

Particle rejuvenation of Rao-Blackwellized Sequential Monte Carlo smoothers for Conditionally Linear and Gaussian models

Ngoc Minh Nguyen*

Sylvain Le Corff†

Eric Moulines‡

Abstract

This paper focuses on Sequential Monte Carlo approximations of smoothing distributions in conditionally linear and Gaussian state spaces. To reduce Monte Carlo variance of smoothers, it is typical in these models to use Rao-Blackwellization: particle approximation is used to sample sequences of hidden regimes while the Gaussian states are explicitly integrated conditional on the sequence of regimes and observations, using variants of the Kalman filter / smoother. The first successful attempt to use Rao-Blackwellization for smoothing extends the Bryson-Frazier smoother for Gaussian linear state space models using the generalized two-filter formula together with Kalman filters / smoothers. More recently, a forward backward decomposition of smoothing distributions mimicking the Rauch-Tung-Striebel smoother for the regimes combined with backward Kalman updates has been introduced. This paper investigates the benefit of introducing additional rejuvenation steps in all these algorithms to sample at each time instant new regimes conditional on the forward and backward particles. This defines particle based approximations of the smoothing distributions whose support is not restricted to the set of particles sampled in the forward or backward filter. These procedures are applied to commodity markets which are described using a two factor model based on the spot price and a convenience yield for crude oil data.

1 Introduction

State space models are bivariate stochastic processes $\{(Y_i, Z_i)\}_{i \geq 1}$ where the state sequence $(Z_i)_{i \geq 1}$ is a Markov chain which is only partially observed through the sequence $(Y_i)_{i \geq 1}$. Conditionally on the state sequence $(Z_i)_{i \geq 1}$ the observations are independent and for all $\ell \geq 1$ the conditional distribution of Y_ℓ given $(Z_i)_{i \geq 1}$ depends on Z_ℓ only. These models are used in a large variety of disciplines such as financial econometrics, biology, signal processing, see [DM13] and the references therein. In general state space models, bayesian filtering and smoothing problems, i.e. the computation of the posterior distributions of a sequence of states (Z_i, \dots, Z_p) for $1 \leq i \leq p \leq \ell$ given observations (Y_1, \dots, Y_ℓ) , are challenging tasks. Filtering refers to the estimation of the distributions of the hidden state Z_i given the observations (Y_1, \dots, Y_i) up to time i , while fixed-interval smoothing stands for the estimation of the distribution of sequence of states (Z_i, \dots, Z_p) given observations (Y_1, \dots, Y_ℓ) with $1 \leq i \leq p < \ell$.

*LTCL, CNRS and Télécom ParisTech, 46 rue Barrault 75634 Paris Cedex 13, France.

†Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

‡Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.

When the state and observation models are linear and Gaussian, filtering can be solved explicitly using the Kalman filter [Kal60]. Exact solutions of the fixed-horizon smoothing problem can be obtained using either the Rauch-Tung-Striebel smoother [RST65] or the Bryson-Frazier two-filter smoother [BF63]. This paper focuses on Conditionally Linear and Gaussian Models (CLGM) given for $i \geq 2$ by:

$$Z_i = d_{a_i} + T_{a_i} Z_{i-1} + H_{a_i} \varepsilon_i, \quad (1)$$

where:

- $(\varepsilon_i)_{i \geq 2}$ is a sequence of independent and identically distributed (i.i.d.) m -dimensional Gaussian vectors with zero mean and identity covariance.
- $(a_i)_{i \geq 1}$ is a homogeneous Markov chain taking values in a finite space $\{1, \dots, J\}$, called *regimes*, with initial distribution π and transition matrix Q .
- $(H_j)_{1 \leq j \leq J}$ are $m \times m$ positive-definite matrices, $(d_j)_{1 \leq j \leq J}$ m -dimensional vectors and $(T_j)_{1 \leq j \leq J}$ $m \times m$ positive-definite matrices.
- Z_1 is a m -dimensional Gaussian random variable with mean μ_1 and variance Σ_1 independent of $(\varepsilon_i)_{i \geq 2}$.

Let n be the number of observations. At each time step $1 \leq i \leq n$, the observation Y_i is given by:

$$Y_i = c_{a_i} + B_{a_i} Z_i + G_{a_i} \eta_i, \quad (2)$$

where:

- $(\eta_i)_{i \geq 1}$ is a i.i.d. sequence of p -dimensional Gaussian vectors, independent of $(\varepsilon_i)_{i \geq 2}$ and Z_1 .
- $(G_j)_{1 \leq j \leq J}$ are $p \times p$ positive-definite matrices, $(c_j)_{1 \leq j \leq J}$ p -dimensional vectors and $(B_j)_{1 \leq j \leq J}$ $p \times m$ matrices.

CLGM play an important role in many applications; see [Sar13] and the references therein for an up-to-date account. A crucial feature of these models is that, conditional on the regime sequence (a_1, \dots, a_n) , both the state equation and the observation equation are linear and Gaussian, which implies that conditional on the sequence of regimes and on the observations, the filtering and the smoothing distributions of the continuous states (Z_1, \dots, Z_n) can be computed explicitly.

To exploit this specific structure, it has been suggested in the pioneering works of [CL00, DGA00] to solve the filtering problem by combining Sequential Monte Carlo (SMC) methods to sample the regimes with the Kalman filter to compute the conditional distribution of the states sequence $(Z_i)_{1 \leq i \leq n}$ conditional on the regimes and on the observations. This is a specific instance of Rao-Blackwellized Monte Carlo filters, often referred to as the *Mixture Kalman Filter*. Improvements of these early filtering techniques have been introduced in [DGK01, SGN05].

The use of Rao-Blackwellization to solve the smoothing problem has been proved to be more challenging and has received satisfactory solutions only recently. The first forward-backward smoother proposed in the literature [FGDW02] was not fully Rao-Blackwellized as it required to sample the hidden linear states in the backward pass. An alternative approach, based on the so-called structural approximation of the model suggested in an early paper by [Kim94], was proposed by [Bar06] to avoid to sample a continuous state in the backward pass. This approximation is rather ad-hoc and the resulting smoother is not consistent when the number of particles goes to infinity. The inaccuracy introduced by the approximation might be difficult to control.

The first fully Rao-Blackwellized SMC smoother which should lead to consistent approximations when the number of particles grows to infinity was proposed by [BDM10] and extends the Bryson-Frazier smoother for Gaussian linear state space models using the generalized two-filter formula with Rao-Blackwellization steps for the forward and the backward filters. This two-filter approach combines a forward filter with a backward information filter which are approximated numerically using SMC for the regime sequence and Kalman filtering techniques for the hidden linear states.

More recently, [SBG12, LBG13, LBS⁺16] introduced a Rao-Blackwellized smoother based on the forward-backward decomposition of the FFBS algorithm with Rao-Blackwellization steps both in the forward and backward time directions. The update of the smoothing distribution of the regime given the observations shares some striking similarities with the Rauch-Tung-Striebel smoothing procedure, which is at the heart of the FFBS procedure. The Rao-Blackwellization requires to update backward in time the smoothing distribution of the states given the regimes and the observations, which is achieved by using an *à la Kalman* backward update.

In this paper, we propose to improve the performance of the algorithms introduced in [BDM10] and in [SBG12, LBG13, LBS⁺16] by using additional Rao-Blackwellization steps which allows to sample new particles in the backward pass. This approach may be seen as an extension of the ideas of [FC03] for Rao-Blackwellized smoothers. In [BDM10], for all $1 \leq i \leq n$, the sampled forward and backward sequences are merged to approximate the posterior distribution of (a_i, z_i) . This provides an approximation whose support is restricted to the particles produced at time i by the backward particle filter. As noted in [FWT10, Section 2.6], these two-filter smoothers are prone to suffer from degeneracy issues when the algorithm associates forward particles at time $i - 1$ with backward particles at time i . We propose to approximate the marginal smoothing distribution of (a_i, z_i) by merging the sampled forward and backward trajectories at times $i - 1$ and $i + 1$ and integrating out all possible paths between time $i - 1$ and time i and between time i and time $i + 1$ instead of sampling random variables. Similarly, in the backward pass of [SBG12, LBG13, LBS⁺16], a regime \tilde{a}_i is sampled at time $1 \leq i \leq n - 1$ using the particles produced by the forward filter at time i . In this case, particle rejuvenation may be introduced by using the forward weighted samples at time $i - 1$ and extending these trajectories at time i with a Kalman filter for all possible values of the regime. Then, \tilde{a}_i is sampled in $\{1, \dots, J\}$ using an appropriately adapted weight.

The paper is organized as follows. The algorithms introduced in [BDM10] and in [SBG12, LBG13, LBS⁺16] as long as the proposed rejuvenation associated with each method are presented in Section 2. The performance of all these methods is illustrated in Section 3 with simulated data. In Section 4, an application to commodity markets is presented; the performance of our procedure is illustrated with crude oil data. A detailed derivation of the algorithms is provided in the Appendix.

2 Rao-Blackwellized smoothing algorithms

This section details the Sequential Monte Carlo algorithms which can be used to approximate the conditional distribution of the states (a_1, \dots, a_n) or the marginal distributions of (a_i, z_i) given the observations (Y_1, \dots, Y_n) . For all $m \times m$ matrix let $|A|$ be the determinant of A . If A is a positive-definite matrix, for all $z \in \mathbb{R}^m$ define

$$\|z\|_A^2 := z' A^{-1} z,$$

where for any vector or matrix z , z' denotes the transpose matrix of z . Let $m(a_i, z_{i-1}; z_i)$ be the probability density of the conditional distribution of Z_i given (a_i, Z_{i-1}) and $g(a_i, z_i; y_i)$ be the

probability density of the conditional distribution of Y_i given (a_i, Z_i) :

$$m(a_i, z_{i-1}; z_i) := |2\pi\overline{H}_{a_i}|^{-1/2} \exp\left\{-\frac{1}{2}\|z_i - d_{a_i} - T_{a_i}z_{i-1}\|_{\overline{H}_{a_i}}^2\right\}, \quad (3)$$

$$g(a_i, z_i; y_i) := |2\pi\overline{G}_{a_i}|^{-1/2} \exp\left\{-\frac{1}{2}\|y_i - c_{a_i} - B_{a_i}z_i\|_{\overline{G}_{a_i}}^2\right\}, \quad (4)$$

where

$$\overline{G}_j := G_j G_j', \quad \overline{H}_j := H_j H_j'.$$

All the algorithms considered in this paper are based on forward-backward or two-filter decompositions of the smoothing distributions and share the same forward filter presented in Section 2.1.

2.1 Forward filter

The SMC approximation $p^N(a_{1:i}, z_i | y_{1:i})$ of $p(a_{1:i}, z_i | y_{1:i})$ may be obtained using a standard Rao-Blackwellized algorithm. The procedure produces a sequence of trajectories $(a_{1:i}^k)_{1 \leq k \leq N}$ associated with normalized importance weights $(\omega_i^k)_{1 \leq k \leq N}$ ($\sum_{k=1}^N \omega_i^k = 1$) used to define the following approximation of $p(a_{1:i}, z_i | y_{1:i})$:

$$p^N(a_{1:i}, z_i | y_{1:i}) = \sum_{k=1}^N \omega_i^k p(z_i | a_{1:i}^k, y_{1:i}) \delta_{a_{1:i}^k}(a_{1:i}), \quad (5)$$

where δ is the Dirac delta function. In this equation, the conditional distribution of the hidden state z_i given the observations $y_{1:i}$ and a trajectory $a_{1:i}^k$ is a Gaussian distribution whose mean μ_i^k and variance P_i^k may be obtained by using the Kalman filter update.

