Performance Assessment of Sequential Bayesian Processors based on Probably Approximately Correct Computation and Information Theory

Article *in* Electronics Letters · January 2018

DOI: 10.1049/el.2017.4159

CITATIONS
0

READS
53

2 authors:

Ismael Jaras
University of Chile
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Marcos E. Orchard
University of Chile
116 PUBLICATIONS **1,652** CITATIONS
SEE PROFILE

Performance assessment of sequential Bayesian processors based on probably approximately correct computation and information theory

I. Jaras✉ and M.E. Orchard

A novel method to characterise the efficacy and efficiency of different sequential Bayesian processor implementations is proposed. This method is based on concepts of probably approximately correct computation and information theory measures. The proposed approach is used to compare the performance of three different Bayesian estimation algorithms (particle filter, unscented Kalman filter (UKF), and UKF with outer feedback correction loops) in the context of lithium-ion battery state-of-charge monitoring.

Introduction: State estimation is a critical task for the characterisation of uncertainty sources in failure prognostic algorithms. The Prognostic and Health Management community has typically used sequential Monte Carlo (SMC) methods [1] to perform this task in cases, where the system is nonlinear or when it undergoes non-Gaussian uncertainty sources [2]. Nevertheless, the computational cost associated with the implementation of SMC methods in embedded systems forces to use simplified versions of these algorithms to ensure real-time operation (in detriment of the quality of obtained results). The use of simplified Bayesian processors justifies the development of a framework to assess and compare the performance of these methods; incorporating the inherent probabilistic nature of different implementations within the analysis. In this regard, we hereby propose a flexible scheme that allows probabilistic assessment of worst-case scenarios in terms of the performance exhibited by a given algorithm implementation. This scheme also helps to analyse the performance of the algorithm as we get arbitrarily close to this worst-case scenario [3]. Information measures are used to quantify the discrepancy, performance wise, between simplified versions of a given Bayesian processor; also helping to compare against full-fledged versions that can be implemented on powerful computers.

Probably approximately correct computation (PACC) analysis: PACC [4] is a methodology born from the statistical learning theory [5] and the study of randomised algorithms [6] that allows estimating the performance loss in algorithms with arbitrary levels of precision and confidence. PACC Analysis denotes a reference algorithm by $f(z)$, $z \in Z$ (Z is the input space), whereas its simplified version is denoted by $f_\Delta(z)$. To characterise the computational accuracy of $f_\Delta(z)$ with respect to $f(z)$, PACC Analysis uses a loss function $u(z,\ \Delta) = u[f(z),\ f_\Delta(z)]$. PACC ensures a $\gamma > 0$ performance level, with probability η, when the following condition holds:

$$\Pr\{u(z,\ \Delta) \leq \gamma\} \geq \eta,\ \forall z \in Z. \tag{1}$$

If we denote $p_\gamma = \Pr\{u[f(z),\ f_\Delta(z)] < \gamma\}$, then PACC analysis would require to characterise p_γ on the whole space Z: a hard problem in terms of computational complexity. To overcome this difficulty, p_γ is approximated by $\hat{p}_N(\gamma) = (1/N)\sum I[f(z),\ f_\Delta(z)]$, where $I[f(z),\ f_\Delta(z)] = 1$ if $u[f(z),\ f_\Delta(z)] < \gamma$, and 0 otherwise.

The convergence of p_γ to p_N depends on the required level of precision ϵ, such that $|p_\gamma - \hat{p}_N| \leq \epsilon$. Since $\hat{p}_N$ is a function of the number of samples N, we require to introduce a confidence level of $(1 - \delta)$. The Chernoff bound [7] characterises the minimum number of samples that allow to guarantee

