Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2020-0151

Vol. **28**, No. 5, pp. 1812-1837
November 2020

Multifidelity Data Fusion via Gradient-Enhanced Gaussian Process Regression

Yixiang Deng[1,2], Guang Lin[3,4] and Xiu Yang[5,*]

[1] *School of Engineering, Brown University, RI 02912, USA.*
[2] *Division of Applied Mathematics, Brown University, RI 02912, USA.*
[3] *Department of Mathematics, Purdue University, IN 47907, USA.*
[4] *School of Engineering, Purdue University, IN 47907, USA.*
[5] *Department of Industrial and Systems Engineering, Lehigh University, PA 18015, USA.*

Received 3 August 2020; Accepted (in revised version) 25 October 2020

Abstract. We propose a data fusion method based on multi-fidelity Gaussian process regression (GPR) framework. This method combines available data of the quantity of interest (QoI) and its gradients with different fidelity levels, namely, it is a Gradient-enhanced Cokriging method (GE-Cokriging). It provides the approximations of both the QoI and its gradients *simultaneously* with uncertainty estimates. We compare this method with the conventional multi-fidelity Cokriging method that does not use gradients information, and the result suggests that GE-Cokriging has a better performance in predicting both QoI and its gradients. Moreover, GE-Cokriging even shows better generalization result in some cases where Cokriging performs poorly due to the singularity of the covariance matrix. We demonstrate the application of GE-Cokriging in several practical cases including reconstructing the trajectories and velocity of an underdamped oscillator with respect to time simultaneously, and investigating the sensitivity of power factor of a load bus with respect to varying power inputs of a generator bus in a large scale power system. Although GE-Cokriging requires slightly higher computational cost than Cokriging in some cases, the comparison of the accuracy shows that this cost is worthwhile.

AMS subject classifications: 60G15, 65D10

Key words: Gaussian process regression, multifidelity Cokriging, gradient-enhanced, integral-enhanced.

---

*Corresponding author. *Email addresses:* yixiang_deng@brown.edu (Y. Deng), guanglin@purdue.edu (G. Lin), xiy518@lehigh.edu (X. Yang)

1 Introduction

Gaussian process (GP) is one of the most well studied stochastic processes in probability and statistics. Given the flexible form of data representation, GP is a powerful tool for classification and regression, and it is widely used in probabilistic scientific computing, engineering design, geostatistics, data assimilation, machine learning, etc. In particular, given a data set comprising input/output pairs of locations and quantity of interest (QoI), *GP regression* (GPR, also known as *Kriging*), can provide a prediction along with a mean squared error (MSE) estimate of the QoI at any location. Alternatively, from the Bayesian perspective, GPR identifies a Gaussian random variable at any location with a posterior mean (corresponding to the prediction) and variance (corresponding to the MSE). Generally speaking, the larger the given data set size is, the closer the GPR's posterior mean is to the ground truth and the smaller the posterior variance is.

In many practical problems, obtaining a large amount of data can be difficult because of the limitation of resources. There are several approaches to augment the data set in different manners. For example, the original Cokriging method exploits the correlation between multiple QoIs in the geostatistical study, e.g., the correlation between temperature and precipitation [1–3], or that between near-surface soil density and the gravity-gradient [4], to improve the accuracy of prediction. Later, the Cokriging method was extended to utilizing correlation between the same QoI from models with different fidelities [5–8]. This GP-based multi-fidelity method is very useful in scientific computing, because low-fidelity models, e.g., coarse-grained molecular dynamics [9, 10], Reynolds-average Navier-Stokes equations [11, 12], numerical simulations on coarse grids, are often used with high-fidelity models, e.g., molecular dynamics, full Navier-Stokes equations, numerical simulations on fine grids [13], in optimization, uncertainty quantification (UQ), control [14], variable-fidelity quantum mechanical calculations of bandgaps of solids [15], etc. In these tasks, the multi-fidelity method leverages low-fidelity models for speedup, while uses a high-fidelity model to establish accuracy and/or convergence guarantees. Moreover, the empirical statistics of simulation results from stochastic scientific computing models can be used to construct single- or multi-fidelity GP models [16–18]. In this work, Cokriging refers to the GP-based multi-fidelity approach.

Another important approach to enlarge the data set is to use gradient information of the QoI. This approach can be categorized as Cokriging because the QoI and its gradients are variables of different species. The idea of incorporating derivatives or gradients to optimize Bayesian prediction was proposed by Morris et al. [19]. The *gradient-enhanced Kriging* (GE-Kriging) method, also referred to as Gradient-based Kriging in some literature, has been widely investigated in areas such as computational fluid dynamics, especially in aerodynamics optimization problems [20–23]. Incorporating gradient information in different ways, this method consists of direct and indirect approaches. The former uses the gradient information through an augmented covariance matrix [24], while the latter approximates the gradient via finite-difference method [23, 25]. The *gradient-enhanced Cokriging* (GE-Cokriging) method in [26] refers to a GE-Kriging method that uses a differ-