On Bayesian Fixed-Interval Smoothing Algorithms

Boujemaa Ait-El-Fquih and François Desbouvries
Institut TELECOM / TELECOM & Management SudParis / dépt. CITI and CNRS UMR 5157, 9 rue Charles Fourier, 91011 Evry, France
tel +33 1 60 76 45 27, fax +33 1 60 76 44 33,
Francois.Desbouvries@it-sudparis.eu

***Abstract*— In this note we revisit fixed-interval Kalman like smoothing algorithms. We have two results. We first unify the family of existing algorithms by deriving them in a common Bayesian framework; as we shall see, all these algorithms stem from forward and/or backward Markovian properties of the state process, involve one (or two) out of four canonical probability density functions, and can be derived from the systematic use of some generic properties of Gaussian variables which we develop in a specific toolbox. On the other hand the methodology we use enables us to complete the set of existing algorithms by five new Kalman like smoothing algorithms, which is our second result.**

I. Introduction

A. Background : Fixed-interval Kalman like smoothing algorithms

Let us consider the state space system

$$\left\{ \begin{array}{lcl} \mathbf{x}_{n+1} & = & \mathbf{F}_n\mathbf{x}_n + \mathbf{G}_n\mathbf{u}_n \\ \mathbf{y}_n & = & \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n \end{array} \right. , \quad (1)$$

in which $\mathbf{x}_n \in \mathbb{R}^{n_\mathbf{x}}$ is the state, $\mathbf{y}_n \in \mathbb{R}^{n_\mathbf{y}}$ the observation, $\mathbf{u}_n \in \mathbb{R}^{n_\mathbf{u}}$ the process noise and $\mathbf{v}_n \in \mathbb{R}^{n_\mathbf{v}}$ the measurement noise. The processes $\mathbf{u} = \{\mathbf{u}_n\}_{n\in\mathbb{N}}$ and $\mathbf{v} = \{\mathbf{v}_n\}_{n\in\mathbb{N}}$ are zero-mean, independent, jointly independent and independent of $\mathbf{x}_0$. Fixed-interval Kalman smoothing aims at estimating $\mathbf{x}_n$ from $\mathbf{y}_{0:N}$ for $0 \leq n \leq N$. In the literature, various algorithms have been derived by using such different methods as calculus of variations [1], maximum a posteriori [2] [3], orthogonal projections [4], the innovations approach [5], the two-filter form [6] [7], complementary models [8] or the Bayesian approach [9] [10] (modern surveys can also be found e.g. in [11, ch. 10] [8] or [12]). The most well-known algorithms are now the Bryson-Frazier algorithm [1], the Rauch-Tung-Streibel (RTS) algorithm [3] and the two-filter algorithm [6] [7].

B. Contributions

In this note we propose a unifying methodology which enables us to gather and extend the family of existing smoothing algorithms. More precisely, we first adopt the Bayesian point of view, and we use both forward and backward Markovian properties of the state process $\mathbf{x} = \{\mathbf{x}_n\}$ in order to derive three families of four smoothing algorithms for general continuous state hidden Markov chains (HMC). We then further particularize to the Gaussian case; our twelve algorithms reduce to seven known Kalman like smoothing algorithms, as well as to five new ones.

Let us give some comments on the originality of our contribution. Of course, as is well known, the Bayesian point of view in Kalman filtering is far from being new [13]. We believe however that our classification of existing smoothing algorithms as specific entries of three two-by-two arrays, which are built from forward and backward Markovian properties of $\mathbf{x}$, and on the use of one (or two) out of four canonical probability density functions (pdf) α_n, β_n, γ_n and δ_n (see Tables III to V), is original. Also, a practical advantage of this new classification is that it provides (as a byproduct) five original smoothing algorithms, which happen to be just specific empty entries in these arrays. So our classification happens to be both a way to unify the existing algorithms as well as a tool for proposing new ones. Finally, the actual algorithms are systematically derived by applying some generic properties of Gaussian variables which we develop for our purpose in a separate toolbox (see the Appendix).

This note is organized as follows. In §II we derive the general smoothing algorithms. In §III we particularize to the Gaussian case and we comment on the algorithms we get. The algorithms we obtain (either new or original) are systematically derived by using some results in Gaussian variables gathered in the appendix; for illustrative purposes §IV is devoted to a worked example.

II. Bayesian smoothing algorithms for continuous state HMC

Let (1) hold. Let $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \cdots, \mathbf{x}_n\}$, $\mathbf{y}_{0:n} = \{\mathbf{y}_0, \cdots, \mathbf{y}_n\}$, and let $p(\mathbf{x}_{0:n})$ (resp. $p(\mathbf{x}_n|\mathbf{y}_{0:n})$), say, denote the pdf (w.r.t. Lebesgue measure) of $\mathbf{x}_{0:n}$ (resp. of $\mathbf{x}_n$ given $\mathbf{y}_{0:n}$); other pdfs of interest are defined similarly. As is well known, model (1) is an HMC, i.e. the following properties hold:

$$p(\mathbf{x}_{n+1}|\mathbf{x}_{0:n}) = p(\mathbf{x}_{n+1}|\mathbf{x}_n); \quad (2)$$

$$p(\mathbf{y}_{0:n}|\mathbf{x}_{0:n}) = \prod_{i=0}^{n} p(\mathbf{y}_i|\mathbf{x}_{0:n}); \quad (3)$$

$$p(\mathbf{y}_i|\mathbf{x}_{0:n}) = p(\mathbf{y}_i|\mathbf{x}_i) \text{ for all } 0 \leq i \leq n. \quad (4)$$

The aim of this section is to compute the smoothing pdf $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ for all n, $0 \leq n \leq N$. The algorithms we propose can be classified into three families :

1) Backward recursive algorithms (see §II-B). These are two-pass algorithms, in which $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ is computed from $p(\mathbf{x}_{n+1}|\mathbf{y}_{0:N})$ (whence the term "backward") via
$$p(\mathbf{x}_n|\mathbf{y}_{0:N}) = \int p(\mathbf{x}_{n+1}|\mathbf{y}_{0:N}) p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N}) d\mathbf{x}_{n+1}, \quad (5)$$
and $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N})$ in (5) is computed in the forward direction (i.e., for increasing values of n);
2) Forward recursive algorithms (see §II-C). These are two-pass algorithms, in which $p(\mathbf{x}_{n+1}|\mathbf{y}_{0:N})$ is computed from $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ via
$$p(\mathbf{x}_{n+1}|\mathbf{y}_{0:N}) = \int p(\mathbf{x}_n|\mathbf{y}_{0:N}) p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{0:N}) d\mathbf{x}_n, \quad (6)$$
and $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{0:N})$ in (6) is computed in the backward direction;
3) Non-recursive algorithms (see §II-D). In these algorithms, $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ is computed from two pdfs; one of them is computed recursively in the forward direction and the other in the backward direction.