Initialization

At time $i = 1$, write, for all $1 \leq j \leq J$,

$$\mu_{1|0}^j = c_j + B_j \mu_1 \quad \text{and} \quad P_{1|0}^j = B_j \Sigma_1 B_j' + \overline{G}_j.$$

$(a_1^k)_{1 \leq k \leq N}$ are sampled independently in $\{1, \dots, J\}$ with probabilities proportional to

$$\pi_j p(a_1 = j | y_1) \propto \pi_j |P_{1|0}^j|^{-1/2} \exp\left\{-(y_1 - \mu_{1|0}^j)'(P_{1|0}^j)^{-1}(y_1 - \mu_{1|0}^j)/2\right\}.$$

Then, μ_1^k and P_1^k are computed using a Kalman filter:

$$\begin{aligned} K_1^k &= \Sigma_1 B_{a_1^k}' \left(B_{a_1^k} \Sigma_1 B_{a_1^k}' + \overline{G}_{a_1^k} \right)^{-1}, \\ \mu_1^k &= \mu_1 + K_1^k \left(Y_1 - c_{a_1^k} - B_{a_1^k} \mu_1 \right), \\ P_1^k &= \left(I_m - K_1^k B_{a_1^k}' \right) \Sigma_1, \end{aligned}$$

where for all positive integer p , I_p is the $p \times p$ identity matrix. Each particle a_1^k is associated with the importance weight $\omega_1^k = 1/N$.

Iterations

Several procedures may be used to extend the trajectories $(a_{1:i-1}^k)_{1 \leq k \leq N}$ at time i . For all sampled trajectories $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and all $1 \leq j \leq J$, [CL00] used the incremental weights:

$$\gamma_i^{j,k} = p(y_i | a_i = j, a_{1:i-1}^k, y_{1:i-1}) Q(a_{i-1}^k, j).$$

The conditional distribution of Y_i given $a_{1:i-1}^k$, a_i and $Y_{1:i-1}$ is a Gaussian distribution with mean $c_{a_i} + B_{a_i} \mu_{i|i-1}^k(a_i)$ and variance $B_{a_i} P_{i|i-1}^k(a_i) B_{a_i}' + \overline{G}_{a_i}$ where

$$\begin{aligned} \mu_{i|i-1}^k(a_i) &= d_{a_i} + T_{a_i} \mu_{i-1}^k, \\ P_{i|i-1}^k(a_i) &= T_{a_i} P_{i-1}^k T_{a_i}' + \overline{H}_{a_i}. \end{aligned}$$

Therefore,

$$\gamma_i^{j,k} \propto Q(a_{i-1}^k, j) |B_j P_{i|i-1}^{j,k} B_j' + \overline{G}_j|^{-1/2} \exp \left\{ -\frac{1}{2} \left\| y_i - c_j - B_j \mu_{i|i-1}^{j,k} \right\|_{B_j P_{i|i-1}^{j,k} B_j' + \overline{G}_j}^2 \right\},$$

where:

$$\mu_{i|i-1}^{j,k} = \mu_{i|i-1}^k(j) = d_j + T_j \mu_{i-1}^k, \quad (6)$$

$$P_{i|i-1}^{j,k} = P_{i|i-1}^k(j) = T_j P_{i-1}^k T_j' + \overline{H}_j. \quad (7)$$

In [CL00], for all $1 \leq k \leq N$, an ancestral path is chosen with probabilities proportional to $(\omega_{i-1}^k)_{1 \leq k \leq N}$. Then, the new regime a_i^k is sampled in $\{1, \dots, J\}$ with probabilities proportional to $(\gamma_i^{j,k})_{1 \leq j \leq J}$. A drawback of this method is that only ancestral paths that have been selected using the importance weights $(\omega_{i-1}^k)_{1 \leq k \leq N}$ are extended at time i . Following [BGM08], this may be improved by considering all the offsprings of all ancestral trajectories $(a_{1:i-1}^k)_{1 \leq k \leq N}$. Each ancestral path has J offsprings at time i , it is thus necessary to choose a given number of trajectories at time i (for instance N) among the NJ possible paths. To obtain the weight associated with each offspring write the following approximation of $p(a_{1:i} | y_{1:i})$ based on the weighted samples at time $i-1$:

$$\begin{aligned} p^N(a_{1:i} | y_{1:i}) &\propto \sum_{k=1}^N \omega_{i-1}^k Q(a_{i-1}^k, a_i) p(y_i | a_{1:i-1}^k, a_i, y_{1:i-1}) \delta_{a_{1:i-1}^k}(a_{1:i-1}), \\ &\propto \sum_{k=1}^N \sum_{j=1}^J \omega_{i-1}^k \gamma_i^{j,k} \delta_{(a_{1:i-1}^k, j)}(a_{1:i}). \end{aligned}$$

Therefore, each ancestral trajectory of the form $(a_{1:i-1}^k, j)$, $1 \leq k \leq N$, $1 \leq j \leq J$, is associated with the normalized weight $\tilde{\omega}_i^{j,k} \propto \omega_{i-1}^k \gamma_i^{j,k}$. Several random selection schemes have been proposed to discard some of the possible offsprings to maintain an average number of N particles at each time step. Following [BGM08], we might choose between the Kullback-Leibler Optimal Selection (KL-OS) or the Chi-Squared Optimal Selection (CS-OS) to associate a new weight to each of the NJ trajectories. If the new weight is 0, then the corresponding particle can be removed.

KL-OS: λ is chosen as the solution of :

$$\sum_{k=1}^N \sum_{j=1}^J \min \left(\tilde{\omega}_i^{j,k} / \lambda, 1 \right) = N.$$

For all $1 \leq j \leq J$ and $1 \leq k \leq N$, if $\tilde{\omega}_i^{j,k} \geq \lambda$ then the new weight $\tilde{\Omega}_i^{j,k}$ is $\tilde{\Omega}_i^{j,k} = \tilde{\omega}_i^{j,k}$ and if $\tilde{\omega}_i^{j,k} < \lambda$:

$$\tilde{\Omega}_i^{j,k} = \begin{cases} \lambda & \text{with probability } \tilde{\omega}_i^{j,k}/\lambda, \\ 0 & \text{with probability } 1 - \tilde{\omega}_i^{j,k}/\lambda. \end{cases}$$

CS-OS: λ is chosen as the solution of :

$$\sum_{k=1}^N \sum_{j=1}^J \min \left(\sqrt{\tilde{\omega}_i^{j,k}/\lambda}, 1 \right) = N.$$

For all $1 \leq j \leq J$ and $1 \leq k \leq N$, if $\tilde{\omega}_i^{j,k} \geq \lambda$ then the new weight $\tilde{\Omega}_i^{j,k}$ is $\tilde{\Omega}_i^{j,k} = \tilde{\omega}_i^{j,k}$ and if $\tilde{\omega}_i^{j,k} < \lambda$:

$$\tilde{\Omega}_i^{j,k} = \begin{cases} \sqrt{\tilde{\omega}_i^{j,k}} \lambda & \text{with probability } \sqrt{\tilde{\omega}_i^{j,k}/\lambda}, \\ 0 & \text{with probability } 1 - \sqrt{\tilde{\omega}_i^{j,k}/\lambda}. \end{cases}$$

Then, in both cases, all particles such that $\tilde{\Omega}_i^{j,k} = 0$ are discarded and for all the other trajectories defined as an ancestral path $(a_{1:i-1}^k)$ extended by $a_i^k = j$, the new corresponding weight ω in (5) is given by the normalized weight $\tilde{\Omega}_i^{j,k}$. In the numerical sections of this paper, the Kullback-Leibler Optimal Selection (KL-OS) scheme is used.

2.2 FFBS based algorithms

2.2.1 FFBS algorithms of [SBG12, LBG13, LBS⁺16]

[SBG12, LBG13, LBS⁺16] proposed a Rao-Blackwellized procedure to sample the regime backward in time following the same steps as in the Forward Filtering Backward Smoothing algorithm [HK98, DGA00]. The algorithm relies on the decomposition given, for all $1 \leq i \leq n-1$, by:

$$p(a_{1:n}|y_{1:n}) = p(a_{1:i}|a_{i+1:n}, y_{1:n})p(a_{i+1:n}|y_{1:n}).$$

This decomposition is similar to the Rauch-Tung-Striebel decomposition of the filtering distribution. The first factor on the right hand side of the previous equation is nevertheless more difficult to handle because it itself relies on all the observations. As noted by [SBG12], this term can be computed recursively by considering the following decomposition:

$$p(a_{1:i}|a_{i+1:n}, y_{1:n}) \propto p(y_{i+1:n}, a_{i+1:n}|a_{1:i}, y_{1:i})p(a_{1:i}|y_{1:i}). \quad (8)$$

The second factor in the last equation may be approximated using the ancestral trajectories $(a_{1:i}^k)_{1 \leq k \leq N}$ and the associated importance weights $(\omega_i^k)_{1 \leq k \leq N}$ produced by the forward filter. Therefore, $p(a_{1:i}|a_{i+1:n}, y_{1:n})$ may be approximated by:

$$p^N(a_{1:i}|a_{i+1:n}, y_{1:n}) = \sum_{k=1}^N \tilde{\omega}_{i|n}^k \delta_{a_{1:i}^k}(a_{1:i}) \quad \text{with} \quad \tilde{\omega}_{i|n}^k \propto \omega_i^k p(y_{i+1:n}, a_{i+1:n}|a_{1:i}^k, y_{1:i}).$$

Then, a trajectory $\tilde{a}_{1:n}$ approximatively distributed according to $p(a_{1:n}|y_{1:n})$ may be sampled following these steps:

- Set $\tilde{a}_n = a_n^k$ with probabilities proportional to $(\omega_n^k)_{1 \leq k \leq N}$.
- For all $1 \leq i \leq n-1$, set $\tilde{a}_i = a_i^k$ with probabilities proportional to $(\tilde{\omega}_{i|n}^k)_{1 \leq k \leq N}$.

This algorithm requires to compute the quantity $p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i})$. This predictive quantity is available analytically using Kalman filtering techniques. However, this has to be done for each trajectory $(a_{1:i}^k)_{1 \leq k \leq N}$, which leads to an algorithm with a prohibitive computational complexity. [LBS⁺16] proposed a procedure computationally less intensive by conditioning with respect to z_i and then marginalizing with respect to this variable:

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i}) = \int p(y_{i+1:n}, a_{i+1:n} | z_i, a_i^k) p(z_i | a_{1:i}^k, y_{1:i}) dz_i. \quad (9)$$

This is similar to the two-filter decomposition of the smoothing distribution, see Section 2.3. By [LBS⁺16],

$$p(y_{i+1:n}, a_{i+1:n} | z_i, a_i) \propto Q(a_i, a_{i+1}) \exp \left\{ - (z_i' \Omega_i(a_{i+1:n}) z_i - 2 \lambda_i'(a_{i+1:n}) z_i) / 2 \right\},$$

where the proportionality is with respect to (a_i, z_i) and

$$p(y_{i:n}, a_{i+1:n} | z_i, a_i) \propto \exp \left\{ - (z_i' \hat{\Omega}_i(a_{i:n}) z_i - 2 \hat{\lambda}_i'(a_{i:n}) z_i) / 2 \right\},$$

where the proportionality is with respect to z_i . These quantities may be computed recursively backward in time with:

$$\begin{aligned} \hat{\Omega}_n(a_n) &= B_{a_n}' \bar{G}_{a_n}^{-1} B_{a_n}, \\ \hat{\lambda}_n(a_n) &= B_{a_n}' \bar{G}_{a_n}^{-1} (y_n - c_{a_n}). \end{aligned}$$

Then, for $1 \leq i \leq n-1$, define $m_{i+1} = \hat{\lambda}_{i+1} - \hat{\Omega}_{i+1} d_{a_{i+1}}$ and $M_{i+1} = H_{a_{i+1}}' \hat{\Omega}_{i+1} H_{a_{i+1}} + I_m$ and write

$$\begin{aligned} \Omega_i(a_{i+1:n}) &= T_{a_{i+1}}' (I_m - \hat{\Omega}_{i+1}(a_{i+1:n}) H_{a_{i+1}} M_{i+1}^{-1} H_{a_{i+1}}') \hat{\Omega}_{i+1}(a_{i+1:n}) T_{a_{i+1}}, \\ \lambda_i(a_{i+1:n}) &= T_{a_{i+1}}' (I_m - \hat{\Omega}_{i+1}(a_{i+1:n}) H_{a_{i+1}} M_{i+1}^{-1} H_{a_{i+1}}') m_{i+1}. \end{aligned}$$

