Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2020-0060

Vol. **29**, No. 4, pp. 1152-1185
April 2021

Cholesky-Based Experimental Design for Gaussian Process and Kernel-Based Emulation and Calibration

Helumt Harbrecht[1], John D. Jakeman[2,*] and Peter Zaspel[3]

[1] *Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland.*
[2] *Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, USA.*
[3] *Computer Science and Electrical Engineering, Jacobs University Bremen, Bremen, Germany.*

Received 27 March 2020; Accepted (in revised version) 30 October 2020

Abstract. Gaussian processes and other kernel-based methods are used extensively to construct approximations of multivariate data sets. The accuracy of these approximations is dependent on the data used. This paper presents a computationally efficient algorithm to greedily select training samples that minimize the weighted L^p error of kernel-based approximations for a given number of data. The method successively generates nested samples, with the goal of minimizing the error in high probability regions of densities specified by users. The algorithm presented is extremely simple and can be implemented using existing pivoted Cholesky factorization methods. Training samples are generated in batches which allows training data to be evaluated (labeled) in parallel. For smooth kernels, the algorithm performs comparably with the greedy integrated variance design but has significantly lower complexity. Numerical experiments demonstrate the efficacy of the approach for bounded, unbounded, multi-modal and non-tensor product densities. We also show how to use the proposed algorithm to efficiently generate surrogates for inferring unknown model parameters from data using Bayesian inference.

AMS subject classifications: 62F15, 62K20, 65D05
Key words: Experimental design, active learning, Gaussian process, radial basis function, uncertainty quantification, Bayesian inference.

1 Introduction

This article considers the approximation of a function $u:\Gamma\to\mathbb{R}$ using kernel-based interpolation. We are particularly interested in constructing approximations $\Pi_V(u)\in V\subset\mathcal{V}_\omega$

*Corresponding author. *Email addresses:* helmut.harbrecht@unibas.ch (H. Harbrecht), jdjakem@sandia.gov (J. D. Jakeman), p.zaspel@jacobs-university.de (P. Zaspel)

of u, which have a small approximation error

$$\epsilon_\omega(u,V,p) := \left(\int_\Gamma |u(\boldsymbol{y}) - (\Pi_V u)(\boldsymbol{y})|^p \omega(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}\right)^{1/p}, \tag{1.1}$$

when measured with respect to a measure $\omega : \Gamma \to \mathbb{R}$, which introduces the weighting on $\mathcal{V}_\omega$. In other words, we aim to minimize the approximation error measured using the ω-weighted $L^p(\Gamma)$-norm

$$\|e\|_{L^p_\omega(\Gamma)} := \left(\int_\Gamma |e(\boldsymbol{y})|^p \omega(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}\right)^{1/p}.$$

While we are in principle not bound to a specific choice of p, we will consider $p=2,4,6$ in this work.

Such weighted function approximation is often essential for uncertainty quantification (UQ). In this setting, u usually corresponds to the parameter-to-solution map of a numerical model – e.g. a partial differential equation (PDE) – the argument $\boldsymbol{y}$ is a finite-dimensional random variable, and ω is its probability density function (PDF). Approximations are built to reduce the number of computationally expensive simulations (i.e. point evaluations of u) needed to estimate statistics.

The techniques discussed in this paper are not limited to UQ of PDEs. The method presented can be utilized to build approximations of any data generating information source. Furthermore, ω need not be a probability measure. For example, ω can be the unnormalized posterior of Bayesian inference, or a biasing measure used for importance sampling that explores important regions of Γ.

Numerous techniques have been developed to build approximations of expensive information sources. Within the computational science and engineering community, some of the most widely adopted methods for approximating models are those based on generalized polynomial chaos expansions [9, 39], sparse grid approximation [23, 38], Gaussian process models [27, 44], low-rank tensor decompositions [12, 24] and neural networks [45, 46]. These methods can be very efficient when building approximations of functions parameterized by independent random variables. However, there is a dearth of algorithmic options for when the variables are dependent.

The objective of this article is to propose a methodology to efficiently generate nested and greedy-pseudo-optimal sample designs that minimize the number of evaluations of u needed to build an approximation with a pre-specified accuracy measured in the L^p_ω norm. The algorithm is pseudo optimal because it minimizes an upper bound on the approximation error. The algorithm is greedy because it both selects from a candidate set which discretizes the design space and chooses points one at a time.

Accurate approximations can be built without tailoring the sampling and approximation strategy to the measure ω. However such approaches, called *domination methods* [2, 14], are sub-optimal and require a larger number of evaluations of u than methods that consider ω [15]. Instead of building an approximation which minimizes the error