As we are going to see, further classification is obtained from two considerations. First, it happens that each of the three families of algorithms above contains one algorithm which only uses the forward HMC transition pdfs (i.e., the forward Markov transition pdf $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$ and the observation transition pdf $p(\mathbf{y}_n|\mathbf{x}_n)$), one algorithm which only uses the backward HMC transition pdfs (i.e., the backward Markov[1] transition pdf $p(\mathbf{x}_n|\mathbf{x}_{n+1})$ and the observation transition pdf $p(\mathbf{y}_n|\mathbf{x}_n)$), and two algorithms which use both. Next, any algorithm out of sections II-B, II-C or II-D makes use of one (or two) out of the four pdfs $\alpha_n \overset{\text{def}}{=} p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$, $\beta_n \overset{\text{def}}{=} p(\mathbf{y}_{n:N}|\mathbf{x}_n)$, $\gamma_n \overset{\text{def}}{=} p(\mathbf{x}_n|\mathbf{y}_{n+1:N})$ and $\delta_n \overset{\text{def}}{=} p(\mathbf{y}_{0:n}|\mathbf{x}_n)$. These pdfs, in turn, can be computed recursively (in the forward direction for α_n and δ_n, in the backward direction for β_n and γ_n) from the (either forward

[1] Since $\mathbf{x}$ is a Markov Chain (MC), $\mathbf{x}$ is also an MC in the backward direction, i.e. $p(\mathbf{x}_n|\mathbf{x}_{n+1:N}) = p(\mathbf{x}_n|\mathbf{x}_{n+1})$.

or backward) HMC transition pdfs; so for sake of clarity we first gather these recursions in §II-A. Proofs of (7)-(22) are obtained from Bayes's rule and (2)-(4).

A. Recursive algorithms for α_n, β_n, γ_n and δ_n

The algorithm described in Proposition 1 (resp. Proposition 2) propagates α_n (resp. δ_n) in the forward direction, β_n (resp. γ_n) in the backward direction, and only uses forward (resp. backward) transition HMC pdfs.

Proposition 1: Assume that (2)-(4) hold, and that we are given $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$ and $p(\mathbf{y}_n|\mathbf{x}_n)$. Then the one-step ahead prediction pdf $\alpha_n = p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ and filtering pdf $\widetilde{\alpha}_n = p(\mathbf{x}_n|\mathbf{y}_{0:n})$ can be propagated from $n = 0$ to N (with $\alpha_0 = p(\mathbf{x}_0)$) as

$$\left\{ \begin{array}{lcl} \widetilde{\alpha}_n & = & \frac{p(\mathbf{y}_n|\mathbf{x}_n)\alpha_n}{\int p(\mathbf{y}_n|\mathbf{x}_n)\alpha_n d\mathbf{x}_n} \\ \alpha_{n+1} & = & \int p(\mathbf{x}_{n+1}|\mathbf{x}_n)\widetilde{\alpha}_n d\mathbf{x}_n \end{array} \right. ; \tag{7}$$

on the other hand, the likelihood functions $\beta_n = p(\mathbf{y}_{n:N}|\mathbf{x}_n)$ and $\widetilde{\beta}_n = p(\mathbf{y}_{n+1:N}|\mathbf{x}_n)$ can be computed from $n = N$ to $n = 0$ (with $\beta_{N+1} = 1$) as

$$\left\{ \begin{array}{lcl} \widetilde{\beta}_n & = & \int p(\mathbf{x}_{n+1}|\mathbf{x}_n)\beta_{n+1}\, d\mathbf{x}_{n+1} \\ \beta_n & = & p(\mathbf{y}_n|\mathbf{x}_n)\, \widetilde{\beta}_n \end{array} \right. . \tag{8}$$

Proposition 2: Assume that (2)-(4) hold, and that we are given $p(\mathbf{x}_n|\mathbf{x}_{n+1})$ and $p(\mathbf{y}_n|\mathbf{x}_n)$. Then the likelihood functions $\delta_n = p(\mathbf{y}_{0:n}|\mathbf{x}_n)$ and $\widetilde{\delta}_{n+1} = p(\mathbf{y}_{0:n}|\mathbf{x}_{n+1})$ can be computed from $n = 0$ to N (with $\widetilde{\delta}_0 = 1$) as

$$\left\{ \begin{array}{lcl} \delta_n & = & p(\mathbf{y}_n|\mathbf{x}_n)\widetilde{\delta}_n \\ \widetilde{\delta}_{n+1} & = & \int p(\mathbf{x}_n|\mathbf{x}_{n+1})\delta_n\, d\mathbf{x}_n \end{array} \right. ; \tag{9}$$

on the other hand, the backward one-step prediction pdf $\gamma_n = p(\mathbf{x}_n|\mathbf{y}_{n+1:N})$ and filtering pdf $\widetilde{\gamma}_{n+1} = p(\mathbf{x}_{n+1}|\mathbf{y}_{n+1:N})$ can be computed from $n = N$ to $n = 0$ (with $\widetilde{\gamma}_N = \frac{p(\mathbf{x}_N,\mathbf{y}_N)}{p(\mathbf{y}_N)}$) as

$$\left\{ \begin{array}{lcl} \gamma_n & = & \int p(\mathbf{x}_n|\mathbf{x}_{n+1})\widetilde{\gamma}_{n+1} d\mathbf{x}_{n+1} \\ \widetilde{\gamma}_n & = & \frac{p(\mathbf{y}_n|\mathbf{x}_n)\gamma_n}{\int p(\mathbf{y}_n|\mathbf{x}_n)\gamma_n d\mathbf{x}_n} \end{array} \right. . \tag{10}$$

B. Backward recursive computation of the smoothing pdf

The aim of this section is to compute the backward conditional transition pdf $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N})$ in (5). From (2)-(4), $\mathbf{y}_{n+1:N}$ and $\mathbf{x}_n$ are independent conditionally on $(\mathbf{x}_{n+1}, \mathbf{y}_{0:n})$, so $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N}) = p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:n})$. Now $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:n})$ can be computed in the forward direction by combining appropriately $(\alpha_n, \widetilde{\alpha}_n)$ or $(\delta_n, \widetilde{\delta}_n)$ with either the forward or backward HMC pdfs, which leads to four different algorithms. Algorithm (11) (resp. (12)) only uses forward (resp. backward) HMC pdfs, and algorithms (13) and (14) use both.

Proposition 3: Assume that (2)-(4) hold, and that we are given the forward and/or the backward HMC pdfs. Then α_n and $\widetilde{\alpha}_n$ (resp. δ_n and $\widetilde{\delta}_n$) can be computed in the forward direction by (7) (resp. (9)), and next $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:n})$ by

$$p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:n}) = \frac{p(\mathbf{x}_{n+1}|\mathbf{x}_n)\widetilde{\alpha}_n}{\int p(\mathbf{x}_{n+1}|\mathbf{x}_n)\widetilde{\alpha}_n d\mathbf{x}_n}, \tag{11}$$

$$= \frac{p(\mathbf{x}_n|\mathbf{x}_{n+1})\delta_n}{\int p(\mathbf{x}_n|\mathbf{x}_{n+1})\delta_n d\mathbf{x}_n}, \tag{12}$$

$$= \frac{\frac{p(\mathbf{x}_n|\mathbf{x}_{n+1})\widetilde{\alpha}_n}{p(\mathbf{x}_n)}}{\int \frac{p(\mathbf{x}_n|\mathbf{x}_{n+1})\widetilde{\alpha}_n}{p(\mathbf{x}_n)} d\mathbf{x}_n}, \tag{13}$$

$$= \frac{p(\mathbf{x}_{n+1}|\mathbf{x}_n)\delta_n p(\mathbf{x}_n)}{\int p(\mathbf{x}_{n+1}|\mathbf{x}_n)\delta_n p(\mathbf{x}_n) d\mathbf{x}_n}. \tag{14}$$

Finally $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ can be computed by (5), initialized by $p(\mathbf{x}_{N+1}|\mathbf{y}_{0:N}) = \alpha_{N+1}$ in case of (11) and (13), or by $\frac{\widetilde{\delta}_{N+1}p(\mathbf{x}_{N+1})}{p(\mathbf{y}_{0:N})}$ with $p(\mathbf{y}_{0:N}) = \int \widetilde{\delta}_{N+1}p(\mathbf{x}_{N+1})d\mathbf{x}_{N+1}$, in case of (12) and (14).