As $p(y_{i:n}, a_{i+1:n} | z_i, a_i) = p(y_i | z_i, a_i) p(y_{i+1:n}, a_{i+1:n} | z_i, a_i)$,

$$\begin{aligned} \hat{\Omega}_i(a_{i:n}) &= \Omega_i(a_{i+1:n}) + B_{a_i}' \bar{G}_{a_i}^{-1} B_{a_i}, \\ \hat{\lambda}_i(a_{i:n}) &= \lambda_i(a_{i+1:n}) + B_{a_i}' \bar{G}_{a_i}^{-1} (y_i - c_{a_i}). \end{aligned}$$

Then, by (9),

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i}) \propto Q(a_i^k, a_{i+1}) |\Lambda_i^k(a_{i+1:n})|^{-1/2} \exp \left\{ - \eta_i^k(a_{i+1:n}) / 2 \right\}, \quad (10)$$

where the proportionality is with respect to $a_{1:i}^k$ and

$$\begin{aligned} \Lambda_i^k(a_{i+1:n}) &= (\Gamma_i^k)' \Omega_i(a_{i+1:n}) \Gamma_i^k + I_m, \\ \eta_i^k(a_{i+1:n}) &= \|\mu_i^k\|_{\Omega_i^{-1}(a_{i+1:n})}^2 - 2 \lambda_i'(a_{i+1:n}) \mu_i^k - \|(\Gamma_i^k)' (\lambda_i(a_{i+1:n}) - \Omega_i(a_{i+1:n}) \mu_i^k)\|_{\Lambda_i(a_{i+1:n})}^2, \end{aligned}$$

where $P_i^k = \Gamma_i^k (\Gamma_i^k)'$. Therefore,

$$\tilde{\omega}_{i|n} \propto \omega_i^k Q(a_i^k, a_{i+1}) |\Lambda_i^k(a_{i+1:n})|^{-1/2} \exp \left\{ -\eta_i^k(a_{i+1:n})/2 \right\}.$$

If $(\tilde{a}_{1:n}^k)_{1 \leq k \leq \tilde{N}}$ are independent copies of $\tilde{a}_{1:n}$, the SMC approximation of [LBS⁺16] of the joint smoothing distribution of the regime is:

$$p_{\tilde{N}}^{\text{Lbscg}}(a_{1:n} | Y_{1:n}) = \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} \delta_{\tilde{a}_{1:n}^k}(a_{1:n}).$$

2.2.2 Particle rejuvenation of FFBS algorithms

The crucial step of the FFBS algorithm is the decomposition (8) which allows to extend a backward trajectory $\tilde{a}_{i+1:n}$ by choosing a particle in the set $(a_i^k)_{1 \leq k \leq N}$ produced by the forward filter (and discarding the states $a_{1:i-1}^k$). An improved version of this FFBS algorithm which is not constrained to sample states in the support $(a_i^k)_{1 \leq k \leq N}$ may be defined for all $2 \leq i \leq n-1$ by writing:

$$\begin{aligned} p(a_{1:i} | a_{i+1:n}, y_{1:n}) &\propto p(y_{i+1:n}, a_{i+1:n} | a_{1:i}, y_{1:i}) p(a_{1:i} | y_{1:i}), \\ &\propto p(y_{i+1:n}, a_{i+1:n} | a_{1:i}, y_{1:i}) \int p(a_{1:i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) \\ &\quad m(a_i, z_{i-1}; z_i) g(a_i, z_i; y_i) dz_{i-1:i}. \end{aligned}$$

Replacing $p(a_{1:i-1}, z_{i-1} | y_{1:i-1})$ in the integral by the particle approximation obtained during the forward pass and using Kalman filtering techniques for each trajectory $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and each $a_i \in \{1, \dots, J\}$ yields:

$$\int p^N(a_{1:i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) m(a_i, z_{i-1}; z_i) g(a_i, z_i; y_i) dz_{i-1:i} \propto \sum_{k=1}^N \omega_{i|i-1}^k(a_i) \delta_{a_{1:i-1}^k}(a_{1:i-1}),$$

where

$$\omega_{i|i-1}^k(a_i) = \omega_{i-1}^k Q(a_{i-1}^k, a_i) |\Sigma_{i|i-1}^k(a_i)|^{-1/2} \exp \left\{ -\frac{1}{2} \|y_i - y_{i|i-1}^k(a_i)\|_{\Sigma_{i|i-1}^k(a_i)}^2 \right\},$$

$$y_{i|i-1}^k(a_i) = c_{a_i} + B_{a_i}(d_{a_i} + T_{a_i} \mu_{i-1}^k) \text{ and } \Sigma_{i|i-1}^k(a_i) = B_{a_i}(T_{a_i} P_{i-1}^k T_{a_i}' + \bar{H}_{a_i}) B_{a_i}' + \bar{G}_{a_i}.$$

On the other hand, for all $1 \leq k \leq N$, $p(y_{i+1:n}, a_{i+1:n} | a_{1:i-1}^k, a_i, y_{1:i})$ is computed as in (10) with all possible values $a_i \in \{1, \dots, J\}$ and not only the regime of the filtering pass $(a_i^k)_{1 \leq k \leq N}$. This means that a Kalman filter must be used for each trajectory $a_{1:i-1}^k$ which may be extended by $a_i \in \{1, \dots, J\}$. Denote by $\mu_{i|i-1}^k(a_i)$ and $P_{i|i-1}^k(a_i)$ the mean and covariance matrix of the law of z_i given $(a_{1:i-1}^k, a_i)$ obtained as in (6) and (7). Then,

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i-1}^k, a_i, y_{1:i}) = Q(a_i, a_{i+1}) |\Lambda_{i|i-1}^k(a_{i:n})|^{-1/2} \exp \left\{ -\eta_{i|i-1}^k(a_{i:n})/2 \right\}, \quad (11)$$

where the proportionality is with respect to $(a_{1:i-1}^k, a_i)$ and

$$\begin{aligned} \Lambda_{i|i-1}^k(a_{i:n}) &= (\Gamma_{i|i-1}^k(a_i))' \Omega_i(a_{i+1:n}) \Gamma_{i|i-1}^k(a_i) + I_m, \\ \eta_{i|i-1}^k(a_{i:n}) &= \|\mu_{i|i-1}^k(a_i)\|_{\Omega_i^{-1}(a_{i+1:n})}^2 - 2\lambda_i'(a_{i+1:n}) \mu_{i|i-1}^k(a_i) \\ &\quad - \|(\Gamma_{i|i-1}^k(a_i))' (\lambda_i(a_{i+1:n}) - \Omega_i(a_{i+1:n}) \mu_{i|i-1}^k(a_i))\|_{\Lambda_i(a_{i+1:n})}^2, \end{aligned}$$

where $\Gamma_{i|i-1}^k(a_i)$ is defined as $P_{i|i-1}^k(a_i) = \Gamma_{i|i-1}^k(a_i)(\Gamma_{i|i-1}^k(a_i))'$. The distribution $p(a_{1:i}|a_{i+1:n}, y_{1:n})$ is then approximated by :

$$p^N(a_{1:i}|a_{i+1:n}, y_{1:n}) \propto \sum_{k=1}^N \omega_{i|i-1}^k(a_i) Q(a_i, a_{i+1}) |\Lambda_{i|i-1}^k(a_{i:n})|^{-1/2} \exp \left\{ -\eta_{i|i-1}^k(a_{i:n})/2 \right\} \delta_{a_{1:i-1}^k}(a_{1:i-1}). \quad (12)$$

By integrating over all possible paths $a_{1:i-1}$, \tilde{a}_i is sampled in $\{1, \dots, J\}$. This particle rejuvenation step allows to explore states which are not in the support of the particles produced by the forward filter and improves the accuracy and the variance of the original FFBS algorithm, see Section 3 for numerical illustrations.

Another modification of the FFBS algorithm based on a Markov chain Monte Carlo (MCMC) sampling step was introduced in [LBS⁺16, Section 5.2]. Instead of sampling from (12), [LBS⁺16, Section 5.2] proposed to draw a forward path $a_{1:i-1}$ in $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and a state a_i in $\{1, \dots, J\}$ according to:

$$\tilde{q}(a_{1:i}|a_{i+1:n}, y_{1:n}) = \sum_{k=1}^N \tilde{\vartheta}_{i-1}^k \tilde{q}(a_i | a_{1:i-1}^k, a_{i+1:n}, y_{1:n}) \delta_{a_{1:i-1}^k}(a_{1:i-1}),$$

where $(\tilde{\vartheta}_{i-1}^k)_{1 \leq k \leq N}$ are adjustment multipliers and $\tilde{q}(a_i | a_{1:i-1}^k, a_{i+1:n}, y_{1:n})$ is a proposal kernel chosen by the user. This means that an ancestral path $a_{1:i-1}^k$ is sampled in $(a_{1:i-1}^k)_{1 \leq k \leq N}$ with weights $(\tilde{\vartheta}_{i-1}^k)_{1 \leq k \leq N}$ and a_i^* is sampled from $\tilde{q}(\cdot | a_{1:i-1}^k, a_{i+1:n}, y_{1:n})$. Then, the proposed sequence $a_{1:i}^*$ is accepted or rejected using the usual Metropolis-Hastings acceptance ratio. The choice of MCMC rejuvenation has interesting practical consequences as the computation of the acceptance ratio only requires to compute the posterior probability (11) for the proposed sequence $a_{1:i}^*$ while our technique is based on the computation of (11) for all combinations of sequences $(a_{1:i}^k)_{1 \leq k \leq N}$ and states $a_i \in \{1, \dots, J\}$. Sampling from (12) is computationally more intensive, especially when N is large, but our method is based on a direct approximation of $p(a_{1:i}|a_{i+1:n}, y_{1:n})$ based on $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and $a_{i+1:n}$ instead of approximate MCMC draws.

2.3 Rao-Blackwellized Two-filter Smoother

2.3.1 Rao-Blackwellized Two-filter Smoother of [BDM10]

Contrary to the previous methods, two-filter based smoothers are designed to compute approximations of marginal smoothing distributions (usually the posterior distribution of one or two consecutive regimes given all the observations). [BDM10] introduced the following decomposition of the smoothing distributions for all $2 \leq i \leq n$:

$$p(a_i, z_i | y_{1:n}) \propto p(a_i, z_i | y_{1:i-1}) p(y_{i:n} | a_i, z_i). \quad (13)$$

The first term on the right hand side may be approximated using the forward filter by noting that:

$$p(a_i, z_i | y_{1:i-1}) = \sum_{a_{i-1}} \int_{z_{i-1}} p(a_{i-1}, z_{i-1} | y_{1:i-1}) m(a_i, z_{i-1}; z_i) Q(a_{i-1}, a_i) dz_{i-1}.$$

In the forward pass described in Section 2.1, a set of possible sequences of regimes $a_{1:i-1}^k$ associated with importance weights ω_{i-1}^k , $1 \leq k \leq N$ are sampled to approximate $p(a_{i-1}, z_{i-1} | y_{1:i-1})$. This provides a normalized approximation $p^N(a_i, z_i | y_{1:i-1})$ of $p(a_i, z_i | y_{1:i-1})$. Define

$$\begin{aligned} \Omega_{i-1}^k(a_i) &= T_{a_i} P_{i-1}^k T_{a_i}' + \overline{H}_{a_i}, \quad \mu_{i-1}^k(a_i) = d_{a_i} + T_{a_i} \mu_{i-1}^k, \quad r_{i-1}^k(a_i) = (\Omega_{i-1}^k(a_i))^{-1} \mu_{i-1}^k(a_i), \\ \omega_{f,i}^k(a_i) &= \omega_{i-1}^k Q(a_{i-1}^k, a_i) |2\pi \Omega_{i-1}^k(a_i)|^{-1/2} \exp \left\{ -\frac{1}{2} \|\mu_{i-1}^k(a_i)\|_{\Omega_{i-1}^k(a_i)}^2 \right\}. \end{aligned}$$

Then,

$$p^N(a_i, z_i | y_{1:i-1}) = \sum_{k=1}^N \omega_{f,i}^k(a_i) \exp \left\{ -\frac{1}{2} \|z_i\|_{\Omega_{i-1}^k(a_i)}^2 + z_i' r_{i-1}^k(a_i) \right\}. \quad (14)$$

