Engineering Applications of Artificial Intelligence 68 (2018) 110–120

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Solving stochastic differential equations through genetic programming and automatic differentiation

Waldir Jesus de Araujo Lobão [a], Marco Aurélio Cavalcanti Pacheco [b], Douglas Mota Dias [c], Ana Carolina Alves Abreu [b,*]

[a] National School of Statistical Sciences of IBGE (Brazilian Institute of Geography and Statistics), Rua André Cavalcanti, 106 - Rio de Janeiro-RJ, 20231-050, Brazil
[b] Department of Electrical Engineering – PUC-Rio (Pontifical Catholic University of Rio de Janeiro), Rua Marquês de São Vicente, 225 - Rio de Janeiro-RJ, 22451-900, Brazil
[c] Department of Electronics and Telecommunications, Faculty of Engineering—UERJ (Rio de Janeiro State University), RuaSão Francisco Xavier, 524 - Rio de Janeiro-RJ, 20550-900, Brazil

A R T I C L E I N F O

MSC:
B0200K
C1100K
C4100K
C7000K
C7400K

Keywords:
Evolutionary algorithm
Genetic programming
Automatic differentiation
Stochastic differential equations
Stochastic calculus
Geometric Brownian motion

A B S T R A C T

This paper investigates the potential of evolutionary algorithms, developed using a combination of genetic programming and automatic differentiation, to obtain symbolic solutions to stochastic differential equations. Using the MATLAB programming environment and based on the theory of stochastic calculus, we develop algorithms and conceive a new methodology of resolution. Relative to other methods, this method has the advantages of producing solutions in symbolic form and in continuous time and, in the case in which an equation of interest is completely unknown, of offering the option of algorithms that perform the specification and estimation of the solution to the equation via a real database. The last advantage is important because it determines an appropriate solution to the problem and simultaneously eliminates the difficult task of arbitrarily defining the functional form of the stochastic differential equation that represents the dynamics of the phenomenon under analysis. The equation for geometric Brownian motion, which is usually applied to model prices and returns from financial assets, was employed to illustrate and test the quality of the algorithms that were developed. The results are promising and indicate that the proposed methodology can be a very effective alternative for resolving stochastic differential equations.

1. Introduction

A large number of research projects in various fields of scientific knowledge utilize mathematical models that are partially or entirely formulated by stochastic differential equations (SDEs). Given the complexity of the specified models, their formulators must almost always address problems that have difficult resolutions and an unknown symbolic solution. In this situation, a typical procedure is to obtain an approximate solution by numeric methods.

Some of these projects aim to generate results that formally represent the temporal dynamics of the phenomena studied. In these cases, the numerical solution is not an appropriate instrument for giving researchers the necessary answers to their evaluations. The solution to the problem in its functional form is required; this solution will facilitate different types of analyses, such as determining the magnitude of partial effects and elasticities, estimating tendencies and volatilities, and realizing simulations and predictions of future scenarios.

The main goal of this study is to investigate the potential of evolutionary computational algorithms, which are developed by combining the methods of genetic programming (GP) and automatic differentiation (AD) and denoted as GPAD, to obtain symbolic solutions to SDE problems. Some algorithms were developed utilizing the programming environment of MATLAB, and a new methodology for a continuous time solution was generated.

The methodology adopted in this paper explores the already proven ability of GP in obtaining symbolic solutions in the form of equations. The GP algorithm looks, in the search space, for the analytical solution that minimizes the sum of the errors' squares between both sides of the chosen differential equation. The AD method is then used for obtaining the derivatives necessary to the minimization problem. It is worth noting that the AD method was chosen for providing the exact value of the punctual derivatives, as the adoption of methods based in finite differences or any other non-exact numerical method would not affect

* Corresponding author.
E-mail addresses: waldir.lobao@ibge.gov.br (W.J.d.A. Lobão), marco@ele.puc-rio.br (M.A.C. Pacheco), douglas.dias@uerj.br (D.M. Dias), carolina@ele.puc-rio.br (A.C.A. Abreu).

https://doi.org/10.1016/j.engappai.2017.10.021
Received 16 February 2017; Received in revised form 15 August 2017; Accepted 29 October 2017
0952-1976/

the accuracy of the equations candidate for solution, impairing the GP's performance during the evolution and the quality of the final obtained solution.

The methodology here proposed is conceptually difficult to compare to the ones based in finite differences and other numerical methods. However, the results given by the different methods can be compared based on error measurements and matching quality, even graphically. Nevertheless, the results generated by GPAD are, in the deterministic case, a differentiable function through traditional calculus, an in the case of this paper, a stochastic process differentiable by Itô's formula or lemma. As such, one of the advantages in the use of this methodology is that the accuracy of the obtained solution can be mathematically and analytically verified.

To compare the developed methodology with existing methods, a bibliographic review was conducted, and a significant number of scientific publications was investigated. It was seen that there is a plethora of papers developing GP applications for symbolic regression, in different areas of the technical and scientific knowledge. From the more recent ones, attention is given to the articles of Qi Chen et al. (2016, 2017) and Xie, et al. (2007). The first two approach the feature selection matter, presenting relevant results and proposing new selection methods, more efficient than the existing ones, for the generalization of genetic programming for high-dimensional symbolic regression problem, contributing for a smaller overfitting probability and lower computational cost.

