APPROXIMATE BAYESIAN FILTERING USING STABILIZED FORGETTING

S. Azizi and A. Quinn

Department of Electronic and Electrical Engineering, Trinity College Dublin, Ireland

ABSTRACT

In this paper, we relax the modeling assumptions under which Bayesian filtering is tractable. In order to restore tractability, we adopt the stabilizing forgetting (SF) operator, which replaces the explicit time evolution model of Bayesian filtering. The principal contribution of the paper is to define a rich class of conditional observation models for which recursive, invariant, finite-dimensional statistics result from SF-based Bayesian filtering. We specialize the result to the mixture Kalman filter, verifying that the exact solution is available in this case. This allows us to consider the quality of the SF-based approximate solution. Finally, we assess SF-based tracking of the time-varying rate parameter (state) in data modelled as a mixture of exponential components.

***Index Terms*—** Approximate Bayesian filtering, stabilized forgetting, exponential family, mixture Kalman filter, exponential mixture

1. INTRODUCTION: BAYESIAN FILTERING

Bayesian filtering (BF) refers to the task of sequential inference, $f(\boldsymbol{\psi}_n|\boldsymbol{X}_n)$, $n = 1, 2, ...$, of state variables, $\boldsymbol{\psi}_n$, given observations, $\boldsymbol{X}_n$. The update from any $n-1$ to n involves two steps:

- The time step:

$$f(\boldsymbol{\psi}_n|\boldsymbol{X}_{n-1}) = \int f(\boldsymbol{\psi}_n|\boldsymbol{\psi}_{n-1}) f(\boldsymbol{\psi}_{n-1}|\boldsymbol{X}_{n-1}) d\boldsymbol{\psi}_{n-1} \tag{1}$$

  $n \geq 2$, where the state evolution model, $f(\boldsymbol{\psi}_n|\boldsymbol{\psi}_{n-1})$, is defined, here, under the first-order Markov property.

- The data step:

$$f(\boldsymbol{\psi}_n|\boldsymbol{X}_n) \quad \propto \quad f(x_n|\boldsymbol{\psi}_n) f(\boldsymbol{\psi}_n|\boldsymbol{X}_{n-1}), \tag{2}$$

  $n \geq 1$, where the observation model, $f(x_n|\boldsymbol{\psi}_n)$, is assumed here to be conditionally independent, known and time-invariant, given $\boldsymbol{\psi}_n$, and where the prior is $f(\boldsymbol{\psi}_1|\boldsymbol{X}_0) \equiv f(\boldsymbol{\psi}_1)$.

The widely known and adopted exact BF solution is the Kalman Filter (KF). Since the KF assumptions are too restrictive in practice (Gaussianity, linearity, and known parameters), it is necessary to relax the assumptions and seek approximations in practice [1–3]. A large research literature has been assembled on deterministic and stochastic approximations for BF in specific modelling contexts, and, indeed, combinations of both types of approximation can achieve good compromises between computational load and accuracy [4–6]. Among the key findings is that deterministic approaches may achieve greater accuracy than particle filtering (PF) techniques (and related sequential Monte Carlo approaches) for a *fixed* computational expenditure [7]. However, deterministic approximations such as variational Bayes (VB) [4] are local approximations, and so the inferential error cannot be bounded in the long-run. Uniquely, the empirical approximation that underlies PF is a global approximation, and so this error is bounded [4]. In this paper, we focus on stabilized forgetting (SF) as a local approximation in BF, and examine the efficient, recursive statistical computations that can be achieved for a rich class of models in this case.

