Optimal Experiment Campaigns under Uncertainty Minimizing Bayes Risk ⋆

Benoît Chachuat [*] **Marco Sandrin** [*,**]
Constantinos C. Pantelides [*,**]

[*] *The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, UK*
[**] *Siemens Process Automation Software, London, UK*

Abstract: Applying model-based design of experiments to compute maximally-informative campaigns with multiple parallel runs is challenging. Herein, we develop a systematic framework for recasting an experiment design problem for model parameter precision as one of discrimination between multiple rival models with different uncertain parameter realizations. We use an algebraic upper bound on the Bayes Risk as information criterion and apply a search procedure that iterates between an effort-based optimization step followed by a gradient-based refinement step. Through the case study of a fed-batch reactor, we show that a Bayes Risk discrimination strategy can provide highly-informative experimental campaigns to improve parameter precision, while being computationally advantageous compared to conventional FIM-based design strategies and capable of handling structurally unidentifiable problems.

Keywords: Optimal experiment design, Model-based design of experiments, Experiment campaign, Uncertainty, Bayes risk

1. INTRODUCTION

Model-based design of experiments (MBDoE) for parameter precision provides a reliable approach to accelerating the development of predictive mechanistic models and has been deployed successfully across various chemical engineering applications (Franceschini and Macchietto, 2008). It can be cast as an optimization problem aiming to maximize the information content of an experimental campaign. A popular choice for expressing the information content in MBDoE considers the Fisher information matrix (FIM) (Fisher, 1971; Atkinson et al., 2007), as a means for decreasing the size of the joint confidence region of the (uncertain) model parameters via the Cramer-Rao lower bound. The high computational tractability of this classical frequentist approach opens the possibility for designing dynamic experiments for models described by differential equations (Espie and Macchietto, 1989; Bauer et al., 2000) and conducting designs sequentially and even online (Galvanin et al., 2009). Much progress has also been made in Bayesian experimental design (BED) in recent years (Rainforth et al., 2024), although applying BED to practically relevant (dynamic) models often remains computationally prohibitive.

⋆ This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955520 (Digitalgaesation). CCP and BC gratefully acknowledge funding by Eli Lilly & Company through the Pharmaceutical Systems Engineering Lab (PharmaSEL) and by the Engineering and Physical Sciences Research Council (EPSRC) as part of its Prosperity Partnership Programme under grant EP/T518207/1. Correspondence should be directed to Benoît Chachuat: b.chachuat@imperial.ac.uk

Nevertheless, using the FIM to estimate the information content of an experimental campaign is also subject to several caveats. First, being a matrix, the FIM may not be used directly as objective function. The use of summary statistics such as the FIM determinant (D-optimality) or trace inverse (A-optimality) may result in suboptimal designs. Many summary statistics are furthermore undefined when the FIM is singular, which either requires projecting the FIM onto a subspace where it is positive definite or applying a regularization (Shahmohammadi and McAuley, 2020). Second, the fact that the elements of the FIM are functions of any parameter that participates nonlinearly in the candidate model can severely undermine its practicality since the experimental campaign is designed to learn about these (unknown or uncertain) parameter values in the first place—the infamous chicken-and-egg problem. Where the FIM is computed using some point estimate for the parameters—usually a maximum likelihood estimate derived from existing data or a previous campaign—only the local properties of the model at that point estimate is exploited, which inevitably leads to suboptimal design decisions. Exploiting a prior distribution of the model parameters to robustify an experimental design, either through averaging the FIM (Pronzato and Walter, 1985) or computing a risk measure (Kusumo et al., 2022), provides a more effective approach. It has been shown that an average FIM-based design can even be cast under the general BED framework with a particular utility function based on the gain in Fisher information (Walker, 2016). However, solving for such robust designs can significantly impair their computational tractability.

The main focus of this article is on developing and assessing a new method, whereby the MBDoE problem for

parameter precision is recast as a discrimination problem. Specifically, we consider a discrete probability model characterized by a finite set of support points and their associated weights to describe the parameter realizations in the candidate model, which allows us to treat experiment design as a problem of classification between these rival model instances. Consistent with a Bayesian approach, the probability of model selection error—also known as *Bayes Risk*—provides a natural information criterion to minimize in this context. And while the Bayes Risk cannot be evaluated in closed form and is thus unsuitable as an optimization criterion in general, upper bounds on the Bayes Risk have been developed that can be computed completely algebraically without requiring multidimensional integration or stochastic analysis (Blackmore et al., 2008). Advantages of the Bayes Risk bound over FIM-based criteria include the fact that it does not become singular when certain model parameter combinations are non-identifiable from the available measurements and is faster to compute since it entails model output evaluations only, as opposed to output sensitivities for the FIM. Despite this, Bayes Risk has seldom been used as an information measure for experiment design purposes to date. Bayard and Neely (2017) recently made several strides in this direction for pharmacokinetic applications but a systematic optimization framework is still lacking.

