

Autonomous X-Ray Scattering Experiments under Uncertainty

Marcus Noack¹, Kevin G. Yager², Masafumi Fukuto³, Gregory S. Doerk², Ruipeng Li³, and James Sethian^{1,4}

¹The Center for Advanced Mathematics for Energy Research Applications (CAMERA), Lawrence Berkeley National Laboratory, Berkeley, CA 94720

²Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973

³National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973

⁴Department of Mathematics, University of California, Berkeley, 94720

A core paradigm in experimental materials science is the iterative exploration of the multi-dimensional parameter spaces that underlie materials makeup, synthesis, and processing. In order to establish synthesis-structure relationships and structure-property relationships, scientists iteratively test different material compositions, processing protocols, and environmental conditions. In each ‘loop’ of this process, the scientist analyzes the trends in the available data, and, exploiting their own domain knowledge, selects useful follow-up measurements. The traditional (manual) mode of data collection is costly in the sense that it consumes the attention of a human expert. Moreover, manual experimentation may be inefficient since there is no guarantee that the measurement selection is optimal, especially if the underlying parameter space is vast and high-dimensional. Finally, a purely manual approach lacks a quantitative metric for deciding when to terminate the investigation; it is left to human judgment whether ‘enough’ data has been collected. This bottleneck is becoming increasingly troublesome, when measured against the rapid advances in instrumentation, data gathering, and analysis tools. Modern scientific instruments feature improved automation (e.g. robotic sample handling), faster data collection (e.g. improved detectors), and faster response (e.g. faster data workup), thereby enabling ever-greater data collection rates. This massive automation allows for more ambitious and complex scientific problems to be investigated, since larger parameter ranges can be studied, these can be sampled more densely, and because the data collection itself is more systematic and reproducible. However, the enormous potential of improved acquisition rates underscore the need for corresponding improvements in the ability of machines to develop and update experimental measurement plans at a speed commensurate to that of the measurement. What is needed are generalizable decision-making algorithms that autonomously drive experiments without human intervention.

In this work, we develop a general approach for autonomous experimentation that automatically selects measurements from the parameter space that defines a given scientific problem. The proposed method, SMART (Surrogate Model Autonomous expeRimentT), creates a surrogate model using Gaussian Process Regression (GPR) to steer autonomous x-ray scattering experiments. GPR defines a surrogate model and a variance function over the domain. The maxima of this variance function are optimal suggestions for where to place subsequent measurements, since measure-

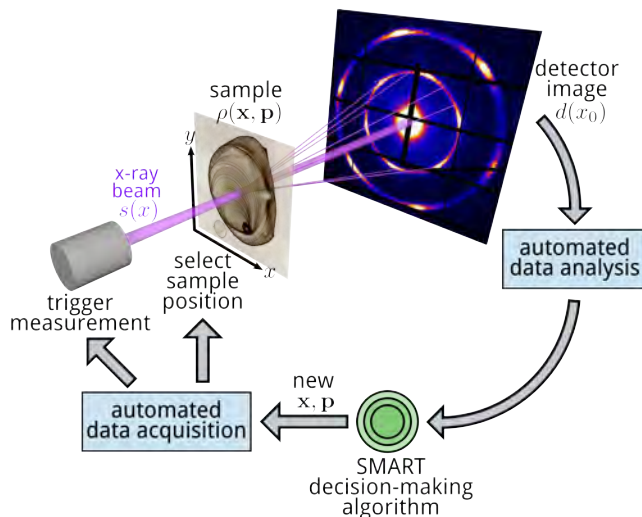


Fig. 1. The autonomous experiment loop. In every iteration the algorithm uses all available data to create a surrogate model and a variance function which is then used to find the optimal location for the next measurement.

ments in high-error parts of the parameter space will necessarily decrease the overall surrogate model uncertainty the most, and are thus of high-value from an information-content point of view. The method we present is agnostic to the underlying meaning of the parameters: any synthesis, processing, or environmental parameter can be included and even combined. Thus, any experiment that can be formulated in terms of an abstract parameter space of possible measurements can be steered with the help of the proposed method, in which the underlying dependency will be captured by the surrogate model. We are also presenting how to treat costs of measurements by defining a local cost function. This can be beneficial when the total number of measurements is not the quantity of interest, but rather time or the amount of a material. Additionally, we discuss how to take advantage of prior knowledge the experimenter might have about the experiment. For instance, knowledge about “steepness” of the model can be included by assessing gradients of the surrogate model. Also, prior knowledge about differentiability and bounds can be included if desired.

The result is an autonomous experiment loop (Fig. 1) that helps explore vast parameter spaces without any human interference, thereby saving the experimenters valuable time and resources, and avoiding unintentional bias.