As the function $(a_i, z_i) \mapsto p(y_{i:n} | a_i, z_i)$ is not a probability density function, approximating the second term of (13) using SMC samples is not straightforward. The backward filter uses artificial densities to introduce a surrogate target density function which may be approximated recursively using SMC methods. Then, the forward and backward weighted samples are combined using (13) to approximate $p(a_i, z_i | y_{1:n})$. Following [BDM10], for any probability densities $(\gamma_i)_{1 \leq i \leq n}$, define the following joint probability densities:

$$\tilde{p}_n(a_n, z_n, y_n) := \gamma_n(a_n, z_n) g(a_n, z_n; y_n), \quad \tilde{p}_n(y_n) := \sum_{a_n=1}^J \int \gamma_n(a_n, z_n) g(a_n, z_n; y_n) dz_n,$$

and, for all $1 \leq i \leq n-1$,

$$\begin{aligned} \tilde{p}_i(a_{i:n}, z_{i:n}, y_{i:n}) &:= \gamma_i(a_i, z_i) p(y_{i:n} | a_{i:n}, z_{i:n}) p(a_{i+1:n}, z_{i+1:n} | a_i, z_i), \\ \tilde{p}_i(y_{i:n}) &:= \sum_{a_{i:n}=1}^J \int \gamma_i(a_i, z_i) p(y_{i:n} | a_{i:n}, z_{i:n}) p(a_{i+1:n}, z_{i+1:n} | a_i, z_i) dz_{i:n}. \end{aligned}$$

Note that this choice differs slightly from [BDM10] where it is advocated to set γ_i as the product of two independent densities $\gamma_i^a(a_i)$ and $\gamma_i^z(z_i)$. As the accuracy of the algorithm relies heavily on a proper tuning of this artificial density, a more general choice of γ_i is considered in this paper. By Lemma 1, these probability densities may be used to approximate the quantities $p(y_{i:n} | a_i, z_i)$, $1 \leq i \leq n$, in (13).

Lemma 1. For all $1 \leq i \leq n-1$,

$$\tilde{p}_i(a_i, z_i | y_{i:n}) = p(y_{i:n} | a_i, z_i) \gamma_i(a_i, z_i) / \tilde{p}_i(y_{i:n}), \quad (15)$$

$$\tilde{p}_i(a_i, z_i | y_{i:n}) = \gamma_i(a_i, z_i) \sum_{a_{i+1:n}=1}^J \frac{\tilde{p}_i(a_{i:n} | y_{i:n}) p(y_{i:n} | a_{i:n}, z_i)}{\int \gamma_i(a_i, z') p(y_{i:n} | a_{i:n}, z') dz'}. \quad (16)$$

Proof. The proof is postponed to Appendix A. \square

By definition of \tilde{p}_i for all $1 \leq i \leq n$,

$$\begin{aligned} \tilde{p}_i(a_{i:n}, z_i | y_{i:n}) &\propto \gamma_i(a_i, z_i) \int p(y_{i:n} | a_{i:n}, z_{i:n}) p(a_{i+1:n}, z_{i+1:n} | a_i, z_i) dz_{i+1:n}, \\ &\propto \gamma_i(a_i, z_i) \left\{ \prod_{u=i}^{n-1} Q(a_u, a_{u+1}) \right\} p(y_{i:n} | z_i, a_{i:n}). \end{aligned}$$

This yields:

$$\tilde{p}_i(a_{i:n}|y_{i:n}) \propto \left\{ \prod_{u=i}^{n-1} Q(a_u, a_{u+1}) \right\} \int \gamma_i(a_i, z_i) p(y_{i:n}|z_i, a_{i:n}) dz_i.$$

A set of weighted trajectories $(\tilde{a}_{i:n}^\ell)_{1 \leq \ell \leq N}$ with importance weights $(\tilde{\omega}_i^\ell)_{1 \leq \ell \leq N}$, $1 \leq i \leq n$, may then be sampled recursively backward in time to produce a SMC approximation of $\tilde{p}(a_{i:n}|y_{i:n})$ as follows.

- For $1 \leq \ell \leq N$, sample $\tilde{a}_n^j \sim \tilde{q}_n(\cdot)$ and set:

$$\tilde{\omega}_n^\ell \propto \frac{\int \gamma_n(\tilde{a}_n^\ell, z') g(\tilde{a}_n^\ell, z'; y_n) dz'}{\tilde{q}_n(\tilde{a}_n^\ell)}.$$

- For all $1 \leq i \leq n-1$, resample the set $(\tilde{a}_{i+1:n}^\ell)_{1 \leq j \leq N}$ using the normalized weights $(\tilde{\omega}_{i+1}^\ell)_{1 \leq j \leq N}$. Then, for $1 \leq \ell \leq N$, sample $\tilde{a}_i^j \sim \tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \cdot)$ and set:

$$\tilde{\omega}_i^\ell \propto \frac{Q(\tilde{a}_i^\ell, \tilde{a}_{i+1}^\ell) \int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n}|\tilde{a}_{i:n}^\ell, z') dz'}{\tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \tilde{a}_i^\ell) \int \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z') p(y_{i+1:n}|\tilde{a}_{i+1:n}^\ell, z') dz'}.$$

To obtain uniformly weighted samples at each time step, in the numerical experiments we use:

$$\tilde{q}_n(\cdot) = \int \gamma_n(\cdot, z') g(\cdot, z'; y_n) dz' \quad \text{and} \quad \tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \cdot) = \frac{Q(\cdot, \tilde{a}_{i+1}^\ell) \int \gamma_i(\cdot, z') p(y_{i:n}|\cdot, \tilde{a}_{i+1:n}^\ell, z') dz'}{\int \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z') p(y_{i+1:n}|\tilde{a}_{i+1:n}^\ell, z') dz'}.$$

By (15) and (16),

$$p(y_{i:n}|a_i, z_i) = \frac{\tilde{p}_i(y_{i:n}) \tilde{p}_i(a_i, z_i|y_{i:n})}{\gamma_i(a_i, z_i)} = \tilde{p}_i(y_{i:n}) \sum_{a_{i+1:n}=1}^J \frac{\tilde{p}_i(a_{i:n}|y_{i:n}) p(y_{i:n}|a_{i:n}, z_i)}{\int \gamma_i(a_i, z') p(y_{i:n}|a_{i:n}, z') dz'},$$

which suggests the following particle approximation $p^N(y_{i:n}|a_i, z_i)$ of $p(y_{i:n}|a_i, z_i)$:

$$p^N(y_{i:n}|a_i, z_i) = \tilde{p}_i(y_{i:n}) \sum_{\ell=1}^N \frac{\tilde{\omega}_i^\ell p(y_{i:n}|\tilde{a}_{i:n}^\ell, z_i)}{\int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n}|\tilde{a}_{i:n}^\ell, z') dz'} \delta_{\tilde{a}_i^\ell}(a_i). \quad (17)$$

The conditional likelihood of the observations given the sequence of states $p(y_{i:n}|a_{i:n}, z_i)$ can be computed explicitly using a Gaussian backward smoother; these computations are summarized in Lemma 2. In the numerical experiments, $\gamma_i(a_i, z_i)$ is set as a mixture of Gaussian distributions. Note that for such a choice, the integral $\int \gamma_i(a_i, z') p(y_{i:n}|a_{i:n}, z') dz'$ may be computed explicitly, see Lemma 3. Then, combining (17) and (14) with (13) provides an approximation of $p(a_i, z_i|y_{1:n})$ by merging the forward particles $(a_{i-1}^k)_{1 \leq k \leq N}$ with the backward particles $(\tilde{a}_{i+1}^k)_{1 \leq k \leq N}$, the support of this SMC approximation of $p(a_i, z_i|y_{1:n})$ being $(\tilde{a}_{i+1}^k)_{1 \leq k \leq N}$.

As noted in [FWT10, Secion 2.6], two-filter smoothers are prone to suffer from degeneracy issues when the algorithm associates forward particles at time $i-1$ with backward particles at time i . The authors illustrate this issue in the case where the hidden state is an AR(2) process. To overcome the weakness of such standard two-filter approaches the particle rejuvenation proposed in Section 2.3.2 follows the idea introduced in [FWT10] where new particles at time i are sampled conditional

on $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and on $(\tilde{a}_{i+1:n}^k)_{1 \leq k \leq N}$ and appropriately weighted. This allows to produce new particles at time i and to obtain a SMC approximation of $p(a_i, z_i | y_{1:n})$ whose support is not restricted to $(\tilde{a}_{i+1}^k)_{1 \leq k \leq N}$. Section 2.3.2 exploits this idea in the specific case of linear and Gaussian models where explicit computations allows to produce an approximation using $(a_{1:i-1}^k)_{1 \leq k \leq N}$ and $(\tilde{a}_{i+1:n}^k)_{1 \leq k \leq N}$ with support $\{1, \dots, J\}$ and without any additional sampling steps.

2.3.2 Particle rejuvenation of two-filter based algorithms

For $2 \leq i \leq n-1$, particle rejuvenation relies on the explicit marginalization:

$$p(a_i, z_i | y_{1:n}) = \sum_{a_{i-1}} \sum_{a_{i+1}} \int_{z_{i-1}} \int_{z_{i+1}} \psi_i^n(a_{i-1:i+1}, z_{i-1:i+1}) dz_{i-1} dz_{i+1}, \quad (18)$$

where $\psi_i^n(a_{i-1:i+1}, z_{i-1:i+1})$ is the smoothing distribution of the hidden regimes and states between time indices $i-1$ and $i+1$. Note that the EM algorithm requires the approximation of $p(a_{i-1}, z_{i-1}, a_{i+1}, z_{i+1} | y_{1:n})$ in the E-step, this may be obtained following the same steps by marginalizing explicitly the linear states at time $i-2$ and $i+1$. Intermediate computations follow the same steps as for the approximation of $p(a_i, z_i | y_{1:n})$. First, ψ_i^n may be decomposed as follows:

$$\begin{aligned} \psi_i^n(a_{i-1:i+1}, z_{i-1:i+1}) &\propto p(y_{i+1:n} | a_{i+1}, z_{i+1}) p(a_{i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) m(a_i, z_{i-1}; z_i) \\ &\quad \times g(a_i, z_i; y_i) Q(a_i, a_{i+1}) m(a_{i+1}, z_i; z_{i+1}), \end{aligned}$$

where the proportionality is with respect to $(a_{i-1:i+1}, z_{i-1:i+1})$. Then, by (18), the smoothing distribution $p(a_i, z_i | y_{1:n})$ may be written as

$$p(a_i, z_i | y_{1:n}) \propto p(a_i, z_i | y_{1:i-1}) g(a_i, z_i; y_i) t_i(a_i, z_i, y_{i+1:n}), \quad (19)$$

where m and g are defined in (3) and (4) and

$$t_i(a_i, z_i, y_{i+1:n}) = \sum_{a_{i+1}} \int_{z_{i+1}} m(a_{i+1}, z_i; z_{i+1}) Q(a_i, a_{i+1}) p(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}. \quad (20)$$

