

Efficient Sampling for Surrogate Modelling

1st Anthony Truelove

Pacific Regional Institute for Marine Energy Discovery (PRIMED)
Institute for Integrated Energy Systems (IESVic) - University of Victoria
Victoria, Canada
wtruelove@uvic.ca

Abstract—[...]

Index Terms—sampling, optimization, surrogate modelling

I. INTRODUCTION

Consider an optimization problem in which the objective function is computationally expensive to evaluate, so much so that applying an optimization algorithm to the objective directly is intractable. For example, in model-based design optimization, the objective function is often high-resolution modelling and simulation software that can be used to assess the performance of candidate designs. However, when individual model runs take anywhere from minutes to days to complete, these models do not scale well in the context of optimization (where algorithms often evaluate the objective thousands of times [or more!] in search of optima). This then motivates the concept of a *surrogate model*: an approximation of a more expensive model that seeks to minimize computational expense while retaining a maximum of model fidelity [1].

Surrogate modelling has taken hold in a variety of domains in recent years. For example, [2] reviews the application of surrogate modelling to sustainable building design. An example is provided in [3] of applying surrogate modelling to the optimal design of combustion systems, while an example is provided in [4] of applying surrogate modelling to the optimal design of wind turbines. Yet another example is provided in [5] of applying surrogate modelling to the optimization of policy and decision-making in the domain of economics. Compound these examples with the utility of machine learning models as surrogates (as mentioned in all of [2], [3], [5]) and it seems that surrogate modelling will remain a useful technique (and hence a relevant topic) for the foreseeable future.

[to build a surrogate model, some amount of expensive sampling is unavoidable, so then ...]

- 1) [robust sampling/surrogate efficiency metric?]
- 2) [a clear winner in terms of sampling scheme?]

II. METHODOLOGY

A. Robust Surrogate Efficiency Metric

The concept of “sampling efficiency” in the context of surrogate modelling is mentioned (but never precisely defined) in all of [6]–[8]. Similarly searching the literature for references to “surrogate efficiency” yields a single relevant result [9] (but again, it is not precisely defined). **Novelty satisfied? I see**

people talking about this, but it’s a bit hand wavy (general benefit vs cost, essentially).

1) Concept:

$$\eta_{SM} \sim \frac{\text{surrogate utility}}{\text{surrogate cost}} \quad (1)$$

2) Surrogate Utility: [...]

$$\text{surrogate utility} \propto \frac{1}{\text{surrogate error}} \quad (2)$$

$$\text{surrogate error} = \mu_{d\text{-MAPE}} + \sigma_{d\text{-MAPE}} \quad (3)$$

where μ and σ are statistics derived from a Monte Carlo approach (for fixed choice of surrogate architecture, sampling scheme, sample size, and problem dimensionality, random combinations of samples of size $S > 0$ and benchmark problems of dimensionality $D > 0$) **I’ll include the code (.py) for these bits, for the sake of reproducibility. As for surrogate architecture, maybe just pick one to keep the degrees of freedom under control? With the benchmark problems I plan on picking (maybe three or four of them), the use case is regression, so perhaps just `sklearn.neural_network.MLPRegressor` to keep it simple(ish)? Pretty much every surrogate is some kind of (prefix)-NN++(etc.) these days anyways ...**

3) Surrogate Cost: [...]

$$\text{surrogate cost} \propto S \quad (4)$$

$S > 0$ is number of sample points (i.e., expensive objective evaluations)

B. Proposed Efficiency Metric Expression

[...]

$$\eta_{SM} = \exp[-S(\mu_{d\text{-MAPE}} + \sigma_{d\text{-MAPE}})] \quad (5)$$

Desirable properties

- $\eta_{SM} \in [0, 1]$ for all $S, \mu_{d\text{-MAPE}}, \sigma_{d\text{-MAPE}} \geq 0$.
- For any $S > 0$, $\eta_{SM} \rightarrow 1$ as $\mu_{d\text{-MAPE}} + \sigma_{d\text{-MAPE}} \rightarrow 0$.
- For any $\mu_{d\text{-MAPE}} + \sigma_{d\text{-MAPE}} > 0$, $\eta_{SM} \rightarrow 0$ as $S \rightarrow \infty$.

C. Sampling Schemes

[...] I think I'll just test a few rather than proposing anything new here. In reviewing the literature, it looks like everyone and their dog is proposing various sampling schemes, but what's lacking is a standardized metric/method for comparing their efficiencies for the purpose of surrogate modelling (as far as I can tell).

D. Benchmark Problems

[...]

Pick some $\mathbb{R}^D \rightarrow \mathbb{R}$ problems, so that results for different problem sizes can be obtained I think I'll just do scalar objectives here, as they're a bit easier to interpret (and tend to be more easily scalable in terms of input dimensionality). I'll punt multi-objective stuff to future work.

III. RESULTS

[...]

IV. DISCUSSION

[...]

V. CONCLUSION

[...]

VI. FUTURE WORK

[...]

REFERENCES

- [1] A. I. J. Forrester, A. Sóbester, and A. J. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons, Ltd, 2008, ISBN: 978-0-47-077080-1.
- [2] P. Westermann and R. Evins, "Surrogate modelling for sustainable building design - a review," *Energy & Buildings*, vol. 198, 2019, DOI: 10.1016/j.enbuild.2019.05.057.
- [3] K. Liu, K. Luo, Y. Cheng, A. Liu, H. Li, J. Fan, and S. Balachandrar, "Surrogate modeling of parameterized multi-dimensional premixed combustion with physics-informed neural networks for rapid exploration of design space," *Combustion and Flame*, vol. 258, 2023, DOI: 10.1016/j.combustflame.2023.113094.
- [4] R. Haghi and C. Crawford, "Surrogate models for the blade element momentum aerodynamic model using non-intrusive polynomial chaos expansions," *Wind Energy Science*, 2022, DOI: 10.5194/wes-7-1289-2022.
- [5] S. van der Hoog, "Surrogate modelling in (and of) agent-based models: A prospectus," *Computational Economics*, vol. 53, 2018, DOI: 10.1007/s10614-018-9802-0.
- [6] W. Gong and Q. Duan, "An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (asmopde)," *Environmental Modelling & Software*, vol. 95, 2017, DOI: 10.1016/j.envsoft.2017.05.005.
- [7] P. Westermann and R. Evins, "Adaptive sampling for building simulation surrogate model derivation using the lola-voronoi algorithm," *Proceedings of the 16th IBPSA Conference*, 2019, DOI: 10.26868/25222708.2019.211232.
- [8] J. Yin, W. Lu, X. Xin, and L. Zhang, "Application of monte carlo sampling and latin hypercube sampling methods in pumping schedule design during establishing surrogate model," *Proceedings of International Symposium on Water Resource and Environmental Protection*, 2011, DOI: .
- [9] T. Casper, U. Römer, and S. Schöps, "Determination of bond wire failure probabilities in microelectronic packages," *Proceedings of 22nd International Workshop on Thermal Investigations of ICs and Systems*, DOI: 10.1109/THERMINIC.2016.7748645.