14th World Congress on Computational Mechanics (WCCM)
ECCOMAS Congress 2020
19–24 July 2020, Paris, France

SEQUENTIAL OPTIMAL EXPERIMENTAL DESIGN USING REINFORCEMENT LEARNING WITH POLICY GRADIENT

Wanggang Shen[1] **and Xun Huan**[2]

[1] University of Michigan, 1231 Beal Ave Room G029, Ann Arbor, MI 48109, USA,
wgshen@umich.edu

[2] University of Michigan, 1231 Beal Ave Room 2033, Ann Arbor, MI 48109, USA,
xhuan@umich.edu, http://rxhuan.com

Key Words: Optimal Experimental Design, Uncertainty Quantification, Reinforcement Learning, Bayesian Inference

Experimental data play a crucial role for developing and refining models in computational mechanics, and well-chosen experiments can provide substantial resource savings. Optimal experimental design (OED) is the research area that seeks to quantify and maximize the value of experiments and their data. When multiple experiments can be conducted in sequence, common current design practices use suboptimal approaches: batch (open-loop) design that chooses all experiments simultaneously with no updates from newly acquired data, and greedy (myopic) design that optimally selects the next experiment without accounting for future consequences and effects. In contrast, the *sequential optimal experimental design (sOED)* formulation is free of these limitations, and achieves the true optimality with respect to the entire design horizon.

With the goal of acquiring data for learning unknown parameters in physical systems, we develop a rigorous Bayesian formulation for sOED using an objective that incorporates a measure of information gain. A sequential design problem thus involves finding good *policies*—that is, functions that instruct design choices depending on the current state. We develop numerical tools for finding the optimal policies, targeting finite horizon design problems while accommodating nonlinear models with continuous parameter, design, and observation spaces. We employ policy gradient methods from reinforcement learning, and directly parameterize the policies and value functions by neural networks—thus adopting an actor-critic approach—and improve them using gradient estimates produced from simulated design and observation sequences. The overall method is demonstrated on an algebraic benchmark and a sensor movement application for contaminant source inversion in a convection-diffusion field. The results provide intuitive insights on the benefits of feedback and lookahead, and indicate substantial computational advantages compared to previous numerical approaches based on approximate dynamical programming.