In Section 1.1, we review the SF operator (local approximation) as a general technique for approximate sequential inference of non-stationary parameters, in the case where there is no explicit parameter evolution model, and we specialize this to the case of BF. In section 2, we present the main result of the paper as a lemma, proposing a rich class of models for which BF remains tractable under the SF operator. In section 3, we define the mixture KF (MKF) and verify that an *exact* BF solution is available in this case. This allows us to bench-mark the performance of the SF-approximated solution against the exact solution in this case, via simulations. In section 4, we examine a mixture of exponential models with time-variant rates. This practical context—though unamenable to exact computation—is one for which the results of our main lemma hold. We consider the performance of the SF solution for this model in simulation. Discussion and conclusions follow in Section 5.

1.1. Stabilized forgetting

When there is no explicit form of state-evolution model (1), then interleaving a forgetting operator with the Bayes operator is quite natural [8]. We emphasize SF because of the attractive trade-offs it can offer between computational load and accuracy. Furthermore, SF has recently been justified as an essential step in sequential local approximation [9]. Finally, SF can be derived as an optimal approximation within the full

This research was partly funded by SFI grant 10/CE/11855 (Lero).

probabilistic design (FPD) framework [10]. The Bayesian interpretation of SF was studied in [7, 8, 11].
In analogy to BF, the SF solution, $\tilde{f}(\psi_n|\boldsymbol{X}_n, \lambda_n)$, is achieved via two steps: (a) The time step, where (1) is approximated by $\tilde{f}(\psi_n|\boldsymbol{X}_{n-1}, \lambda_n)$, as follows:

$$\tilde{f}(\cdot) \in \arg\min_{g} \{\lambda_n D[g(\psi_n|\boldsymbol{X}_{n-1})\|f(\psi_{n-1}|\boldsymbol{X}_{n-1})_{\psi_n}]+ \\ +(1-\lambda_n)D[g(\psi_n|\boldsymbol{X}_{n-1})\|\bar{f}(\psi_n|\boldsymbol{X}_{n-1})]\} \quad (3)$$

where, $0 \leq \lambda_n \leq 1$ is the known forgetting factor at time n [7] and

$$D[g\|f] = \int_{\psi} g(\psi)\ln(\frac{g(\psi)}{f(\psi)})d\psi \quad (4)$$

is the Kullback-Leibler divergence (KLD) [7] from g to f. Also, $\bar{f}(\psi_n|\boldsymbol{X}_{n-1})$ is the alternative distribution in which prior knowledge about the states can also be updated; and $f(.)_{\psi_n}$ denotes the replacement of the argument of $f(.)$ with ψ_n. posterior distribution. Then, the time step of BF (1) is replaced by (3) [11], whose form is

$$\tilde{f}(\psi_n|\boldsymbol{X}_{n-1}, \lambda_n) \propto [f(\psi_{n-1}|\boldsymbol{X}_{n-1})_{\psi_n}]^{\lambda_n}\bar{f}(\psi_n|\boldsymbol{X}_{n-1})^{1-\lambda_n}.$$

(b) The data step (2) is then

$$\tilde{f}(\psi_n|\boldsymbol{X}_n, \lambda_n) \propto f(x_n|\psi_n)\tilde{f}(\psi_n|\boldsymbol{X}_{n-1}, \lambda_n)$$

2. THE MIXTURE OF EXPONENTIAL FAMILY OBSERVATION MODEL

In this section, we aim to relax significantly the restrictive assumptions of the observation model in KF, in such a way that the SF solution is tractable, while the KF solution is no longer available. Recall that the SF solution is computed without an explicit state evolution model. Tractability of SF is defined next.

Definition 2.1 (Tractable SF). *The SF solution is tractable if the filtering distribution, $\tilde{f}(\psi_n|\boldsymbol{X}_n, \lambda_n)$, can be recursively updated via finite, fixed-dimensional, non- sufficient statistics, which can be computed by an invariant algorithm.*

The following lemma provides us with a sufficient condition (set of observation models) for which the SF solution can be computed tractably.