$$\Pr\{|p_\gamma - \hat{p}_N| < \varepsilon\} > (1 - \delta). \tag{2}$$

PACC allows 'to estimate the probability p_γ by means of $\hat{p}_N$ with precision ε' with probability $(1 - \delta)$. In this regard, we can approximate $\overline{\gamma}$, the maximum theoretical error (i.e. the minimum value of $\overline{\gamma}$ satisfying $\Pr[u(z,\ \Delta) \leq \overline{\gamma}] = 1\ \forall z \in Z$), by $\hat{\gamma}$ [maximum empirical error, obtained from $\hat{p}_N(\hat{\gamma})$]. Thus, if $\hat{\gamma}$ is an indicator of the computational accuracy for the simplified algorithm, D_1 (simplified Algorithm 1) is related to performance loss $\hat{\gamma}_1$, and D_2 (simplified Algorithm 2) is related to performance loss $\hat{\gamma}_2$, we can safely state that: the computational accuracy of D_1 is greater than $D_2 \Leftrightarrow \hat{\gamma}_1 < \hat{\gamma}_2$.

Information theory measures integration for PACC analysis in SMC-based algorithms: SMC-based algorithms [8] allow characterising the uncertainty associated with the latent variable x (also referred to as the state vector) in hidden Markov models (HMMs), which are typically represented by the target distribution $\pi_{1:T}(x_{1:T})$ [with marginal $\pi_n(x_n)$ in instant n], using a *posterior* distribution that is conditional on observations $y_{1:T}$. Assuming an ideal representation of the state-space $\pi^*_{1:T}(x_{1:T})$, and an approximation $\hat{\pi}_{1:T}(x_{1:T})$ that is generated by a simplified Bayesian processor, we can denote (in the context of PACC analysis) $f(z) = \pi^*_{1:T}(x_{1:T})$ as the outcome of a reference (i.e.; 'ground truth') algorithm and $f_\Delta(z) = \hat{\pi}_{1:T}(x_{1:T})$ as the outcome of a simplified algorithm version. A measure of the performance loss associated with the simplified algorithm implementation $u[\pi^*_{1:T}(x_{1:T}),\ \hat{\pi}_{1:T}(x_{1:T})]$ can be provided, based on Definitions 1 and 2.

Definition 1: (Average loss of information) $\overline{D}_{\mathrm{KL}}[\pi^*_{1:T}(x_{1:T}) \,||\, \hat{\pi}_{1:T}(x_{1:T})] = (1/T)\sum_{k=1}^{T} D_{KL}[\pi^*_k(x_k)||\hat{\pi}_k(x_k)]$.

Definition 2: (Absolute difference in average uncertainty) $\Delta\{H[\pi^*_{1:T}(x_{1:T}),\ \hat{\pi}_{1:T}(x_{1:T})]\} = |H[\pi^*_{1:T}(x_{1:T})] - H[\hat{\pi}_{1:T}(x_{1:T})]|$, where T is the number of available measurements. The *average uncertainty* $\overline{H}[\pi_{1:T}(x_{1:T})]$ is defined as $\overline{H}[\pi_{1:T}(x_{1:T})] = (1/T)\sum_{k=1}^{T} H[\pi_k(x_k)]$, D_{KL} is the Kullback–Lieber divergence, and H is the entropy [9].

At this point, we can integrate information theory measures for SMC-based algorithms in the PACC analysis, estimating the performance loss associated with different (and simplified) filtering strategies by approximating both the maximum *average loss of information* and the maximum *absolute difference in average uncertainty*. To illustrate this process, we will focus on a specific case study, where we seek to implement real-time state-of-charge (SoC) lithium-ion battery monitoring algorithms.