C. Forward recursive computation of the smoothing pdf

This section is parallel to §II-B. Our aim here is to compute $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{0:N})$ in (6). From (2)-(4), $\mathbf{y}_{0:n}$ and $\mathbf{x}_{n+1}$ are independent conditionally on $(\mathbf{x}_n, \mathbf{y}_{n+1:N})$, so $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{0:N}) = p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{n+1:N})$. Now $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{n+1:N})$ can be computed in the backward direction by combining appropriately $(\beta_n, \widetilde{\beta}_n)$ or $(\gamma_n, \widetilde{\gamma}_n)$ with either the forward or backward HMC pdfs, which leads to four different algorithms. The algorithm (15) (resp. (16)) only uses forward (resp. backward) HMC pdfs, and the algorithms (17) and (18) use both.

Proposition 4: Assume that (2)-(4) hold, and that we are given the forward and/or the backward HMC pdfs. Then β_n and $\widetilde{\beta}_n$ (resp. γ_n and $\widetilde{\gamma}_n$) can be computed in the backward direction by (8) (resp. (10)), and next $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{n+1:N})$ by

$$p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{n+1:N}) = \frac{p(\mathbf{x}_{n+1}|\mathbf{x}_n)\beta_{n+1}}{\int p(\mathbf{x}_{n+1}|\mathbf{x}_n)\beta_{n+1} d\mathbf{x}_{n+1}}, \tag{15}$$

$$= \frac{p(\mathbf{x}_n|\mathbf{x}_{n+1})\widetilde{\gamma}_{n+1}}{\int p(\mathbf{x}_n|\mathbf{x}_{n+1})\widetilde{\gamma}_{n+1} d\mathbf{x}_{n+1}}, \tag{16}$$

$$= \frac{\frac{p(\mathbf{x}_{n+1}|\mathbf{x}_n)\widetilde{\gamma}_{n+1}}{p(\mathbf{x}_{n+1})}}{\int \frac{p(\mathbf{x}_{n+1}|\mathbf{x}_n)\widetilde{\gamma}_{n+1}}{p(\mathbf{x}_{n+1})} d\mathbf{x}_{n+1}}, \tag{17}$$

$$= \frac{p(\mathbf{x}_n|\mathbf{x}_{n+1})\beta_{n+1}p(\mathbf{x}_{n+1})}{\int p(\mathbf{x}_n|\mathbf{x}_{n+1})\beta_{n+1}p(\mathbf{x}_{n+1})d\mathbf{x}_{n+1}} \tag{18}$$

Finally $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ can be computed in the forward direction by (6), initialized by $p(\mathbf{x}_0|\mathbf{y}_{0:N}) = \frac{\beta_0 p(\mathbf{x}_0)}{p(\mathbf{y}_{0:N})}$ with $p(\mathbf{y}_{0:N}) = \int \beta_0 p(\mathbf{x}_0)d\mathbf{x}_0$ in case of (15) and (18), or by $\widetilde{\gamma}_0$ in case of (16) and (17).

D. Non recursive computation of the smoothing pdf

Let us finally see that $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ can be computed as a (normalized) product of $(\alpha_n, \widetilde{\alpha}_n)$ (or $(\delta_n, \widetilde{\delta}_n)$) and $(\beta_n, \widetilde{\beta}_n)$ (or $(\gamma_n, \widetilde{\gamma}_n)$), which leads to four algorithms. The algorithm (19) (resp. (20)) implicitely uses forward (resp. backward) HMC pdfs only, and (21) and (22) use both.

Proposition 5: Assume that (2)-(4) hold, and that we are given the forward and/or backward HMC pdfs . Then $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ can be computed as

$$p(\mathbf{x}_n|\mathbf{y}_{0:N}) = \frac{\alpha_n\beta_n}{\int \alpha_n\beta_n d\mathbf{x}_n} = \frac{\widetilde{\alpha}_n\widetilde{\beta}_n}{\int \widetilde{\alpha}_n\widetilde{\beta}_n d\mathbf{x}_n} \tag{19}$$

$$= \frac{\gamma_n\delta_n}{\int \gamma_n\delta_n d\mathbf{x}_n} = \frac{\widetilde{\gamma}_n\widetilde{\delta}_n}{\int \widetilde{\gamma}_n\widetilde{\delta}_n d\mathbf{x}_n}, \tag{20}$$

$$= \frac{\frac{\alpha_n\widetilde{\gamma}_n}{p(\mathbf{x}_n)}}{\int \frac{\alpha_n\widetilde{\gamma}_n}{p(\mathbf{x}_n)} d\mathbf{x}_n} = \frac{\frac{\widetilde{\alpha}_n\gamma_n}{p(\mathbf{x}_n)}}{\int \frac{\widetilde{\alpha}_n\gamma_n}{p(\mathbf{x}_n)} d\mathbf{x}_n}, \tag{21}$$

$$= \frac{\delta_n\widetilde{\beta}_n p(\mathbf{x}_n)}{\int \delta_n\widetilde{\beta}_n p(\mathbf{x}_n)d\mathbf{x}_n} = \frac{\widetilde{\delta}_n\beta_n p(\mathbf{x}_n)}{\int \widetilde{\delta}_n\beta_n p(\mathbf{x}_n)d\mathbf{x}_n}, \tag{22}$$

in which α_n (resp. δ_n) is computed in the forward direction by (7) (resp. (9)), and $(\widetilde{\beta}_n, \beta_n)$ (resp. $(\widetilde{\gamma}_n, \gamma_n)$) is computed in the backward direction by (8) (resp. (10)).

III. THE GAUSSIAN CASE

In this section (2)-(4) still hold, but we now further assume that the state-space model is Gaussian, i.e. that

$$\mathbf{x}_0 \sim \mathcal{N}(\widehat{\mathbf{x}}_0, \mathbf{P}_0), \mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n) \text{ and } \mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n). \tag{23}$$

Then all the pdfs in §II are Gaussian. Let us set

$$p(\mathbf{x}_n) \sim \mathcal{N}(\widehat{\mathbf{x}}_n, \mathbf{P}_n), \tag{24}$$

$$p(\mathbf{x}_n|\mathbf{y}_{i:j}) \sim \mathcal{N}(\widehat{\mathbf{x}}_{n|i:j}, \mathbf{P}_{n|i:j}), \tag{25}$$

$$\mu_{n|i:j} = \mathbf{P}^{-1}_{n|i:j}\widehat{\mathbf{x}}_{n|i:j} \tag{26}$$

for all n, i, j with $0 \leq n \leq N$ and $0 \leq i \leq j \leq N$ ($\mu_{n|i:j}$ and $\mathbf{P}_{n|i:j}^{-1}$ are respectively the information vector and matrix associated with $p(\mathbf{x}_n|\mathbf{y}_{i:j})$).

