ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Ultra-Orthogonal Forward Regression Algorithms for the Identification of Non-Linear Dynamic Systems

Yuzhu Guo [a,b,*], L.Z. Guo [a,b], S.A. Billings [a,b], Hua-Liang Wei [a,b]

[a] Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK
[b] INSIGNEO Institute for in silico Medicine, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form
21 July 2015
Accepted 3 August 2015

Keywords:
Orthogonal forward regression
System identification
Ultra-least squares
Ultra-Orthogonal Forward Regression
Ultra-Orthogonal Least Squares

ABSTRACT

A new ultra-least squares (ULS) criterion is introduced for system identification. Unlike the standard least squares criterion which is based on the Euclidean norm of the residuals, the new ULS criterion is derived from the Sobolev space norm. The new criterion measures not only the discrepancy between the observed signals and the model prediction but also the discrepancy between the associated weak derivatives of the observed and the model signals. The new ULS criterion possesses a clear physical interpretation and is easy to implement. Based on this, a new Ultra-Orthogonal Forward Regression (UOFR) algorithm is introduced for nonlinear system identification, which includes converting a least squares regression problem into the associated ultra-least squares problem and solving the ultra-least squares problem using the orthogonal forward regression method. Numerical simulations show that the new UOFR algorithm can significantly improve the performance of the classic OFR algorithm.

1. Introduction

System identification plays a more important role in revealing the unknown mechanisms and rules underlying complex phenomena [1]. System identification includes the detection of the model structure and estimation of the associated parameters. A system identification problem can often be thought of as an optimisation problem where the optimal model is searched from a large predefined candidate model dictionary, given a criterion. The criterion is used to evaluate the performance of each model by measuring the discrepancy between the observed data and the model predictions. The candidate model dictionary is often chosen to be large enough to include the unknown correct model. Hence an exhaustive search algorithm is often infeasible in these kinds of applications because of the large solution space. Even an evolutionary algorithm which can greatly reduce the search process can still be very computationally intensive. Hence an algorithm which can efficiently find the optimal solution is desired. However, a fast algorithm often dictates an optimal substructure; otherwise the search may converge to a suboptimal solution. Many efforts have been made to improve the search process under a certain specific loss function or performance index, for example, the simulated annealing algorithm, particle swarm optimisation, and so on. In this paper, a different and new methodology will be introduced. Instead of improving the search method, a new and effective criterion will be introduced to describe the objective of the regression more accurately. Under the new criterion, the solution space has a better structure and a fast algorithm is more likely to find the optimal solution.

System identification aims to identify a model from observed data based on a criterion. A good criterion results in not only better parameter estimation but also a good search path along which the search process converges quickly to the optimal solution. Over the years, different criteria have been used in system identification such as the L^2 norm in least squares regression, the L^1 norm in least absolute value regression [2,3], and zero-norm minimisation [4]. Among these criteria, the least squares criterion is the most used because of its excellent properties, for example, least squares estimation can be configured to give estimates which are unbiased and efficient when the noise satisfies some basic assumptions. Least squares problem have analytic solutions and can easily be solved using the QR decomposition technique, and least squares regression produces unique and numerically robust solutions. Consequently a large number of system identification algorithms based on the least squares criterion have been developed [5–8].

However, the standard least squares method only reveals part of the information in the observed data. The least squares criterion, which considers the datum points individually, discards the connections among the datum points especially for the identification of dynamic systems where the data set are time series which are

* Corresponding author.
E-mail address: yuzhu.guo@sheffield.ac.uk (Y. Guo).

http://dx.doi.org/10.1016/j.neucom.2015.08.022

samples of continuous functions of time. These individual datum points are time dependent and connected with each other through the derivatives of the time continuous functions, for example, an ordinary differential equation. Many important characteristics of a system can be determined by these interconnections. An absence of this information may lead to over-fitted models in least squares regression, which can be seen in the motivational example described in Fig. 1 and discussed in the next section.

The standard least squares regression investigates the problem of model fitting on the space $L^2([0,T])$, where $[0,T]$ represents the time span of a signal. The associated Residual Sum of Squares (RSS), which is the square of the L^2 norm of the residual, is used to measure the fitness of the model. When the model structure is known, the standard least squares algorithm produces the best parameters with which the model will be optimal in the sense of RRS. Considering different model structures, there are plenty of very different models which give the same fitness for a set of observed data in the sense of the RSS criterion. In this paper, an alternative criterion, called ultra-least squares (ULS) criterion will be introduced to characterise the model fitness more accurately. Unlike the least squares criterion considers the model fitting on the space L^2, the ULS criterion considers the model fitting in a smaller space, more specifically, the Sobolev space $H^m([0,T])$ [9]. The norm defined on this space will be modified and used as the ULS criterion for system identification, where not only residuals but also the associated weak derivatives will be used to measure the model fitness.

Usage of the derivatives of the data in system identification has been studied, especially the identification of continuous time models [1,10,11]. However, as far as the authors are aware this is the first study in which the weak derivatives have been combined with the least squares criterion to build a completely new metric for the prediction errors and which used the new metric to improve the model structure detection in non-linear system identification.

