are displayed in Figures 10–12 for each test case. The two methods offer overall similar performance with regard to the performance metrics investigated for the banana and neutronic cases. For the bimodal case, however, our methods seem more reliable. Besides, the IP-SUR is backed by a convergence guarantee, which is not the case for the method of Sinsbeck et al. For the latter, the supermartingale property seems unreachable.

5. Conclusion

This work presents two new sequential design strategies to build efficient Gaussian process surrogate models in Bayesian inverse problems. These strategies are especially important for

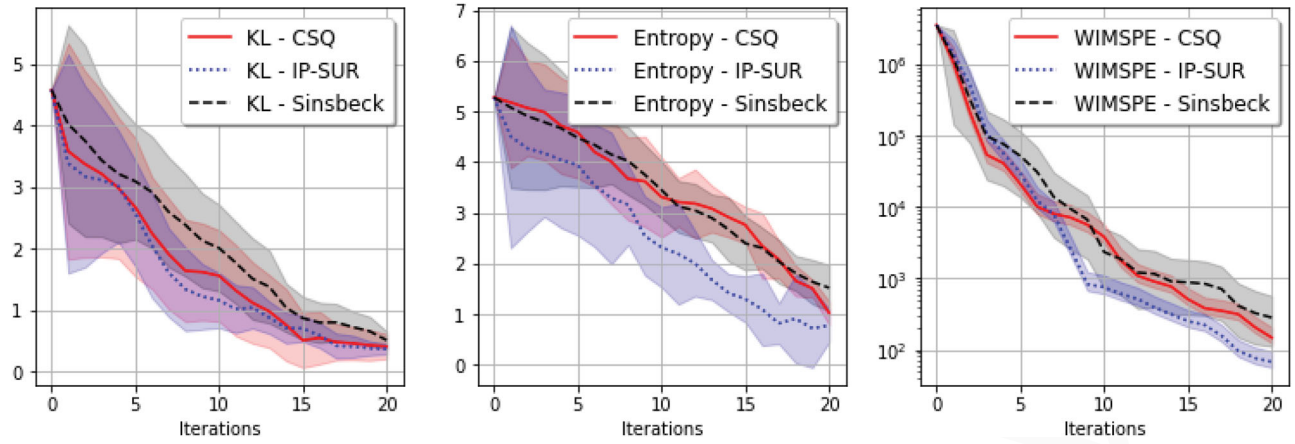

Figure 12. Mean evolution and 95% confidence intervals of the performance metrics over 20 iterations of the CSQ, IP-SUR and Sinsbeck et al. strategy—Neutronic test case.

cases where the posterior distribution in the inverse problem has thin support or is high-dimensional, in which case space-filling designs are not as competitive. The IP-SUR strategy introduced in this work is shown to be tractable and is supported by a theoretical guarantee of almost sure convergence of the weighted integrated mean square prediction error to zero. This method is compared to a simpler CSQ strategy, which is adapted from MMSE designs, and to a strategy based on the minimization of the Bayes risk with respect to the variance of the likelihood estimate. While both methods perform better than MMSE and I-optimal strategies, the IP-SUR method seems to provide better performance than CSQ for higher dimensions while not relying on the choice of a hyperparameter, all the while being grounded on strong theoretical foundations. It is also comparable to the Bayes risk minimization for all test cases and even superior for the bimodal test case. The latter strategy also does not display a convergence guarantee. The IP-SUR criterion developed in this work can be related to the traditional IMSPE criterion by a tempered variant of the design strategy (Neal 2001; Hu and Zidek 2002; Del Moral, Doucet, and Jasra 2006) in which we could introduce a tempering parameter $\beta \in [0, 1]$ and define a new design strategy based on the functional $H_{n,\beta} = \frac{1}{\kappa_{n,\beta}} \int_{\mathcal{X}} |\mathbf{K}_n(\mathbf{x}')| L_n(\mathbf{Y}|\mathbf{x}')^{\beta} p(\mathbf{x}') d\mathbf{x}'$ with $\kappa_{n,\beta} = \int_{\mathcal{X}} L_n(\mathbf{Y}|\mathbf{x}')^{\beta} p(\mathbf{x}') d\mathbf{x}'$. The IMSPE criterion can be obtained for $\beta = 0$ while the IP-SUR strategy is obtained for $\beta = 1$. The almost sure convergence can be verified with the same approach as for $(H_n)_{n\geq 0}$.

The extension of this design strategy is yet to be explored for other types of surrogate models. Gaussian process surrogate models have the advantage of keeping the SUR criterion tractable, though one could consider a variant of this method with approximation methods to evaluate the SUR criterion, with the risk of losing the guarantee of convergence. The robustness of such an approach would also have to be investigated.

