Journal of Computational Physics 335 (2017) 736–746

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Inferring solutions of differential equations using noisy multi-fidelity data

Maziar Raissi [a], Paris Perdikaris [b], George Em Karniadakis [a,*]

[a] Division of Applied Mathematics, Brown University, Providence, RI, USA
[b] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 17 July 2016
Received in revised form 21 January 2017
Accepted 25 January 2017
Available online 2 February 2017

Keywords:
Machine learning
Integro-differential equations
Multi-fidelity modeling
Uncertainty quantification

ABSTRACT

For more than two centuries, solutions of differential equations have been obtained either analytically or numerically based on typically well-behaved forcing and boundary conditions for well-posed problems. We are changing this paradigm in a fundamental way by establishing an interface between probabilistic machine learning and differential equations. We develop data-driven algorithms for general linear equations using Gaussian process priors tailored to the corresponding integro-differential operators. The only observables are scarce noisy multi-fidelity data for the forcing and solution that are not required to reside on the domain boundary. The resulting predictive posterior distributions quantify uncertainty and naturally lead to adaptive solution refinement via active learning. This general framework circumvents the tyranny of numerical discretization as well as the consistency and stability issues of time-integration, and is scalable to high-dimensions.

1. Introduction

Nearly two decades ago a visionary treatise by David Mumford anticipated that stochastic methods will transform pure and applied mathematics in the beginning of the third millennium, as probability and statistics will come to be viewed as the natural tools to use in mathematical as well as scientific modeling [1]. Indeed, in recent years we have been witnessing the emergence of a data-driven era in which probability and statistics have been the focal point in the development of disruptive technologies such as probabilistic machine learning [2,3]. Only to verify Mumford's predictions, this wave of change is steadily propagating into applied mathematics, giving rise to novel probabilistic interpretations of classical deterministic scientific methods and algorithms. This new viewpoint offers an elegant path to generalization and enables computing with probability distributions rather than solely relying on deterministic thinking. In particular, in the area of numerical analysis and scientific computing, the first hints of this paradigm shift were clearly manifested in the thought-provoking work of Diaconis [4], tracing back to Poincaré's courses on probability theory [5]. This line of work has recently inspired resurgence in probabilistic methods and algorithms [6–8] that offer a principled and robust handling of uncertainty due to model inadequacy, parametric uncertainties, and numerical discretization/truncation errors. In particular, several statistical inference techniques have been reported in [9–12] for constructing probabilistic time-stepping schemes for systems of ordinary differential equations (e.g., systems arising after a partial differential equation is discretized in space). In the same spirit, the work of [13–16] has highlighted the possibility of solving linear partial differential equations and quantifying parameter and

* Corresponding author.
E-mail addresses: maziar_raissi@brown.edu (M. Raissi), george_karniadakis@brown.edu (G.E. Karniadakis).

http://dx.doi.org/10.1016/j.jcp.2017.01.060
0021-9991/

discretization uncertainty using Gaussian process priors. These developments are defining a new area of scientific research in which probabilistic machine learning and classical scientific computing coexist in unison, providing a flexible and general platform for Bayesian reasoning and computation. In this work, we exploit this interface by developing a novel Bayesian inference framework that enables learning from (noisy) data and equations in a synergistic fashion.