The main contribution of this article, therefore, is a systematic framework for minimizing Bayes Risk for designing optimal experiment campaigns under uncertainty (Sec. 2). We extend a recent methodology (Sandrin et al., 2024a) that iterates between (i) the solution of an effort-based exact design subproblem operating on a discretized experimental design space (Sandrin et al., 2024b) and (ii) a conventional gradient-based subproblem to refine the experiments that were selected by the previous step out of the set of discrete candidates, in order to enable Bayes Risk in either steps (Sec. 3). We also propose a scenario-reduction approach for more tractable Bayes Risk bound computation with large number and/or wide domain of uncertain parameters. We compare the proposed approach with an average D-optimal approach on a case study involving a dynamic experiment design in a semi-batch reactor, both in terms of information content and computational efficiency (Sec. 4).

2. PROBLEM DEFINITION

Consider a system with n_x experimental controls $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^{n_x}$ and n_y measured responses $\boldsymbol{y} \in \mathcal{Y} \subset \mathbb{R}^{n_y}$,

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\theta}, \boldsymbol{x}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\theta} \subset \mathbb{R}^{n_\theta}$ are uncertain parameters in the mathematical model $\boldsymbol{\eta}$. For simplicity, assume the measurements to be independent and the measurement error $\boldsymbol{\epsilon} \in \mathbb{R}^{n_y}$ to have zero mean $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ with uncorrelated and homoscedastic covariance Σ_y. Although written in closed form in (1), the model $\boldsymbol{\eta}$ may also be defined implicitly via a set of algebraic and/or differential equations without the loss of generality.

We consider an experimental campaign comprising N_t experimental runs for the purpose of generating data for the estimation of the model parameters $\boldsymbol{\theta}$. Since such campaigns often consist of repeated runs with identical experimental controls (replicates), it is convenient to denote an experimental design ξ as

$$\xi \doteq \left\{ \begin{matrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_{N_c} \\ \omega_1 & \dots & \omega_{N_c} \end{matrix} \right\}, \tag{2}$$

where $0 < N_c \leq N_t$ is the number of distinct runs; and $\omega_i \in (0, N_t]$ is the number of replications—or *effort*—of the i-th experimental candidate with controls $\boldsymbol{x}_i$ for each $i = 1, \dots, N_c$, so that the ω_i's add up to N_t. The set $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_c}\}$ is called the *support* of the experimental design ξ, denoted by $\text{supp}(\xi)$.

Our focus herein is on designs having integral efforts, $\omega_i \in \mathbb{Z}_+$, also called *exact* designs. In contrast, *continuous* designs allow ω_i to vary continuously in the standard simplex $\sum_i \omega_i = N_t$ with $\omega_i > 0, \forall i$. Although computationally advantageous, the latter typically result in fractional values for the optimized efforts and require the application of an a posteriori rounding procedure to recover integral efforts. Such rounding can introduce large suboptimality, most noticeably in campaigns where the number N_t of experimental runs is small (Sandrin et al., 2024b).

An optimal design ξ^* is one that maximizes some scalar information criterion ϕ,

$$\xi^* \in \arg\max_{\xi} \phi(\xi). \tag{3}$$

Determining an optimal design entails simultaneously searching over all possible number of supports N_c, the experimental controls $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_c} \in \mathbb{R}^{n_x}$, and the corresponding efforts $\omega_1, \dots, \omega_{N_c} \in \{0, \dots, N_t\}$.

2.1 FIM-based Optimal Designs

In classical experiment design, the scalar information criterion ϕ in (3) is expressed as a function of the Fisher information matrix (FIM), $\mathbf{M} \in \mathbb{R}^{n_\theta \times n_\theta}$, which under the assumption of uncorrelated homoscedastic measurement errors in (1) is given by (Atkinson et al., 2007)

$$\mathbf{M}(\xi, \boldsymbol{\theta}) \doteq \sum_{i=1}^{N_c} \omega_i \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}) \tag{4}$$

$$\text{with} \quad \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}) \doteq \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{\theta})^\intercal \, \Sigma_y^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{5}$$

The D-optimality criterion is a widely used summary statistic, aiming at minimizing the volume of confidence ellipsoids for the parameters,

$$\phi(\xi) \doteq \log\det\left(\mathbf{M}(\xi, \boldsymbol{\theta}^*)\right). \tag{6}$$

In principle, $\boldsymbol{\theta}^*$ in (6) should be the 'true' parameter value: this value is unknown and, in fact, the main reason for doing experiment design in the first place. Only for a linear model $\boldsymbol{\eta}$ in $\boldsymbol{\theta}$ does the dependence of ϕ on $\boldsymbol{\theta}$ vanish.