The general algorithms of Propositions 3 to 5 compute $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ from α_n (or δ_n) and/or γ_n (or β_n). In the Gaussian case, this amounts to computing the parameters of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ from the parameters of α_n (or δ_n) and/or γ_n (or β_n). More precisely, (7) to (22) reduce to equations which compute $\arg\max_{\mathbf{x}_n} p(\mathbf{x}_n|\mathbf{y}_{0:N})$ (i.e., $\widehat{\mathbf{x}}_{n|0:N}$), and the associated covariance matrix, from $\arg\max_{\mathbf{x}_n} \alpha_n = \widehat{\mathbf{x}}_{n|0:n-1}$ (or $\arg\max_{\mathbf{x}_n} \delta_n$) and/or $\arg\max_{\mathbf{x}_n} \gamma_n = \widehat{\mathbf{x}}_{n|n+1:N}$ (or $\arg\max_{\mathbf{x}_n} \beta_n$), as well as the associated covariance matrice(s). In practice, these equations can be derived by systematically applying some simple results for Gaussian variables which are gathered in the Appendix (see Propositions 7 to 11); each one of the twelve general algorithms in Propositions 3, 4 and 5 then reduces to a particular Kalman smoothing algorithm. The main result of this paper is that some of these algorithms already exist, but to our best knowledge some others are original. For want of space, we shall not write down all of them down explicitly (however for illustrative purposes, we give the algorithm (20) in §IV). Let us nevertheless comment on how to get them, and on their originality (the comments in §III-A to III-D below are also summarized in Tables I to V.

A. Recursive algorithms for α_n, β_n, γ_n and δ_n

- Using Proposition 7 ((52)-(53), covariance (resp. information) form) in (7) provides the Kalman filter in covariance [14] (resp. information [15] [11]) form;
- Using Propositions 8 and 9 in (8) provides the backward algorithm in the two-filter smoother obtained by Mayne [6] (see also [11, eqs. (10.4.14)-(10.4.15)]);
- Using Propositions 8 and 9 in (9) provides equations (33)-(37) in section IV. A variant of this algorithm has been already introduced by using complementary models [8, §3.2];
- Using Proposition 7 ((52)-(53), covariance (resp. information) form) in (10) provides the backward Kalman filter in covariance [11, §9.8] (resp. information) form.

B. Backward recursive computation of the smoothing pdf

In the Gaussian case, the backward recursive propagation (5) of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ reduces to the backward recursive propagation of its mean $\widehat{\mathbf{x}}_{n|0:N}$ and covariance $\mathbf{P}_{n|0:N}$. We have $p(\mathbf{x}_{n+1}|\mathbf{y}_{0:N}) \sim \mathcal{N}(\widehat{\mathbf{x}}_{n+1|0:N}, \mathbf{P}_{n+1|0:N})$ and $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N}) \sim \mathcal{N}(\mathbf{A}_{n+1}^{\Psi}\mathbf{x}_{n+1} + \mathbf{b}_{n+1}^{\Psi}, \mathbf{P}_{n+1}^{\Psi})$ for some matrices $\mathbf{A}_{n+1}^{\Psi}$, $\mathbf{b}_{n+1}^{\Psi}$ and $\mathbf{P}_{n+1}^{\Psi}$. From Prop. 7 ((52), covariance form[2]), (5) reduces to :

$$\widehat{\mathbf{x}}_{n|0:N} = \mathbf{A}_{n+1}^{\Psi}\widehat{\mathbf{x}}_{n+1|0:N} + \mathbf{b}_{n+1}^{\Psi}, \quad (27)$$

$$\mathbf{P}_{n|0:N} = \mathbf{A}_{n+1}^{\Psi}\mathbf{P}_{n+1|0:N}(\mathbf{A}_{n+1}^{\Psi})^T + \mathbf{P}_{n+1}^{\Psi}. \quad (28)$$

It remains to compute $\mathbf{A}_{n+1}^{\Psi}$, $\mathbf{b}_{n+1}^{\Psi}$ and $\mathbf{P}_{n+1}^{\Psi}$. Equations (27) and (28) can be written under five different forms, depending on the way $p(\mathbf{x}_n|\mathbf{x}_{n+1}, \mathbf{y}_{0:N})$ (or equivalently $\mathbf{A}_{n+1}^{\Psi}$, $\mathbf{b}_{n+1}^{\Psi}$ and $\mathbf{P}_{n+1}^{\Psi}$) is computed (i.e., via equation (11), (12), (13) or (14)). More precisely:

- Using Prop. 7 ((53), covariance (resp. information) form) in (11) provides the RTS algorithm[3] which was derived by using the Maximum a posteriori approach [3];
- Using Proposition 7 ((53), information form) in (12) we get an algorithm similar to that obtained in [8, p. 40] by using complementary models;
- Using Proposition 10 (resp. Prop. 11) in (13) (resp. (14)) we get two other backward smoothing algorithms. To our best knowledge, these algorithms are original.

[2]one could also use ((52), information form) in order to compute the information parameters $\mu_{n|0:N}$ and $\mathbf{P}_{n|0:N}^{-1}$ of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$, which results to variations of the algorithm (the same remark also holds later for the forward computation of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$).

[3]As is well known, the Bryson and Frazier algorithm [1] is closely related to the RTS algorithm and can actually be derived from it. However it cannot be derived from Bayesian considerations.

C. Forward recursive computation of the smoothing pdf

As in §III-B, the forward recursive propagation (6) of $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ reduces to the forward recursive propagation of its parameters, i.e. to the equations

$$\widehat{\mathbf{x}}_{n+1|0:N} = \mathbf{A}_n^{\Phi}\widehat{\mathbf{x}}_{n|0:N} + \mathbf{b}_n^{\Phi}, \quad (29)$$

$$\mathbf{P}_{n+1|0:N} = \mathbf{A}_n^{\Phi}\mathbf{P}_{n|0:N}(\mathbf{A}_n^{\Phi})^T + \mathbf{P}_n^{\Phi} \quad (30)$$

for some matrices $\mathbf{b}_n^{\Phi}$, $\mathbf{A}_n^{\Phi}$ and $\mathbf{P}_n^{\Phi}$. Now, eqs. (29) and (30) can be written under five different forms, depending on how $p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{y}_{0:N})$ is computed (i.e., via (15), (16), (17) or (18)) :

- Using Prop. 7 ((53), information form) in (15) we get an algorithm which is similar to that introduced in [16] [8, p. 35] by using complementary models;
- Using Proposition 7 ((53), covariance or information form) in (16) leads to two algorithms which were partially derived in [11, pp. 401, Exs. 10.12 & 10.14];
- To our best knowledge, using Proposition 10 in (17) and Proposition 11 in (18) leads to two original algorithms.

D. Non recursive computation of the smoothing pdf

- In the discrete case, (19) is nothing but the Forward-Backward or BCJR algorithm [17] [18]. In the Gaussian case, using Proposition 7 ((53), information form) in (19) provides the two-filter algorithm by Mayne [6] (see also [7]);
- (20) reduces to an algorithm (given explicitly in Prop. 6, see §IV) which to our best knowledge is original;
- Using Proposition 10 in (21) provides the General two-filter algorithm (already derived from the innovations approach [11, Theorem 10.4.1]);
- Finally, using Proposition 11 in (22) provides an algorithm which is similar to that obtained in [8, §3.3] by using complementary models.