The backward pass described in Section 2.3.1 produces a sequence of states $\tilde{a}_{i+1:n}^\ell$ associated with importance weights $\tilde{\omega}_{i+1}^\ell$, $1 \leq \ell \leq N$ which are used to approximate $p(y_{i+1:n} | a_{i+1}, z_{i+1})$. Plugging this approximation into (20) provides an approximation $t_i^N(a_i, z_i, y_{i+1:n})$ of $t_i(a_i, z_i, y_{i+1:n})$ integrating over all possible choices (a_{i+1}, z_{i+1}) . These steps are then combined to form a non normalized SMC approximation of $p(a_i, z_i | y_{1:n})$ using (19). The normalization of the SMC approximation of $p(a_i, z_i | y_{1:n})$ is obtained by integrating over the states a_i, z_i , when $p(a_i, z_i | y_{1:i-1})$ and $t_i(a_i, z_i, y_{i+1:n})$ are replaced by $p^N(a_i, z_i | y_{1:i-1})$ and $t_i^N(a_i, z_i, y_{i+1:n})$ in (19). Our procedure allows to construct sequence of regimes with non-degenerated importance weights in the combination step. This procedure improves significantly [BDM10] where no marginalization of $p(a_i, z_i | y_{1:n})$ over the states at times $i-1$ and $i+1$ is performed and where the proposed forward and backward paths are directly merged. This method often leads to importance weights which are close to be numerically degenerated. By Lemma 2, the SMC approximation $p^N(y_{i:n} | a_i, z_i)$ of $p(y_{i:n} | a_i, z_i)$ is then given by:

$$p^N(y_{i:n} | a_i, z_i) = \tilde{p}_i(y_{i:n}) \sum_{\ell=1}^N \frac{\delta_{\tilde{a}_i^\ell}(a_i) \tilde{\omega}_i^\ell}{\int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n} | \tilde{a}_{i:n}^\ell, z') dz'} \exp \left\{ -\frac{1}{2} \|z_i\|_{\tilde{P}_i^\ell}^2 + z_i' \tilde{V}_i^\ell - \frac{1}{2} \tilde{C}_i^\ell \right\}, \quad (21)$$

where $(\tilde{P}_i^\ell)^{-1} := \tilde{P}_i^{-1}(\tilde{a}_{i:n}^\ell)$, $\tilde{\nu}_i^\ell := \tilde{\nu}_i(\tilde{a}_{i:n}^\ell)$ and $\tilde{c}_i^\ell := \tilde{c}_i(\tilde{a}_{i:n}^\ell)$ are defined in Lemma 2. Define

$$\begin{aligned}\Delta_{i+1}^\ell &:= \left(I_m + H'_{\tilde{a}_{i+1}^\ell} (\tilde{P}_{i+1}^\ell)^{-1} H_{\tilde{a}_{i+1}^\ell} \right)^{-1}, \\ \delta_{i+1}^\ell &:= \tilde{\nu}_{i+1}^\ell + \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} (d_{\tilde{a}_{i+1}^\ell} + T_{\tilde{a}_{i+1}^\ell} z_i).\end{aligned}$$

Then, by (20), the SMC approximation $t_i^N(a_i, z_i, y_{i+1:n})$ of $t_i(a_i, z_i, y_{i+1:n})$ is given by:

$$\begin{aligned}t_i^N(a_i, z_i, y_{i+1:n}) &= \sum_{a_{i+1}=1}^J \int_{z_{i+1}} m(a_{i+1}, z_i; z_{i+1}) Q(a_i, a_{i+1}) p^N(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}, \\ &= \tilde{p}_{i+1}(y_{i+1:n}) \sum_{\ell=1}^N C_i^{-1}(\tilde{a}_{i+1:n}^\ell) Q(a_i, \tilde{a}_{i+1}^\ell) \tilde{\omega}_{i+1}^\ell |\overline{H}_{\tilde{a}_{i+1}^\ell}|^{-1/2} |H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell}|^{1/2} \\ &\quad \times \exp \left\{ \frac{1}{2} (\delta_{i+1}^\ell)' H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell} \delta_{i+1}^\ell - \frac{1}{2} \left\| d_{\tilde{a}_{i+1}^\ell} + T_{\tilde{a}_{i+1}^\ell} z_i \right\|_{\overline{H}_{\tilde{a}_{i+1}^\ell}}^2 \right\}, \\ &= \sum_{\ell=1}^N \tilde{\omega}_{b,i}^\ell(a_i) \exp \left\{ -\frac{1}{2} \|z_i\|_{\tilde{S}_{i+1}^\ell}^2 + z_i' \tilde{s}_{i+1}^\ell \right\},\end{aligned}\tag{22}$$

where

$$\begin{aligned}C_i(\tilde{a}_{i+1:n}^\ell) &:= \exp \{ \tilde{c}_{i+1}^\ell / 2 \} \int_{z_{i+1}} \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z) \tilde{p}(y_{i+1:n} | \tilde{a}_{i+1:n}^\ell, z) dz, \\ \tilde{\omega}_{b,i}^\ell(a_i) &= \tilde{p}_{i+1}(y_{i+1:n}) C_i(\tilde{a}_{i+1:n}^\ell)^{-1} Q(a_i, \tilde{a}_{i+1}^\ell) \tilde{\omega}_{i+1}^\ell |\overline{H}_{\tilde{a}_{i+1}^\ell}|^{-1/2} |H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell}|^{1/2} \\ &\quad \times \exp \{ -d_{\tilde{a}_{i+1}^\ell}' \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell} / 2 \} \exp \{ (\tilde{\nu}_{i+1}^\ell + \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell})' H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell} (\tilde{\nu}_{i+1}^\ell + \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}) / 2 \}, \\ (\tilde{S}_{i+1}^\ell)^{-1} &= T'_{\tilde{a}_{i+1}^\ell} \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} (T_{\tilde{a}_{i+1}^\ell} - H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell} \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} T_{\tilde{a}_{i+1}^\ell}), \\ \tilde{s}_{i+1}^\ell &= T'_{\tilde{a}_{i+1}^\ell} \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} (H_{\tilde{a}_{i+1}^\ell} \Delta_{i+1}^\ell H'_{\tilde{a}_{i+1}^\ell} (\tilde{\nu}_{i+1}^\ell + \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}) - d_{\tilde{a}_{i+1}^\ell}).\end{aligned}$$

In the numerical experiments, $\gamma_i(a_i, z_i)$ is set as a mixture of Gaussian distributions. As explained in Section 2.3.1, the integral $\int_{z_{i+1}} \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z) \tilde{p}(y_{i+1:n} | \tilde{a}_{i+1:n}^\ell, z) dz$ may be computed explicitly, see Lemma 3.

3 Simulated data

This section highlights the improvements brought by the additional Rao-Blackwellization steps for the two-filter and the FFBS approximations of the marginal smoothing distributions in the case where the number of states is $J = 2$. The transition matrix Q is such that the probability of switching from one regime to the other is small, as expected for the WTI crude oil data, see Section 4. First, the algorithms are applied to a simple one-dimensional model with:

$$\begin{aligned}\pi_1 &= \pi_2 = 0.5; \quad d_1 = 0.5 \quad d_2 = 0; \quad c_1 = 0.1 \quad c_2 = 0, \\ Q &= \begin{pmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{pmatrix} \quad T_1 = T_2 = 1 \quad \overline{H}_1 = \overline{H}_2 = 0.1,\end{aligned}$$

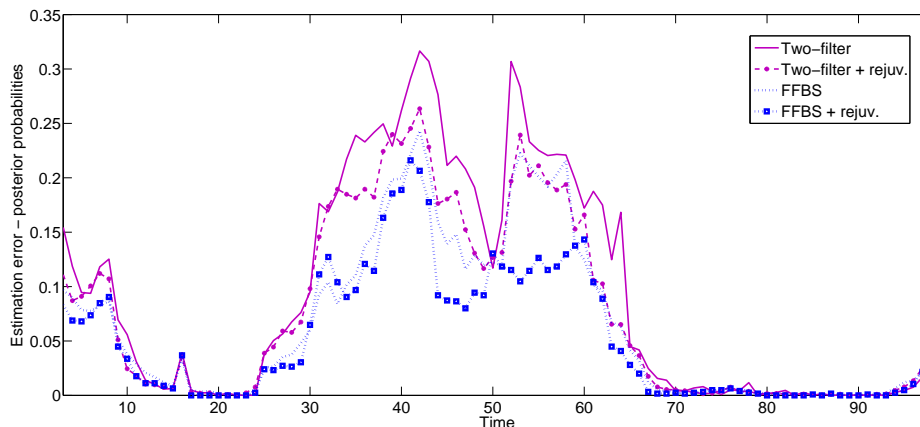


Figure 1: Posterior probabilities estimation error for all algorithms.

$$B_1 = B_2 = 1 \quad \bar{G}_1 = 0.3 \quad \bar{G}_2 = 0.1.$$

The original FFBS algorithm of [LBS⁺16] and the FFBS algorithm with rejuvenation proposed in this paper are used with $N = \bar{N} = 25$. For comparable computational costs, the two-filter method of [BDM10] and the method with rejuvenation are run with $N = 100$. The artificial distributions are chosen as $\gamma_i(a_i, z_i) = p^N(a_i, z_i | y_{1:i-1})$ where $p^N(a_i, z_i | y_{1:i-1})$ is defined by (14). All these algorithms are compared to the estimation obtained with the proposed FFBS algorithm with rejuvenation and 5000 particles considered as a benchmark value. Figure 1 displays the mean estimation error over 100 independent Monte Carlo runs. The estimation error is defined as the absolute difference between the benchmark value and the estimations given by all algorithms. In addition, Figure 2 displays the empirical variance of the estimation for each algorithm. Figure 1 and Figure 2 illustrate that in both cases the additional rejuvenation step improves the accuracy and the variability of SMC smoothers. In addition, even with a sharp choice for the artificial distributions γ_i , $1 \leq i \leq n$, FFBS based methods outperform two-filter based smoothers for this model.

4 Application to CME crude oil (WTI)

4.1 Model

Modeling commodity prices is a crucial step to value contingent claims related to energy markets and to optimize storage or extraction strategies. In [GE90, Sch97], the authors proposed a model where the spot price of a commodity ($S_t, t \geq 0$) depends on a second factor ($\delta_t, t \geq 0$), referred to as the instantaneous convenience yield. This factor plays the role of dividends in equity markets and models the benefit of holding the physical commodity or the storage and maintenance costs required to keep the commodity. In this model, this convenience yield is described as an Ornstein-Uhlenbeck process:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \eta dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho dt, \end{aligned}$$

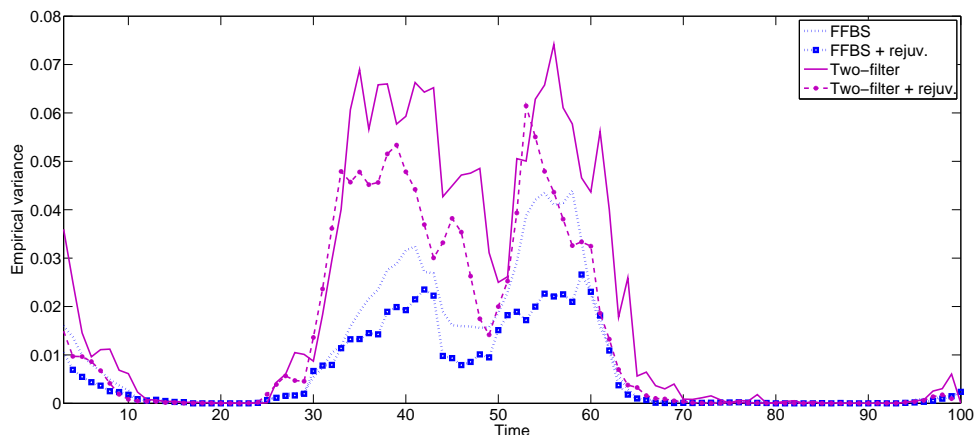


Figure 2: Empirical variances of the estimation of $\mathbb{P}(a_k = 1 | Y_{1:n})$ for all algorithms.