Lemma 2.1. *A sufficient condition for tractable SF is that the observation model be in the form*

$$f(x_n|\psi_n, p) = a_p(\psi_n)b_p(x_n)\exp[\langle\phi_p(\psi_n), h_p(x_n)\rangle] \\ n = 1, 2, ... \text{ and } p = 1, ..., p_u \quad (5)$$

where $a_p(.)$ and $b_p(.)$ are p-conditional known, scalar kernels, $\langle ., . \rangle$ is a known two-argument scalar operator which is distributive across addition in its second argument, and $\phi_p(.)$, $h_p(.)$ are p-conditional known kernels with compatible dimensions. p is an unknown time-invariant index (model pointer) with p_u states.

Proof. Under the conditions of the lemma, each p-indexed component is a member of the exponential family (EF), and therefore possesses a conjugate (CEF) distribution that is invariant under Bayes' rule [7]. Let us consider the CEF distribution and its alternative, at time $n-1$, for the p^{th} component:

$$f(\psi_{n-1}|p, \boldsymbol{X}_{n-1}) = f(\psi_{n-1}|p, S_{n-1}) \\ \propto a_p^{\tilde{\nu}_{n-1}^{(p)}}(\psi_{n-1})\exp[\langle\phi_p(\psi_{n-1}), \tilde{V}_{n-1}^{(p)})\rangle] \quad (6)$$

where $S_{n-1} = (\tilde{V}_{n-1}, \tilde{\nu}_{n-1})$, and

$$\bar{f}(\psi_n|p, \bar{S}_{n-1}) \propto a_p^{\bar{\nu}_{n-1}^{(p)}}(\psi_n)\exp[\langle\phi_p(\psi_n), \bar{V}_{n-1}^{(p)})\rangle]$$

where $\bar{S}_{n-1} = (\bar{V}_{n-1}^{(p)}, \bar{\nu}_{n-1}^{(p)})$. The time step of SF can be effected as

$$f(\psi_n|p, \boldsymbol{X}_{n-1}, \lambda_n) = f(\psi_n|p, S_{n-1}, \bar{S}_{n-1}, \lambda_n) \\ \propto a_p^{\tilde{\nu}_{n|n-1}^{(p)}}(\psi_n)\exp[\langle\phi_p(\psi_n), \tilde{V}_{n|n-1}^{(p)}\rangle] \quad (7)$$

where

$$\tilde{V}_{n|n-1}^{(p)} = \lambda_n\tilde{V}_{n-1}^{(p)} + (1-\lambda_n)\bar{V}_{n-1}^{(p)}$$
$$\tilde{\nu}_{n|n-1}^{(p)} = \lambda_n\tilde{\nu}_{n-1}^{(p)} + (1-\lambda_n)\bar{\nu}_{n-1}^{(p)}$$

Therefore, the data step of SF is effected as

$$f(\psi_n|p, \boldsymbol{X}_n, \lambda_n) = f(\psi_n|p, S_n, \lambda_n) \\ \propto a_p^{\tilde{\nu}_n^{(p)}}(\psi_n)\exp[\langle\phi_p(\psi_n), \tilde{V}_n^{(p)}\rangle] \quad (8)$$

where,

$$\tilde{V}_n^{(p)} = \tilde{V}_{n|n-1}^{(p)} + h_p(\boldsymbol{x}_n) \quad (9)$$
$$\tilde{\nu}_n^{(p)} = \tilde{\nu}_{n|n-1}^{(p)} + 1 \quad (10)$$

