A Combined MAP and Bayesian Scheme for Finite Data and/or Moving Horizon Estimation ⋆

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

[a] *School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*

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.

Key words: Moving Horizon Estimation; Nonlinear Filtering.

1 Introduction

Finite data estimation arise 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 (Rao, Rawlings & Lee 2001, Rao 2000, Başar & Bernhard 2008, Verdu & Poor 1987, Rao, Rawlings & Mayne 2003, Alessandri, Baglietto & Battistelli 2008, Rawlings & Bakshi 2006). MHE transforms filtering, smoothing and prediction problems into a standard constrained optimisation 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" [1] .

MHE has several advantages compared with other schemes. These advantages arise due to the transformation of the problem into a standard optimisation 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. (Rawlings & Mayne 2009, Diehl, Ferreau & Haverbeke 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. (Verdu & Poor 1987, Başar & Bernhard 2008, Rao et al. 2003). However, the formulation of a statistically well posed arrival cost remains an open problem

⋆ This paper was not presented at any IFAC meeting. Corresponding author R. A. Delgado. Tel. 61-2-49217378. Fax +61-2-49601712.

Email addresses: ramon.delgado@uon.edu.au (Ramón A. Delgado), graham.goodwin@newcastle.edu.au (Graham C. Goodwin).

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

Peer-reviewed author's copy of:
R.A. Delgado and G.C. Goodwin. **A Combined MAP and Bayesian Scheme for Finite Data and/or Moving Horizon Estimation**. *Automatica*, Vol 50(4):1116-1121, April 2014.
Available at http://dx.doi.org/10.1016/j.automatica.2014.02.001

(Haseltine & Rawlings 2005). To address this problem various approximate arrival cost strategies have been proposed, see e.g (Rao et al. 2001, Rao 2000, Alessandri et al. 2008, Zavala 2010, Ungarala 2009). One such strategy expresses the arrival cost as a simple quadratic function of the 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 [2] 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 1 *From 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,

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

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 non linear system, then applies 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 focussed 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 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}.

Our proposal has 4 components.

(i) MAP estimation: In the combined MAP and Bayesian estimation, we first use MAP estimation to obtain a single trajectory estimate $\hat{x}_1, \ldots, \hat{x}_N$. Note that we do not produce a MAP estimate for x_0, instead we marginalize over the point distribution form of $p(x_0)$, i.e.

$$\begin{aligned} &\hat{x}_1, \ldots, \hat{x}_N \\ &= \arg \max_{x_1,\ldots,x_N} \sum_{s=1}^{N_x} p(\bar{x}_0(s), x_1, \ldots, x_N | y_1, \ldots, y_N). \end{aligned} \tag{9}$$

The estimated trajectory $\hat{x}_1, \ldots, \hat{x}_N$ is the "most likely" trajectory in the state space. However, there is no information about the accuracy of estimate.

(ii) EKF for breadth estimation: The EKF provides a computationally efficient way to obtain a measure of the covariance of the estimates. The EKF uses the standard Kalman Filter covariance update given by

$$\Pi_{k|k-1} = A_{k-1}\Pi_{k-1|k-1}A_{k-1}^T + Q \tag{10}$$

$$L_k = \Pi_{k|k-1}C_k^T(C_k\Pi_{k|k-1}C_k^T + R)^{-1} \tag{11}$$

$$\Pi_{k|k} = \Pi_{k|k-1} - L_k C_k \Pi_{k|k-1} \tag{12}$$

where, in our approach, we linearize the nonlinear system about the MAP estimate $\hat{x}_1, \ldots, \hat{x}_N$, i.e. we use

$$A_k = \left.\frac{\partial f(x)}{\partial x^T}\right|_{x=\hat{x}_k} \qquad C_k = \left.\frac{\partial h(x)}{\partial x^T}\right|_{x=\hat{x}_k} \tag{13}$$