Case study (SoC lithium-ion battery monitoring): This case study is focused on real-time estimation of the SoC in lithium-ion batteries, which are a measure of the energy that is stored in the battery at any time instant. In [10], Pola *et al.* presented a state-space model for the lithium-ion battery cell that is compatible with the formulation of an HMM. Thus, the system is characterised by state transition (3) and (4), and the observation equation below:

$$x_1(k+1) = x_1(k) + w_1(k) \tag{3}$$

$$x_2(k+1) = x_2(k) - \frac{\hat{v}[x_1(k),\ x_2(k),\ u(k)]\,u(k)\,T_{\mathrm{s}}}{E_{\mathrm{crit}}} + w_2(k) \tag{4}$$

$$y(k) = \hat{v}[x_1(k),\ x_2(k),\ u(k)] + \eta(k), \tag{5}$$

where $\hat{v}[x_1(k),\ x_2(k),\ u(k)]$ is defined as

$$\begin{aligned}\hat{v}[x_1(k),\ x_2(k),\ u(k)] = {} & v_L + (v_0 - v_L)\,\mathrm{e}^{\gamma[x_2(k)-1]} \\ & + \alpha\, v_L\,[x_2(k) - 1] - u(k)\,x_1(k) \\ & + (1-\alpha)\,v_L\,[\mathrm{e}^{-\beta} - \mathrm{e}^{-\beta\sqrt{x_2(k)}}]\end{aligned} \tag{6}$$

$w_1(k) \sim \mathcal{N}(0,\ \sigma_1)$ and $w_2(k) \sim \mathcal{N}(0,\ \sigma_2)$ are Gaussian additive noises, whereas $\eta(k) \sim \mathcal{N}(0,\ \sigma_{\mathrm{obs}})$ represents the uncertainty in the measurements. The sampling time is T_{s}[sec], $u(k) = i(k)$[A] is the discharge current (input variable to the system), $y(k) = v(k)$[V] is the battery voltage. The state x_1 represents the internal impedance of the battery, whereas x_2 is the SoC of the battery (we will denote it as x^{SoC}). All the other parameters of the battery can be determined according to the procedure described in [10].

For this case study, we use voltage and current data from *Federal Urban Driving Schedule*, appropriately scaled to the case of a single lithium-ion cell [10], and use the proposed performance assessment method to evaluate the performance loss of simplified algorithms: particle filter (PF), unscented Kalman filter (UKF) [11], and UKF with outer feedback correction loops (OFCLs) [12]. A forward-filter–backward-smoothing (FF-BS) [8] algorithm represents the reference algorithm in this case (i.e.; 'the ground truth') since it allows to estimate the marginal distribution at any particular time $\pi^*_n(x_n)$ given the measurements that will be acquired up to some *later* time (via the backward-smoothing subroutine).

It is important to note that the FF-BS algorithm, as well as all other three filter implementations, considers (statistically) the same initial conditions for the system state

- $x_1(0) \sim \mathcal{N}(0.10,\ 0.05)$.
- $x_2(0) \sim \mathcal{N}(0.85,\ 0.10)$.

According to the Chernoff bound, we require $N = 1060$ realisations of any simplified Bayesian processor to estimate the probability associated with the maximum theoretical error with precision $\varepsilon = 0.05$ and confidence $(1 - \delta) = 0.99$, where a 'realisation' corresponds to an algorithm execution that uses the same set of acquired measurements and initial conditions sampled from identical distributions. Results from each realisation of simplified Bayesian processors are compared with the 'ground truth' $\pi^*_{1:T}(x_{1:T})$, all to obtain the corresponding evaluation of the loss function. After computing N realisations of this loss function, we can approximate the probability p_γ by $\hat{p}_N(\hat{\gamma})$, such that $\hat{p}_N(\hat{\gamma}) = 1$, $\forall \gamma \geq \hat{\gamma}$. The obtained parameter $\hat{\gamma}$ will correspond to the maximum empirical error for the simplified algorithm.

If we use 90% of the battery data (during any given discharge cycle, and starting with a fully charged cell), then models (1)–(6) allows the implementation of Bayesian processors to generate a statistical characterisation for the evolution of the SoC in time. Our proposed methodology will help to assess the performance of the aforementioned Bayesian processors (PF, UKF, and UKF with OFCL); estimating the maximum *average loss of information* and the maximum *absolute difference in average uncertainty*. Indeed, Fig. 1 shows the D_{KL} and its rate at each time instant, depicting the evolution of the information loss for each one of the three simplified algorithms with respect to the 'ground truth'.

