Characterizing the Pareto Optimal Trade-Off Between Model-Based Information Content and Measurements Cost

Jialu Wang and Alexander W. Dowling*
Department of Chemical and Biomolecular Engineering, University of Notre Dame
Notre Dame, IN 46556

Abstract
Model-based design of experiments (MBDoE) leverages science-based models to maximize information gain from experiments while minimizing time and resource costs. When not enough or unsuitable measurements are available, MBDoE may suggest uninformative or infeasible optimal experiments. We propose a novel framework to maximize the experimental information content. This framework identifies measurement campaigns for multi-response systems by leveraging the continuous-effort design concept to retain a convex optimization problem regardless of model structure. Sensitivity matrix computed by Pyomo.

Keywords
Applied statistics, data science, convex optimization, design of experiments, Fisher information matrix

Introduction
Design of experiments (DoE) is an essential tool for the building and validation of predictive mathematical models. It links the experimental and the modelling world by suggesting experiments that yield the most informative data from an experimental apparatus. Model-based DoE (MBDoE) leverages science-based mathematical models constructed from the underlying physical principles of the system (Franceschini and Macchietto, 2008). Taking advantage of the prior knowledge of the experimental system, MBDoE can discriminate between scientific hypotheses, posed as mathematical models, and facilitate nonconvex optimization with state-of-the-art algorithms that exploit 1st and 2nd derivative information. MBDoE has a rich history of success at the intersection of chemical engineering, applied statistics, and mathematical programming research communities including chemical kinetics (Waldron et al., 2020), heat/mass transfer modeling (Balsa-Canto et al., 2007), and biological modeling (Chakrabarty et al., 2013). Recent open-source general-purpose software including Pyomo.DE (Wang and Dowling, 2022) and Pydex (Kusumo et al., 2022) help reduce barriers by making MBDoE accessible to Python users.

Pyomo.DE implements a two-stage stochastic program to solve the MBDoE optimization problem
\[ \max_{\phi} \psi(M) \]
(1) where \( \phi \) is the vector of experimental design variables, \( M \) is Fisher information matrix (FIM). The design criteria \( \psi \) can be the classical alphabetic criteria such as A-optimality (trace), D-optimality (determinant), or user-defined expression.

Pydex

\[ M(\phi) \] is the FIM of this experiment. Unlike Eq. (1) which leverages process models to evaluate FIM, Eq. (2) precomputes FIM and how many of them should be repeated. In this way, Eq. (2) circumvents the challenges of

* To whom all correspondence should be addressed, adowling@nd.edu
solving the usually large-scale and highly-nonlinear MBDoE problem memory requirements to store sensitivity data. Figure 1 shows model calibration and uncertainty quantification workflow in Pyomo (Bynum et al., 2021) ecosystem. Building Pyomo models with prior knowledge and preliminary data, Pyomo.DE evaluates the FIM with a specific set of measurements the identifiability of the model, suggests new experiments to provide data for parameter estimation. ureang and Dowling 2022 shows with a fixed-bed case study that optimal designs may be uninformative or unidentified when certain measurements are unavailable. Considering the possibility and difficulty to set up and the usually high costs to add measurements, choosing a experimental measurement, a.k.a. hat, when, and how to measure, critical balance the trade-off between high information content provided by a high cost. We highlight the importance of choosing such a measurement and s

$$\max \psi(M)$$ (a)
$$\sum_{k \in K} c_k \cdot m_k$$ (b)

where the design criteria $\psi$ can be the classical alphabetic criteria or -defined . Classical information criteria for improving the parameter accuracy include A-, D-, E- optimality, which are the trace, determinant, and the minimal eigen value of the Fisher information matrix (FIM) $M$ (Wang and Dowling, 2022). FIM is approximated by the dynamic sensitivity matrix and inverse of the error variance-covariance matrix. is the FIM from prior experiments, prior information. is the total cost $y$ The sensitivity matrix $Q_r$ for measurement $y_r$ is defined as:

$$Q_r = \begin{bmatrix}
\frac{\partial y_r}{\partial x_1} | t_1 & \cdots & \frac{\partial y_r}{\partial x_{1}} | t_1 \\
\vdots & \ddots & \vdots \\
\frac{\partial y_r}{\partial x_r} | t_r & \cdots & \frac{\partial y_r}{\partial x_{r}} | t_r
\end{bmatrix}$$ (c)

here $\{t_1, ..., t_r\}$ are time points for measurement $y_r$. The dynamic sensitivity matrix $Q$ combines all measurements:

$$Q = \begin{bmatrix}
Q_1 \\
\vdots \\
Q_{N_y}
\end{bmatrix}$$

Eq. (c)Local sub-optimal designs can be avoided CED problem and convex optimization . An obvious drawback of the CED method is that the size of the optimization problem scales with the number of measurements. But the sensitivity matrix can be precomputed prior to solving the optimization problem, which reduces the computational burden. In this work, all process models are constructed on Pyomo he precomputed sensitivity matrix are obtained by Pyomo.DE. The convex CED formulation is solved numerically using MOSEK (Andersen and Andersen, 2000) interfaced through cvxpy (Diamond and Boyd, 2016) with a Macbook Pro (15-inch, 2018) with a 2.6 GHz 6-core Intel Core i7 processor in seconds.

