Two-phase approaches to optimal model-based design of experiments: how many experiments and which ones?

Charlie Vanaret[1], Philipp Seufert[1], Jan Schwientek[1],
Gleb Karpov[2], Gleb Ryzhakov[2], Ivan Oseledets[2],
Norbert Asprion[3], Michael Bortz[1]

[1] Fraunhofer ITWM, Kaiserslautern, Germany
charlie.vanaret@itwm.fraunhofer.de
[2] Skoltech, Moscow, Russia
[3] BASF, Ludwigshafen, Germany

Keywords: *model-based experimental design, approximation strategies, equivalence theorem*

1 Motivation

Model-based experimental design is attracting increasing attention in chemical process engineering. Typically, an iterative procedure is pursued: an approximate model is devised, prescribed experiments are then performed and the resulting data is exploited to refine the model. To help reduce the cost of trial-and-error approaches, strategies for model-based design of experiments suggest experimental points where the expected gain in information for the model is the largest. From a technical perspective, it requires the resolution of a large nonlinear, generally nonconvex, optimization problem, whose solution may greatly depend on the starting point.

2 Contribution

The approaches of statistical and linear experimental designs [2] may be exploited for the initialization and multistarting of nonconvex experimental design optimization to increase the odds of finding the global optimum. Since Newton-based optimization techniques, such as SQP or interior point methods, converge quadratically close to a local optimum, finding accurate initial points is a crucial task. We adopt a two-phase approach:

- phase 1: an initial point may be generated:
  - using a problem-independent pattern-based strategy ;
  - solving a problem-dependent approximation of the experimental design problem for a discrete set of candidate experiments.
- phase 2: the phase-1 solution is used to initialize the original experimental design problem.

When the number of experiments is optimal, the reliability of the model with respect to its parameters increases when the number of repetitions of some of the experiments is carefully tuned. However, in the nonlinear case, the questions of how many experiments to pick, which ones, and which experiments to repeat are non trivial and have not been addressed so far. This is one major contribution of this article. We also implemented a verification test to assess the optimality of the phase-1 and phase-2 solutions. It exploits the Kiefer-Wolfowitz equivalence theorem [3]. We present two approximation strategies that can assist the experimenter in setting the number of relevant experiments and performing an optimal selection, and we compare them against two pattern-based strategies that are independent of the problem.

3 Numerical results

The validity of the approach is demonstrated experimentally on an academic example and two test problems from chemical engineering, namely vapor liquid equilibrium [1] and reaction kinetics. Figure 1 represents the two-phase approach for the vapor liquid equilibrium initialized with a factorial design with 6 experiments: the phase-1 (dashed) and phase-2 (colored) solutions are represented by disks whose sizes are proportional to the relevance of the experiments. Two of the experiments were deemed irrelevant during phase 2. Figure 2a represents the phase-1 and phase-2 solutions computed by both approximation strategies on a grid of 90 candidates. The phase-1 solution is very close to the phase-2 solution, which turns out to be the global minimizer. The visualization in Figure 2b, based on the Kiefer-Wolfowitz equivalence theorem, indicates that the phase-1 solution is indeed globally optimal for the given discretization. This example shows that our approximation techniques answer the key interrogations in experimental design: how many experiments should be picked, and which ones.

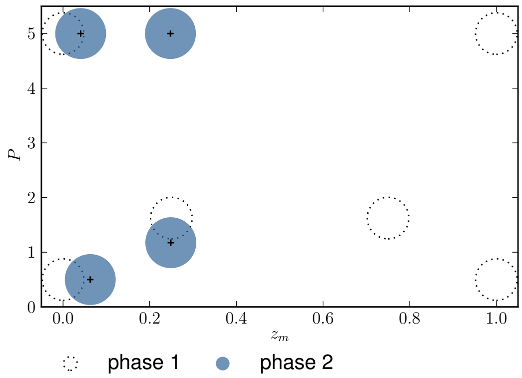
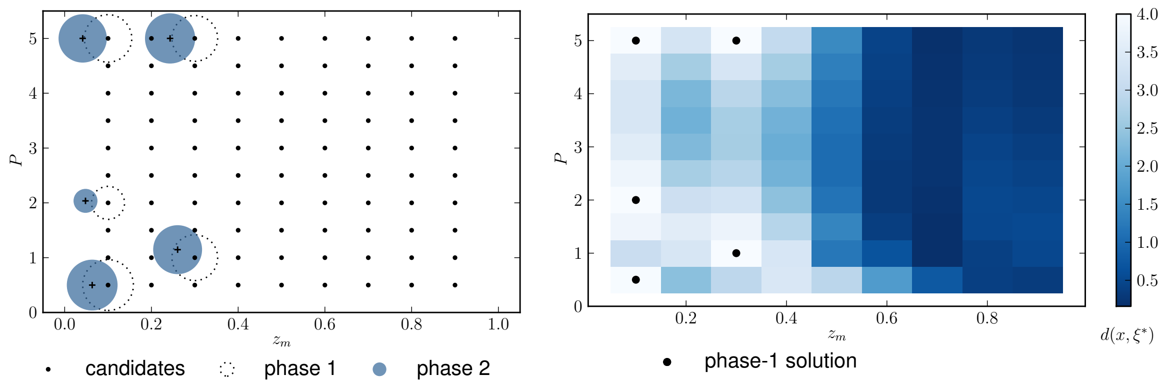


FIG. 1: Flash distillation: factorial design with 6 experiments



(a) Phase-1 and phase-2 solutions

(b) Verification of optimal phase-1 design

FIG. 2: Flash distillation: approximation strategies with 90 candidates

References

[1] N. Asprion, R. Boettcher, J. Mairhofer, M. Yliruka, J. Hoeller, J. Schwientek, C. Vanaret, and M. Bortz. Implementation and application of model-based design of experiments in a flowsheet simulator. *J. Chem. Eng. Data (Accepted)*, 2019.

[2] V. V. Fedorov and S. L. Leonov. *Optimal Design for Nonlinear Response Models.* CRC Press, Boca Raton, 2014.

[3] J. Kiefer. Optimum Experimental Designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(2):272–304, jul 1959.