In this paper, the ULS criterion will be combined with the well known Orthogonal Forward Regression (OFR) algorithm [7] to construct a new Ultra-Orthogonal Forward Regression (UOFR) algorithm for nonlinear system identification. The proposed UOFR algorithm is shown to be very powerful for model structure detection in many modelling tasks and is more likely to produce an optimal model.

The remainder of the paper is organised as follows: Section 2 briefly reviews some main results on the Lebesgue space L^2 and the Sobolev space H^m. The ULS criterion will be presented by modifying the H^m norm in Section 3. The associated solution to the ultra-least squares problem is then defined, and the new UOFR algorithm is described in Section 4. Three benchmark examples are discussed in Section 5 to illustrate the efficiency of the new UOFR algorithm. Conclusions are finally drawn in Section 6.

2. Problems of least squares regression and model fitting in Sobolev space

In this section, a motivational example is first given to show the problems that can arise while using a standard least square criterion. The reasons which cause these problems will then be discussed in detail and an alternative criterion will be proposed.

Consider the time series fitting problem shown in Fig. 1. In this example, three models were identified from an observed signal y which is represented by a thick solid line in Fig. 1(a). The reproduced signals by the three models are represented by the curves y_1, y_2, and y_3 in Fig. 1(a) respectively. Fig. 1(b) shows the different measurements of the model fitness of the three models: the L^2 norm and the H^m $(m=1,2,3)$ norms of the residuals.

From Fig. 1(b), it can be observed that the three models give the same fitness in the sense of the least squares criterion, which is presented by the line with the circle marks along the abscissa in Fig. 1(b), although the reproduced signal y_1 looks significantly different from y_2 andy_3 in Fig. 1(a).

Fig. 1(b) also shows the measurements of the errors in the sense of H^mnorms when $m=1,\ 2,\ 3$. It can be observed that the performances of the three models under the H^m norms are significantly different. Model 3 fitted the signal y better than models 1 and 2 did. The system identification problems consists of finding the function on $H^m([0,T])$ which best fits the observed data $\{y_n\}$, $n=1, 2, \ldots, N$, where both the data points and the interconnections among the datum points (described by the weak derivatives) are considered.

This example shows that the least squares criterion which was defined on the L^2 space neglects some very important information in the observations. This information is crucial for identifying a correct model. Alternatively, the model fitness can more accurately be characterised on a smaller space, the Sobolev space H^m, which consists of all the functions which are L^2 integrable and the where up to mth weak derivatives exist and are also L^2 integrable. The new introduced ULS criterion is a realisation of the H^m norm based on the observations.

The generic least squares regression problem includes determining the structure of a linear-in-the-parameters model and estimating the associated coefficients

$$y=\sum_{i=1}^{\kappa}\theta_i x_i+e \tag{1}$$

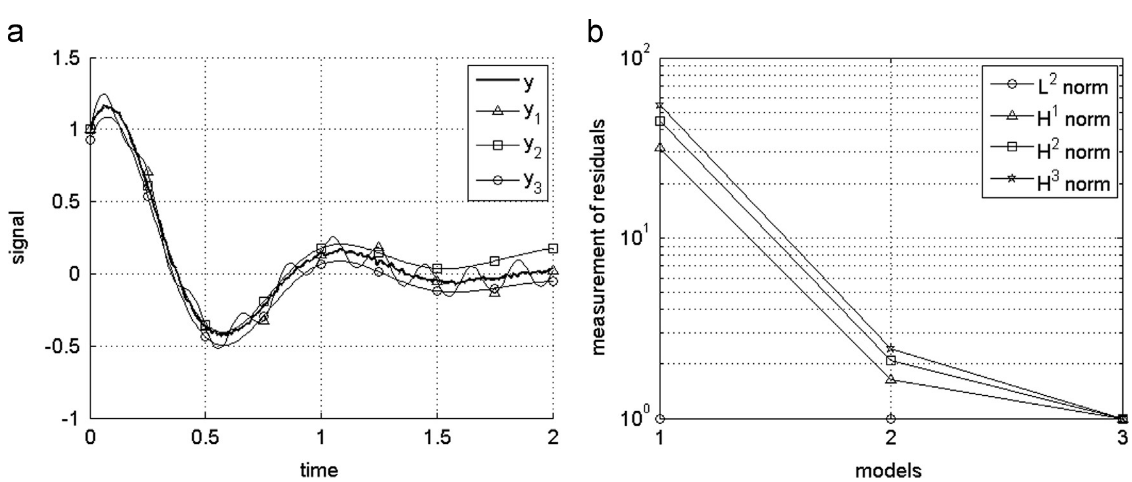

Fig. 1. A motivational example for model fitting of a noisy signal. (a) Observed data and reproduced signals for three different models (b) measurement of the fitness of the models using different criteria.

Please cite this article as: Y. Guo, et al., Ultra-Orthogonal Forward Regression Algorithms for the Identification of Non-Linear Dynamic Systems, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.022