Adopting the uniform prior distribution for p,

$$Pr\{L = p\} \equiv \frac{1}{p_u}, p = 1, \ldots, p_u,$$

then the posterior distribution of p is given by

$$\tilde{\alpha}_n^{(p)} = \Pr(L = p|\boldsymbol{x}_n, \lambda_n) \\ = \frac{\zeta_{\psi_n}^{(p)}(\tilde{V}_n^{(p)}, \tilde{\nu}_n^{(p)})}{\sum_{d=1}^{p_u}\zeta_{\psi_n}^{(d)}(\tilde{V}_n^{(d)}, \tilde{\nu}_n^{(d)})}, \quad p = 1, \ldots, p_u \quad (11)$$

where $\zeta_{\psi_n}^{(p)}(.)$ is the (available) normalizing constant of the recovered p-conditioned CEF form (8). Finally, the filtering distribution can be computed by marginalizing over p and using the chain rule:

$$f(\psi_n|\boldsymbol{X}_n, \lambda_n) = \sum_{p=1}^{p_u} f(\psi_n|p, \boldsymbol{X}_n, \lambda_n)\Pr(L = p|\boldsymbol{X}_n, \lambda_n) \\ \propto \sum_{p=1}^{p_u}\tilde{\alpha}_n^{(p)}a_p^{\tilde{\nu}_n^{(p)}}(\psi_n)\exp[\langle\phi_p(\psi_n), \tilde{V}_n^{(p)}\rangle] \quad (12)$$

□

Remark 2.2. *If $p_u = 1$ in (5), then the lemma specializes to the exponential family observation model, for which the SF BF solution follows from (8):*

$$f(\psi_n|\boldsymbol{X}_n, \lambda_n) \propto a^{\tilde{\nu}_n}(\psi_n) \exp[\langle \phi(\psi_n), \tilde{V}_n \rangle]$$

with

$$\tilde{V}_n = \tilde{V}_{n|n-1} + h(x_n)$$
$$\tilde{\nu}_n = \tilde{\nu}_{n|n-1} + 1$$

3. THE MIXTURE KALMAN FILTER

In this section, we present an extension of the KF model constraints for which BF remains tractable. This mixture KF (MKF) model involves a *time-invariant* index into one of p_u conditional observation models, each of which satisfies the KF observation model constraints. Hence, (13) is specialized to

$$f(x_n|\psi_n, p) \equiv N_{x_n}(C_n^{(p)}\psi_n, r_x^{(p)}) \tag{13}$$

where $C_n^{(p)}, r_x^{(p)}, \quad n = 1, 2, \cdots$ are known scalars. Recall that an explicit state evolution model (it may also be indexed by p) is adopted by the KF, as follows:

$$f(\psi_n|\psi_{n-1}) \equiv N_{\psi_n}(A_n\psi_{n-1}, r_\psi) \tag{14}$$

where $A_n, r_\psi, \quad n = 1, 2, \cdots$ are known scalars. Therefore, the p-conditional BF updates are of the KF type:

$$f(\psi_n|p, \boldsymbol{X}_n) \quad \propto \quad N_{\psi_n}(\mu_n^{(p)}, \sigma_n^{(p)2}) \tag{15}$$

where,

$$\mu_n^{(p)} = A_n\mu_{n-1} + \{\frac{C_n^{(p)}(r_\psi + A_n^2\sigma_{n-1}^{(p)2})}{C_n^{(p)2}(r_\psi + A_n^2\sigma_{n-1}^{(p)2}) + r_x^{(p)}}\} .[x_n - C_n^{(p)}A_n\mu_{n-1}^{(p)}] \tag{16}$$

$$\sigma_n^{(p)2} = (r_\psi + A_n^2\sigma_{n-1}^{(p)2})\{1 - \frac{C_n^{(p)2}(r_\psi + A_n^2\sigma_{n-1}^{(p)2})}{C_n^{(p)2}(r_\psi + A_n^2\sigma_{n-1}^{(p)2}) + r_x^{(p)}}\} \tag{17}$$

Using the insight of (12), we can write down the MKF solution (15) as follows:

$$f(\psi_n|\boldsymbol{X}_n) \propto \sum_{p=1}^{p_u} \alpha_n^{(p)}.\, N_{\psi_n}(\mu_n^{(p)}, \sigma_n^{(p)2})$$

where,

$$\alpha_n^{(p)} = \frac{\sigma_n^{(p)2}}{\sum_{d=1}^{p_u}\sigma_n^{(d)2}}, \ \ p = 1, \ldots, p_u$$