where $\Pi_{0|0}$ is computed as

$$\Pi_{0|0} = \sum_{s=1}^{N_x} p_s^0 (\bar{x}_0(s) - \hat{x}_0)(\bar{x}_0(s) - \hat{x}_0)^T \tag{14}$$

$$\hat{x}_0 = \sum_{s=1}^{N_x} p_s^0 \bar{x}_0(s) \tag{15}$$

(iii) Gridding: (Here we shift and scale the point distribution (8)). Given the MAP estimates $\hat{x}_1, \ldots, \hat{x}_N$ in (9) and the covariances $\Pi_{1|1}, \ldots, \Pi_{N|N}$ computed using (12), we have a Gaussian approximation for the distribution of x_k, $k = 1, \ldots, N$. Next, for each $k = 1, \ldots, N$ we allocate a set of N_x points by using the following affine transformation of $M(x)$, i.e. for $l = 1, \ldots, N_x$, we define

$$\bar{x}_k(l) = \hat{x}_k + \Pi_{k|k}^{1/2} u(l). \tag{16}$$

where $M(x)$ and the corresponding points $u(l)$ are defined by (8), and $\hat{x}_k$ as in (9).

Remark 1 *Note, that the probability masses q_l, $l = 1, \ldots, N_x$ of $M(x)$ are not used in the method, since the probability masses are computed in step **(iv)**.* $\triangledown\triangledown\triangledown$

(iv) Approximate Bayesian update: The previous step provide a grid of points $\bar{x}_k(l)$, $k = 1, \ldots, N$, $l = 1, \ldots, N_x$ that is allocated in a part of the state space where the a-posteriori distribution is concentrated. In this step, the corresponding probability masses p_l^k are computed. To compute p_l^k we carry out a full recursive Bayesian computation over the grid. For $l = 1, \ldots, N_x$ beginning with $k = 1$ compute

$$\begin{aligned} p_l^k &= p(\bar{x}_k(l)|y_1, \ldots, y_k) \\ &= c \sum_{j=1}^{N_x} p_j^{k-1} p(\bar{x}_k(l)|\bar{x}_{k-1}(j)) \cdot p(y_k|\bar{x}_k(l)) \end{aligned} \quad (17)$$

where c is a normalizing constant, when k is updated then p_l^k is computed for $l = 1, \ldots, N_x$ using (17). The procedure is repeated until p_N^l is computed for $l = 1, \ldots, N_x$.

The proposed method is summarized in Algorithm 1

Algorithm 1 Combined MAP and Bayesian estimation algorithm

- Initialize:
  - ⋆ Provide a point distribution for the initial state as in (7).
  - ⋆ Provide a point distribution $M(x)$ that approximates a multivariate standard Gaussian distribution, as in (8).
- Obtain a trajectory $\hat{x}_1, \ldots, \hat{x}_N$ by solving (9).
- Compute the covariance of x_0, $\Pi_{0|0}$, using (14), and then compute $\Pi_{1|1}, \ldots, \Pi_{N|N}$ using (10)-(13).
- Allocate a grid of points, $\bar{x}_k(l)$, $k = 1, \ldots, N$; $l = 1, \ldots, N_x$ on the state space by using (16).
- Perform full Bayesian update over the grid of points, i.e. for $k = 1, \ldots, N$
  - ⋆ for $l = 1 : N_x$, compute the point masses p_l^k by using (17)
  - ⋆ $k = k + 1$
- End Result: The pair $(\bar{x}_k(l), p_l^k)$ for $k = 1, \ldots, N$; $l = 1, \ldots, N_x$ that approximates the a-posteriori distribution (4).