For a nonlinear model, using a nominal parameter value $\boldsymbol{\theta}_0$ is risk-inclined since it ignores the uncertainty in the model parameters. To mitigate the risk of uninformative experiments, one can adopt the Bayesian paradigm and describe the model parameter uncertainty with a probability distribution $\pi(\boldsymbol{\theta})$. The average design (AD) approach maximizes the expected value of information content over the model uncertainty (Pronzato and Walter, 1985),

$$\phi^{\text{AD}}(\xi) \doteq \int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta}) \log\det\left(\mathbf{M}(\xi, \boldsymbol{\theta})\right) d\boldsymbol{\theta}. \tag{7}$$

Other robust approaches are possible, such as focusing on those model uncertainty scenarios corresponding to a given lower percentile of the information content, for instance using a conditional-value-at-risk (Kusumo et al., 2022).

2.2 Bayes Risk Optimal Designs

In order to recast the experiment design problem for parameter precision as one of discrimination, we assume that the uncertain parameters $\boldsymbol{\theta}$ in the mathematical model $\boldsymbol{\eta}$ are in the discrete set $\Theta \doteq \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{N_\pi}\}$, with Bayesian prior probabilities $\pi(\boldsymbol{\theta}_j), j = 1 \dots N_\pi$. In practice, these samples may be obtained using Bayesian inference with an appropriate likelihood function and prior, or simply drawn from a frequentist confidence region.

The posterior probability $\pi(\boldsymbol{\theta}_j \mid \mathbf{Y}, \mathbf{X})$ of the parameter scenario $\boldsymbol{\theta}_j$, for a given vector of experimental controls $\mathbf{X} \doteq [\cdots \boldsymbol{x}_i^\intercal \cdots]^\intercal$ with corresponding responses $\mathbf{Y} \doteq [\cdots \boldsymbol{y}_i^\intercal \cdots]^\intercal$, can be calculated using Bayes rule as

$$\pi(\boldsymbol{\theta}_j \mid \mathbf{Y}, \mathbf{X}) = \frac{\pi(\mathbf{Y} \mid \boldsymbol{\theta}_j, \mathbf{X})\, \pi(\boldsymbol{\theta}_j)}{\pi(\mathbf{Y} \mid \mathbf{X})}\,.$$

Then, the most probable parameter value $\boldsymbol{\theta}^*$ in an N_π-category classification problem is simply selected as

$$\boldsymbol{\theta}^* \in \underset{\boldsymbol{\theta}_j \in \Theta}{\arg\max}\ \pi(\boldsymbol{\theta}_j \mid \mathbf{Y}, \mathbf{X}). \tag{8}$$

But the selection rule in (8), which is often referred to as the *Bayes optimal classifier*, has a finite probability of selecting an incorrect parameter value. To express this probability, known as the *Bayes Risk*, it is convenient to subdivide the space $\mathcal{Y}$ of possible observations into (up to) N_π regions defined by

$$\begin{aligned}\mathcal{Y}_j &\doteq \{\mathbf{Y} \in \mathcal{Y} \mid \forall k \neq j, \pi(\boldsymbol{\theta}_j \mid \mathbf{Y}, \mathbf{X}) > \pi(\boldsymbol{\theta}_k \mid \mathbf{Y}, \mathbf{X})\} \\ &= \{\mathbf{Y} \in \mathcal{Y} \mid \forall k \neq j, \pi(\boldsymbol{\theta}_j, \mathbf{Y} \mid \mathbf{X}) > \pi(\boldsymbol{\theta}_k, \mathbf{Y} \mid \mathbf{X})\}\,,\end{aligned}$$

where the latter is simply obtained by multiplying both sides of the inequalities by $\pi(\mathbf{Y} \mid \mathbf{X})$.

The Bayes Risk, $\mathcal{R}(\mathbf{X})$, is calculated by summing over all possible ways that a Bayes optimal classifier can make a classification error; that is, by summing over the probabilities of selecting $\boldsymbol{\theta}_k$ when $\boldsymbol{\theta}_j$ would be the correct value,

$$\begin{aligned}\mathcal{R}(\mathbf{X}) &\doteq \sum_{\substack{j,k=1,\\ j\neq k}}^{N_\pi} \pi(\boldsymbol{\theta}_j, \mathbf{Y} \in \mathcal{Y}_k \mid \mathbf{X}) \\ &= \sum_{\substack{j,k=1,\\ j\neq k}}^{N_\pi} \int_{\mathcal{Y}_k} \pi(\mathbf{Y} \mid \boldsymbol{\theta}_j, \mathbf{X})\ \pi(\boldsymbol{\theta}_j)\ d\mathbf{Y}.\end{aligned} \tag{9}$$

Computing the Bayes risk via (9) is generally arduous. Following Bayard and Neely (2017), we instead calculate a tractable upper bound on the Bayes Risk,

$$\mathcal{R}(\mathbf{X}) \leq \sum_{\substack{j,k=1,\\ j<k}}^{N_\pi} \sqrt{\pi(\boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_k)}\, \exp\left(-\sum_i \rho_{j,k}(\boldsymbol{x}_i)\right) \tag{10}$$

$$\text{with}\quad \rho_{j,k}(\boldsymbol{x}_i) \doteq \frac{1}{8}[\boldsymbol{\eta}(\boldsymbol{\theta}_j, \boldsymbol{x}_i) - \boldsymbol{\eta}(\boldsymbol{\theta}_k, \boldsymbol{x}_i)]^\intercal \Sigma_y^{-1} [\boldsymbol{\eta}(\boldsymbol{\theta}_j, \boldsymbol{x}_i) - \boldsymbol{\eta}(\boldsymbol{\theta}_k, \boldsymbol{x}_i)]. \tag{11}$$

The validity of this bound, here written in the special case of Gaussian measurement noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma_y)$, was first established by Blackmore et al. (2008, Theorem 1).