E. Comments and remarks

- The algorithms of §III-A to III-D hold under simple conditions which however vary from one algorithm to another. Let us begin with §III-A. For all four algorithms we assume that $\mathbf{R}_n$ is positive definite ($> \mathbf{0}$). Furthermore the information Kalman filter (resp. the Gaussian forms of (9) and (10), in which we first compute the parameters of the backward model) can only be derived if $\mathbf{P}_{n+1|0:n} > \mathbf{0}$ (resp. $\mathbf{P}_n > \mathbf{0}$) for all n. Both conditions, in turn, hold if $\mathbf{P}_0 > \mathbf{0}$ and $\mathbf{F}_n$ is invertible for all n, or if $\mathbf{P}_0 > \mathbf{0}$ and $\overline{\mathbf{Q}}_n = \mathbf{G}_n\mathbf{Q}_n\mathbf{G}_n^T > \mathbf{0}$ for all n. Finally the Gaussian form of (8) requires that $\overline{\mathbf{Q}}_n > \mathbf{0}$.
  Though the smoothing algorithms of sections III-B to III-D rely on the algorithms of section III-A some further restrictions may apply. For instance the information form of algorithm (11) relies on the information Kalman filter, but its derivation from Proposition 7 holds if $\overline{\mathbf{Q}}_n > \mathbf{0}$. The results are summarized in Tables I to V.[4].

[4]these results apply to the algorithms directly obtained from Propositions 7 to 11. Other variants (related via matrix inversion lemmas, if they can be applied) may also hold, possibly under different sufficient conditions.

- Comparing the computational cost of one algorithm w.r.t. another depends on the position of $n_{\mathbf{x}}$ vs. $n_{\mathbf{u}}$ and $n_{\mathbf{y}}$. However if both forms are available, the computational cost of an information form algorithm always exceeds that of the covariance version. On the other hand, the forward and backward models are theoretically equivalent, but if $p(\mathbf{x}_n|\mathbf{x}_{n+1})$ is not available computing its parameters from (1) and (23) costs $\mathcal{O}(4n_{\mathbf{x}}^3 + \frac{n_{\mathbf{x}}^2 n_{\mathbf{u}}}{2} + n_{\mathbf{x}} n_{\mathbf{u}}^2)$ elementary operations (see §IV-A below); this point explains, for instance, the difference between the computational cost of the information form of (9) (resp. (10)) w.r.t. that of (7) (resp. (8)).
- As is well known (see e.g. [12]) the maximum a posteriori and maximum likelihood estimators coincide in the so-called non informative case, which corresponds to an (improper) flat prior distribution. This can be seen from the information form of (53) (see Proposition 7), which relates the information parameters associated with the a posteriori pdf $p(\mathbf{x}|\mathbf{y})$ to those of the prior $p(\mathbf{x})$ and of the likelihood $p(\mathbf{y}|\mathbf{x})$. As a consequence if $\mathbf{P}_0^{-1}$ is set equal to $\mathbf{0}$ then some of the Gaussian algorithms (in information form) above coincide. More precisely, (10) (resp. (7)) reduces to (8) (resp. (9)); and consequently the backward algorithms (11), (12) (13) and (14) coincide, the forward algorithms (15), (16), (17) and (18) coincide, and the non recursive algorithms (19), (20), (21) and (22) coincide.

IV. A WORKED EXAMPLE

For illustrative purposes we address the actual computation of (20). To compute $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ via (20) we need to propagate γ_n in the backward direction (via 10) and δ_n in the forward direction (via 9). Both algorithms use the backward HMC parameters, so we first have to compute $p(\mathbf{x}_n|\mathbf{x}_{n+1})$. Let us address these different points.

A. Computing the backward HMC transition pdfs

From (1) and (23) we have $p(\mathbf{x}_{n+1}|\mathbf{x}_n) \sim \mathcal{N}(\mathbf{F}_n\mathbf{x}_n, \mathbf{G}_n\mathbf{Q}_n\mathbf{G}_n^T)$ and $p(\mathbf{x}_n) \sim \mathcal{N}(\widehat{\mathbf{x}}_n, \mathbf{P}_n)$, with $\widehat{\mathbf{x}}_{n+1} = \mathbf{F}_n\widehat{\mathbf{x}}_n$ and $\mathbf{P}_{n+1} = \mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T + \mathbf{G}_n\mathbf{Q}_n\mathbf{G}_n^T$ for all $n \geq 0$. Using Proposition 7 ((53), covariance form), the backward Markov transition pdf $p(\mathbf{x}_n|\mathbf{x}_{n+1})$ is Gaussian with :

$$\begin{aligned} p(\mathbf{x}_n|\mathbf{x}_{n+1}) &\sim \mathcal{N}(\widetilde{\mathbf{F}}_{n+1}\mathbf{x}_{n+1} + \mathbf{c}_{n+1}, \widetilde{\mathbf{Q}}_{n+1}), \quad (31)\\ \widetilde{\mathbf{F}}_{n+1} &= \mathbf{P}_n\mathbf{F}_n^T\mathbf{P}_{n+1}^{-1},\\ \mathbf{c}_{n+1} &= [\mathbf{I}_{n_{\mathbf{x}}} - \mathbf{P}_n\mathbf{F}_n^T\mathbf{P}_{n+1}^{-1}\mathbf{F}_n]\widehat{\mathbf{x}}_n,\\ \widetilde{\mathbf{Q}}_{n+1} &= \mathbf{P}_n - \mathbf{P}_n\mathbf{F}_n^T\mathbf{P}_{n+1}^{-1}\mathbf{F}_n\mathbf{P}_n, \end{aligned}$$

in which $\mathbf{I}_{n_{\mathbf{x}}}$ is the $n_{\mathbf{x}} \times n_{\mathbf{x}}$ identity matrix. On the other hand,

$$p(\mathbf{y}_{n+1}|\mathbf{x}_{n+1}) \sim \mathcal{N}(\mathbf{H}_{n+1}\mathbf{x}_{n+1}, \mathbf{R}_{n+1}). \quad (32)$$

B. Propagating δ_n in the forward direction

The forward computation of δ_n (resp $\tilde{\delta}_n$) reduces to a forward maximum likelihood algorithm which consists in propagating the information vector ν_n^{δ} (resp. $\nu_n^{\tilde{\delta}}$) and matrix Γ_n^{δ} (resp. $\Gamma_n^{\tilde{\delta}}$) of δ_n (resp. $\tilde{\delta}_n$), see Prop. 7, eqs. (50) and (51). From Prop. 9 (resp. Prop. 8) as well as (50) and (51), the first (resp. second) equation of (9) reduces to (33)-(34) (resp. (35)-(37)) (initialized with $(\nu_0^{\tilde{\delta}}\ ,\ \Gamma_0^{\tilde{\delta}}) = (\mathbf{0}\ ,\ \mathbf{0})$) :

$$\begin{aligned} \Gamma_n^{\delta} &= \Gamma_n^{\tilde{\delta}} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n, & (33)\\ \nu_n^{\delta} &= \nu_n^{\tilde{\delta}} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{y}_{n+1}, & (34)\\ \mathbf{L}_{n+1} &= [\Gamma_n^{\delta} + \widetilde{\mathbf{Q}}_{n+1}^{-1}]^{-1}, & (35)\\ \Gamma_{n+1}^{\tilde{\delta}} &= \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}[\widetilde{\mathbf{Q}}_{n+1} - \mathbf{L}_{n+1}]\widetilde{\mathbf{Q}}_{n+1}^{-1}\widetilde{\mathbf{F}}_{n+1}, & (36)\\ \nu_{n+1}^{\tilde{\delta}} &= \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}\mathbf{L}_{n+1}(\nu_n^{\delta} - \Gamma_n^{\delta}\mathbf{c}_{n+1}). & (37) \end{aligned}$$

