Hyperparameter Optimization of Deep Neural Networks With Application in Medical Device Manufacturing

Gautham Sunder[1], Thomas Albrecht[2], Christopher J. Nachtsheim[1]
[1]*Carlson School of Management, University of Minnesota, Minneapolis, MN, United States*
[2]*Atlassian Corporation, United States*

Hyperparameter Optimization (HPO)

"Identifying the optimal hyperparameter values that minimize the validation loss of a Deep Neural Network (DNN)"

Characteristics:

- Choice of hyperparameter is critical
- Function complexity is unknown
- Active hyperparameters are unknown
- Noisy response

Response Surface Optimization(RSO)

RSO methods, specifically, Bayesian Optimization (BO) is a popular HPO strategy

Assumptions:

BO, through the choice of kernels, assumes that the response surface is nonlinear and complex

Gaps:

- The assumptions on response surface complexity is not validated
  - When second-order, classical response surface optimization (C-RSO) is demonstrably more efficient
- BO based HPO strategies tend to overfit the validation data

Compound-RSO

	C-RSO	Compound-RSO	BO
Response Surface Complexity	Quadratic	Quadratic with validations on complexity assumptions	Complex
Starting Design	Optimal Design	Compound Uniform Design	Space-filling
Optimization Strategy	Estimate the function over the entire design space	Use Lack of Fit test to choose between BO and C-RSO	Estimate the function with high precision in the region of optimum

Stage 1
Definitive Screening Design (DSD)

↓

Stage 2
Compound Uniform Designs

↓

Lack of Fit Test

Second-Order → Stage 3
Adaptive-RSO

Complex → Stage 3
Bayesian Optimization

Compound Uniform Design

$$R_{opt} = wD_{eff} + (1-w)U_{eff}$$

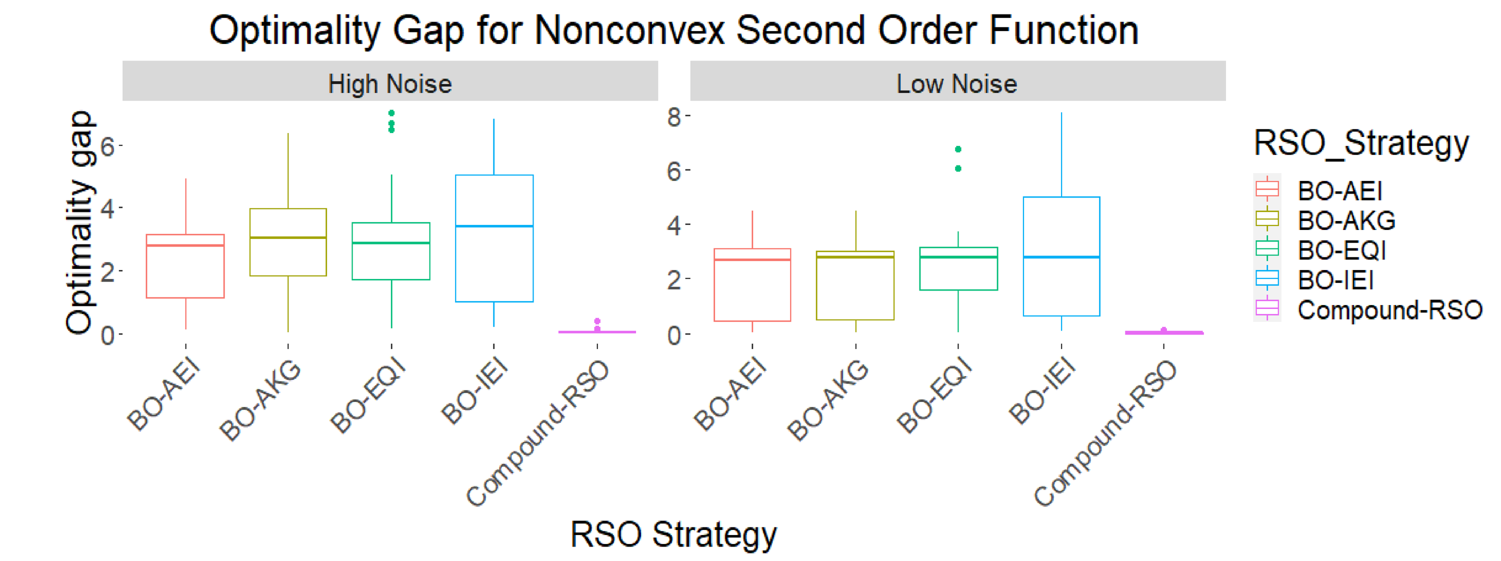
D_{eff}: Bayesian D-efficiency
U_{eff}: Uniform efficiency
w: weighting function