Finally, our formulation of the Bayes Risk information criterion accounts for the number of replications ω_i of each support $\boldsymbol{x}_i$, $i = 1 \dots N_c$ in the experimental campaign ξ,

$$\phi^{\mathrm{BR}}(\xi) \doteq -\log \sum_{\substack{j,k=1,\\ j<k}}^{N_\pi} \sqrt{\pi(\boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_k)}\, \exp\left(-\sum_{i=1}^{N_c} \omega_i\, \rho_{j,k}(\boldsymbol{x}_i)\right), \tag{12}$$

where we use log-scaling to avoid numerical issues when the Bayes Risk becomes very small, and the negative sign for consistency with the maximization in (3). A possible interpretation of ϕ^{BR} is to maximally discriminate between the uncertainty scenarios by maximizing the distance between responses of any scenario pair, weighted by their respective probabilities of occurrence.

3. COMPUTATIONAL METHODOLOGY

Solving the experiment design problem (3), with either the average D-optimality criterion (7) or the Bayes Risk criterion (12), by simultaneously searching over all possible number of supports N_c, the experimental controls $\boldsymbol{x}_{1\dots N_c}$, and the corresponding efforts $\omega_{1\dots N_c}$ is intractable in general. Instead, we adopt a decomposition approach that iterates between an effort-based optimization step and a gradient-based refinement step (Vanaret et al., 2021; Sandrin et al., 2024a).

Following an initial discretization of the experimental design space, the values of the efforts are first optimized to determine which candidate experiments should be included in the experimental campaign as well as the (integer) number of replications. The selected experiments are refined in the second step using gradient-based search to further increase the information content. These refined supports are appended to the set of experiment candidates, before updating the selected experiments and corresponding efforts by repeating the effort-based design, until no further improvement is obtained.

This approach takes advantage of the convexity of the effort-based optimization subproblem to prevent a suboptimal effort selection over the discretized experimental design space. The selected supports can then be used to warm-start the gradient-based refinement subproblem, a nonconvex optimization in general. Although this iterative procedure cannot guarantee finding the best possible experimental campaign, it is effective in practice and typically terminates after a small number of iterations. In a variant of this procedure (Yang et al., 2013; Schmid et al., 2024), the refinement step can also be computed via the solution of a maximum optimality violator subproblem, before returning to the effort-based subproblem.

3.1 Effort-based Optimization Step

The iterative procedure starts with an initial discretization of the experimental design space into a finite collection of experiment candidates, denoted by $\mathcal{X}^{(0)} \doteq \{\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_{N_s}\} \subset \mathcal{X}$ with $N_s \gg N_t$. In the case of a simple experimental design space $\mathcal{X}$, a sample can be obtained via gridding or the application of low-discrepancy sequences such as Sobol' sampling (Sobol', 1967).

This discretization recasts the search over ξ in (3) into a more tractable search over the experimental efforts ω_i associated with each experiment candidate $\hat{\boldsymbol{x}}_i \in \mathcal{X}^{(k-1)}$, where $\mathcal{X}^{(k-1)}$ is the set of experiment candidates at the start of iteration $k \geq 1$. The resultant optimized efforts $(\hat{\omega}_1, \ldots, \hat{\omega}_{N_s})$ are allowed to take a value of zero here—and most of them do when $N_s \gg N_t$.

In particular, an average D-optimal (DOPT) design is computed by solving the following pure integer nonlinear program (INLP),

$$\max_{\boldsymbol{\omega}} \ \sum_{j=1}^{N_\pi} \pi(\boldsymbol{\theta}_j) \ \log\det\left(\sum_{i=1}^{N_s} \omega_i \mathbf{A}(\hat{\boldsymbol{x}}_i, \boldsymbol{\theta}_j)\right) \tag{13}$$

$$\text{s.t.} \ \sum_{i=1}^{N_s} \omega_i = N_t, \ \omega_i \in \mathbb{Z}_+, \ \forall i, \tag{14}$$

where the objective function is an estimator of the average design criterion (7) using a sampled average approximation (SAA) based on the discretized model uncertainty set Θ.