C. Propagating γ_n in the backward direction

The backward computation of γ_n reduces to the backward propagation of $\widehat{\mathbf{x}}_{n|n+1:N}$ and $\mathbf{P}_{n|n+1:N}$ (in covariance form), or of $\mu_{n|n+1:N}$ and $\mathbf{P}_{n|n+1:N}^{-1}$ (in information form). Let us write the information algorithm. Let $\nu_{n+1}^{\mathbf{x}|\mathbf{x}}$ (resp. $\nu_{n+1}^{\mathbf{y}|\mathbf{x}}$) be the information vector, and $\Gamma_{n+1}^{\mathbf{x}|\mathbf{x}}$ (resp. $\Gamma_{n+1}^{\mathbf{y}|\mathbf{x}}$) be the information matrix of the likelihood $p(\mathbf{x}_n|\mathbf{x}_{n+1})$ (resp. $p(\mathbf{y}_{n+1}|\mathbf{x}_{n+1})$). From (31), (32) and (50)-(51) (see Prop. 7) we get

$$\begin{aligned} \Gamma_{n+1}^{\mathbf{x}|\mathbf{x}} &= \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}\widetilde{\mathbf{F}}_{n+1}, & (38)\\ \nu_{n+1}^{\mathbf{x}|\mathbf{x}} &= \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}(\mathbf{x}_n - \mathbf{c}_{n+1}), & (39)\\ \Gamma_{n+1}^{\mathbf{y}|\mathbf{x}} &= \mathbf{H}_{n+1}^T\mathbf{R}_{n+1}^{-1}\mathbf{H}_{n+1}, & (40)\\ \nu_{n+1}^{\mathbf{y}|\mathbf{x}} &= \mathbf{H}_{n+1}^T\mathbf{R}_{n+1}^{-1}\mathbf{y}_{n+1}. & (41) \end{aligned}$$

Now, by applying Proposition 7 ((52), information form) and using (38), the first equation of (10) reduces to (42)-(44); by applying Proposition 7 ((53), information form) to the second equation of (10) and by using (40) and (41), we obtain (45)-(46) :

$$\begin{aligned} \mathcal{L}_n &= [\mathbf{P}_{n+1|n+1:N}^{-1} + \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}\widetilde{\mathbf{F}}_{n+1}]^{-1}, & (42)\\ \mu_{n|n+1:N} &= \widetilde{\mathbf{Q}}_{n+1}^{-1}[\widetilde{\mathbf{F}}_{n+1}\mathcal{L}_n(\mu_{n+1|n+1:N} - \widetilde{\mathbf{F}}_{n+1}^T\widetilde{\mathbf{Q}}_{n+1}^{-1}\mathbf{c}_{n+1})\\ &\quad +\mathbf{c}_{n+1}], & (43)\\ \mathbf{P}_{n|n+1:N}^{-1} &= \widetilde{\mathbf{Q}}_{n+1}^{-1}[\widetilde{\mathbf{Q}}_{n+1} - \widetilde{\mathbf{F}}_{n+1}\mathcal{L}_n\widetilde{\mathbf{F}}_{n+1}^T]\widetilde{\mathbf{Q}}_{n+1}^{-1}, & (44)\\ \mu_{n|n:N} &= \mu_{n|n+1:N} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{y}_n, & (45)\\ \mathbf{P}_{n|n:N}^{-1} &= \mathbf{P}_{n|n+1:N}^{-1} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n. & (46) \end{aligned}$$

D. Computing $p(\mathbf{x}_n|\mathbf{y}_{0:N})$ via (20)

It remains to apply Prop. 7 ((53), information form). Finally (9), (10) and (20) reduce to the following algorithm :

Proposition 6: Let (1) and (23) hold. Then ν_n^{δ} and Γ_n^{δ} (resp. $\mu_{n|n+1:N}$ and $\mathbf{P}_{n|n+1:N}^{-1}$) can be computed in the forward (resp. backward) direction by (33)-(37) (resp. (42)-(46)), and finally $\mu_{n|0:N}$ and $\mathbf{P}_{n|0:N}^{-1}$ can be computed by

$$\begin{aligned} \mu_{n|0:N} &= \mu_{n|n+1:N} + \nu_n^{\delta}, & (47)\\ \mathbf{P}_{n|0:N}^{-1} &= \mathbf{P}_{n|n+1:N}^{-1} + \Gamma_n^{\delta}. & (48) \end{aligned}$$

V. CONCLUSION

Our aim was twofold. We first unified some existing Kalman like fixed-interval smoothing algorithms by deriving them from a common Bayesian point of view. The first step in our derivations consisted in appropriately exploiting Markovian properties of the state-space model for proposing algorithms for general continuous state HMC. They all involve one (or two) out of four canonical pairs of pdfs $(\alpha_n,\tilde{\alpha}_n)$, $(\beta_n,\tilde{\beta}_n)$, $(\gamma_n,\tilde{\gamma}_n)$ and $(\delta_n,\tilde{\delta}_n)$. The second one consisted in obtaining the actual Kalman like algorithms by further injecting the Gaussian assumption; in order to facilitate the derivations we developed a specific toolbox of generic properties of Gaussian variables which were used recurrently and systematically in the derivations. Moreover, the methodology we introduced enabled us to fill some gaps by completing the set of existing solutions by five new Kalman like smoothing algorithms.

VI. APPENDIX

The algorithms of §III are directly obtained from Prop. 7 to 11. Prop. 8 to 11 can be derived from Prop. 7 which is the basic result; detailed proofs are omitted due to lack of space.

Proposition 7: Let $p(\mathbf{x}) \sim \mathcal{N}(\widehat{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$ and $p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{P}_{|\mathbf{x}})$. Then the following holds :

