Discussion of "Design augmentation for response optimization and model estimation"

David C. Woods

University of Southampton, Southampton, UK

In their paper, Drs Nachtsheim and Jones revisit some old questions in sequential response surface design. Assuming the aim is to maximise the response, the standard approach identifies a path of steepest ascent from a first-order linear model fitted using data from a phase 1 (fractional) factorial screening experiment. After a sequence of experimental runs have been performed along this path, a phase 2 experiment is performed, typically centred on a best-guess of the location of the maximum, designed to allow estimation of a second-order model. Such a strategy was proposed in the seminal papers of Box and Wilson (1951) and Box and Hunter (1957) and has been described in numerous text books (e.g. Box et al., 2005, Wu and Hamada, 2009 and Montgomery, 2017).

The methodology suggested and assessed by the authors follows this basic recipe but with some important differences. Firstly, a phase 1 design is used capable of allowing estimation of a second-order model. Secondly, the quadratic response surface model is fitted, and hence the direction of the optimum is identified using the ridge, or optimization, trace. Thirdly, a D-optimal phase 2 design is employed, in place of the more common central composite design. While individual elements of this strategy have been suggested before, I am not aware of all these elements being combined and, importantly, assessed via comparisons to steepest ascent. It was therefore with great interest, and enjoyment, that I read this paper. I commend the authors for providing new insights into an important problem, and for providing much food for thought. Below I provide a few comments on their work, particularly highlighting areas that might be fruitful for extension and further study.

1 Shift in the design space and extrapolation

The decision of how to shift the design space is a key component of the steepest ascent method but is perhaps one of the less researched aspects (particularly in comparison to the choice of designs), as discussed by the authors. However, this issue is not really addressed in the augmentation strategy discussed in the paper, with a shift of 1.5 to 2 coded units chosen only to ensure some overlap between the phase 1 and phase 2 design spaces. This seems particularly limiting in light of the fact that several of the example response surfaces have maxima outside of these allowed phase 2 design spaces. Perhaps one argument for a fixed shift in design space is the use of a definitive screening design and second-order model at phase 1 of the experimentation. With this approach, it may be possible to assume wider factor ranges as it is not necessary to assume a purely linear approximation to the response surface.

The authors' method does not require the choice of step size, unlike steepest ascent, as only evaluations of the fitted model along the optimization trace are used. Of course, this means you are relying on model extrapolations, which, as is well known, can be dangerous. The authors mention lack of precision for extrapolated predictions. I think bias beyond the phase 1 design space is of equal, if not greater, importance (Box and Draper, 1959). This is actually another reason for limiting the size of the shift to the phase 2 design space. It is also perhaps another explanation of why use of the optimization trace beats use of the path of steepest ascent in the authors' first simulation study. Although the authors state they "do not consider

Contact: David Woods; D.Woods@southampton.ac.uk; Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ UK

possible model bias", that is likely to be exactly one reason extrapolation with the first-order model fails, and hence is an advantage of using an optimization trace from a second-order model. First principle (physical or mechanistic) modeling is often presented as a safer option for extrapolation (c.f. accelerated stability testing in the pharmaceutical industry, Waterman, 2011) but these models are usually still estimated, or tuned, using data from a particular design space. The reliability of these calibrated models should still be questioned and checked, especially if any effort has been made to estimate the discrepancy between the model and reality (Brynjarsdóttir and O'Hagan, 2014). Extrapolated use of such models in response-surface type experiments is a possible area for future research.

2 Comparisons, and why is the proposed method better?

The selected baseline of a first-order phase 1 design followed by experimentation along the path of steepest ascent but no phase 2 design or more detailed modeling seems something of a straw man. Such a strategy would rarely, if ever be adopted in practice, where experimentation along the path of steepest ascent would usually be followed by a phase 2 design (as briefly mentioned by the authors).

A stronger comparison would be to a full "response surface" strategy of fractional factorial phase 1 design, experimentation along the path of steepest ascent, and a follow-up phase 2 design. Clearly this approach requires more experimentation. To help understand the impact of experiment size, the existing design augmentation strategy (matching the number of phase 2 runs to the steepest ascent stage) could also have been included in the comparison.

Also, matching the number of runs for design augmentation to the number used for steepest ascent may have resulted in some unusual, and perhaps unrealistically small, numbers of runs for efficient estimation of a second-order model; the distribution of the number of runs is not reported in the paper. All of the performance measures reported in Section 4.3 are relative, so we do not know if the absolute performance of the authors' augmentation strategy was satisfactory.