Supplementary Materials

Supplementary file: this file contains an in-depth presentation of the SUR paradigm (Section 1), additional proofs (Section 2), MCMC diagnostics (Section 3), and an introduction to the neutronic forward model (Section 4).

Python files: the Python files to recreate the toy case example from Section 3.1.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Paul Lartaud http://orcid.org/0000-0001-6114-5762

References

Agrell, C., and Dahl, K. R. (2021), "Sequential Bayesian Optimal Experimental Design for Structural Reliability Analysis," *Statistics and Computing*, 31, 27. [1]

Antognini, A. B., and Zagoraiou, M. (2010), "Exact Optimal Designs for Computer Experiments via Kriging Metamodelling," *Journal of Statistical Planning and Inference*, 140, 2607–2617. [4]

Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2021), "Adaptive Design of Experiments for Conservative Estimation of Excursion Sets," *Technometrics*, 63, 13–26. [1]

Beck, J., and Guillas, S. (2016), "Sequential Design with Mutual Information for Computer Experiments (mice): Emulation of a Tsunami Model," *SIAM/ASA Journal on Uncertainty Quantification*, 4, 739–766. [7]

Bect, J., Bachoc, F., and Ginsbourger, D. (2019), "A Supermartingale Approach to Gaussian Process based Sequential Design of Experiments," *Bernoulli*, 25, 2883–2919. [4,6,7]

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012), "Sequential Design of Computer Experiments for the Estimation of a Probability of Failure," *Statistics and Computing*, 22, 773–793. [3]

Bect, J., Li, L., and Vazquez, E. (2017), "Bayesian Subset Simulation," *SIAM/ASA Journal on Uncertainty Quantification*, 5, 762–786. [4]

Bohachevsky, I. O., Johnson, M. E., and Stein, M. L. (1986), "Generalized Simulated Annealing for Function Optimization," *Technometrics*, 28, 209–217. [5]

Bonilla, E. V., Chai, K., and Williams, C. (2007), "Multi-Task Gaussian Process Prediction," in *Advances in Neural Information Processing Systems* (Vol. 20). [2,7]

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995), "A Limited memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, 16, 1190–1208. [3]

Chen, J., Kang, L., and Lin, G. (2021), "Gaussian Process Assisted Active Learning of Physical Laws," *Technometrics*, 63, 329–342. [1]

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014), "Fast Parallel Kriging-based Stepwise Uncertainty reduction with Application to the Identification of an Excursion Set," *Technometrics*, 56, 455–465. [4]

Cole, D. A., Gramacy, R. B., Warner, J. E., Bomarito, G. F., Leser, P. E., and Leser, W. P. (2023), "Entropy-based Adaptive Design for Contour Finding and Estimating Reliability," *Journal of Quality Technology*, 55, 43–60. [1]

Consonni, G., Fouskakis, D., Liseo, B., and Ntzoufras, I. (2018), "Prior Distributions for Objective Bayesian Analysis," *Bayesian Analysis*, 13, 627–679. [2]

Crombecq, K., Gorissen, D., Deschrijver, D., and Dhaene, T. (2011), "A Novel Hybrid Sequential Design strategy for Global Surrogate Modeling of Computer Experiments," *SIAM Journal on Scientific Computing*, 33, 1948–1974. [7]

Dashti, M., and Stuart, A. M. (2017), *The Bayesian Approach to Inverse Problems*, pp. 311–428, Cham: Springer. [2]

de Aguiar, P. F., Bourguignon, B., Khots, M., Massart, D., and Phan-Than-Luu, R. (1995), "D-Optimal Designs," *Chemometrics and Intelligent laboratory Systems*, 30, 199–210. [1]

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society*, Series B, 68, 411–436. [12]

Drovandi, C. C., Moores, M. T., and Boys, R. J. (2018), "Accelerating Pseudo-Marginal MCMC Using Gaussian Processes," *Computational Statistics & Data Analysis*, 118, 1–17. [1]

Du, X., and Chen, W. (2002), "Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 36223), pp. 871–880. [1]

Dubourg, V., Sudret, B., and Deheeger, F. (2013), "Metamodel-based Importance Sampling for Structural Reliability Analysis," *Probabilistic Engineering Mechanics*, 33, 47–57. [1]

Ezzat, A. A., Pourhabib, A., and Ding, Y. (2018), "Sequential Design for Functional Calibration of Computer Models," *Technometrics*, 60, 286–296. [1]

Frangos, M., Marzouk, Y., Willcox, K., and van Bloemen Waanders, B. (2010), "Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems," in *Large-Scale Inverse Problems and Quantification of Uncertainty*, eds. L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, pp. 123–149, Chichester: Wiley. [1]