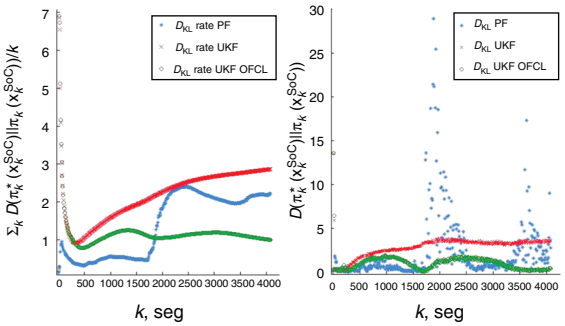

Fig. 1 *Evolution of Kullback–Lieber divergence for one realisation of estimation process, for each of algorithms included in this Letter*

Table 1 shows the obtained results in the analysis for the three simplified Bayesian processors. In this regard, the main results are summarised as follows:

- The loss function *average information loss* $(\bar{D}_{\mathrm{KL}}(\pi^*_{1:T}||\hat{\pi}_{1:T}))$ is a measure that indicates the average information lost when representing the target distribution $\pi^*_{1:T}(x_{1:T})$ using an approximate distribution $\hat{\pi}_{1:T}(x_{1:T})$. In this regard, the distribution $\hat{\pi}_{1:T}(x^{\mathrm{SoC}}_{1:T})$ obtained by the UKF, with OFCL, obtains the lowest value for the maximum theoretical error. According to this measure

$$\hat{\gamma}^{\mathrm{UKF_{ofcl}}}_{D_{\mathrm{KL}}} < \hat{\gamma}^{\mathrm{UKF}}_{D_{\mathrm{KL}}} < \hat{\gamma}^{\mathrm{PF}}_{D_{\mathrm{KL}}}$$

Thus, it can be stated that the UKF algorithm with OFCL obtains the highest computational accuracy, considering the amount of average information that is lost.

- Considering the *average information loss* $[\bar{D}_{\mathrm{KL}}(\pi^*_{1:T}||\hat{\pi}_{1:T})]$ and *absolute difference in average uncertainty* $\{\Delta[\bar{H}(\pi_{1:T}, \hat{\pi}_{1:T})]\}$, we can state that OFCL has a positive impact on the UKF algorithm. This routine is able to reduce the uncertainty associated with the state vector, without violating constraints imposed by the 'ground truth' algorithm. Thus, the inclusion of OFCLs has a positive effect on the computational accuracy and effectiveness of the algorithm.
- Finally, when solely consider the *absolute difference in average uncertainty* $\{\Delta[\bar{H}(\pi^*_{1:T}, \hat{\pi}_{1:T})]\}$, which represents a measure regarding the handling of the uncertainty by the algorithms, we have that

$$\hat{\gamma}^{\mathrm{UKF_{ofcl}}}_{\Delta(\bar{H})} < \hat{\gamma}^{\mathrm{PF}}_{\Delta(\bar{H})} < \hat{\gamma}^{\mathrm{UKF}}_{\Delta(\bar{H})}$$

This can lead to hasty conclusions about computational accuracy since you may think that PF handles uncertainty more effectively than the UKF. However, it must be noted that the maximum value for *absolute difference in average uncertainty* in PF (see Table 1) is obtained when the average entropy of the PF algorithm is lower than the one provided by the FF-BS (reference algorithm), whereas this never happens for UKF. The fact that the maximum value of *absolute difference in average uncertainty* for PF is reached when its average entropy is lower than that of the reference algorithm should be considered as a loss of important information of the approximation since it is representing the state of the system with less uncertainty than the referential algorithm. The measure *average information loss* is able to incorporate this loss in information by penalising more strongly a representation with lower entropy than the target distribution.