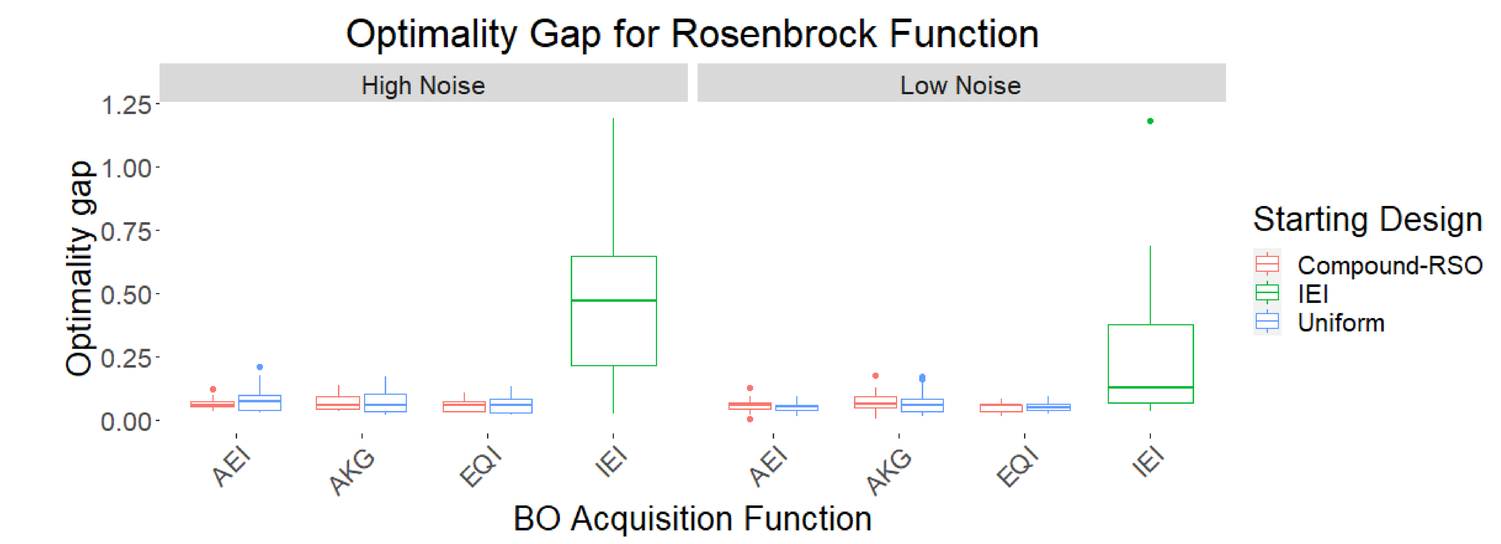
Adaptive-RSO

$$A_{opt} = wD_{eff} + (1-w)I(\lambda)_{eff}$$

D_{eff}: Bayesian D-efficiency
$I(\lambda)_{eff}$: I (λ) efficiency
w: weighting function

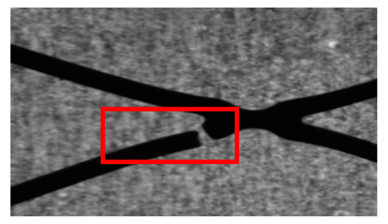
Simulation Study





HPO of DNN Case Study

Boston Scientific
Advancing science for life™



HPO Setting	Loss	Defect 1	Defect 2	Defect 3	Defect 4
Predicted Optimum 1	0.115	99.8%	98.7%	98.34%	94.16%
Predicted Optimum 2	0.130	99.36%	98.7%	98.88%	92.86%
Poor	5.533	76.5%	76.2%	79.7%	76.7%

Conclusion

Compound-RSO outperforms Standard-BO

+ Second Order: Compound-RSO is superior to BO
+ Complex: Compound-RSO is comparable to BO

Contributions

+ Exact robust-supersaturated design for a full second-order polynomial function
+ A principled RSO strategy which estimates the response surface complexity