Automatica 50 (2014) 1116–1121

Contents lists available at ScienceDirect

Automatica

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

Brief paper

A combined MAP and Bayesian scheme for finite data and/or moving horizon estimation☆

Ramón A. Delgado [1], Graham C. Goodwin

School of Electrical Engineering and Computer Science, The University of Newcastle, Australia

ARTICLE INFO

Article history:
Received 17 August 2012
Received in revised form
6 November 2013
Accepted 19 December 2013
Available online 20 February 2014

Keywords:
Moving horizon estimation
Nonlinear filtering

ABSTRACT

Finite data and moving horizon estimation schemes are increasingly being used for a range of practical problems. However, both schemes suffer from potential conceptual difficulties. In the case of finite data, most of the methods in common use, excluding Bayesian strategies, depend upon asymptotic results. On the other hand, in the case of moving horizon estimation, there are two associated problems, namely (i) estimation error quantification is typically not available as a part of the solution and (ii) one needs to provide some form of prior state estimate (the so-called arrival cost). The current paper proposes a combined MAP–Bayesian scheme which, inter alia, addresses the finite data and moving horizon problems described above. The scheme combines MAP and Bayesian strategies. The efficacy of the method is illustrated via numerical examples.

1. Introduction

Finite data estimation arises in many practical problems. A well-known example is parameter estimation when only a small amount of data is available. It is common practice to use schemes such as Prediction Error Methods (PEM) (Ljung, 1999) for finite data parameter estimation. These generally perform well but suffer from conceptual problems. For example, the usual quantification of the accuracy in PEM depends upon asymptotic results. This has motivated several authors to develop alternative schemes for parameter estimation with finite data (Campi & Weyer, 2002; Weyer & Campi, 2002). Of course, full Bayesian methods also provide a solution to the finite data problem but these suffer from other difficulties as we will discuss below.

A closely related problem to finite data estimation occurs in Moving Horizon Estimation (MHE). MHE combines a sequence of finite data problems. It has received increasing attention over the last decade (Alessandri, Baglietto, & Battistelli, 2008; Başar & Bernhard, 2008; Rao, 2000; Rao, Rawlings, & Lee, 2001; Rao, Rawlings, & Mayne, 2003; Rawlings & Bakshi, 2006; Verdu & Poor, 1987). MHE transforms filtering, smoothing and prediction problems into a standard constrained optimization problem over a finite horizon. In order to limit the size of the problem, MHE requires that the range of data used for estimation be small. This means that, when new data arrives, the oldest data is summarized by a, so called, arrival cost.[2]

MHE has several advantages compared with other schemes. These advantages arise due to the transformation of the problem into a standard optimization problem. One advantage is that it allows one to incorporate constraints, for example, on the states of the system (e.g. a tank cannot be more than full or less than empty). Also, standard tools developed for Model Predictive Control can be applied to MHE (see e.g. Diehl, Ferreau, & Haverbeke, 2009, Rawlings & Mayne, 2009).

On the other hand, there are difficulties associated with the usual MHE scheme. For example, the impact of past data needs to be summarized in the form of an a-priori distribution. This is typically achieved by adding an arrival cost (Başar and Bernhard (2008) and Rao et al. (2003) Verdu and Poor (1987)). However, the formulation of a statistically well posed arrival cost remains an open problem (Haseltine & Rawlings, 2005). To address this problem various approximate arrival cost strategies have been proposed; see e.g. Alessandri et al. (2008), Rao (2000), Rao et al. (2001), Ungarala (2009) and Zavala (2010). One such strategy expresses the arrival cost as a simple quadratic function of the

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Tongwen Chen under the direction of Editor Ian R. Petersen.

E-mail addresses: ramon.delgado@uon.edu.au (R.A. Delgado), graham.goodwin@newcastle.edu.au (G.C. Goodwin).

[1] Tel.: +61 2 49217378; fax: +61 2 49601712.

[2] This has also been called "entry cost" in the literature.

http://dx.doi.org/10.1016/j.automatica.2014.02.001
0005-1098/

difference between the current initial state and the propagation of the initial state estimate from the previous horizon (see e.g. Alessandri et al., 2008, Alessandri, Baglietto, Battistelli, & Zavala, 2010).

MHE can be given two interpretations. If one adopts probabilistic models, then MHE can be interpreted as computing the Maximum A Posteriori (MAP) estimate. Alternatively, one can interpret the MHE method as simply a procedure for comparing a measured trajectory with a model trajectory via a suitable cost function. Whichever interpretation one uses, MHE requires optimization for its solution. A fundamental issue of relevance to the current paper is that only a point estimate is obtained. In the sequel we adopt the probabilistic interpretation.

By way of contrast, Bayesian estimation computes the complete a-posteriori distribution. However, Bayesian estimation also suffers from disadvantages. In particular, Bayesian estimation is generally computationally expensive. Moreover, the size of the problem typically grows exponentially with the number of data points. Hence some form of simplification is usually required. In practice, this is achieved by using approximate schemes e.g. deterministic gridding algorithms, particle filtering or other resampling methods (Chen, 2003).