Remark 3.1 (Kalman filter)**.** *When $p_u = 1$, MKF collapses to the exact KF. The exact KF solution can be similarly obtained from (15)-(17).*

We can derive the SF solution for the MKF context by replacing the state evolution model (14) with the forgetting operator . We call this solution the mixture SF (MSF), and aim to compareit against the exact MKF solution. We consider the conditional observation model (5) to be univariate normal with known variance $r_x^{(p)}$:

$$f(x_n|\psi_n, p) \equiv N_{x_n}(\psi_n, r_x^{(p)}), \ \ n = 1, 2, ... \tag{18}$$

Therefore, using (7) and (8), the SF solution at time n is conjugate normal as,

$$f(\psi_n|p, X_n, \lambda_n) \equiv N_{\psi_n}(\tilde{\mu}_n^{(p)}, \tilde{\sigma}_n^{(p)2}) \tag{19}$$

Replacing $h_p(\boldsymbol{x}_n)$ in (9) with $[x_n \quad 1]^T[x_n \quad 1]$, $\tilde{V}_n^{(p)}$ can be updated. Also, we partition $\tilde{V}_n^{(p)} \in R^{2\times 2}$ (the extended information matrix [7]) into blocks as

$$\tilde{V}_n^{(p)} = \begin{bmatrix} \tilde{V}_{n,11}^{(p)} & \vdots & \tilde{V}_{n,21}^{(p)T} \\ \cdots & \cdots & \cdots \\ \tilde{V}_{n,21}^{(p)} & \vdots & \tilde{V}_{n,22}^{(p)} \end{bmatrix}, \tag{20}$$

which accumulates data. Using (20), the shaping parameters can be computed as follows:

$$\tilde{\mu}_n^{(p)} = \frac{\tilde{V}_{n,21}^{(p)}}{\tilde{V}_{n,22}^{(p)}}, \quad \tilde{\sigma}_n^{(p)2} = \frac{r_x^{(p)}}{\tilde{V}_{n,22}^{(p)}}$$

Therefore, substituting (19) into (12), the SF solution in this special case is attained as follows:

$$f(\psi_n|\boldsymbol{X}_n, \lambda_n) \propto \sum_{p=1}^{p_u} \tilde{\alpha}_n^{(p)}.\, N_{\psi_n}(\tilde{\mu}_n^{(p)}, \tilde{\sigma}_n^{(p)2})$$

where,

$$\tilde{\alpha}_n^{(p)} = \frac{\tilde{\sigma}_n^{(p)2}}{\sum_{d=1}^{p_u}\tilde{\sigma}_n^{(d)2}}, \ \ p = 1, \ldots, p_u$$

Further specialization of (5) can be considered when $p_u = 1$ in (18). The resulting SF solution is attained similarly, as follows:

$$f(\psi_n|\boldsymbol{X}_n, \lambda_n) \propto N_{\psi_n}(\tilde{\mu}_n, \tilde{\sigma}_n^2)$$

3.1. Simulation study

We examine the performance of the MSF solution against the exact MKF solution. Once again, in MSF, there is no explicit state evolution model, while the Gaussian distribution is adopted as the state evolution model in MKF. Firstly, we compare KF solution versus SF solution in Figure 2. Here, the results are attained in 1000 Monte Carlo runs. Figure 1 shows the performance of the exact KF solution versus the SF solution (with three values of forgetting factor $\lambda = 0.7, 0.8, 0.9$) in the Gaussian case, with