2. Problem setup

We consider general linear integro-differential equations of the form

$$\mathcal{L}_x u(x) = f(x), \tag{1}$$

where x is a D-dimensional vector that includes spatial or temporal coordinates, $\mathcal{L}_x$ is a linear operator, $u(x)$ denotes an unknown solution to the equation, and $f(x)$ represents the external force that drives the system. We assume that $f_L := f$ is a complex, expensive to evaluate, "black-box" function. For instance, f_L could represent force acting upon a physical system, the outcome of a costly experiment, the output of an expensive computer code, or any other unknown function. We assume limited availability of high-fidelity data for f_L, denoted by $\{\boldsymbol{x}_L, \boldsymbol{y}_L\}$, that could be corrupted by noise $\boldsymbol{\epsilon}_L$, i.e., $\boldsymbol{y}_L = f_L(\boldsymbol{x}_L) + \boldsymbol{\epsilon}_L$. In many cases, we may also have access to supplementary sets of less accurate models $f_\ell,\ \ell = 1, \ldots, L-1$, sorted by increasing level of fidelity, and generating data $\{\boldsymbol{x}_\ell, \boldsymbol{y}_\ell\}$ that could also be contaminated by noise $\boldsymbol{\epsilon}_\ell$, i.e., $\boldsymbol{y}_\ell = f_\ell(\boldsymbol{x}_\ell) + \boldsymbol{\epsilon}_\ell$. Such data may come from simplified computer models, inexpensive sensors, or uncalibrated measurements. In addition, we also have a small set of data on the solution u, denoted by $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$, perturbed by noise $\boldsymbol{\epsilon}_0$, i.e., $\boldsymbol{y}_0 = u(\boldsymbol{x}_0) + \boldsymbol{\epsilon}_0$, sampled at scattered spatio-temporal locations, which we call anchor points to distinguish them from boundary or initial values. Although they could be located on the domain boundaries as in the classical setting, this is not a requirement in the current framework as solution data could be partially available on the boundary or in the interior of either spatial or temporal domains. Here, we are not primarily interested in estimating f. We are interested in estimating the unknown solution u that is related to f through the linear operator $\mathcal{L}_x$. For example, consider a bridge subject to environmental loading. In a two-level of fidelity setting (i.e., $L = 2$), suppose that one could only afford to collect scarce but accurate (high-fidelity) measurements of the wind force $f_2(x)$ acting upon the bridge at some locations. In addition, one could also gather samples by probing a cheaper but inaccurate (low-fidelity) wind prediction model $f_1(x)$ at some other locations. How could this noisy data be combined to accurately estimate the bridge displacements $u(x)$ under the laws of linear elasticity? What is the uncertainty/error associated with this estimation? How can we best improve that estimation if we can afford another observation of the wind force? Quoting Diaconis [4], "once we allow that we don't know f, but do know some things, it becomes natural to take a Bayesian approach".

3. Solution methodology

The basic building blocks of the Bayesian approach adopted here are Gaussian process (GP) regression [17,18] and auto-regressive stochastic schemes [19,21]. This choice is motivated by the Bayesian non-parametric nature of GPs, their analytical tractability properties, and their natural extension to the multi-fidelity settings that are fundamental to this work. In particular, GPs provide a flexible prior distribution over functions, and, more importantly, a fully probabilistic workflow that returns robust posterior variance estimates which enable adaptive refinement and active learning [22–24]. The framework we propose is summarized in Fig. 1 and is outlined in the following.

Inspired by [19,21], we will present the framework considering two-levels of fidelity (i.e. $L = 2$), although generalization to multiple levels is straightforward. Let us start with the auto-regressive model $u(x) = \rho u_1(x) + \delta_2(x)$, where $\delta_2(x)$ and $u_1(x)$ are two independent Gaussian processes [17–19,21] with $\delta_2(x) \sim \mathcal{GP}(0, g_2(x, x'; \theta_2))$ and $u_1(x) \sim \mathcal{GP}(0, g_1(x, x'; \theta_1))$. Here, $g_1(x, x'; \theta_1), g_2(x, x'; \theta_2)$ are covariance functions, θ_1, θ_2 denote their hyper-parameters, and ρ is a cross-correlation parameter to be learned from the data (see Sec. 3.1). Then, one can trivially obtain

$$u(x) \sim \mathcal{GP}(0, g(x, x'; \theta)), \tag{2}$$

with $g(x, x'; \theta) = \rho^2 g_1(x, x'; \theta_1) + g_2(x, x'; \theta_2)$, and $\theta = (\theta_1, \theta_2, \rho)$. The key observation here is that the derivatives and integrals of a Gaussian process are still Gaussian processes. Therefore, given that the operator $\mathcal{L}_x$ is linear, we obtain

$$f(x) \sim \mathcal{GP}(0, k(x, x'; \theta)), \tag{3}$$

with

$$k(x, x'; \theta) = \mathcal{L}_x \mathcal{L}_{x'} g(x, x'; \theta). \tag{4}$$

Similarly, we arrive at the auto-regressive structure $f(x) = \rho f_1(x) + \gamma_2(x)$ on the forcing, where $\gamma_2(x) = \mathcal{L}_x \delta_2(x)$, and $f_1(x) = \mathcal{L}_x u_1(x)$ are consequently two independent Gaussian processes with $\gamma_2(x) \sim \mathcal{GP}(0, k_2(x, x'; \theta_2))$, $f_1(x) \sim \mathcal{GP}(0, k_1(x, x'; \theta_1))$. Furthermore, for $\ell = 1, 2$, $k_\ell(x, x'; \theta_\ell) = \mathcal{L}_x \mathcal{L}_{x'} g_\ell(x, x'; \theta_\ell)$.