Likewise, a Bayes Risk optimal (BROPT) design is computed by solving the following INLP,

$$\max_{\boldsymbol{\omega}} \ -\log \sum_{\substack{j,k=1,\\ j<k}}^{N_\pi} \sqrt{\pi(\boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_k)} \exp\left(-\sum_{i=1}^{N_s} \omega_i \, \rho_{j,k}(\hat{\boldsymbol{x}}_i)\right) \tag{15}$$

$$\text{s.t.} \ \sum_{i=1}^{N_s} \omega_i = N_t, \ \omega_i \in \mathbb{Z}_+, \ \forall i. \tag{16}$$

Since both the average D-optimality criterion (7) and the Bayes Risk criterion (12) are concave functions of the efforts $\boldsymbol{\omega}$, the corresponding INLPs (13)–(14) and (15)–(16) can be solved to global optimality using a standard outer-approximation algorithm (Fletcher and Leyffer, 1994); see Sandrin et al. (2024b) for a recent investigation concluding that this approach is highly tractable.

3.2 Gradient-based Refinement Step

After solving either effort-based subproblem (13)–(14) or (15)–(16), the support of the optimized design $\hat{\xi}$ is *recovered* as the collection of experiment candidates $\hat{\boldsymbol{x}}_i \in \mathcal{X}^{(k-1)}$ that have an effort $\hat{\omega}_i > 0$,

$$\text{supp}(\hat{\xi}) \doteq \{\hat{\boldsymbol{x}}_i \in \mathcal{X}^{(k-1)} : \hat{\omega}_i > 0\}, \tag{17}$$

with the corresponding optimized number of supports $\hat{N}_c \doteq |\text{supp}(\hat{\xi})|$.

A refined DOPT design is computed by solving the following (nonconvex) nonlinear program (NLP),

$$\max_{\boldsymbol{x}_i \in \mathcal{X}} \ \sum_{j=1}^{N_\pi} \pi(\boldsymbol{\theta}_j) \ \log\det\left(\sum_{i=1}^{\hat{N}_c} \hat{\omega}_i \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}_j)\right) \tag{18}$$

while a refined BROPT design is computed similarly by solving the following (nonconvex) NLP

$$\max_{\boldsymbol{x}_i \in \mathcal{X}} -\log \sum_{\substack{j,k=1,\\ j<k}}^{N_\pi} \sqrt{\pi(\boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_k)} \exp\left(-\sum_{i=1}^{\hat{N}_c} \hat{\omega}_i \, \rho_{j,k}(\boldsymbol{x}_i)\right) \tag{19}$$

using $\text{supp}(\hat{\xi})$ in (17) as initial guesses for the $\boldsymbol{x}_i$'s.

Finally, the resultant refined supports are appended to the set of experiment candidates $\mathcal{X}^{(k-1)}$ to form the augmented set $\mathcal{X}^{(k)}$ at the next iteration. This procedure is interrupted as soon as the effort-based search over $\mathcal{X}^{(k)}$ does not yield any further improvement.

3.3 Scenario Reduction in Bayes Risk Upper Bound

A prerequisite of the Bayes Risk minimization approach is that the discretization set Θ of the model parametric uncertainty should be fine enough to create sufficient overlap among all the responses corresponding to the various uncertainty realizations—otherwise, the Bayes Risk upper bound in (10)–(11) could become vanishingly small and the experiment design optimization would be pointless. In particular, the problem is exacerbated with an increasing number of uncertainty parameters or a wider uncertainty support set.

However, increasing the uncertainty discretization size N_π can impose a significant computational overhead, both in terms of the initial candidate experiment evaluation and the function and gradient evaluations in the optimization subproblems. In the effort-based subproblem (15)–(16), for instance, the number of summands in the cost functions scales as $\mathcal{O}(N_\pi^2 N_s n_y)$.

In response to this, we propose a simple scenario reduction strategy in a preprocessing step, that leverages the initial discretization of the experimental design space $\mathcal{X}^{(0)}$ with equal efforts. Specifically, for each uncertainty scenario $j \in \{1, \ldots, N_\pi\}$, we determine the M nearest-neighbours $\hat{\jmath}_1, \ldots, \hat{\jmath}_M$ which minimize the following distance

$$d_{j,\hat{\jmath}_k} \doteq \exp\left(-\sum_{i=1}^{N_s} \rho_{j,\hat{\jmath}_k}(\hat{\boldsymbol{x}}_i)\right), \quad k = 1 \ldots M, \ \hat{\boldsymbol{x}}_i \in \mathcal{X}^{(0)}$$

and we sum over these selected scenario pairs only when evaluating the criteria in (15) & (19) and their gradients.

3.4 Software Implementation

For the effort-based optimization step, the master ILP subproblems of the outer-approximation algorithm are solved using the solver GUROBI (v11.0.1), [1] with relative and absolute convergence tolerances of 10^{-6} and 10^{-9}, respectively. The NLP from the initial continuous relaxation is solved using the sparse nonlinear solver SNOPT (v7.7), [2] with optimality tolerance set to 10^{-6}. The outer-approximation iterations are terminated when the relative gap between the master solution value and the incumbent is below 10^{-5}. For the gradient-based refinement step, the NLP subproblems are also solved using SNOPT within the same tolerance of 10^{-6}.