Remark 2 *If the prior distribution is continuous, then one can obtain $p(x_0)$ as in* (7) *by random sampling or by use of vector quantization methods (see e.g. (Gersho & Gray 1992)). This step is computationally expensive but will be performed* off-line *before running the algorithm. The same techniques can also be used to obtain the point distribution $M(x)$.* ▽▽▽

Remark 3 *The idea behind the proposed approach is to provide additional information about the estimate obtained by MAP estimation. However the proposed approach can also be used to improve the accuracy information provided by other methods, such as the Extended Kalman Filter, or the Unscented Kalman Filter, among others.* ▽▽▽

Remark 4 *The current exposition of the combined MAP-Bayesian estimation, suggests that the single trajectory $\hat{x}_1, \ldots, \hat{x}_N$ is obtained by using MAP estimation. Nevertheless, the proposed approach can also be applied when the single trajectory is obtained using other optimization based methods.* ▽▽▽

In the next section we will show how the above finite data scheme can be deployed in the moving horizon framework.

4 Moving Horizon Estimation

In Moving Horizon Estimation (MHE), the estimation is carried out online by successively solving a sequence of finite data problems. A new issue that arises in the MHE framework is how to incorporate the effect of the past data (outside the current estimation interval). In the combined MAP-Bayesian framework, the a-posteriori distribution for the last state can be stored and reused at a later time as the prior distribution.

There are several versions of the moving horizon algorithm depending upon whether estimates are required only after each block of N samples or whether estimates are needed after each data point is received. We refer to these cases as Block estimation and Sample-by-Sample estimation, respectively.

(i) Block estimation. Here one simply uses the a-posteriori distribution for the final state obtained in the previous block as prior distribution for the next block (see e.g (Rawlings & Mayne 2009, page 353))

(ii) Sample-by-Sample estimation with filtered update. Here one needs to place the a-posteriori distribution for the final state obtained from the current block in a "rotating store". This scheme can be seen as N Block estimators running in parallel.

(ii) Sample-by-Sample estimation with smoothed update: in this update scheme the second value of the previous solution is used as the initial condition.

An analysis of the different Sample-by-Sample update schemes can be found in (Findeisen 1997). Here, we focus on the Block estimation scheme. However, there is no conceptual difficulties to use the proposed approach with any of the MHE schemes described above.

Regarding the stability and convergence of the approach, the proposed MHE scheme described here, can be seen as a special case of the algorithm described in (Rao et al. 2003). Hence, the stability and convergence results presented in (Rao et al. 2003) can also be adapted to the method proposed here.

5 Numerical examples

In this section we present two numerical examples. In the first example, we consider a two state batch reactor. For the second example, we consider a single state water tank, with two different radii as a function of the height.

5.1 Batch reactor

Consider a well-mixed semibatch chemical reactor where the material balances for the two components are

$$\dot{c}_A = -2\,\kappa\, c_A^2 + \frac{Q_f}{V} c_{Af} \tag{18}$$

$$\dot{c}_B = \kappa\, c_A^2 + \frac{Q_f}{V} c_{Bf} \tag{19}$$

with parameter values

$$\frac{Q_f}{V} = 0.4; \;\; \kappa = 0.16; \;\; c_{Af} = 1; \;\; c_{Bf} = 0$$

The initial conditions are $c_A(0) = 3$ and $c_B(0) = 1$. The measurements correspond to the total pressure, which is the sum of the two states. The sample period is $\Delta = 0.1$. State estimation is carried out considering that x_0 is normal distributed with mean $[3\ 1]^T$ and variance I_2. We consider that in (18)-(19) a disturbance is introduced, which is zero-mean Gaussian distributed with variance $Q \cdot \Delta$, where $Q = 0.01 I_2$.

We are given $N = 100$ measurements, that are contaminated with additive zero-mean Gaussian white noise, with variance $\Delta \cdot R$, with $R = 0.01$.

