CHARACTERIZING THE PARETO OPIMAL 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 , MBDoE may suggest uninformative or infeasible optimal experiments. We proposed a novel to the measurementmaximiz the experimental information content. This framework identifies measurement campaigns for multi-response systems variance and covariances between responses. We leveraged 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, posted 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 user.

Pyomo.DE implements a two-stage stochastic program tosolve the MBDoE optimization problem

$$\max_{\varphi} \psi(\mathbf{M})$$

(1) where φ is the vector of experimental design variables, $\mathbf{M}$ is Fisher information matrix (FIM). The design criteria ψ can be the classical alphabetic criteria such a A-optimality (trace), D-optimality (determinant), or user-defined expression.

Pydex

$\mathbf{M}(\varphi_i)$ 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 repeatedIn this way, Eq. (2) circumvents the challenges of

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 unidentifiable 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 s the high costs . We highlight the importance of choosing such a measurement and s
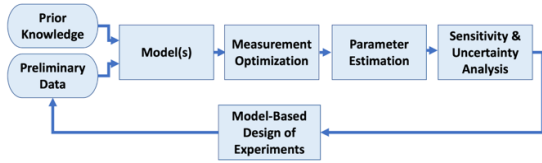

Figure 1. Parameter estimation, sensitivity and uncertainty analysis, and MBDoE are combined into an iterative framework to select, refine, and calibrate science-based mathematical models with quantified uncertainty.

Method

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

$$f(\boldsymbol{\theta}, \boldsymbol{\varphi}, \boldsymbol{y}) = 0 \qquad ()$$

where $\boldsymbol{y} \in \boldsymbol{\mathcal{Y}} \subset \mathbb{R}^{n_y}$ is the vector of state variables, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{n_\theta}$ is the vector of parameters $\boldsymbol{f}$ is the process model, $\boldsymbol{\varphi} \in \boldsymbol{\Phi} \subset \mathbb{R}^{n_\varphi}$ is the vector of A set of data is collected from the experimental system by measuring part of or all the state variables $\boldsymbol{y}_r$ at a set of specific timepoint $\boldsymbol{t}_r$. $y_{r,t}$ means the dynamical measurement $\boldsymbol{y}_r \in \boldsymbol{\mathcal{Y}} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_{N_y}]$ is measured at a specific time point $t \in \boldsymbol{t}_r = \{t_1, \dots, t_r\}$. 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 :

$$\max_I \psi(\mathbf{M}) \qquad \text{(a}$$

$$\text{(4b)} \quad \sum_{k \in \mathcal{K}} c_k \cdot m_k$$

(c)

where the design criteria ψ can be the classical alphabetic criteria or -defined . Classical information criteria for improving the parameter accuracy include A-, D-, and E-optimality, which are the trace, determinant, and the minimal eigen value of the Fisher information matrix (FIM) $\mathbf{M}$(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 .$\mathcal{Y}$ The sensitivity matrix $\mathbf{Q}_r$ for measurement $\boldsymbol{y}_r$ is defined as:

$$\mathbf{Q}_r = \begin{bmatrix} \frac{\partial y_r}{\partial \theta_1}|_{t_1} & \cdots & \frac{\partial y_r}{\partial}|_{t_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_r}{\partial \theta_1}|_{t_r} & \cdots & \frac{\partial y_r}{\partial}|_{t_r} \end{bmatrix} \qquad ()$$

here $\{t_1, \dots, t_r\}$ are time points for measurement $\boldsymbol{y}_r$. The dynamic sensitivity matrix $\mathbf{Q}$ combines all measurements:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_{N_y} \end{bmatrix} \qquad ()$$

Eq. ()sLocal 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)(The candidate measurement space includes three time-varying state variables, C_A, C_B, C_c,, the concentrations of the species A, B, C ach 9 timepoints $\boldsymbol{t}$ 0 to 1 hr. The observation error variance are assumed to be different for A, B, C, $\sigma_{A,t_i}^2 = 1, \sigma_{B,t_i}^2 = 4, \sigma_{C,t_i}^2 = 8$ mol^2 L^{-2} , $i \in \boldsymbol{t}$, and there are covariance between measurements at the same time point, $\sigma_{(A,B),t_i} = 0.1, \sigma_{(A,C),t_i} = 0.1, \sigma_{(B,C),t_i} = 0.5, i \in \boldsymbol{t}$ 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

experimental information content with the budget. The left axis shows the logarithm of the D-optimality, a.k.a. the determinant of FIM, theinformation content. The right axis shows the number of measurements that can be selected. With the budget increasing from \$50 to \$180, he number of increases from 7 to all 24, and the information content increases by around 3 orders of magnitude. The optimization formulation provid 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 CO_2 adsorption system apply -- .

Acknowledgments

References

Andersen, E. D., & Andersen, K. D. (2000). The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In High performance optimization (pp. 197-232). Springer, Boston, MA.

Balsa-Canto, E., Rodriguez-Fernandez, M., & Banga, J. R. (2007). Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *Journal of Food Engineering*, 82(2), 178-188.

Bynum, M. L., Hackebeil, G. A., Hart, W. E., Laird, C. D., Nicholson, B. L., Siirola, J. D., ... & Woodruff, D. L. (2021). *Pyomo-optimization modeling in python* (Vol. 67). Springer.

Chakrabarty, A., Buzzard, G. T., & Rundell, A. E. (2013). Model-based design of experiments for cellular processes. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 5(2), 181-203.

Diamond, S., & Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. The Journal of Machine Learning Research, 17(1), 2909-2913.

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

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19), 4846-4872.

Kusumo, K. P., Kuriyan, K., Vaidyaraman, S., García-Muñoz, S., Shah, N., & Chachuat, B. (2022). Risk mitigation in model-based experiment design: A continuous-effort approach to optimal campaigns. Computers & Chemical Engineering, 159, 107680.

Nicholson, B., Siirola, J. D., Watson, J. P., Zavala, V. M., & Biegler, L. T. (2018). pyomo. dae: a modeling and automatic discretization framework for optimization with differential and algebraic equations. *Mathematical Programming Computation*, 10(2), 187-223.

Thierry, D. (2019). Nonlinear Optimization based frameworks for Model Predictive Control, State-Estimation, Sensitivity Analysis, and Ill-posed Problems (Doctoral dissertation, Carnegie Mellon University).

Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., & Gavriilidis, A. (2020). Model-based design of transient flow experiments for the identification of kinetic parameters. *Reaction Chemistry & Engineering*, 5(1), 112-123.

Wang, J., & Dowling, A. W. Pyomo. DOE: An Open-Source Package for Model-Based Design of Experiments in Python. *AIChE Journal*, e17813.