where the parameter $(r, \sigma, \kappa, \alpha, \eta, \rho)$ are constant and $((W_t^1, W_t^2), t \geq 0)$ are standard Brownian motions. This model appears to be too restrictive as energy markets are not likely to revert to a single equilibrium value. This assumption is relaxed using Markov switching models to allow several possible regimes for the spot price and the convenience yield. Following [Alm16], the spot price and convenience yield are described in this paper as:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma_{a_t} S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha_{a_t} - \delta_t)dt + \eta_{a_t} dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

where $(a_t)_{t \geq 0}$ is a finite state space Markov process. This model allows to exhibit fundamental features of commodity future prices, which typically display different regimes of volatility and/or convenience yield. A two-regime model is already sufficient to produce stylized effects such as contango (increase of future prices) and backwardation (decrease of future prices). Assuming that the switching rate between regimes is negligible compared to the inverse of the discretization period, the discretized version of the spot price and convenience yield $Z_i = (\ln S_i, \delta_i)$ (the sampling period is taken to be 1) is modeled as a CLGM. The explicit integration of this SDE detailed in Lemma 4 yields the following discrete time model for $(Z_i)_{i \geq 2}$:

$$Z_i = d_{a_{i-1}} + TZ_{i-1} + H_{a_{i-1}} \varepsilon_i,$$

where (with $\bar{H}_{a_{i-1}} := H_{a_{i-1}} H'_{a_{i-1}}$ and $\tau = t_i - t_{i-1}$):

$$d_{a_{i-1}} := \begin{pmatrix} \left[\mu - \alpha_{a_{i-1}} - \sigma_{a_{i-1}}^2 / 2 \right] \tau + \alpha_{a_{i-1}} [1 - e^{-\kappa\tau}] / \kappa \\ \alpha_{a_{i-1}} [1 - e^{-\kappa\tau}] \end{pmatrix},$$

$$T := \begin{pmatrix} 1 & -[1 - e^{-\kappa\tau}] / \kappa \\ 0 & e^{-\kappa\tau} \end{pmatrix},$$

$$\bar{H}_{a_{i-1}}(1, 1) = \sigma_{a_{i-1}}^2 \tau + \eta_{a_{i-1}}^2 \left\{ \tau + (1 - e^{-2\kappa\tau}) / (2\kappa) - 2(1 - e^{-\kappa\tau}) / \kappa \right\} / \kappa^2$$

$$\quad - 2\rho_{a_{i-1}} \eta_{a_{i-1}} \sigma_{a_{i-1}} \left\{ \tau t_i - (1 - e^{-\kappa\tau}) / \kappa \right\} / \kappa,$$

$$\bar{H}_{a_{i-1}}(1, 2) = \left(\rho_{a_{i-1}} \eta_{a_{i-1}} \sigma_{a_{i-1}} - \eta_{a_{i-1}}^2 / \kappa \right) (1 - e^{-\kappa\tau}) / \kappa + \eta_{a_{i-1}}^2 (1 - e^{-2\kappa\tau}) / (2\kappa^2),$$

$$\bar{H}_{a_{i-1}}(2, 1) = \bar{H}_{a_{i-1}}(1, 2), \quad \bar{H}_{a_{i-1}}(2, 2) = \eta_{a_{i-1}}^2 (1 - e^{-2\kappa\tau}) / (2\kappa).$$

The observations are Wednesday future contracts of the West Texas Intermediate crude oil (WTI) traded in the Chicago Mercantile Exchange (CME) from 11 January 1995 to 13 November 2013. The contracts are numbered F_1, F_2, \dots, F_{36} where F_1 (or front month) is the earliest delivery future contract, F_2 is the second earliest delivery future contract and so on. Among these 36 contracts, the four future contracts: F_1, F_4, F_6, F_{13} are used since their trading volumes and their impacts on the Term Structures are the most important (F_1 is the most liquid contract, F_{13} characterizes the gap between prices over a one year period, F_4 and F_6 are intermediate future contracts that are mostly traded). As in [Alm16], we consider that each future contract has a fixed time to maturity: F_1, F_4, F_6, F_{13} have time to maturity 4, 16, 26, 56 weeks. Our time series contains $n = 975$ weekly data with 534 in backwardation and 441 in contango (the backwardation effect is more frequent with crude oil data). At each time $t_i = i\tau$, with $\tau = 0.0192$, the observations of the $p = 4$ future prices are $Y_i := (\ln(F_{i\tau t, m_1}^{(market)}), \dots, \ln(F_{i\tau t, m_p}^{(market)}))'$, where $F_{t_i, m}$ is the future price at t_i for a maturity m weeks. A closed form solution for $F_{t_i, m}$ may be written:

$$F_{t_i, m} := \exp(\mathbf{A}_m(a_i) + \mathbf{B}_m Z_i),$$

where $\mathbf{B}_0 = (1 \ 0)$ and $\mathbf{B}_m = \mathbf{B}_{m-1} T$ so that $\mathbf{B}_m = (1 \ -(1 - e^{-\kappa m \tau}) / \kappa)$ and for all $1 \leq j \leq J$, $\mathbf{A}_0(j) = 0$, and

$$\mathbf{A}_m(j) = \ln \left(\sum_{k=1}^J Q(j, k) \exp(\mathbf{A}_{m-1}(k)) \right) + \mathbf{B}_{m-1} d_j + \frac{1}{2} \mathbf{B}_{m-1} \bar{H}_j \mathbf{B}'_{m-1}.$$

Therefore, the observations of the logfuture prices are given, for all $1 \leq i \leq n$, by:

$$Y_i = c_{a_i} + \mathbf{B} Z_i + \mathbf{G} \eta_i,$$

where η_i is a standard multivariate Gaussian random variable and:

$$c'_j = [\mathbf{A}_{m_1}(j), \dots, \mathbf{A}_{m_p}(j)], \quad \mathbf{B}' = [\mathbf{B}'_{m_1}, \dots, \mathbf{B}'_{m_p}], \quad \mathbf{G} = \text{diag}(g_1, \dots, g_d).$$

The model depends of the parameters:

$$\theta := \{\pi, Q, \mu_1, \Sigma_1, \kappa, (\alpha_j)_{1 \leq j \leq J}, (\sigma_j)_{1 \leq j \leq J}, (\eta_j)_{1 \leq j \leq J}, (\rho_j)_{1 \leq j \leq J}, (g_\ell)_{1 \leq \ell \leq d}\}.$$

The aim of this section is to estimate θ and the posterior probabilities $\mathbb{P}(a_k = j|Y_{1:n})$, $1 \leq k \leq n$, $1 \leq j \leq J$. Given the observations $Y_{1:n}$, the EM algorithm introduced in [DLR77] maximizes the incomplete data log-likelihood $\theta \mapsto \ell_\theta^n$ defined by

$$\ell_\theta^n(Y_{1:n}) := \log \left(\sum_{a_1=1}^J \cdots \sum_{a_n=1}^J \int p_\theta(a_{1:n}, z_{1:n}, Y_{1:n}) dz_{1:n} \right),$$

where the complete data likelihood p_θ is given by

$$p_\theta(a_{1:n}, z_{1:n}, Y_{1:n}) := \pi(a_1) \phi_{\mu_1, \Sigma_1}(z_1) g_\theta(a_1, z_1; Y_1) \prod_{i=2}^n Q(a_{i-1}, a_i) m_\theta(a_i, z_{i-1}; z_i) g_\theta(a_i, z_i; Y_i).$$

Denote by $\mathbb{E}_\theta[\cdot|Y_{1:n}]$ the conditional expectation given $Y_{1:n}$ when the parameter value is set to θ . The EM algorithm iteratively builds a sequence $\{\theta_p\}_{p \geq 0}$ of parameter estimates following the two steps:

1. **E-step:** compute $\theta \mapsto Q(\theta, \theta_p) := \mathbb{E}_{\theta_p}[\log p_\theta(a_{1:n}, Z_{1:n}, Y_{1:n})|Y_{1:n}]$;
2. **M-step:** choose θ_{p+1} as a maximizer of $\theta \mapsto Q(\theta, \theta_p)$.

All the conditional expectations involved in $Q(\theta, \theta_p)$ are approximated using our two-filter algorithm with rejuvenation to define the SMC approximation $\theta \mapsto Q^N(\theta, \theta_p)$ of $\theta \mapsto Q(\theta, \theta_p)$. As the function $\theta \mapsto Q^N(\theta, \theta_p)$ cannot be maximized analytically, the M-step is performed numerically using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) introduced in [HO01]. This derivative-free optimization procedure is known to perform well in complex multimodal optimization settings, see e.g. [HK04].

4.2 Numerical results

The initial transition probability in CMA-ES is chosen as $Q(1, 1) = 0.98$, $Q(2, 2) = 0.97$ and $\pi_1 = \pi_2 = 0.5$ where the number 1 represents the backwardation regime and 2 represents the contango regime. The other parameters are initialized as shown in Table 1. The number of particles

κ	α_1	α_2	σ_1	σ_2	η_1	η_2	ρ_1	ρ_2	g_1	g_2	g_3	g_4
5.0	0.1	-0.05	0.4	0.4	0.5	0.5	0.75	0.65	0.1	0.1	0.1	0.1

Table 1: Initial values for the EM algorithm.

is set to $N = 100$, $\tau = 1/52$. The interest rate is $r = 0.0296$ as in [Alm16]. The initial guess for the mean and variance of the initial state are

$$\mu_1 = \left(\ln F_{1,4}^{(market)} \quad r - \frac{\ln F_{1,16}^{(market)} - \ln F_{1,4}^{(market)}}{(16-4)\tau} \right) \quad \text{and} \quad \Sigma_1 = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.05 \end{pmatrix}.$$

The CMA-ES algorithm is used with an initial standard deviation for the parameters $\sigma_{cmaes} = 0.005$, a number of selected search points $\mu_{cmaes} = 20$ and a population size $\lambda_{cmaes} = 100$. The algorithm is stopped after 10000 iterations.

In Gibson-Schwartz model [GE90], a stronger backwardation effect implies a greater value for α for the same values of the other parameters. For the CME WTI Crude Oil, backwardation effect is more frequent than contango effect so that α_1 should be greater than α_2 . Therefore, this condition is imposed for all simulations in the CMA-ES algorithm. The results after 2500 iterations of the EM algorithm are given in Table 2. The estimated values and standard deviations are obtained with 50 independent runs of the algorithm. As expected, we obtain $\sigma_1 \geq \sigma_2$, $\alpha_1 \geq \alpha_2$, $\eta_1 \geq \eta_2$

Parameter	κ	σ_1	σ_2	η_1	η_2	ρ_1	ρ_2
Value	2.6378	0.3733	0.3485	0.5892	0.3814	0.8709	0.6761
Std. Dev	0.1999	0.005438	0.002884	0.0483	0.0349	0.0064	0.0052

Parameter	α_1	α_2	g_1	g_2	g_3	g_4	$Q(1, 1)$	$Q(2, 2)$
Value	0.0889	-0.0281	2.3e-2	1.0e-4	3.0e-4	2.3e-2	0.9917	0.9880
Std. Dev	0.004248	0.00149	1.9e-4	2.6e-4	2.3e-4	2.1e-04	6.7e-4	9.6e-4

Table 2: Final estimates after 2500 iterations.

and $\rho_1 \geq \rho_2$ at convergence of the EM algorithm. Moreover, $Q(1, 1) > Q(2, 2)$ corresponds to the prediction that we did from the data description. The fact that $\sigma_1 \geq \sigma_2$ and $\alpha_1 \geq \alpha_2$ indicates the first regime (backwardation) characterized by a higher value in both volatility and equilibrium level of convenience yield, and the second regime (contango) characterized by a lower value in both volatility and equilibrium convenience yield level. This in accordance with the theory of storage that the volatility of the commodity spot price is high when the inventory is low, and the convenience yield is all the higher as inventory is low.

Figure 3 compares the evolution of future 1M (the nearest contracts) to the *term structure* observed from CME WTI crude oil, defined as the difference of future 13M and future 1M (to avoid seasonality). The figure shows that it is not necessary to have an inverse relationship between the price of the nearest contract and the term structure. But when a significant drop in the price of the nearest contract occurs, the term structure increases (i.e. in contango).

The correlation between the spot price and the convenience yield is positive and high in both two regimes. This is an accordance to what as been observed in most commodity market, see [GE90]. The slope of future curve decreases in function of maturity.

Figure 4 and 5 display the the estimated posterior probabilities of the regimes and the observed future slope. When the future curve is in backwardation (resp. contango), the model is expected to be in the first regime (resp. second regime), except for the period where the slope of the future curve is too small and in the period from December 2008 to April 2009 (beginning of the crisis).