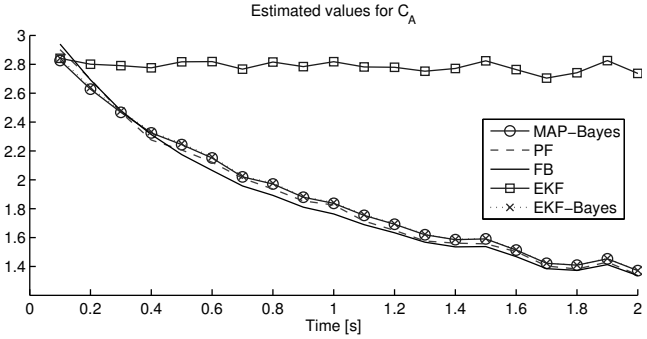
We compare the proposed approach (MAP-Bayes) with $N_x = 13$ against a Particle Filter (PF) (Gordon et al. 1993) with $N_{pf} = 100$, the Extended Kalman Filter (EKF) (see e.g. (Jazwinski 1970)) and an approximation to Full Bayesian Filtering via a fine deterministic grid (FB). Also we consider a variant of our algorithm, where instead of using MAP to obtain the single trajectory, we use EKF.

Figures 1a-1b show the estimated mean value obtained by each method. All the estimated mean values are close to the mean obtained by FB, save for the mean value provided by EKF which is far from the FB's estimate. Note that the proposed EKF-Bayes approach provides satisfactory results, even when EKF fails. Thus the combined strategy remedies the failure of EKF even though EKF is used as part of the combined algorithm. The failure of EKF in the batch reactor example has been previously reported in (Haseltine & Rawlings 2005).

Figure 2 shows the tradeoff in the MAP-Bayes approach between performance and computational load. This tradeoff is managed by choosing the number of points to be allocated, N_x. The performance is measured by using the Hellinger divergence (see e.g. (Liese & Vajda 2006)), where we use the distribution provided by FB as a reference. The computational load is measured by the average execution time per sample.

Fig. 1. Mean value of the distributions provided by each method.

(a) C_A (Concentration of A)



(b) C_B (Concentration of B)

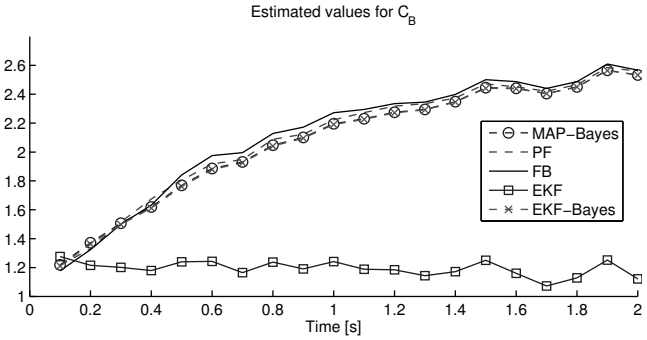
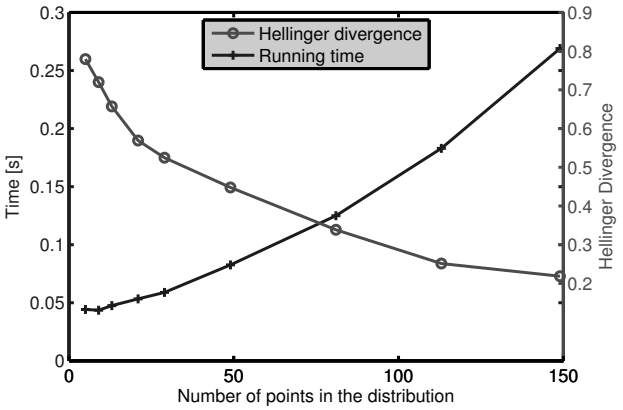


Fig. 2. Trade off between computational load and performance in MAP-Bayes method.



5.2 Water Tank

Consider a water tank having two areas: at some height, h_c, the transversal area increases by a factor of 10. We

consider that the water level is measured with noise, and we are interested in estimating the volume of water in the tank.