The third paper uses GP for high-dimensional symbolic regression, to predict price inflation over short and medium terms for the New Zealand consumers' price index. The results obtained are significant, as the prediction model estimated by GP shows predictions with precision comparable to the official prediction system, developed and disclosed by the Reserve Bank of New Zealand (RBNZ).

Beyond the papers previously mentioned, Dr. Sara Silva's recent work on GP for symbolic regression was also consulted, as found on the following link: http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/SaraSilva.html.

It is important to acknowledge the existence of a great number of papers, with different kinds and applications of Genetic Algorithms, as well as significant results, mainly in the electrical and mechanical engineering areas. Between the consulted ones, attention is called to Güllü (2012, 2013, 2014, 2016, 2017) and Canakci et al. (2009).

Although extensive literature on GP and AD exists, the majority of the studies separately discuss and apply these methods. Among the studies that develop GPADs, those on the solution of differential equations are rare, and in the specific case of SDEs, no study was identified. Among these studies, the most notable research includes studies by Imae et al. (2004) and Tsoulos and Lagaris (2006).

In Imae et al. (2004), a GPAD algorithm is proposed to solve Hamilton–Jacobi–Bellman (HJB) equations, Hamilton–Jacobi–Isaacs (HJI) equations and Francis–Byrnes–Isidori (FBI) equations in nonlinear dynamical systems of optimal control problems. The methodology uses the conventional GP technique combined with an AD method developed by the authors. The algorithm is applied to equations that contain known symbolic solutions and generates approximate solutions with low estimate errors for every simulated case. The conclusions of the study suggest that the methodology presents promising results and is a suitable alternative for solving optimal control problems.

Among the research, the study by Tsoulos and Lagaris (2006) presents the most diverse and highest number of solution examples for differential equations using GPAD algorithms. Solutions are presented for first- and second-degree ODEs and partial differential equations (PDEs)—both linear and nonlinear equations. Exact symbolic solutions are identified for the vast majority of the addressed problems. The developed algorithm employs the grammatical evolution of GP and different AD methods. The study concludes that the developed methodology is adequate for solving ODEs and PDEs and that the results are very encouraging.

Few similarities between this study and the studies of Imae et al. (2004) and Tsoulos and Lagaris (2006) were identified. They are only related by the fact that they develop GPAD algorithms, and the goals and problems that are addressed differ. Consequently, attempts to solve SDEs require programs with very different technical details than those of the programs that are developed for solving ODEs and PDEs. Although some similarities are assumed, the differences among the studies are substantially greater.

Although the literature does not identify a study that employs a GPAD algorithm to solve SDEs, other methodologies and studies specifically address these equations. Numerous articles and books address this subject; the mathematical development of the theory of stochastic integration of the stochastic calculus of Itô (Itô, 1951) has produced general solutions for various types of SDEs. However, exact solutions for other important types or classes remain open. A diversity of methods have been developed with the aim of obtaining solutions for SDEs; most of the utilized methods are based on statistical estimation, computational numeric methods and stochastic simulation methods. Among the bibliographic references that theoretically informed this study, the most notable studies are Itô (1951), Arnold (1974), Schuss (1980), Ikeda and Watanabe (1989), Kloeden and Platen (1992), Campbell et al. (1997) and Aït-Sahalia (2002).

Therefore, the methodology that is proposed is an alternative for resolving the problem with the advantages of producing solutions in symbolic form and in continuous time and, in a case in which the equation of interest is completely unknown, of offering the option of algorithms that perform the specification and estimation of the solution to this equation via a real database. The last advantage is important because it determines an appropriate solution to the problem and simultaneously eliminates the difficult task of arbitrarily defining the functional form of the SDE that represents the dynamics of the phenomenon. The equation for geometric Brownian motion (GBM), which is usually applied to model prices and returns from financial assets, was used to illustrate and test the quality of the developed algorithms. The results are promising and indicate that the proposed methodology can be a very effective alternative for resolving SDEs.

In addition to this introduction, the paper is organized into four sections. The second section describes the methodological procedures that were adopted in the development of the GPAD algorithms. In the third section, we present the results of the application of the proposed methodology in the solution of a SDE and GBM. The fourth section summarizes and presents the study's conclusions.

2. Methodology

In this section, the theoretical aspects and modeling procedures that are adopted in the development of the general methodology of this study are presented. They are elaborated with the intention of obtaining symbolic solutions for SDEs via GPADs. This section is divided into three subsections: the first section describes the mathematical problem of interest; the second section presents the methodology for the specification and estimation of the SDE model; and the third section describes the structure and stages of the functioning of the basic algorithm of GPAD.

2.1. Relevant mathematical aspects

According to the objective of this study, the mathematical problem that the developed methodology proposes to solve, with the aid of GPAD algorithms, are the SDEs in a traditional form, which are specified according to the following expression:

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t); t \in [0, T], \quad (1)$$

where $\{X(t); t \in [0, T]\}$: stochastic process for the solution of SDE (1), which is defined in the space of real probability $(\Omega, \mathcal{F}, P)$; $a(t, X(t))$: tendency coefficient (drift); $b(t, X(t))$: diffusion coefficient (volatility); and $W(t)$: Wiener process (or standard Brownian motion).