Mechanical Systems and Signal Processing 52-53 (2015) 133–146

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Bayesian system identification of a nonlinear dynamical system using a novel variant of Simulated Annealing

P.L. Green

Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form
20 March 2014
Accepted 9 July 2014
Available online 4 August 2014

Keywords:
Bayesian model updating
Nonlinear system identification
Markov chain Monte Carlo
Simulated Annealing
Deviance Information Criterion

ABSTRACT

This work details the Bayesian identification of a nonlinear dynamical system using a novel MCMC algorithm: 'Data Annealing'. Data Annealing is similar to Simulated Annealing in that it allows the Markov chain to easily clear 'local traps' in the target distribution. To achieve this, training data is fed into the likelihood such that its influence over the posterior is introduced gradually - this allows the annealing procedure to be conducted with reduced computational expense. Additionally, Data Annealing uses a proposal distribution which allows it to conduct a local search accompanied by occasional long jumps, reducing the chance that it will become stuck in local traps. Here it is used to identify an experimental nonlinear system. The resulting Markov chains are used to approximate the covariance matrices of the parameters in a set of competing models before the issue of model selection is tackled using the Deviance Information Criterion.

1. Introduction

This paper is concerned with the system identification of a nonlinear dynamical system using experimentally obtained training data. A probabilistic, Bayesian approach is utilised throughout. Such an approach is now well established in the structural dynamics community – relatively recent advances include the use of Bayesian methods in structural health monitoring [1], modal identification [2], state-estimation [3] (through use of the particle filter), the sensitivity analysis of large bifurcating nonlinear models [4] as well as an interesting study investigating the relations between frequentist and Bayesian approaches to probabilistic parameter estimation [5].

The identification problem detailed herein is one of *model selection* as well as *parameter estimation* such that, using experimental data $\mathcal{D}$, one must endeavor to find the optimum model $\mathcal{M}$ from a set of competing model structures as well as estimate the parameter vector $\boldsymbol{\theta}$ of that particular model. Using Bayes' theorem a measure of the plausibility of a parameter vector $\boldsymbol{\theta}$, given experimental data $\mathcal{D}$ and assumed model structure $\mathcal{M}$, is given by

$$P(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})P(\boldsymbol{\theta}|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})} \tag{1}$$

where $P(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ is the posterior probability density function (PDF) which one wishes to evaluate, $P(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ is termed the likelihood, $P(\boldsymbol{\theta}|\mathcal{M})$ the prior and $P(\mathcal{D}|\mathcal{M})$ the evidence. The likelihood represents the probability that the experimental training data $\mathcal{D}$ was witnessed according to the model $\mathcal{M}$ with parameters $\boldsymbol{\theta}$. Defining the likelihood requires the selection

E-mail address: p.l.green@sheffield.ac.uk

http://dx.doi.org/10.1016/j.ymssp.2014.07.010
0888-3270

of an error-prediction model which describes the uncertainties present in the measurement and modelling processes (see [6] for a detailed discussion of error-prediction models). The prior is a PDF which represents one's parameter estimates for model $\mathcal{M}$ before the training data was known. The evidence is a normalising constant which ensures that the posterior PDF integrates to one.

This paper makes two main contributions. Firstly, a novel variant of Simulated Annealing (referred to as Data Annealing) is proposed and applied to a real system identification problem. It is shown to be computationally cheap and easy to tune. Secondly, it is shown that the issue of model selection of a real nonlinear dynamical system can be addressed using the Deviance Information Criterion (DIC). For the sake of readability the remainder of the introduction is split into two sections. The first outlines the motivation for the Data Annealing algorithm while the second focuses on the issue of model selection.

1.1. Motivation for the data annealing algorithm

For the case where one is attempting to identify N_D parameters (such that $\boldsymbol{\theta} \in \mathbb{R}^{N_D}$), the evidence is given by

$$P(\mathcal{D}|\mathcal{M}) = \int \cdots \int P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}) P(\boldsymbol{\theta}|\mathcal{M})\, d\theta_1 \cdots d\theta_{N_D}. \tag{2}$$

This integral is usually intractable and its multidimensional nature makes it too computationally expensive to evaluate numerically (if $N_D > 2$). Relatively early papers such as [7] made use of the property that the *maximum a posteriori* (MAP) parameter vector remains the same regardless of whether the posterior distribution has been normalised such that, through locating the MAP, a Taylor series expansion of the log posterior could be used to approximate the posterior PDF as a Gaussian.[1] Since then, an increase in computing power has allowed the adoption of Markov chain Monte Carlo (MCMC) methods. These involve the creation of an ergodic Markov chain whose stationary distribution is equal to the posterior PDF such that, once converged, the Markov chain is generating samples from $P(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})$ (see [9] for more information on the convergence of Markov chains). This can be achieved without having to evaluate the evidence term. While many MCMC methods are available in the literature (Hamiltonian Monte Carlo for example [10]), by far the most popular is the Metropolis algorithm. Although well-established, a brief description of the Metropolis algorithm is given here as it helps to establish the motivation for the Data Annealing algorithm presented in Section 2 of this work.

Essentially, the aim of MCMC methods is to generate a sequence of samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots\}$ from a target PDF $\pi(\boldsymbol{\theta})/Z$ (where Z is a normalising constant). In the context of this paper, $\pi(\boldsymbol{\theta})$ represents the unnormalised posterior PDF and Z represents the evidence term. Initialising the Metropolis algorithm from parameter vector $\boldsymbol{\theta}^{(i)}$, a new state $\boldsymbol{\theta}'$ is proposed using a user-defined proposal PDF. The proposal PDF is conditional on the current state $\boldsymbol{\theta}^{(i)}$. For example, in the case where a Gaussian proposal is used then the new state is generated according to

$$\boldsymbol{\theta}' \sim \mathcal{N}(\boldsymbol{\theta}^{(i)}, \Sigma) \tag{3}$$

(where Σ is a user-defined covariance matrix). The new state is then accepted with probability:

$$a = \min\left\{1, \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(i)})}\right\}. \tag{4}$$

If accepted then $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}'$ else $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$. This has the property that if the proposed state $\boldsymbol{\theta}'$ is in a region of higher probability density than the current state then it is always accepted. However, the Markov chain is also able to move into regions of lower probability density. One of the benefits of using such an acceptance rule is that the acceptance probability a can be computed without having to evaluate the evidence term. It can be shown that such an acceptance rule allows the chain to generate samples from $\pi(\boldsymbol{\theta})$ (for more information references [8,11] are recommended).

The advantages of using MCMC are numerous. Recalling that the purpose of system identification is usually to establish a reliable model which can be used to accurately and robustly predict the system's future response then, using the notation outlined in [12], one may want to predict a structural quantity of interest $h(\boldsymbol{\theta})$ using

$$R = \int \cdots \int h(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})\, d\theta_1 \cdots d\theta_{N_D}. \tag{5}$$

While evaluating Eq. (5) is difficult (for the same reason it is difficult to evaluate the evidence term), if one has used an MCMC algorithm to generate samples $\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}\}$ from the posterior parameter distribution then Eq. (5) can be approximated by

$$R \approx \frac{1}{M} \sum_{i=1}^{M} h\left(\boldsymbol{\theta}^{(i)}\right). \tag{6}$$

Additionally, it has been shown that important information with regard to parameter correlations can be realised through the use of MCMC methods [13] (this is also demonstrated in Section 4 of the present work). However, MCMC also has its disadvantages. Before samples from the target distribution can be drawn in an effective manner, the Markov chain must

[1] For more information the reader may wish to consult the description of the Laplace approximation given in reference [8]