Ghosh, M. (2011), "Objective Priors: An Introduction for Frequentists," *Statistical Science*, 26, 187–202. [2]

Gramacy, R. B. (2020), *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, New York: Chapman and Hall/CRC. [2]

Gramacy, R. B., and Lee, H. K. (2009), "Adaptive Design and Analysis of Supercomputer Experiments," *Technometrics*, 51, 130–145. [1,2]

Gramacy, R. B., and Lee, H. K. H. (2008), "Bayesian Treed Gaussian Process Models with an Application to Computer Modeling," *Journal of the American Statistical Association*, 103, 1119–1130. [7]

Haario, H., Saksman, E., and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, 223–242. [7]

Hu, F., and Zidek, J. V. (2002), "The Weighted Likelihood," *Canadian Journal of Statistics*, 30, 347–371. [12]

Imani, M., and Ghoreishi, S. F. (2020), "Bayesian Optimization Objective-based Experimental Design," in *2020 American Control conference (ACC)*, pp. 3405–3411, IEEE. [1]

Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186, 453–461. [2]

Jin, R., Chen, W., and Sudjianto, A. (2002), "On Sequential Sampling for Global Metamodeling in Engineering Design," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 36223), pp. 539–548. [2]

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131–148. [1]

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

Kaipio, J., and Somersalo, E. (2006), *Statistical and computational Inverse Problems* (Vol. 160), New York: Springer. [1]

Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society*, Series B, 63, 425–464. [1]

Kleijnen, J. P., and Beers, W. V. (2004), "Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodelling," *Journal of the Operational Research Society*, 55, 876–883. [1]

Kullback, S., and Leibler, R. A. (1951), "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86. [7]

Lartaud, P., Humbert, P., and Garnier, J. (2023), "Multi-Output Gaussian Processes for Inverse Uncertainty Quantification in Neutron Noise Analysis," *Nuclear Science and Engineering*, 197, 1928–1951. [3]

Lee, T. H., and Jung, J. J. (2008), "A Sampling Technique Enhancing Accuracy and Efficiency of Metamodel-based RBDO: Constraint Boundary Sampling," *Computers & Structures*, 86, 1463–1476. [1]

Li, J., and Marzouk, Y. M. (2014), "Adaptive Construction of Surrogates for the Bayesian Solution of Inverse Problems," *SIAM Journal on Scientific Computing*, 36, A1163–A1186. [2]

Li, Y., Wang, Y., and Yan, L. (2023), "Surrogate Modeling for Bayesian Inverse Problems based on Physics-Informed Neural Networks," *Journal of Computational Physics*, 475, 111841. [1]

Lin, Y., Mistree, F., Allen, J. K., Tsui, K.-L., and Chen, V. C. (2004), "A Sequential Exploratory Experimental Design Method: Development of Appropriate Empirical Models in Design," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 46946), pp. 1021–1035. [7]

Liu, H., Ong, Y.-S., and Cai, J. (2018), "A Survey of Adaptive Sampling for Global Metamodeling in Support of Simulation-based Complex Engineering Design," *Structural and Multidisciplinary Optimization*, 57, 393–416. [2]

Locatelli, M. (2000), "Simulated Annealing Algorithms for Continuous Global Optimization: Convergence Conditions," *Journal of Optimization Theory and Applications*, 104, 121–133. [5]

MacKay, D. J. (1992), "Information-based Objective Functions for Active Data Selection," *Neural Computation*, 4, 590–604. [2]

Mai, C., Konakli, K., and Sudret, B. (2017), "Seismic Fragility Curves for Structures Using Non-parametric Representations," *Frontiers of Structural and Civil Engineering*, 11, 169–186. [1]

Mitchell, T. J. (2000), "An Algorithm for the Construction of "D-optimal" Experimental Designs," *Technometrics*, 42, 48–54. [1]

Neal, R. M. (2001), "Annealed Importance Sampling," *Statistics and Computing*, 11, 125–139. [12]

Osio, I. G., and Amon, C. H. (1996), "An Engineering Design Methodology with Multistage Bayesian Surrogates and Optimal Sampling," *Research in Engineering Design*, 8, 189–206. [1]

Pázsit, I., and Pál, L. (2007), *Neutron Fluctuations: A Treatise on the Physics of Branching Processes*, Amsterdam: Elsevier. [8]

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., and Kim, N. (2010), "Adaptive Designs of Experiments for Accurate Approximation of Target Regions," *Journal of Mechanical Design*, 132, 071008. [2]