The scenario generation and evaluation for the jointly discretized experimental design space and model uncertainty, the effort-based optimization step, and the gradient-based refined step all together are coordinated from a new class named MBDOESLV in the C++ library CANON (v4.0) [3]. CANON builds on the library MC++ (v4.0) [4] to evaluate

[1] https://www.gurobi.com/solutions/gurobi-optimizer/
[2] https://ccom.ucsd.edu/~optimizers/solvers/snopt/
[3] https://github.com/omega-icl/canon
[4] https://github.com/omega-icl/mcpp

and differentiate expression trees, along with the library CRONOS (v4.0) [5] for numerical integration and sensitivity analysis of ODE systems by interfacing with the solver CVODES of the library SUNDIALS (v7.1.1) [6]. Linear algebra calculations, including matrix rank and determinant, Cholesky decomposition and triangular system solve, are all carried out using the library Armadillo (v12.6). [7] Concurrent evaluation of the model responses or the FIMs corresponding to different uncertainty scenarios is enabled on multiple threads, both as part of the initial scenario generation or the gradient-based refinement step.

4. CASE STUDY

The case study considers a fed-batch reactor hosting the model reaction $\mathrm{A} \longrightarrow \nu\mathrm{B}$ (Kusumo et al., 2022). A mechanistic model of the system is given by

$$\dot{c}_\mathrm{A}(t) = \frac{q_\mathrm{in}(t)}{V(t)}\left(c_\mathrm{A}^\mathrm{in} - c_\mathrm{A}(t)\right) - k(T)\,c_\mathrm{A}(t)^\alpha \tag{20}$$

$$\dot{c}_\mathrm{B}(t) = -\frac{q_\mathrm{in}(t)}{V(t)}c_\mathrm{B}(t) + \nu\,k(T)\,c_\mathrm{A}(t)^\alpha \tag{21}$$

$$\dot{V}(t) = q_\mathrm{in}(t) \tag{22}$$

$$k(T) \doteq \exp\left(\theta_0 + \theta_1\left(\frac{T - T^\mathrm{ref}}{T}\right)\right) \tag{23}$$

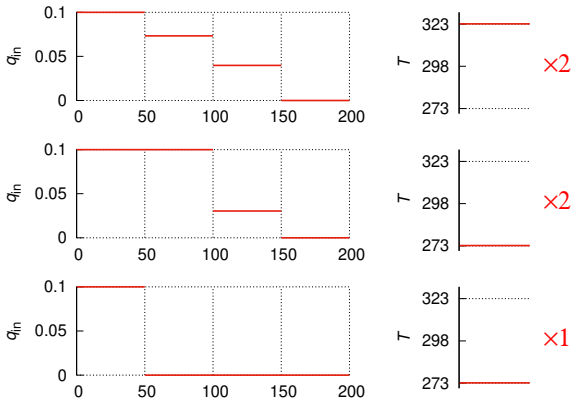
with c_i ($\mathrm{mol\,L^{-1}}$) denoting the concentration of species $i \in \mathrm{A}, \mathrm{B}$, t (min) time, V (L) the reaction mixture volume, and $T^\mathrm{ref} = 273.15\,\mathrm{K}$ the reference temperature.

The duration of an experimental batch is set to 200 minutes. The experimental setup allows for measurements of c_A and c_B to be taken at 25 minute intervals during each batch, leading to 8 sampling times at $t = 25, 50, \ldots, 200$ min, with IID measurement errors with $0.2\,\mathrm{mol\,L^{-1}}$ standard deviation. The experimental controls comprise the time-varying inlet volumetric flowrate $q_\mathrm{in}(t) \in [0.0, 0.1]$ $\mathrm{L\,min^{-1}}$, parameterized as a piecewise constant function with fixed switching times at $t = 50, 100, 150$ min, and the time-invariant reaction temperature $T \in [273.15, 323.15]$ K. The inlet concentration of A is kept constant at $c_\mathrm{A}^\mathrm{in} = 10\,\mathrm{mol\,L^{-1}}$, and the initial concentrations and volume are set to $c_\mathrm{A}(0) = 5\ \mathrm{mol\,L^{-1}}$, $c_\mathrm{B}(0) = 0\ \mathrm{mol\,L^{-1}}$ and $V(0) = 1$ L.