- *maximum likelihood :* Let $\mathcal{X}(\mathbf{y}) \stackrel{\text{def}}{=} \underset{\mathbf{x}}{\operatorname{argmax}}\ p(\mathbf{y}|\mathbf{x})$. Then $\mathcal{X}$ satisfies
$$(\mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}\mathbf{A})\mathcal{X} = \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}(\mathbf{y}-\mathbf{b}). \quad (49)$$
By analogy with (26)[5] we thus define the information matrix $\Gamma_{\mathbf{x}}$ and information vector $\nu_{\mathbf{x}} = \Gamma_{\mathbf{x}}\mathcal{X}$ associated with $p(\mathbf{y}|\mathbf{x})$ as :
$$\nu_{\mathbf{x}} = \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}(\mathbf{y}-\mathbf{b}), \quad (50)$$
$$\Gamma_{\mathbf{x}} = \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}\mathbf{A}; \quad (51)$$
- *joint pdf parameters :*
$$p(\mathbf{x},\mathbf{y}) \sim \mathcal{N}(\begin{bmatrix} \widehat{\mathbf{x}} \\ \widehat{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{\mathbf{x}} & \mathbf{P}_{\mathbf{x}}\mathbf{A}^T \\ \mathbf{A}\mathbf{P}_{\mathbf{x}} & \mathbf{P}_{\mathbf{y}} \end{bmatrix}); \quad (52)$$
Covariance form :
$$\widehat{\mathbf{y}} = \mathbf{A}\widehat{\mathbf{x}} + \mathbf{b}, \mathbf{P}_{\mathbf{y}} = \mathbf{P}_{|\mathbf{x}} + \mathbf{A}\mathbf{P}_{\mathbf{x}}\mathbf{A}^T;$$
Information form :
$$\mathbf{P}_{\mathbf{y}}^{-1}\widehat{\mathbf{y}} = \mathbf{P}_{|\mathbf{x}}^{-1}[\mathbf{A}(\mathbf{P}_{\mathbf{x}}^{-1}+\Gamma_{\mathbf{x}})^{-1}(\mathbf{P}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}} - \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}\mathbf{b}) + \mathbf{b}]$$
$$\mathbf{P}_{\mathbf{y}}^{-1} = \mathbf{P}_{|\mathbf{x}}^{-1}[\mathbf{P}_{|\mathbf{x}} - \mathbf{A}(\mathbf{P}_{\mathbf{x}}^{-1}+\Gamma_{\mathbf{x}})^{-1}\mathbf{A}^T]\mathbf{P}_{|\mathbf{x}}^{-1};$$
- *a posteriori pdf parameters :*
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \sim \mathcal{N}(\widehat{\mathbf{x}}_{|\mathbf{y}}, \mathbf{P}_{|\mathbf{y}}); \quad (53)$$
- *Covariance form :* $\mathbf{K} = \mathbf{P}_{\mathbf{x}}\mathbf{A}^T\mathbf{P}_{\mathbf{y}}^{-1}$, $\widehat{\mathbf{x}}_{|\mathbf{y}} = \widehat{\mathbf{x}} + \mathbf{K}(\mathbf{y}-\widehat{\mathbf{y}})$, $\mathbf{P}_{|\mathbf{y}} = \mathbf{P}_{\mathbf{x}} - \mathbf{K}\mathbf{P}_{\mathbf{y}}\mathbf{K}^T$;
- *Information form :* $\mathbf{P}_{|\mathbf{y}}^{-1}\widehat{\mathbf{x}}_{|\mathbf{y}} = \mathbf{P}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}} + \nu_{\mathbf{x}}$, $\mathbf{P}_{|\mathbf{y}}^{-1} = \mathbf{P}_{\mathbf{x}}^{-1} + \Gamma_{\mathbf{x}}$.

Proposition 8: Let $(\mathbf{x},\mathbf{y},\mathbf{z})$ be Gaussian in which conditionally on $\mathbf{z}$, $\mathbf{y}$ and $\mathbf{x}$ are independent. Let $p(\mathbf{z}|\mathbf{x}) \sim \mathcal{N}(\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{P}_{|\mathbf{x}})$ and let $\nu_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ (resp. $\nu_{\mathbf{z}}$ and $\Gamma_{\mathbf{z}}$) be the information parameters of $p(\mathbf{y}|\mathbf{x})$ (resp. $p(\mathbf{y}|\mathbf{z})$) (see Prop. 7). Then
$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{z}|\mathbf{x})p(\mathbf{y}|\mathbf{z})d\mathbf{z};$$
$$\nu_{\mathbf{x}} = \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}[\mathbf{P}_{|\mathbf{x}}^{-1}+\Gamma_{\mathbf{z}}]^{-1}(\nu_{\mathbf{z}} - \Gamma_{\mathbf{z}}\mathbf{b}),$$
$$\Gamma_{\mathbf{x}} = \mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}[\mathbf{P}_{|\mathbf{x}} - [\mathbf{P}_{|\mathbf{x}}^{-1}+\Gamma_{\mathbf{z}}]^{-1}]\mathbf{P}_{|\mathbf{x}}^{-1}\mathbf{A}.$$

Proposition 9: Let $(\mathbf{x}, \underbrace{(\mathbf{y}_1,\mathbf{y}_2)}_{\mathbf{y}})$ be Gaussian. Then the information parameters (see Prop. 7) $(\nu_{\mathbf{x}},\Gamma_{\mathbf{x}})$ (resp. $(\nu_{\mathbf{x}_1},\Gamma_{\mathbf{x}_1})$, $(\nu_{\widetilde{\mathbf{x}}_2},\Gamma_{\widetilde{\mathbf{x}}_2})$) of $p(\mathbf{y}|\mathbf{x})$ (resp. $p(\mathbf{y}_1|\mathbf{x})$, $p(\mathbf{y}_2|\mathbf{x},\mathbf{y}_1)$) in factorization $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}_1|\mathbf{x})p(\mathbf{y}_2|\mathbf{x},\mathbf{y}_1)$ are related as
$$\nu_{\mathbf{x}} = \nu_{\mathbf{x}_1} + \nu_{\widetilde{\mathbf{x}}_2},$$
$$\Gamma_{\mathbf{x}} = \Gamma_{\mathbf{x}_1} + \Gamma_{\widetilde{\mathbf{x}}_2}.$$

Proposition 10: Let $(\mathbf{x}, \underbrace{(\mathbf{y}_1,\mathbf{y}_2)}_{\mathbf{y}})$ be Gaussian in which conditionally on $\mathbf{x}$, $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent. Let $p(\mathbf{x}) \sim \mathcal{N}(\widehat{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$ and $p(\mathbf{x}|\mathbf{y}_i) \sim \mathcal{N}(\widehat{\mathbf{x}}_{|\mathbf{y}_i}, \mathbf{P}_{|\mathbf{y}_i})$. Then
$$p(\mathbf{x}|\mathbf{y}) = \frac{\frac{p(\mathbf{x}|\mathbf{y}_1)p(\mathbf{x}|\mathbf{y}_2)}{p(\mathbf{x})}}{\int \frac{p(\mathbf{x}|\mathbf{y}_1)p(\mathbf{x}|\mathbf{y}_2)}{p(\mathbf{x})}d\mathbf{x}} \sim \mathcal{N}(\widehat{\mathbf{x}}_{|\mathbf{y}}, \mathbf{P}_{|\mathbf{y}});$$
$$\mathbf{P}_{|\mathbf{y}}^{-1}\widehat{\mathbf{x}}_{|\mathbf{y}} = \mathbf{P}_{|\mathbf{y}_1}^{-1}\widehat{\mathbf{x}}_{|\mathbf{y}_1} + \mathbf{P}_{|\mathbf{y}_2}^{-1}\widehat{\mathbf{x}}_{|\mathbf{y}_2} - \mathbf{P}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}},$$
$$\mathbf{P}_{|\mathbf{y}}^{-1} = \mathbf{P}_{|\mathbf{y}_1}^{-1} + \mathbf{P}_{|\mathbf{y}_2}^{-1} - \mathbf{P}_{\mathbf{x}}^{-1}.$$

[5]This analogy results from the fact that the covariance matrix $\mathbf{P}_{\mathcal{X}} = E((\mathcal{X}(\mathbf{y})-\mathbf{x})(\mathcal{X}(\mathbf{y})-\mathbf{x})^T)$, if invertible, is equal to $(\mathbf{A}^T\mathbf{P}_{|\mathbf{x}}^{-1}\mathbf{A})^{-1}$. (Note however that the algorithms in this paper do not require $\mathbf{P}_{\mathcal{X}}$ to be invertible.)