Results

We consider a reaction kinetics system containing two first-order liquid phase reactions in a batch reactor from Wang and Dowling 2022. The system is dynamic algebraic equations (DAE) (see the candidate measurement space includes three time-varying state variables, $C_A, C_B, C_C$, the concentrations of the species A, B, C at 9 timepoints $t \in [0, 1]$ hr. The observation error variance are assumed to be different for $A, B, C$, $\sigma^2_{A,i} = 1, \sigma^2_{B,i} = 4, \sigma^2_{C,i} = 8$ mol$^2$ L$^{-2}$, $i \in t$, and there are covariance between measurements at the same time point, $\sigma_{(A,B),i} = 0.1, \sigma_{(A,C),i} = 0.1, \sigma_{(B,C),i} = 0.5$, $i \in t$ in mol$^2$ L$^{-2}$. The cost at one time point $10$ for A $6$ for B and C. Figure 2. Figure 2 shows the optimal measurement strategy and its

---

**Figure 1.** Parameter estimation, sensitivity and uncertainty analysis, and MBDoE are combined into an iterative framework to select, refine, and calibrate science-based mathematical models with quantified uncertainty.

**Method**

Consider a multi-input multi-response model of a process given by:

$$f(\theta, \varphi, y) = 0$$ (d)

where $y \in Y \subset \mathbb{R}^{n_y}$ is the vector of state variables, $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ is the vector of parameters $\varphi$ is the process model, $\varphi \in \Phi \subset \mathbb{R}^{n_\varphi}$ is the vector of A set of data is collected from the experimental system by measuring part of or all the state variables $y$, at a set of specific timepoint $t$. $y_{r,t}$ means the dynamical measurement $y_r \in Y = \{y_1, ..., y_{N_y}\}$ is measured at a specific time point $t \in t_r = \{t_1, ..., t_r\}$. $N_y$ is the number of dynamical measurements. The $p_{r,t} \in \{0,1\}$ indicates if the data of $y_{r,t}$ is measured. Thes

An optimal measurement maximizes the experimental information content indicated by experimental design criteria while satisfying constraints to the total cost of this measurement:

$$\max \psi(M)$$ (a)
$$\sum_{k \in K} c_k \cdot m_k$$ (b)

where the design criteria $\psi$ can be the classical alphabetic criteria or -defined . Classical information criteria for improving the parameter accuracy include A-, D-, E- optimality, which are the trace, determinant, and the minimal eigen value of the Fisher information matrix (FIM) $M$. Considering the possibility and difficulty to set up and the usually high costs to add measurements, choosing a specific set of measurements the identifiability of the model, suggests new experiments to provide data for parameter estimation. Considering the possibility and difficulty to set up and the usually high costs to add measurements, choosing a specific set of measurements the identifiability of the model, suggests new experiments to provide data for parameter estimation.
experimental information content with the budget. The left axis shows the logarithm of the D-optimality, a.k.a. the determinant of FIM, the information content. The right axis shows the number of measurements that can be selected. With the budget increasing from $50 to $180, the number of increases from 7 to all 24, and the information content increases by around 3 orders of magnitude. The optimization formulation provide the most informative measurement campaign the budget.

Conclusions and Future Outlook

Many challenges faced in MBDoE for nonlinear process models are rooted in the selection of measurement and. When an unsuitable measurement is used, computed optimal experiments may be uninformative or infeasible. In this work, we proposed a novel method to optimize the measurement for maximizing the experimental information content. This framework effective measurement campaigns for multi-response systems with different variance and covariances between responses. We implemented this framework in combination with the continuous-effort design concept to retain a convex optimization problem regardless of model structure. With the sensitivity matrix precomputed by Pyomo.DOE, the optimization problem can be scaled to more measurements with less computational burden.

We the methodology with a reaction kinetics example. We that the method can quickly solve the global optimum in the measurement space. a fixed-bed CO₂ adsorption system apply → .

Acknowledgments

This work was supported as part of the Carbon Capture Simulation for Industry Impact (CCSI²) project, funded through the U.S. DOE Office of Fossil Energy and Carbon Management by the National Energy Technology Laboratory, an agency of the United States Government, through a support contract.

References


Fedorov, V. V., & Leonov, S. L. (2013). Optimal design for nonlinear response models. CRC Press.