The volume of water in the tank is governed by

$$\frac{dV}{dt} = V_{in} - V_{out} \tag{20}$$

where

$$V_{out} = \begin{cases} \pi r_0^2 \sqrt{2g\frac{V}{\pi r_0^2}} & \text{if } V/(\pi r_0^2) \le h_c \\ \\ \pi r_0^2 \sqrt{2g(\frac{V}{\pi r_1^2} + h_c(1 - \frac{r_0^2}{r_1^2}))} & \text{if } V/(\pi r_0^2) > h_c \end{cases} \tag{21}$$

The input water volume is given by $V_{in} = q_{in} + w$, with q_{in} constant and w an i.i.d. Gaussian distributed process, with zero mean and variance $\Delta \cdot Q$. The sampling period is $\Delta = 0.1[s]$. We measure the water level, that is related to the volume by

$$h = \begin{cases} \frac{V}{\pi r_0^2} & \text{if } h \le h_c \\ \\ \frac{V}{\pi r_1^2} + h_c(1 - \frac{r_0^2}{r_1^2}) & \text{if } h > h_c \end{cases} \tag{22}$$

and the measurements are contaminated with an additive zero-mean Gaussian white noise, with variance $\Delta \cdot R$. The parameter values are chosen to be

$$\begin{array}{lll} r_0 = 1/\pi; & r_1 = \sqrt{10}/\pi & h_c = 1/(\pi r_0^2); \\ g = 9.80665; & q_{in} = 0.447 & \end{array}$$

We are given $N = 100$ measurements. We assume that the initial condition for the water volume is distributed uniformly between 0 and 2. We take $R = 0.01$ and $Q = 0.1$.

We compare the proposed approach (MAP-Bayes) against the Particle Filter (PF), the Extended Kalman Filter (EKF) and an approximation to Full Bayesian Filtering via a fine deterministic grid (FB). Also we consider a variant of our algorithm, where instead of using MAP to obtain the single trajectory, we use EKF.

For the MAP-Bayes method we use rolling horizon length $N_h = 1$, and use $N_x = 19$ points to approximate the standard Normal distribution. In PF, we use $N_{pf} = 100$ particles. For the EKF we approximate the initial distribution as Normal distributed with mean 1 and variance 1/3.

Table 1 shows the average running time per sample. We see that the proposed MAP-Bayes and EKF-Bayes have the same order of magnitud running time. For this example PF requires the least computational effort.

Table 2 shows the Mean Square Error (MSE) (between the "true value" as computed by FB and the given method) of the mean, mode, variance and skewness estimated for each method. For the estimated mean, variance and skewness PF provides the best estimates, but produce a poor estimate for the mode value. Next, EKF-Bayes produces mean, variance and skewness estimates with the same order of magnitude error as PF, nevertheless the estimated mode value is much better than that provided by PF. The MAP-Bayes method produces slightly worse estimates than EKF-Bayes, however this could be problem dependant, since in MAP-Bayes the points are allocated around the MAP estimate, and in the EKF-Bayes method the points are allocated around the mean value. EKF has the poorest performance over all the methods.

Table 1
Average running time per sample for each method.

	MAP-Bayes	PF	FB	EKF	EKF-Bayes
Time	0.046	0.003	3.660	0.002	0.016

Table 2
Mean Square Error of the mean, variance, mode and skewness of each method, compared to that obtained by Full Bayesian filtering.

Method	mean	mode	variance	skewness
MAP-Bayes	0.0245	0.0092	0.0029	0.5433
PF	0.0014	0.1341	0.0003	0.1545
EKF	0.1701	0.0856	0.0127	0.6341
EKF-Bayes	0.0030	0.0039	0.0004	0.3013

6 Conclusions

This paper has described a finite data and Moving Horizon state estimation scheme which allows one to combine MAP and Bayesian strategies. The scheme resolves two difficulties that are inherent in the usual (MAP based) Moving Horizon state estimation schemes, namely (i) how to obtain a meaningful "arrival cost" and (ii) how to quantify the accuracy of the resulting estimates. The performance of the scheme has been compared with other common schemes via simulation examples. The examples show that the new scheme gives good performance at reasonable computational cost. In practice, the trade-off in computational effort depends on the relative effort directed at finding the maximum of a function and exploring the full a posteriori distribution. The appropriate balance is problem dependent. We thus encourage others to use the ideas presented here. We provide access to preliminary software, see http://db.tt/b459SW2e.