In the second simulation study, the authors confound the choice of phase 1 design (two-level orthogonal design or definitive screening design) with the augmentation strategy (experimentation along the path of steepest ascent or D-optimal augmentation). Hence, it is difficult from this simulation study to decide what is the major reason for the advantage displayed by the authors' approach. The first simulation study assessed various combinations of phase 1 designs and phase 2 optimal augmentations (although not the steepest ascent strategy). This study suggests that the use of a second-order phase 1 design, model and optimization trace is a key advantage of the proposed method, agreeing with my comments above on the dangers of extrapolation with very simple models. However, the authors do not really address this question.

3 Choice of phase 2 design

Interestingly, I was a little unclear about which data was used to estimate the second-order model at phase 2: (a) all the data from phases 1 and 2; (b) just the data from the phase 2 experiment; or (c) the phase 2 data combined with any phase 1 data that lies in the phase 2 design space? I would imagine that often the phase 2 data alone would be insufficient to estimate the second-order model, and that some combination of phase 1 and phase 2 data would be necessary. In the standard response-surface strategy I discussed above, the choice of what data to use to estimate the phase 2 model would probably be based on the size of the shift in the phase 2 design space. A large shift, leading to no overlap with the phase 1 design space, would result in a completely separate phase 2 design and analysis. The authors exclude this case from their simulation studies.

The authors find that choosing the phase 2 design space as the confidence cone or region for the path of steepest ascent or optimization path is less effective than using a shifted cuboid. This does not surprise me, at least in terms of D-efficiency, as in the absence of structural model inadequacy (e.g. the true response surface being more complex than the assumed quadratic), the D-optimal design will aim to include points at the corners of a cuboidal region. For the authors' simulation studies, where the true response surface is quadratic, there may actually be a benefit in terms of estimating the model in finding points within an extended design space that encompasses both the phase 1 and phase 2 spaces. The effectiveness in practice of such an approach would be determined, in part, by the complexity of the true response surface.

The authors' briefly mention other optimality criteria that could be employed, particularly I-optimality, which seeks a design that will enable precise predictions. Here, the design region in which good prediction is required could be tailored towards the path of steepest ascent or optimization trace, and hence restricting the phase 2 design space to the confidence cone/region may be advantageous. It may also be interesting to consider designs tailored to estimation of the location of the maximum response (a nonlinear combination of the model parameters) or criteria that incorporate both variance and bias (see again Box and Draper, 1959).

4 Modern optimization problems

I'll finish by discussing another possible area of future research within this field. Modern optimization problems increasingly seek not just a maximum but a region of the design space within which there is an assurance of meeting specification, for example, the response exceeding a threshold. An important application here is "quality by design" within the pharmaceutical industry[1].

A common approach is to develop a statistical model that allows probabilistic statements concerning the predicted response, and then use this model to identify regions within which the predicted response has a high probability of meeting specification (e.g. Peterson and Yahyah, 2009). Reliable identification of such regions requires an estimated model that allows accurate uncertainty quantification. Building such models requires data from experiments tailored to the estimation problem and hence the authors' strategy may have considerable utility here.

References

Box, G. E. P. and N. R. Draper (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association 54*, 622–654.

Box, G. E. P. and J. S. Hunter (1957). Multi-factor designs for exploring response surfaces. *The Annals of Mathematical Statistics 28*, 195–241.

Box, G. E. P., J. S. Hunter, and W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation and Discovery* (2nd ed.). Hoboken, NJ: Wiley.

Box, G. E. P. and K. B. Wilson (1951). On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society B 13*, 1–45.

Brynjarsdóttir, J. and A. O'Hagan (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems 30*, 114007.

Montgomery, D. C. (2017). *Design and Analysis of Experiments* (9th ed.). Hoboken, NJ: Wiley.

Peterson, J. J. and M. Yahyah (2009). A Bayesian design space approach to robustness and system suitability for pharmaceutical assays and other processes. *Statistics in Biopharmaceutical Research 1*, 441–449.

Waterman, K. C. (2011). The application of the accelerated stability assessment program (ASAP) to quality by design (QbD) for drug product stability. *AAPS PharmSciTech 12*, 932–937.

Wu, C. F. J. and M. S. Hamada (2009). *Experiments: Planning, Analysis and Optimization* (2nd ed.). Hoboken, NJ: Wiley.

[1] Food and Drug Administration Q8(R2) Pharmaceutical Development (`https://www.fda.gov/downloads/drugs/guidances/ucm073507.pdf`)