Ranjan, P., Bingham, D., and Michailidis, G. (2008), "Sequential Experiment Design for Contour Estimation from Complex Computer Codes," *Technometrics*, 50, 527–541. [1]

Sacks, J., Schiller, S. B., and Welch, W. J. (1989), "Designs for Computer Experiments," *Technometrics*, 31, 41–47. [1]

Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003), *The Design and Analysis of Computer Experiments* (Vol. 1), New York: Springer. [1]

Sauer, A., Gramacy, R. B., and Higdon, D. (2023), "Active Learning for Deep Gaussian Process Surrogates," *Technometrics*, 65, 4–18. [1]

Scales, J. A., and Tenorio, L. (2001), "Prior Information and Uncertainty in Inverse Problems," *Geophysics*, 66, 389–397. [2]

Schobi, R., Sudret, B., and Wiart, J. (2015), "Polynomial-Chaos-based Kriging," *International Journal for Uncertainty Quantification*, 5, 171–193. [3]

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015), "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE*, 104, 148–175. [1]

Shewry, M. C., and Wynn, H. P. (1987), "Maximum Entropy Sampling," *Journal of Applied Statistics*, 14, 165–170. [1]

Singh, P., Deschrijver, D., and Dhaene, T. (2013), "A Balanced Sequential Design Strategy for Global Surrogate Modeling," in *2013 Winter Simulations Conference (WSC)*, pp. 2172–2179, IEEE. [7]

Sinsbeck, M., Cooke, E., and Nowak, W. (2021), "Sequential Design of Computer Experiments for the Computation of Bayesian Model Evidence," *SIAM/ASA Journal on Uncertainty Quantification*, 9, 260–279. [2]

Sinsbeck, M., and Nowak, W. (2017), "Sequential Design of Computer Experiments for the Solution of Bayesian Inverse Problems," *SIAM/ASA Journal on Uncertainty Quantification*, 5, 640–664. [2,10]

Snoek, J., Larochelle, H., and Adams, R. P. (2012), "Practical Bayesian Optimization of Machine Learning Algorithms," in *Advances in Neural Information Processing Systems* (Vol. 25). [4]

Stein, M. (1987), "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, 29, 143–151. [1]

Stroh, R., Bect, J., Demeyer, S., Fischer, N., Marquis, D., and Vazquez, E. (2022), "Sequential Design of Multi-Fidelity Computer Experiments: Maximizing the Rate of Stepwise Uncertainty Reduction," *Technometrics*, 64, 199–209. [1]

Stuart, A. M. (2010), "Inverse Problems: A Bayesian Perspective," *Acta Numerica*, 19, 451–559. [1]

Sürer, Ö., Plumlee, M., and Wild, S. M. (2024), "Sequential Bayesian Experimental Design for Calibration of Expensive Simulation Models," *Technometrics*, 66, 157–171. [1,8]

Villemonteix, J., Vazquez, E., and Walter, E. (2009), "An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions," *Journal of Global Optimization*, 44, 509–534. [3]

Wan, H.-P., and Ren, W.-X. (2016), "Stochastic Model Updating Utilizing Bayesian Approach and Gaussian Process Model," *Mechanical Systems and Signal Processing*, 70, 245–268. [1]

Wang, H., and Li, J. (2018), "Adaptive Gaussian Process Approximation for Bayesian Inference with Expensive Likelihood Functions," *Neural Computation*, 30, 3072–3094. [1]

Wang, Y., Myers, R. H., Smith, E. P., and Ye, K. (2006), "D-Optimal Designs for Poisson Regression Models," *Journal of Statistical Planning and Inference*, 136, 2831–2845. [1]

Williams, C. K., and Rasmussen, C. E. (2006), *Gaussian Processes for Machine Learning* (Vol. 2), Cambridge, MA: MIT Press. [2]

Xiang, Y., Sun, D., Fan, W., and Gong, X. (1997), "Generalized Simulated Annealing Algorithm and its Application to the Thomson Model," *Physics Letters A*, 233, 216–220. [5]

Yan, L., and Zhou, T. (2019a), "Adaptive Multi-Fidelity Polynomial Chaos Approach to Bayesian Inference in Inverse Problems," *Journal of Computational Physics*, 381, 110–128. [1]

——— (2019b), "An Adaptive Surrogate Modeling based on Deep Neural Networks for Large-Scale Bayesian Inverse Problems," arXiv preprint arXiv:1911.08926. [1]

Zhang, R., Lin, C. D., and Ranjan, P. (2019), "A Sequential Design Approach for Calibrating Dynamic Computer Simulators," *SIAM/ASA Journal on Uncertainty Quantification*, 7, 1245–1274. [6]