5 Conclusions

This paper presents Rao-Blackwellized Sequential Monte Carlo methods to approximate smoothing distributions in conditionally linear and Gaussian state spaces in a common unifying framework. It also provides different techniques that could be used in the forward filtering pass to improve significantly the usual mixture Kalman filter. The filtering distributions are approximated at each time step by considering all possible offsprings of all ancestral trajectories before discarding degenerated paths instead of resampling the ancestral paths before propagating them at the next time step. The paper investigates the benefit of additional Rao-Blackwellization steps to sample new

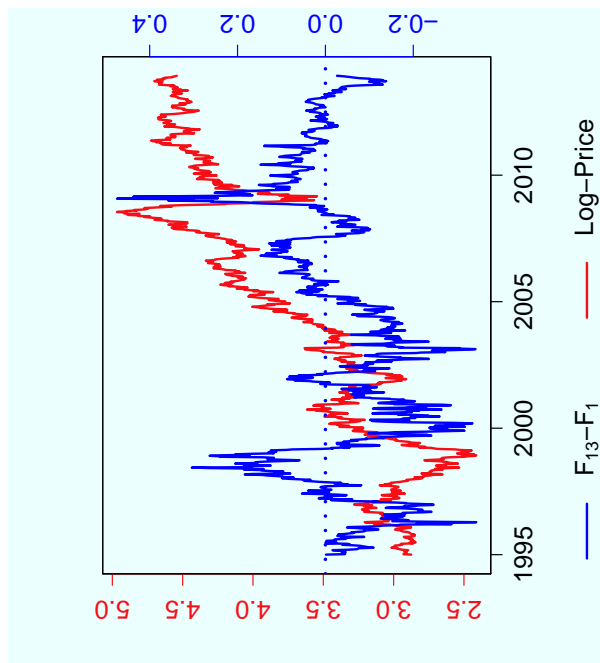


Figure 3: Log-price (red line) and slope of future curves (blue line).

regimes at each time step conditional on the forward and backward particles. This rejuvenation step uses explicit integration of the hidden linear states before merging the forward and backward filters for two-filter based algorithms or before sampling new states backward in time for FFBS based methods. The paper displays some Monte Carlo experiments with simulated data to illustrate that this additional rejuvenation step improves the performance of the smoothing algorithms with no substantial additional computational costs. They are also applied to commodity markets using WTI crude oil data.

A Technical lemmas

Lemmas 1, 2 and 3 are close to [BDM10, Proposition 5, Proposition 6]. The proofs are detailed in this appendix for completeness.

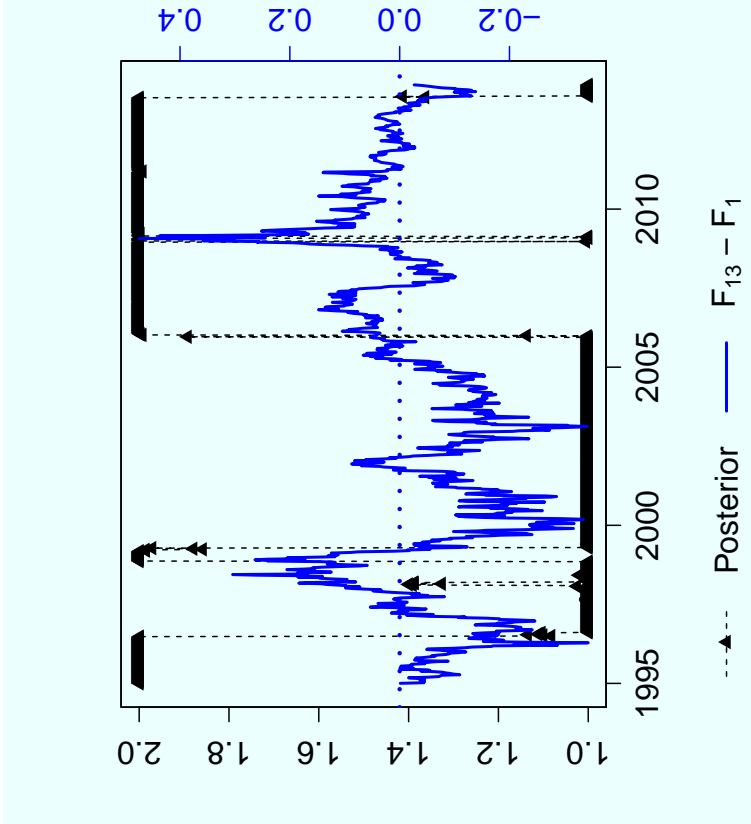


Figure 4: Posterior probability (triangle black) and slope of future curves (blue line).

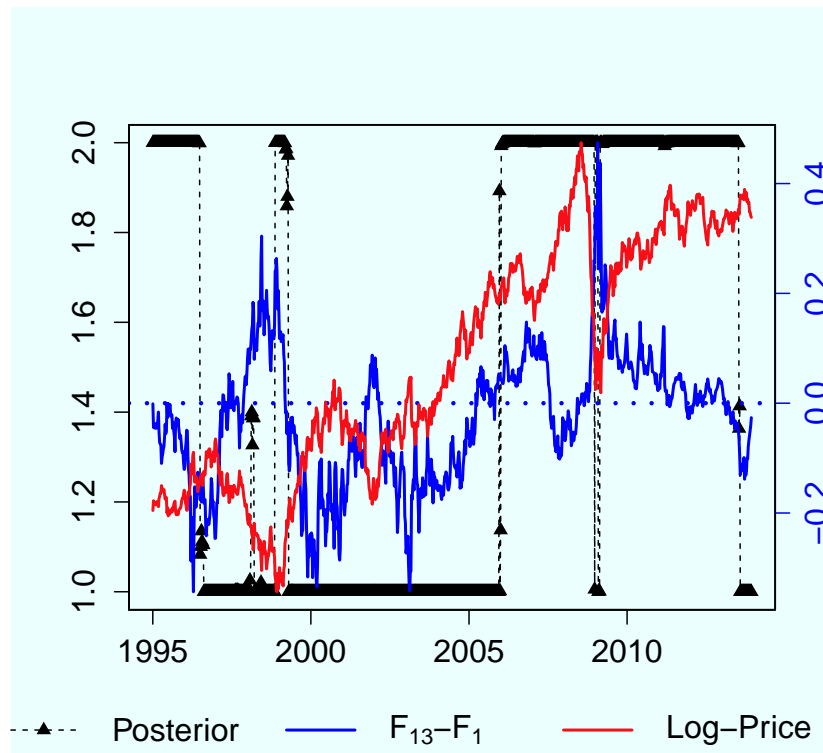


Figure 5: Posterior probability (triangle black) and slope of future curves (blue line).

Proof of Lemma 1. For all $1 \leq i \leq n-1$,

$$\begin{aligned}
p(y_{i:n}|a_i, z_i) &= \sum_{a_{i+1:n}} \int p(y_{i:n}, a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
&= \sum_{a_{i+1:n}} \int p(a_{i+1:n}, z_{i+1:n}|a_i, z_i) p(y_{i:n}|a_{i:n}, z_{i:n}) dz_{i+1:n}, \\
&= \frac{\tilde{p}_i(y_{i:n})}{\gamma_i(a_i, z_i)} \sum_{a_{i+1:n}} \int \frac{\gamma_i(a_i, z_i)}{\tilde{p}_i(y_{i:n})} p(a_{i+1:n}, z_{i+1:n}|a_i, z_i) p(y_{i:n}|a_{i:n}, z_{i:n}) dz_{i+1:n}, \\
&= \frac{\tilde{p}_i(y_{i:n})}{\gamma_i(a_i, z_i)} \sum_{a_{i+1:n}} \int \tilde{p}_i(a_{i:n}, z_{i:n}|y_{i:n}) dz_{i+1:n}, \\
&= \frac{\tilde{p}_i(y_{i:n})}{\gamma_i(a_i, z_i)} \tilde{p}_i(a_i, z_i|y_{i:n}),
\end{aligned}$$

which concludes the proof of (15). To prove (16) write,

$$\begin{aligned}
\tilde{p}_i(a_{i:n}, z_i|y_{i:n}) &= \frac{\gamma_i(a_i, z_i)}{\tilde{p}_i(y_{i:n})} \int p(y_{i:n}|a_{i:n}, z_{i:n}) p(a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
&= \frac{\gamma_i(a_i, z_i)}{\tilde{p}_i(y_{i:n})} \int \frac{p(y_{i:n}|a_{i:n}, z_i) p(z_{i+1:n}|y_{i:n}, a_{i:n}, z_i)}{p(z_{i+1:n}|a_{i:n}, z_i)} p(a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
&= \frac{\gamma_i(a_i, z_i)}{\tilde{p}_i(y_{i:n})} p(y_{i:n}|a_{i:n}, z_i) p(a_{i+1:n}|a_i).
\end{aligned}$$

Therefore,

$$\tilde{p}_i(a_i, z_i|y_{i:n}) = \frac{\gamma_i(a_i, z_i)}{\tilde{p}_i(y_{i:n})} \sum_{a_{i+1:n}} p(y_{i:n}|a_{i:n}, z_i) p(a_{i+1:n}|a_i)$$

and the proof is completed upon noting that

$$\tilde{p}_i(a_{i:n}|y_{i:n}) = \frac{p(a_{i+1:n}|a_i)}{\tilde{p}_i(y_{i:n})} \int \gamma_i(a_i, z) p(y_{i:n}|a_{i:n}, z) dz.$$

□

Lemma 2. For all $1 \leq i \leq n$,

$$p(y_{i:n}|a_{i:n}, z_i) = \exp \left\{ -\frac{1}{2} \tilde{c}_i(a_{i:n}) - \frac{1}{2} \|z_i\|_{\tilde{P}_i(a_{i:n})}^2 + z_i' \tilde{\nu}_i(a_{i:n}) \right\}, \quad (23)$$

where

$$\tilde{c}_n(a_n) = \mathfrak{p} \log(2\pi) + \log |\overline{G}_{a_n}| + \|y_n - c_{a_n}\|_{\overline{G}_{a_n}}^2, \quad (24)$$

$$\tilde{P}_n^{-1}(a_n) = B_{a_n}' \overline{G}_{a_n}^{-1} B_{a_n}, \quad (25)$$

$$\tilde{\nu}_n(a_n) = B_{a_n}' \overline{G}_{a_n}^{-1} (y_n - c_{a_n}) \quad (26)$$

and, for all $1 \leq i \leq n-1$,

$$\tilde{c}_i(a_{i:n}) = \tilde{c}_{i|i+1}(a_{i+1:n}) + \mathfrak{p} \log(2\pi) + \log |\overline{G}_{a_i}| + \|y_i - c_{a_i}\|_{\overline{G}_{a_i}}^2, \quad (27)$$

$$\tilde{P}_i^{-1}(a_{i:n}) = \tilde{P}_{i|i+1}^{-1}(a_{i+1:n}) + B_{a_i}' \overline{G}_{a_i}^{-1} B_{a_i}, \quad (28)$$

$$\tilde{\nu}_i(a_{i:n}) = \tilde{\nu}_{i|i+1}(a_{i+1:n}) + B_{a_i}' \overline{G}_{a_i}^{-1} (y_i - c_{a_i}), \quad (29)$$

with

$$\Delta_{i+1}(a_{i+1:n}) = \left(I_m + H_{a_{i+1}}' \tilde{P}_{i+1}^{-1}(a_{i+1:n}) H_{a_{i+1}} \right)^{-1},$$

$$\tilde{r}_{i|i+1}(a_{i+1:n}) = \tilde{\nu}_{i+1}(a_{i+1:n}) + \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}},$$

$$\tilde{c}_{i|i+1}(a_{i+1:n}) = \tilde{c}_{i+1}(a_{i+1:n}) + \log |\overline{H}_{a_{i+1}}| + d_{a_{i+1}}' \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}} - \log |H_{a_{i+1}} \Delta_i(a_{i+1:n}) H_{a_{i+1}}'| \\ - \tilde{r}_{i|i+1}'(a_{i+1:n}) H_{a_{i+1}} \Delta_i(a_{i+1:n}) H_{a_{i+1}}' \tilde{r}_{i|i+1}(a_{i+1:n}),$$

$$\tilde{P}_{i|i+1}^{-1}(a_{i+1:n}) = T_{a_{i+1}}' \left(I_m - \overline{H}_{a_{i+1}}^{-1} H_{a_{i+1}} \Delta_i(a_{i+1:n}) H_{a_{i+1}}' \right) \overline{H}_{a_{i+1}}^{-1} T_{a_{i+1}},$$

$$\tilde{\nu}_{i|i+1}(a_{i+1:n}) = T_{a_{i+1}}' \overline{H}_{a_{i+1}}^{-1} \left[-d_{a_{i+1}} + H_{a_{i+1}} \Delta_i(a_{i+1:n}) H_{a_{i+1}}' \left(\tilde{\nu}_{i+1}(a_{i+1:n}) + \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}} \right) \right].$$