Proposition 11: Let $(\mathbf{x}, \underbrace{(\mathbf{y}_1,\mathbf{y}_2)}_{\mathbf{y}})$ be Gaussian in which conditionally on $\mathbf{x}$, $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent. Let $p(\mathbf{x}) \sim \mathcal{N}(\widehat{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$ and let $\nu_{\mathbf{x}_i}$ and $\Gamma_{\mathbf{x}_i}$ be the information parameters of $p(\mathbf{y}_i|\mathbf{x})$. Then
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}_1|\mathbf{x})p(\mathbf{y}_2|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}_1|\mathbf{x})p(\mathbf{y}_2|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \sim \mathcal{N}(\widehat{\mathbf{x}}_{|\mathbf{y}}, \mathbf{P}_{|\mathbf{y}});$$
$$\mathbf{P}_{|\mathbf{y}}^{-1}\widehat{\mathbf{x}}_{|\mathbf{y}} = \nu_{\mathbf{x}_1} + \nu_{\mathbf{x}_2} + \mathbf{P}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}},$$
$$\mathbf{P}_{|\mathbf{y}}^{-1} = \Gamma_{\mathbf{x}_1} + \Gamma_{\mathbf{x}_2} + \mathbf{P}_{\mathbf{x}}^{-1}.$$

	Forward computation of		
	α_n : (7) + Prop. 7		δ_n : (9) + Props. 8, 9
	Cov. form	Info. form	
Fwd HMC	KF [14] $\mathbf{R}_n > \mathbf{0}$	KF Info. [15] $\mathbf{R}_n > \mathbf{0}$, $\mathbf{P}_{n+1\|0:n} > \mathbf{0}$	×
Bwd HMC	×	×	(33)-(37) [8, §3.2] $\mathbf{R}_n$ and $\mathbf{P}_n > \mathbf{0}$

TABLE I

RECURSIVE ALGORITHMS FOR α_n AND δ_n : THE GAUSSIAN CASE

	Backward computation of		
	β_n : (8) + Props. 8, 9	γ_n : (10) + Prop.7	
		Cov. form	Info. form
Fwd HMC	[11, Th. 10.4.3] $\mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	×	×
Bwd HMC	×	Bwd KF [11, §9.8] $\mathbf{R}_n$ and $\mathbf{P}_n > \mathbf{0}$	(42)-(46) $\mathbf{R}_n$ and $\mathbf{P}_n > \mathbf{0}$

TABLE II

RECURSIVE ALGORITHMS FOR β_n AND δ_n : THE GAUSSIAN CASE

	Backward computation (27)-(28)	
	α_n	δ_n
Fwd HMC	(11) + Prop. 7 RTS [3] $\mathbf{P}_0, \mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	(14) + Prop. 11 original $\mathbf{P}_0, \mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$
Bwd HMC	(13) + Prop. 10 original $\mathbf{R}_n, \mathbf{P}_n$ and $\mathbf{P}_{n+1\|0:n} > \mathbf{0}$	(12) + Prop. 7 [8, p. 40] $\mathbf{R}_n$ and $\mathbf{P}_n > \mathbf{0}$

TABLE III

BACKWARD COMPUTATION OF $p(\mathbf{x}_n|\mathbf{y}_{0:N})$: THE GAUSSIAN CASE

	Forward computation (29)-(30)	
	β_n	γ_n
Fwd HMC	(15) + Prop. 7 [16] $\mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	(17) + Prop. 10 original $\mathbf{P}_0$, $\mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$
Bwd HMC	(18) + Prop. 11 original $\mathbf{P}_0$, $\mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	(16) + Prop. 7 [11, p. 401] $\mathbf{P}_0$, $\mathbf{R}_n$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$

TABLE IV

FORWARD COMPUTATION OF $p(\mathbf{x}_n|\mathbf{y}_{0:N})$: THE GAUSSIAN CASE

REFERENCES

[1] A.E. Bryson and M. Frazier, "Smoothing for linear and non-linear dynamic systems," *TDR 63-119, Aeronautical Systems Division, Wrigth-Patterson Air Force Base*, pp. 353–364, February 1963.

	β_n	γ_n
α_n	(19) + Prop. 7 Two-filter [6] $\mathbf{R}_n$, $\mathbf{P}_0$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	(21) + Prop. 10 General Two-filter [11, Thm.10.4.1] $\mathbf{R}_n$, $\mathbf{P}_n$ and $\mathbf{P}_{n+1\|0:n} > \mathbf{0}$
δ_n	(22) + Prop. 11 [8, §3.3] $\mathbf{R}_n$, $\mathbf{P}_0$ and $\overline{\mathbf{Q}}_n > \mathbf{0}$	(20) + Prop. 7: (47)-(48) original (see §IV) $\mathbf{R}_n$ and $\mathbf{P}_n > \mathbf{0}$

TABLE V

NON-RECURSIVE COMPUTATION OF $p(\mathbf{x}_n|\mathbf{y}_{0:N})$: THE GAUSSIAN CASE

[2] H.E. Rauch, "Solutions to smoothing problem," *IEEE Transactions on Automatic Control*, vol. 8, pp. 371–372, October 1963.

[3] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, August 1965.

[4] J. S. Meditch, "Orthogonal projection and discrete optimal linear smoothing," *SIAM Journal on Control*, vol. 5, pp. 74–89, January 1967.

[5] T. Kailath and P. Frost, "An innovations approach to least squares estimation, part II: Linear smoothing in additive white noise," *IEEE Transactions on Automatic Control*, vol. 13, pp. 655–660, December 1968.

[6] D. Q. Mayne, "A solution to the smoothing problem for linear dynamical systems," *Automatica*, vol. 4, pp. 73–92, 1966.

[7] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Transactions on Automatic Control*, vol. 7, pp. 387–90, Aug. 1969.

[8] H. L. Weinert, *Fixed interval smoothing for state-space models*, The Kluwer international series in engineering and computer science. Kluwer Academic Publishers, Boston, 2001.

[9] R. C. K. Lee, "Optimal estimation: Identification and control," 1964, pp. 69–73, Cambridge, MA, M.I.T. Press.

[10] M. Askar and H. Derin, "A recursive algorithm for the Bayes solution of the smoothing problem," *IEEE Transactions on Automatic Control*, vol. 26, pp. 558–560, 1981.

[11] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, Prentice Hall Information and System Sciences Series. Prentice Hall, Upper Saddle River, NJ, 2000.

[12] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer-Verlag, 2005.

[13] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, pp. 333–339, October 1964.

[14] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng., Trans. ASME, Series D*, vol. 82, no. 1, pp. 35–45, 1960.

[15] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[16] U. Desai, H. Weinert, and G. Yusypchuk, "Discrete-time complementary models and smoothing algorithms : The correlated noise case," *IEEE Transactions on Automatic Control*, vol. 28, no. 4, pp. 536–539, April 1983.

[17] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–63, 1966.

[18] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–87, March 1974.