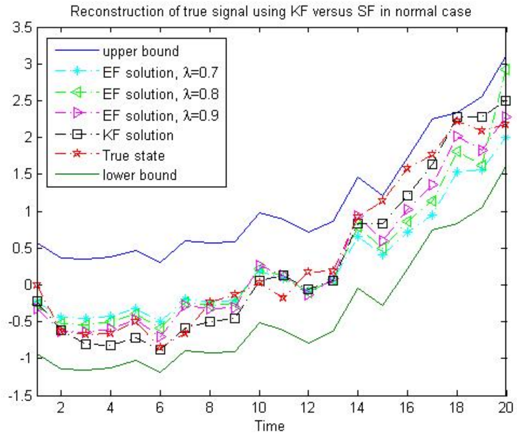

Fig. 1. State reconstruction using SF versus KF in normal case $r_x = 1, r_\psi = 1$.

$r_x = r_\psi = 1$, and considering diffuse prior $\tilde{\mu}_1 = 0$, $\tilde{\sigma}_1^2 = 1$. Figure 1 shows reconstruction of ψ_n. It is clear that the exact KF solution performs best in tracking ψ_n (red $\star$), while the SF solution with $\lambda = 0.9$ also performs well. This is because the variation of state r_ψ is small (slowly time-varying). The upper and lower Bayesian interval, $[E(\psi_n|\boldsymbol{X}_n) - \sqrt{\text{var}(\psi_n|\boldsymbol{X}_n)}, E(\psi_n|\boldsymbol{X}_n) + \sqrt{\text{var}(\psi_n|\boldsymbol{X}_n)}]$, around the SF solution with $\lambda = 0.9$ is also displayed. Figure 2 illustrates the performance of the MKF solution versus MSF solution in the Gaussian case with $r_\psi = 2$, $p_u = 5$ and $C^{(p)} = \{1.2, 1.4, 1.6, 1.8, 2\}$. Here, observations are generated from $N(\psi_n, 2)$ and the state evolution model is $N(2\psi_{n-1}, 1)$. Also diffuse prior $\tilde{\mu}_1 = 0$, $\tilde{\sigma}_1^2 = 1$ is considered. MSF with the highest λ performs best. This is because the state is again slow-varying, and so minimal forgetting (large λ) best matches the data in this case. Nevertheless, most of inferred state trajectories remain inside the standard interval for all tested values of λ.

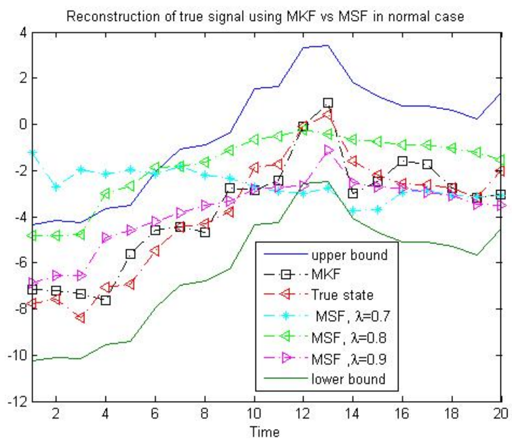

Fig. 2. State reconstruction using MKF vs MSF (normal case).

4. EXPONENTIAL MIXTURE WITH TIME-VARIANT RATE

We now investigate the exponential observation model as a special case of the (5). It models, for example, the inter-arrival times of particles released by a fixed mass of a radio-isotope, known to be one of p_u possible species ($p = 1, ..., p_u$):

$$f(x_n|\psi_n, p) \overset{cid}{\equiv} \mathcal{E}xp(\psi_n^{(p)}), \ \ x_n > 0, \ \psi_n^{(p)} > 0$$

$$a_p(\psi_n) = \phi_p(\psi_n) = \psi_n^{(p)}, \quad h_p(x_n) = x_n.$$

Here, *cid* denotes 'conditionally independently distributed', $\psi_n^{(p)}$ is the time-varying rate parameter of the exponential observation model for the p^{th} species, and $a_p(.)$ and $\phi_p(.)$ are the exponential family kernels (5). The CEF distribution (6) at time $n-1$ is Ga$(\tilde{\nu}_{n-1}^{(p)}, \tilde{V}_{n-1}^{(p)})$, and the alternative distribution is also chosen as Ga(.). Therefore using (8), the SF solution is