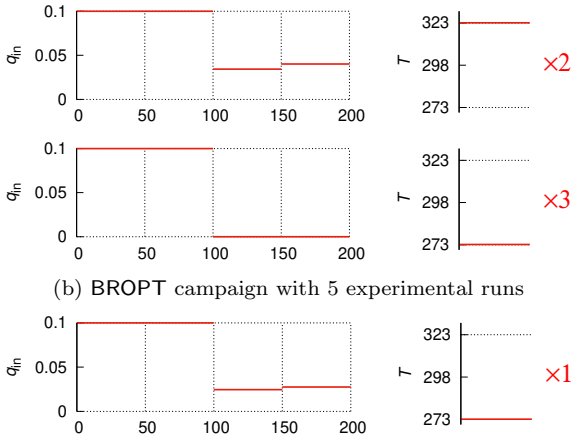
The order of reaction α, stoichiometric ratio ν, and dimensionless pre-exponential factor θ_0 and activation energy θ_1 are all uncertain parameters. The former is assumed to follow a Bernoulli distribution with 75% and 25% probability of being, respectively, either a first-order ($\alpha = 1$) or second-order ($\alpha = 2$) reaction. The other three parameters are assumed to be uniformly distributed as $\theta_0 \sim \mathcal{U}(-5.87, -0.54)$, $\theta_1 \sim \mathcal{U}(0.45, 4.39)$, and $\nu \sim \mathcal{U}(0.3, 0.7)$.

4.1 Comparison of Experiment Designs

The average D-optimal (DOPT) and Bayes Risk optimal (BROPT) designs for a campaign with $N_t = 5$ experimental runs are shown in Figs. 1a & 1b, respectively. Both designs are computed with an initial set of $N_s = 500$ candidate experiments and $N_\pi = 1000$ uncertainty scenarios,



(a) DOPT campaign with 5 experimental runs



(b) BROPT campaign with 5 experimental runs

(c) BROPT campaign with a single experimental run

Fig. 1. Comparison of optimal experimental campaigns.

generated using Sobol' sampling (Sobol', 1967), and convergence within a single refinement iteration. The DOPT campaign comprises 3 supports, while the BROPT only has 2. Despite this difference, the optimal experiments have in common that the flowrate profiles are initially maximal before being turned down or off, and the temperature is set to either its minimal or maximal level. This similarity is also reflected in the values of the information criteria, with $\phi^\mathrm{AD} = 25.43$ and $\phi^\mathrm{BR} = 2.32$ in the DOPT campaign, and $\phi^\mathrm{AD} = 25.39$ and $\phi^\mathrm{BR} = 2.35$ in the BROPT campaign.

Next, we test the effectiveness of the design criteria by simulating both the DOPT and BROPT campaigns for another 1000 scenarios of the uncertain model parameters—different from the scenarios used for the optimization—and computing the confidence intervals that would be obtained for the parameters after recalibrating the model in each scenario. The box plots on Fig. 2 show summary statistics for the distribution of the confidence interval width for each parameter (as a percentage of the nominal parameter value) over these 1000 scenarios. This comparison confirms that the DOPT and BROPT campaigns perform similarly in terms of improving parameter precision.

Notice also that a campaign with at least $N_t = 2$ experiments is required to conduct a DOPT design for the model (20)–(23), otherwise the parameters θ_0 and θ_1 are not structurally identifiable and the FIM is singular. In contrast, a BROPT design can be conducted even for a campaign with a single experimental run, as illustrated

[5] https://github.com/omega-icl/cronos
[6] https://github.com/LLNL/sundials/tree/main
[7] https://arma.sourceforge.net/docs.html

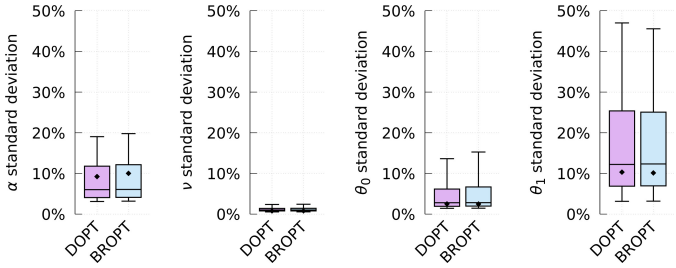


Fig. 2. Effect of optimal experimental campaigns on parameter precision. The bar, box and whiskers indicate the median, interquartile range and 10-90$^{\text{th}}$ percentile range, respectively, and the dot the average value.

Table 1. Computational comparison between optimal experiment designs (with wall-time in seconds).† The scenario reduction selects M nearest neighbors (see Sec. 3.3).

	DOPT design		BROPT design		
Uncertainty scenarios	1000	250	1000	250	1000
Scenario reduction	–	–	–	–	20
Sample evaluation	446	88	55	17	79
Effort-based step	126	34	957	94	22
Gradient-based step	1344	333	94	53	77
Total wall-time	1916	455	1106	163	178

†Lenovo ThinkPad X1 Carbon Gen 10 with 12$^{\text{th}}$ Gen Intel® Core™ i7-1260P × 16, 32.0 GiB memory, Ubuntu 22.04 operating system

in Fig. 1c. This experiment picks the lower reactor temperature level and a feedrate profile that is intermediate between the two experimental runs in Fig. 1b.