Here, we propose an alternative approach to finite data and/or moving horizon estimation that combines MAP and Bayesian techniques. It provides a solution to both the entry cost and error quantification problems.

The layout of the remainder of the paper is as follows: in Section 2, we present the problem formulation. In Section 3 we outline the combined MAP–Bayesian scheme for finite data problems. In Section 4 we explain the extension to Moving Horizon Estimation. In Section 5, we present several examples. Conclusions are presented in Section 6.

2. Problem formulation

Consider a nonlinear system described by a state space model of the form

$$x_{t+1} = f(x_t) + w_t \tag{1}$$

$$y_t = h(x_t) + v_t \tag{2}$$

where $x_t \in \mathbb{R}^{n_x}$, $y_t \in \mathbb{R}^{n_y}$. For simplicity[3] we assume that

$$\begin{bmatrix} w_t \\ v_k \end{bmatrix} \sim N\left(\begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\right). \tag{3}$$

Our goal is to estimate the states $x_0, \ldots, x_N$, given observations $y_1, \ldots, y_N$. We also assume that a prior distribution is available for x_0.

Two general approaches for solving this problem are MAP and Bayesian estimation. These two approaches are based on the common element of the a-posteriori distribution. An expression for the a-posteriori distribution is given in Lemma 1 below:

Lemma 1. *For the system (1)–(2), the a-posteriori distribution for the states $x_0, \ldots, x_N$, given the observations $y_1, \ldots, y_N$ is*

$$p(x_0, x_1, \ldots, x_N | y_1, \ldots, y_N) \propto \prod_{i=1}^{N} p(x_i|x_{i-1}) p(y_i|x_i) p(x_0) \tag{4}$$

where $\propto$ denotes "modulo a normalizing constant".

Proof. From the Bayes rule,

$$p(x_0, x_1, \ldots, x_N | y_0, \ldots, y_{N-1}) \propto p(y_1, \ldots, y_N | x_0, \ldots, x_N) p(x_0, \ldots, x_N). \tag{5}$$

The results then follows by using the Markov property inherent in (1), (2). □

MAP and Bayesian estimation can then be described as follows:

Maximum A Posteriori (MAP) estimation provides a point estimate corresponding to the maximum of the a-posteriori distribution, i.e.

$$\hat{x}_0, \ldots, \hat{x}_N = \arg \max_{x_0, \ldots, x_N} p(x_0, x_1, \ldots, x_N | y_1, \ldots, y_N). \tag{6}$$

Note that the associated algorithm only explores the a-posteriori distribution in so far as is necessary to reach the maximum.

On the other hand, Bayesian estimation is aimed at computing (at least approximately) the complete a-posteriori distribution as in (4). From this distribution, one can extract any desired point estimate (e.g. mean, MAP, etc.). Information about the accuracy of any particular estimate is automatically available.

Unfortunately, the computation of the complete a-posteriori distribution is, in general, intractable. However there are very special cases, such as unconstrained linear Gaussian problems, where the Kalman Filter provides an exact representation of the a-posteriori distribution. Hence, for most problems, approximate methods are typically used in practice. For example, the Extended Kalman Filter (EKF), see e.g. Jazwinski (1970), linearizes the nonlinear system, and then applies the standard Kalman Filter to propagate the mean and covariance of the estimates. Alternatively, one could use a deterministic grid on the state space. A related approach is Minimum Distortion Filtering (MDF) (Goodwin, Feuer, & Müller, 2010), which uses a grid for the a-posteriori distribution that is focused on the most likely parts of the state space. Another commonly used method is Particle Filtering (PF) (Gordon, Salmond, & Smith, 1993). This method draws a set of random samples from the disturbance distribution.

Here we propose a strategy which combines MAP and Bayesian methods. The core idea is explained in the next section.

3. Combined MAP and Bayesian estimation

We begin by describing the algorithm in the context of finite data estimation. (Note that this is a necessary precursor to the moving horizon case.)

Initialization: we assume that we are given the prior distribution $p(x_0)$ and data $y_1, \ldots, y_N$. Also, we assume that $p(x_0)$ is well approximated by a point distribution of the form

$$p(x_0) = \sum_{s=1}^{N_x} p_s^0 \delta(x_0 - \bar{x}_0(s)) \tag{7}$$

where $p_1^0, \ldots, p_{N_x}^0$ denote point probability masses at $\bar{x}_0(1), \ldots, \bar{x}_0(N_x)$ respectively, and N_x is the number of points in the point distribution.

We also assume that we are given a point distribution $M(x)$ with N_x points,

$$M(x) = \sum_{l=1}^{N_x} q_l \delta(x - u(l)) \tag{8}$$

that approximates a multivariate standard Gaussian distribution in $\mathbb{R}^{n_x}$, i.e. zero mean and diagonal unitary variance I_{n_x}.

[3] The extension to more general models and noise distributions presents no additional conceptual difficulties.