Acknowledgements

We thank the anonymous reviewers for their constructive comments.

References

Alessandri, A., Baglietto, M. & Battistelli, G. (2008), 'Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes', *Automatica* **44**(7), 1753–1765.

Alessandri, A., Baglietto, M., Battistelli, G. & Zavala, V. (2010), Advances in moving horizon estimation for nonlinear systems, *in* '49th IEEE Conference on Decision and Control (CDC)', pp. 5681–5688.

Başar, T. & Bernhard, P. (2008), H_∞ *Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser Boston, Boston, MA.

Campi, M. C. & Weyer, E. (2002), 'Finite sample properties of system identification methods', *IEEE Transactions on Automatic Control* **47**, 1329–1334.

Chen, Z. (2003), 'Bayesian filtering: From Kalman filters to particle filters, and beyond', *Adaptive Systems Lab., McMaster University., Hamilton, Ontario, Canada.* .

Diehl, M., Ferreau, H. & Haverbeke, N. (2009), Efficient Numerical Methods for Nonlinear MPC and Moving Horizon Estimation, *in* L. Magni, D. Raimondo & F. Allgöwer, eds, 'Nonlinear Model Predictive Control', Vol. 384 of *Lecture Notes in Control and Information Sciences*, Springer Berlin / Heidelberg, pp. 391–417.

Findeisen, P. (1997), Moving horizon state estimation of discrete time systems, Master's thesis, University of Wisconsin-Madison.

Gersho, A. & Gray, R. (1992), *Vector Quantization and Signal Compression*, Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers.

Goodwin, G., Feuer, A. & Müller, C. (2010), 'Sequential bayesian filtering via minimum distortion filtering', *Control: Three Decades of Progress: Dedicated to Chris Byrnes and Anders Lindquist* .

Gordon, N. J., Salmond, D. & Smith, A. (1993), 'Novel approach to nonlinear/non-gaussian bayesian state estimation', *Radar and Signal Processing, IEEE Proceedings F* **140**(2), 107–113.

Haseltine, E. & Rawlings, J. (2005), 'Critical evaluation of extended Kalman filtering and moving-horizon estimation', *Industrial & engineering chemistry research* **44**(8), 2451–2460.

Jazwinski, A. (1970), *Stochastic processes and filtering theory*, Academic Press.

Liese, F. & Vajda, I. (2006), 'On divergences and informations in statistics and information theory', *Information Theory, IEEE Transactions on* **52**(10), 4394–4412.

Ljung, L. (1999), *System Identification: Theory for the user*, 2nd edn, Prentice Hall.

Rao, C. (2000), Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems, PhD thesis, University of Wisconsin-Madison, Madison, WI.

Rao, C., Rawlings, J. & Lee, J. H. (2001), 'Constrained linear state estimation – a moving horizon approach', *Automatica* **37**, 1619–1628.

Rao, C., Rawlings, J. & Mayne, D. (2003), 'Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations', *Automatic Control, IEEE Transactions on* **48**(2), 246–258.

Rawlings, J. & Bakshi, B. (2006), 'Particle filtering and moving horizon estimation', *Computers & chemical engineering* **30**(10-12), 1529–1541.

Rawlings, J. & Mayne, D. (2009), *Model Predictive Control: Theory and Design*, Nob Hill Publishing.

Ungarala, S. (2009), 'Computing arrival cost parameters in moving horizon estimation using sampling based filters', *Journal of Process Control* **19**(9), 1576 – 1588.

Verdu, S. & Poor, H. V. (1987), 'Abstract dynamic programming models under commutativity conditions', *SIAM J. Control Optim.* **25**, 990.

Weyer, E. & Campi, M. (2002), 'Non-asymptotic confidence ellipsoids for the least-squares estimate', *Automatica* **38**(9), 1539 – 1547.

Zavala, V. (2010), 'Stability analysis of an approximate scheme for moving horizon estimation', *Computers & Chemical Engineering* **34**(10), 1662–1670.