Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 304 (2016) 81–101
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Data-driven computational mechanics

T. Kirchdoerfer, M. Ortiz*

Graduate Aerospace Laboratories, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA, 91125, USA

Received 10 September 2015; received in revised form 19 January 2016; accepted 1 February 2016
Available online 8 February 2016

Abstract

We develop a new computing paradigm, which we refer to as data-driven computing, according to which calculations are carried out directly from experimental material data and pertinent constraints and conservation laws, such as compatibility and equilibrium, thus bypassing the empirical material modeling step of conventional computing altogether. Data-driven solvers seek to assign to each material point the state from a prespecified data set that is closest to satisfying the conservation laws. Equivalently, data-driven solvers aim to find the state satisfying the conservation laws that is closest to the data set. The resulting data-driven problem thus consists of the minimization of a distance function to the data set in phase space subject to constraints introduced by the conservation laws. We motivate the data-driven paradigm and investigate the performance of data-driven solvers by means of two examples of application, namely, the static equilibrium of nonlinear three-dimensional trusses and linear elasticity. In these tests, the data-driven solvers exhibit good convergence properties both with respect to the number of data points and with regard to local data assignment. The variational structure of the data-driven problem also renders it amenable to analysis. We show that, as the data set approximates increasingly closely a classical material law in phase space, the data-driven solutions converge to the classical solution. We also illustrate the robustness of data-driven solvers with respect to spatial discretization. In particular, we show that the data-driven solutions of finite-element discretizations of linear elasticity converge jointly with respect to mesh size and approximation by the data set.

Keywords: Data science; Big data; Approximation theory; Scientific computing

1. Introduction

Boundary-value problems in science and engineering typically combine two types of equations: (i) *Conservation laws*, which derive from universal principles such as conservation of momentum or energy and are, therefore, uncertainty-free; and (ii) *material laws*, formulated through physical modeling based on experimental observation, that are, therefore, empirical and uncertain. The prevailing classical computational paradigm has been to calibrate empirical material models using observational data and then use the calibrated material model in calculations. This process of modeling *a fortiori* adds error and uncertainty to the solutions, especially in systems with high-dimensional phase spaces and complex behavior. This modeling error and uncertainty arise from imperfect knowledge of the

* Correspondence to: 1200 E. California Blvd., MC 105-50, Pasadena, CA, 91125, USA.
E-mail address: ortiz@caltech.edu (M. Ortiz).

http://dx.doi.org/10.1016/j.cma.2016.02.001
0045-7825/

functional form of the material laws, the phase space in which they are defined, and from scatter and noise in the experimental data. Furthermore, often the models used to fit the data are *ad hoc*, without a clear basis in physics or a mathematical criterion for their selection, and thus the process of modeling is mired in empiricism and arbitrariness. Indeed, the entire process of empirical material modeling, and model validation thereof, is open-ended and no rigorous mathematical theory exists to date that makes it precise and quantitative.

Previous work has been carried out with a view to incorporating observational data into boundary-value problem solution methodologies, but typically with the aim of parametric identification, or augmenting and automating, rather than replacing, the use and generation of material models. Material informatics uses database techniques to first identify parameters of correlation and then use machine-learning regression techniques [1] to ultimately provide predictive quantitative models [2]. Principal-component analysis provides methods of dimensional reduction that allow such modeling techniques to be applied [3]. These approaches have been extended to the generation of multi-scale modeling correlations between macroscopic and microscopic constitutive properties [4–8].

These efforts, and others like them, may be understood as instances of *Data Science*, the extraction of 'knowledge' from large volumes of unstructured data [9,10]. Data science often requires sorting through big-data sets and extracting 'insights' from these data. Data science uses data management, statistics and machine learning to derive mathematical models for subsequent use in decision making. Data Science currently influences primarily fields such as marketing, advertising, finance, social sciences, security, policy, medical informatics, whereas the full potential of Data Science as it relates to high-performance scientific computing is yet to be realized. Despite these limitations, reference to Data Science does effectively serve the purpose of bringing data and artificial intelligence considerations to the forefront.

In this work, we propose a new and different paradigm, which we refer to as *data-driven computing*, consisting of formulating calculations *directly* from experimental material data and pertinent essential constraints and conservation laws, thus bypassing the empirical material modeling step of conventional computing altogether. In this new computing paradigm, essential constraints and conservation laws such as compatibility and equilibrium remain unchanged, as do all the numerical schemes used in their discretization, such as finite elements, time-integrators, *etc*. Such conservation laws confer mathematical structure to the calculations, and this mathematical structure carries over to the present data-driven paradigm. However, in sharp contrast to conventional computing, in data-driven computing the experimental material-data points are used directly in calculations *in lieu* of an empirical material model. In this manner, material modeling empiricism, error and uncertainty are eliminated entirely and no loss of experimental information is incurred. Specifically, data-driven solvers seek to assign to each material point the state from a prespecified data set that is closest to satisfying the conservation laws. Equivalently, data-driven solvers aim to find the state satisfying the conservation laws that is closest to the data set. The resulting data-driven problem thus consists of the minimization of a distance function to the data set in phase space subject to the satisfaction of essential constraints and conservation laws.

We provide an efficient implementation of data-driven computing and demonstrate the practicality of the approach by means of two examples of application, namely, the static equilibrium of nonlinear three-dimensional trusses and linear elasticity. In these tests, the data-driven solvers exhibit good convergence properties both with respect to the number of data points and with regard to local data assignment. The variational structure of the data-driven problem also renders it amenable to analysis. We show that, as the data set approximates increasingly closely a classical material law in phase space, the data-driven solutions converge to the classical solutions. We also illustrate the robustness of data-driven solvers with respect to spatial discretization. In particular, we show that the data-driven solutions of finite-element discretizations of linear elasticity converge jointly with respect to mesh size and approximation by the data set. The mathematical analysis is also suggestive of a number of generalizations and extensions of the data-driven computing paradigm.

2. Truss structures

We proceed to introduce and motivate the general approach with the aid of a simple non-linear elastic truss problem. Trusses are assemblies of articulated bars that deform in uniaxial tension or compression. Therefore, the material behavior of a bar is characterized by a particularly simple relation between uniaxial strain ε and uniaxial stress σ. We refer to the space of pairs (ε, σ) as *phase space*. We assume that the behavior of the material of each bar $e = 1, \ldots, m$, where m is the number of bars in the truss, is characterized by – possibly different – sets E_e of pairs (ε, σ), or *local*