4.2 Comparison of Computational Performance

The computation comparison in Table 1 shows a breakdown in terms of sample evaluation (including scenario reduction), effort-based optimization, and gradient-based refinement wall-times. With $N_\pi = 1000$ uncertainty scenarios, using BROPT about halves the computational time compared to DOPT. However, there is a noticeable burden shift between the effort- and gradient-based steps. The effort-based step becomes dominant with BROPT because of the very large number of summands (c. 4 billion) in the Bayes Risk upper bound expression in (15), as discussed in Sec. 3.3. A significant improvement in computational time can be obtained by reducing the number of uncertainty scenarios from $N_\pi = 1000$ to 250. By doing so, there is nevertheless a risk that the performance of the experiment design will be degraded, especially with BROPT where sufficient overlap between the model responses is critical for a meaningful optimization. Alternatively, the proposed scenario reduction strategy—here preselecting the 20 nearest-neighbours to each uncertainty scenario—provides a similar computational improvement, yet without impairing the accuracy of the Bayes Risk bound.

5. CONCLUSIONS

Through this paper, we investigated an approach for recasting an experiment design problem for parameter precision as a discrimination problem. We used an algebraic upper bound on the Bayes Risk to design experimental campaigns that maximally discriminate between a set of rival models corresponding to different model parameter uncertainty realizations. Our case study results showed that a Bayes Risk minimization strategy can provide highly-informative experimental campaigns to improve parameter precision in mathematical process models, while being computationally advantageous compared to conventional FIM-based design strategies and capable of handling structurally unidentifiable problems. Future work will entail testing the methodology on additional case studies.

REFERENCES

Atkinson, A.C., Donev, A.N., and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.

Bauer, I., Bock, H.G., Körkel, S., and Schlöder, J.P. (2000). Numerical methods for optimum experimental design in DAE systems. *J Comput Appl Math*, 120(1), 1–25.

Bayard, D. and Neely, M. (2017). Experiment design for nonparametric models based on minimizing Bayes Risk: application to voriconazole. *J Pharmacokinet Pharmacodyn*, 44, 95–111.

Blackmore, L., Rajamanoharan, S., and Williams, B.C. (2008). Active estimation for jump Markov linear systems. *IEEE Trans Automat Contr*, 53(10), 2223–2236.

Espie, D. and Macchietto, S. (1989). The optimal design of dynamic experiments. *AIChE J*, 35(2), 223–229.

Fisher, R.A. (1971). *The Design of Experiments*. Macmillan, 9$^{\text{th}}$ edition.

Fletcher, R. and Leyffer, S. (1994). Solving mixed integer nonlinear programs by outer approximation. *Math Program*, 66(1), 327–349.

Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci*, 63(19), 4846–4872.

Galvanin, F., Barolo, M., and Bezzo, F. (2009). Online model-based redesign of experiments for parameter estimation in dynamic systems. *Ind Eng Chem Res*, 48(9), 4415–4427.

Kusumo, K.P., Kuriyan, K., Vaidyaraman, S., García-Muñoz, S., Shah, N., and Chachuat, B. (2022). Risk mitigation in model-based experiment design: A continuous-effort approach to optimal campaigns. *Comput Chem Eng*, 159, 107680.

Pronzato, L. and Walter, E. (1985). Robust experiment design via stochastic approximation. *Math Biosci*, 75(1), 103–120.

Rainforth, T., Foster, A., Ivanova, D.R., and Bickford Smith, F. (2024). Modern Bayesian experimental design. *Stat Sci*, 39(1), 100–114.

Sandrin, M., Chachuat, B., and Pantelides, C.C. (2024a). Integrating effort-and gradient-based approaches in optimal design of experimental campaigns. *Comput Aided Chem Eng*, 53, 313–318.

Sandrin, M., Kusumo, K.P., Pantelides, C.C., and Chachuat, B. (2024b). Solving for exact designs in optimal experiment campaigns under uncertainty. *IFAC-PapersOnLine*, 58(14), 658–663.

Schmid, J., Seufert, P., and Bortz, M. (2024). Adaptive discretization algorithms for locally optimal experimental design. *arXiv*, math.OC, 2406.01541.

Shahmohammadi, A. and McAuley, K.B. (2020). Using prior parameter knowledge in model-based design of experiments for pharmaceutical production. *AIChE J*, 66(11), e17021.

Sobol', I.M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math & Math Phys*, 7(4), 86–112.

Vanaret, C., Seufert, P., Schwientek, J., Karpov, G., Ryzhakov, G., Oseledets, I., Asprion, N., and Bortz, M. (2021). Two-phase approaches to optimal model-based design of experiments: how many experiments and which ones? *Comput Chem Eng*, 146, 107218.

Walker, S.G. (2016). Bayesian information in an experiment and the Fisher information distance. *Stat Probab Lett*, 112, 5–9.

Yang, M., Biedermann, S., and Tang, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *J Am Stat Assoc*, 108(504), 1411–1420.