$$f(\psi_n|p, \boldsymbol{X}_n, \lambda_n) \propto [\psi_n^{(p)}]^{\tilde{\nu}_n^{(p)}} \exp[\langle \psi_n^{(p)}, \tilde{V}_n^{(p)} \rangle] \tag{21}$$

$$\tilde{V}_n^{(p)} = \tilde{V}_{n|n-1}^{(p)} + x_n$$

$$\tilde{\nu}_n^{(p)} = \tilde{\nu}_{n|n-1}^{(p)} + 1$$

Furthermore, substituting (21) into (12), the MSF solution can be deduced as

$$f(\psi_n|\boldsymbol{X}_n, \lambda_n) \propto \sum_{p=1}^{p_u} \{\tilde{\alpha}_n^{(p)} [\psi_n^{(p)}]^{\tilde{\nu}_n^{(p)}} . \exp[\langle \psi_n^{(p)}, \tilde{V}_n^{(p)} \rangle]\}$$

where $\tilde{\alpha}_n^{(p)}$ is obtained by substituting the normalizing constant, $\zeta_{\psi_n}^{(p)}(\tilde{V}_n^{(p)}, \tilde{\nu}_n^{(p)}) = \frac{\Gamma(\tilde{\nu}_n^{(p)})}{(\tilde{V}_n^{(p)})^{\tilde{\nu}_n^{(p)}}}$, into (11).

4.1. Simulation study

Figures 3 and 4 illustrate the reconstruction of the true state, ψ_n, using the SF solution in the case of slow- and fast-varying state variables, respectively. In Figure 3, $f(\ln(\psi_n)|\ln(\psi_{n-1})) \sim N(0, 0.5)$, for which larger λ is chosen. On the other hand, in Figure 4, $f(\ln(\psi_n)|\ln(\psi_{n-1})) \sim N(0, 2)$ for which small λ is selected. We infer (reconstruct) the state using $\lambda = 0.9$ in the former case, and using $\lambda = 0.2$ in the latter case. Once again, the true state trajectory is dominantly contained within the inferred Bayesian interval estimate of the state.

5. DISCUSSION

Lemma 2.1 remains true for a time-variant forgetting factor, λ_n, and there is interest in the design of a schedule for λ_n in this case. In particular, a sequential, data-driven assignment rule, $\lambda_n \equiv \lambda(\mathbf{X}_{n-1})$ is of interest, and several proposals are available in the Bayesian literature [9, 12]. In this paper, we have confined our attention to fixed-λ forgetting,

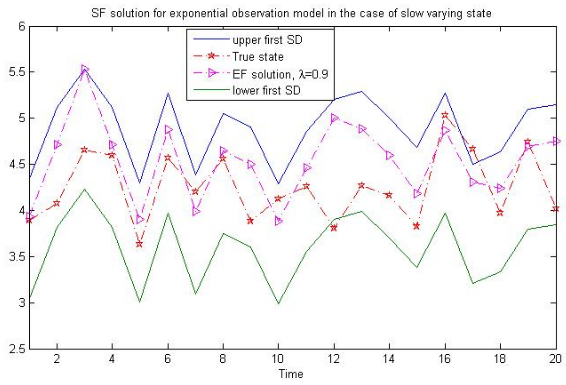

Fig. 3. State reconstruction using SF for the exponential observation model in the slowly-varying state case.