Table 1: Maximum empirical error for 90% of data ($N = 1060$)

Maximum empirical error ($\hat{\gamma}$) 90%	FF-BS	PF	UKF	UKF OFCL
$\bar{H}(\hat{\pi}_{1:T})\ \frac{[\mathrm{bits}]}{[\mathrm{seg}]}$	1.42	1.82	4.74	1.69
$\bar{D}_{KL}(\pi^*_{1:T}\|\|\hat{\pi}_{1:T})\ \frac{[\mathrm{bits}]}{[\mathrm{seg}]}$	0	44.87	2.86	0.99
$\Delta\{\bar{H}[(\pi^*_{1:T},\ \hat{\pi}_{1:T})]\}\ \frac{[\mathrm{bits}]}{[\mathrm{seg}]}$	0	0.82 (*)	3.32	0.26

Conclusion: A novel approach for evaluating sequential Bayesian processors is proposed. The methodology allows, on the one hand, generating an adequate framework that facilitates the calculation of the maximum error in probability, and therefore allows to assess the performance of simplified versions of Bayesian processors in terms of computational accuracy. On the other hand, we surmise that it does not suffice to compare the moments of the distribution to evaluate the performance of a given Bayesian processor since it is also important to quantify the performance in terms of information loss. In this regard, the development of novel methods to incorporate information measures for the assessment of estimation algorithms should be encouraged.

Acknowledgments: This work has been supported by the CONICYT FONDECYT 1170044, CONICYT PIA ACT 1405, and CONICYT Basal Project FB0008.

Submitted: *7 November 2017*
doi: 10.1049/el.2017.4159
One or more of the Figures in this Letter are available in colour online.

I. Jaras and M.E. Orchard (*Department of Electrical Engineering, University of Chile, Santiago, Chile*)

✉ E-mail: ismael.jaras@ug.uchile.cl

References

1 Doucet, A., de Freitas, N., and Gordon, N.: 'Sequential Monte Carlo methods in practice' (Springer, New York, NY, USA, 2001)
2 Orchard, M., Kacprzynski, G., Goebel, K., *et al.*: 'Advances in uncertainty representation and management for particle filtering applied to prognostics'. Proc. Int. Conf. Prognostics and Health Management, New York, NY, USA, November 2008
3 Bai, Er-Wei: 'Worst-case properties of the uniform distribution and randomized algorithms for robustness analysis'. American Control Conf., Albuquerque, NM, USA, July 1997
4 Alippi, C.: 'A probably approximately correct framework to estimate performance degradation in embedded systems', *Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 2002, **21**, pp. 749–762
5 Vidyasagar, M.: 'Statistical learning theory and randomized algorithms for control', *Control Syst.*, 1998, **7**, pp. 287–310
6 Alippi, C.: 'Randomized algorithms: a system-level, poly-time analysis of robust computation', *Trans. Comput.*, 2002, **51**, pp. 740–749
7 Chernoff, H.: 'A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations', *Ann. Math. Stat.*, 1952, **23**, pp. 493–507
8 Godsill, S., Doucet, A., and West, M.: 'Monte Carlo smoothing for nonlinear time series', *J. Am. Stat. Assoc.*, 2004, **99**, pp. 438–449
9 Cover, T., and Thomas, J.: 'Elements of information theory' (Wiley-Interscience, Hoboken, NJ, USA, 1991)
10 Pola, D., Navarrete, H., Orchard, M., *et al.*: 'Particle-filtering-based discharge time prognosis for lithium-ion batteries with a statistical characterization of use profiles', *Trans. Reliab.*, 2015, **64**, pp. 1–11
11 Wan, E.A., and Van der Merwe, R.: 'The unscented kalman filter for nonlinear estimation', Proc. of IEEE Symposium 2000 (AS-SPCC), Lake Louise, Alberta, Canada, October 2000
12 Orchard, M., Tobar, F., and Vachtsevanos, G.: 'Outer feedback correction loops in particle filtering-based prognosis algorithms: statistical performance comparison', *Stud. Inf. Controls*, 2009, **18**, pp. 295–304