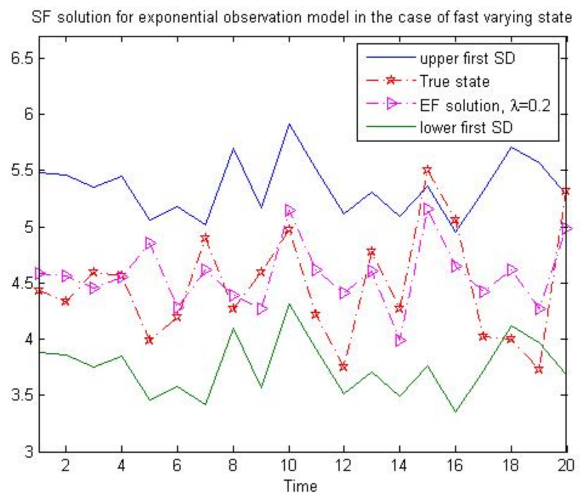

Fig. 4. State reconstruction using SF for the exponential observation model in the fast-varying state case.

and have noted the dependence of the optimal choice on the bandwidth of the state. Nevertheless, the state inference typically remains within the standard interval of the optimal λ for a range of these forgetting factors. The results in this paper highlight the elegant and computationally efficient recursive computations that are preserved by SF for a rich class of models, achieving good performance. Indeed, it provides an attractive trade-off between accuracy and computational load, potentially favourably in comparison to Monte Carlo-based techniques such as PF. In addition to the application in Section 4, Lemma 2.1 can also be applied to certain signal-dependent-noise problems [13]. It can also be applied to the problem described in [14], where the user of a wireless channel is drawn from p_u possible users, and where $\psi_n^{(p)}$ is the time-variant capacity, indexed by the active (pth) user.

The lemma does not claim to provide a necessary condition for tractability of SF-approximated Bayesian filtering, and, indeed, we have explored further extensions to time-variant parameter models for which tractability is preserved. This work will be reported in future publications.

REFERENCES

[1] A. H. Jazwinski, *Stochastic processes and filtering theory*, Courier Dover Publications, 1979.

[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series)*, The MIT press, 2000.

[3] F. E. Daum, "Nonlinear filters: beyond the Kalman filter," in *Aerospace and Electronic Systems Magazine, IEEE*, 2005, pp. 57–69.

[4] V. Šmídl and A. Quinn, "Variational Bayesian filtering," in *Signal Processing, IEEE Transactions on, 56(10)*, 2008, pp. 5020–5030.

[5] C. Nemeth, P. Fearnhead, and L. Mihaylova, "Sequential Monte Carlo methods for state and parameter estimation in abruptly changing environments," in *Signal Processing, IEEE Transactions on, 62(5)*, 2014, pp. 1245–1255.

[6] D. Xu, C. Shen, and F. Shen, "A robust particle filtering algorithm with non-Gaussian measurement noise using Student-t distribution," in *IEEE Signal Processing Letters, 21*, 2014, pp. 30–34.

[7] V. Šmídl and A. Quinn, *The variational Bayes method in signal processing*, Springer, 2006.

[8] R. Kulhavý and M. B. Zarrop, "On a general concept of forgetting," in *International Journal of Control, 58(4)*, 1993, pp. 905–924.

[9] M. Kárný, "Approximate Bayesian recursive estimation," in *Information Sciences, 285*, 2014, pp. 100–111.

[10] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesar, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, 2006.

[11] R. Kulhavý and F. J. Kraus, "On duality of regularized exponential and linear forgetting," in *Automatica, 32(10)*, 1996, pp. 1403–1415.

[12] V. Šmídl and A. Quinn, "Bayesian estimation of non-stationary ar model parameters via an unknown forgetting factor," in *In proceedings of the IEEE on signal processing*, 2004, pp. 100–105.

[13] S. Azizi and A. Quinn, "A data-driven forgetting factor for stabilized forgetting in approximate Bayesian filtering," in *26th Irish Signals and Systems Conference, accepted for publication*, 2015.

[14] S. Shakkottai and A. L. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," in *Translations of the American Mathematical Society-Series 2, 207*, 2002, pp. 185–202.