Proof. The result is proved by backward induction. (24), (25) and (26) follow directly from (2). Assume that for a given $1 \leq i \leq n-1$, $p(y_{i+1:n}|a_{i+1:n}, z_{i+1})$ is given by (23). Write

$$p(y_{i:n}|a_{i:n}, z_i) = \int m(a_{i+1}, z_i; z_{i+1}) g(a_i, z_i; y_i) p(y_{i+1:n}|a_{i+1:n}, z_{i+1}) dz_{i+1},$$

with

$$m(a_{i+1}, z_i; z_{i+1}) = \exp \left\{ -\frac{\mathfrak{m}}{2} \log(2\pi) - \frac{1}{2} \log |\overline{H}_{a_{i+1}}| - \frac{1}{2} \|z_{i+1} - d_{a_{i+1}} - T_{a_{i+1}} z_i\|_{\overline{H}_{a_{i+1}}}^2 \right\},$$

$$g(a_i, z_i; y_i) = \exp \left\{ -\frac{\mathfrak{p}}{2} \log(2\pi) - \frac{1}{2} \log |\overline{G}_{a_i}| - \frac{1}{2} \|y_i - c_{a_i} - B_{a_i} z_i\|_{\overline{G}_{a_i}}^2 \right\},$$

$$p(y_{i+1:n}|a_{i+1:n}, z_{i+1}) = \exp \left\{ -\frac{1}{2} c_{i+1}(a_{i+1:n}) - \frac{1}{2} \|z_{i+1}\|_{\tilde{P}_{i+1}(a_{i+1:n})}^2 + z_{i+1}' \tilde{\nu}_{i+1}(a_{i+1:n}) \right\}.$$

Let Δ_{i+1} and δ_{i+1} be given by:

$$\Delta_{i+1}(a_{i+1:n}) := \left(I_m + H_{a_{i+1}}' \tilde{P}_{i+1}^{-1}(a_{i+1:n}) H_{a_{i+1}} \right)^{-1},$$

$$\delta_{i+1}(a_{i+1:n}) := \nu_{i+1}(a_{i+1:n}) + \overline{H}_{a_{i+1}}^{-1} (d_{a_{i+1}} + T_{a_{i+1}} z_i).$$

Then, $\overline{H}_{a_{i+1}}^{-1} + \tilde{P}_{i+1}^{-1}(a_{i+1:n}) = \left(H_{a_{i+1}} \Delta_{i+1}(a_{i+1:n}) H_{a_{i+1}}' \right)^{-1}$ and (27), (28) and (29) follows from

$$\int \exp \left\{ -\frac{1}{2} \|z_{i+1}\|_{H_{a_{i+1}} \Delta_{i+1}(a_{i+1:n}) H_{a_{i+1}}'}^2 + z_{i+1}' \delta_{i+1}(a_{i+1:n}) \right\} dz_{i+1} \\ = \exp \left\{ \frac{1}{2} \log(2\pi) + \frac{1}{2} \log |H_{a_{i+1}} \Delta_{i+1}(a_{i+1:n}) H_{a_{i+1}}'| \right\} \\ \times \exp \left\{ \frac{1}{2} \delta_{i+1}'(a_{i+1:n}) H_{a_{i+1}} \Delta_{i+1}(a_{i+1:n}) H_{a_{i+1}}' \delta_{i+1}(a_{i+1:n}) \right\}. \quad \square$$

Lemma 3. For all $1 \leq i \leq n$,

$$\int \phi_{\mu_i, \Sigma_i}(z_i) p(y_{i:n} | a_{i:n}, z_i) dz_i = \exp \left\{ -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} \mu_i' \Sigma_i^{-1} \mu_i \right\} \\ \times \exp \left\{ -\frac{1}{2} \tilde{c}_i(a_{i:n}) + \frac{1}{2} \log |\tilde{\Omega}_i(a_{i:n})| + \frac{1}{2} z_i'(a_{i:n}) \tilde{\Omega}_i(a_{i:n}) z_i(a_{i:n}) \right\},$$

where $\phi_{\mu, \Sigma}$ is the probability density function of a m dimensional Gaussian random variable with mean μ and variance matrix Σ and

$$\tilde{\Omega}_i(a_{i:n}) := \left(\Sigma_i^{-1} + \tilde{P}_i^{-1}(a_{i:n}) \right)^{-1} \quad \text{and} \quad \tilde{z}_i(a_{i:n}) := \Sigma_i^{-1} \mu_i + \tilde{\nu}_i(a_{i:n})$$

and where c_i , \tilde{P}_i and ν_i are given in Lemma 2.

Proof. By Lemma 2,

$$\phi_{\mu_i, \Sigma_i}(z_i) p(y_{i:n} | a_{i:n}, z_i) = \exp \left\{ -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} \|z_i - \mu_i\|_{\Sigma_i}^2 \right\} \\ \times \exp \left\{ -\frac{1}{2} c_i(a_{i:n}) - \frac{1}{2} \|z_i\|_{\tilde{P}_i(a_{i:n})}^2 + z_i' \nu_i(a_{i:n}) \right\}.$$

The proof is completed noting that

$$\int \exp \left\{ -\frac{1}{2} z_i' \tilde{\Omega}_i^{-1}(a_{i:n}) z_i + z_i' (\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n})) \right\} dz_i \\ = \exp \left\{ \frac{m}{2} \log(2\pi) + \frac{1}{2} \log |\tilde{\Omega}_i(a_{i:n})| + \frac{1}{2} [\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n})]' \tilde{\Omega}_i(a_{i:n}) [\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n})] \right\}.$$

□

Lemma 4. Let $(X_t, \delta_t)_{t \geq 0}$ be solutions to the following SDE:

$$dX_t = (\mu - \delta_t - \sigma^2/2) dt + \sigma dW_t^1, \\ d\delta_t = \kappa(\alpha - \delta_t) dt + \eta dW_t^2,$$

$(W_t^1)_{t \geq 0}$ and $(W_t^2)_{t \geq 0}$ are standard Brownian motions such that $d\langle W_t^1, W_t^2 \rangle = \rho dt$. Then, for all $t \geq 0$ and $h > 0$,

$$\begin{pmatrix} X_{t+h} \\ \delta_{t+h} \end{pmatrix} = d_h + T_h \begin{pmatrix} X_t \\ \delta_t \end{pmatrix} + H_h \varepsilon,$$

where ε is a standard 2-dimensional Gaussian random variable and (with $\bar{H}_h := H_h' H_h$),

$$d := \begin{pmatrix} [\mu - \alpha - \sigma^2/2] h + \alpha[1 - e^{-\kappa h}]/\kappa \\ \alpha[1 - e^{-\kappa h}] \end{pmatrix}, \quad T_h := \begin{pmatrix} 1 & -[1 - e^{-\kappa h}]/\kappa \\ 0 & e^{-\kappa h} \end{pmatrix},$$

$$\bar{H}_h(1, 1) := \sigma^2 h + \eta^2 \{ h + (1 - e^{-2\kappa h})/(2\kappa) - 2(1 - e^{-\kappa h})/\kappa \} / \kappa^2 \\ - 2\rho\eta\sigma \{ h - (1 - e^{-\kappa h})/\kappa \} / \kappa,$$

$$\bar{H}_h(1, 2) := (\rho\eta\sigma - \eta^2/\kappa) (1 - e^{-\kappa h}) / \kappa + \eta^2 (1 - e^{-2\kappa h}) / (2\kappa^2),$$

$$\bar{H}_h(2, 1) := \bar{H}_h(1, 2),$$

$$\bar{H}_h(2, 2) := \eta^2 (1 - e^{-2\kappa h}) / (2\kappa).$$

Proof. For all $t \geq 0$,

$$X_t = X_0 + (\mu - \sigma^2/2)t - \int_0^t \delta_s ds + \sigma W_t^1$$

and, as $(\delta_t)_{0 \leq t \leq T}$ is an Ornstein-Uhlenbeck process,

$$\delta_t = \delta_0 e^{-\kappa t} + \alpha(1 - e^{-\kappa t}) + \int_0^t \eta e^{\kappa(s-t)} dW_s^2.$$

Then,

$$\begin{aligned} \int_0^t \delta_s ds &= (\delta_0 - \alpha)(1 - e^{-\kappa t})/\kappa + \alpha t + \eta \int_0^t \int_0^s e^{\kappa(u-s)} dW_u^2 ds, \\ &= (\delta_0 - \alpha)(1 - e^{-\kappa t})/\kappa + \alpha t + (\eta/\kappa) \int_0^t (1 - e^{-\kappa(t-s)}) dW_s^2. \end{aligned}$$

Defining $\tilde{W}_t^1 := -(\eta/\kappa) \int_0^t (1 - e^{-\kappa(t-s)}) dW_s^2 + \sigma W_t^1$ and $\tilde{W}_t^2 := \int_0^t \eta e^{\kappa(s-t)} dW_s^2$, this yields:

$$\begin{aligned} X_t &= X_0 + (\mu - \sigma^2/2)t + (\alpha - \delta_0)(1 - e^{-\kappa t})/\kappa - \alpha t + \tilde{W}_t^1, \\ \delta_t &= \delta_0 e^{-\kappa t} + \alpha(1 - e^{-\kappa t}) + \tilde{W}_t^2. \end{aligned}$$

The proof is concluded upon noting that \tilde{W}_t^1 and \tilde{W}_t^2 are centered Gaussian random variables such that:

$$-\text{Var} [\tilde{W}_t^1] = \sigma^2 t + \eta^2 \{t + (1 - e^{-2\kappa t})/(2\kappa) - 2(1 - e^{-\kappa t})/\kappa\} / \kappa^2 - 2\rho\eta\sigma \{t - (1 - e^{-\kappa t})/\kappa\} / \kappa,$$

$$-\text{Var} [\tilde{W}_t^2] = \eta^2(1 - e^{-2\kappa t})/(2\kappa),$$

$$-\text{Cov} [\tilde{W}_t^1, \tilde{W}_t^2] = (\rho\eta\sigma - \eta^2/\kappa) (1 - e^{-\kappa t}) / \kappa + \eta^2 (1 - e^{-2\kappa t}) / (2\kappa^2).$$

□

References

- [Alm16] A. Almansour. Convenience yield in commodity price modeling: A regime switching approach. *Energy Economics*, 53:238–247, 2016.
- [Bar06] D. Barber. Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning Research*, 7:2515–2540, 2006.
- [BDM10] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61–89, 2010.
- [BF63] A.E. Bryson and M. Frazier. Smoothing for linear and nonlinear dynamic systems. *Proceedings of the optimum system synthesis conference*, 1963.
- [BGM08] S. Barembuch, A. Garivier, and E. Moulines. On optimal sampling for particle filtering in digital communication. *IEEE 9th Workshop on Signal Processing Advances in Wireless*, pages 634–638, 2008.

- [CL00] R. Chen and J.S. Liu. Mixture kalman filters. *Journal of the Royal Statistical Society B*, 62:493–508, 2000.
- [DGA00] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10:197–208, 2000.
- [DGG01] A. Doucet, N. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38 (with discussion), 1977.
- [DM13] P. Del Moral. *Mean field simulation for Monte Carlo integration*. Chapman & Hall / CRC Monographs on Statistics & Applied Probability, 2013.
- [FC03] P. Fearnhead and P. Clifford. On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society: Series B*, 65:887–899, 2003.
- [FGDW02] W. Fong, S.J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing*, 50(2):438–449, 2002.
- [FWT10] P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97:447–464, 2010.
- [GE90] R. Gibson and Schwartz E.S. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of finance*, 45(3):959–976, 1990.
- [HK98] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics*, 7:175–193, 1998.
- [HK04] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Eighth International Conference on Parallel Problem Solving from Nature*, 72:337–354, 2004.
- [HO01] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [Kal60] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.
- [Kim94] C.J. Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22, 1994.
- [LBGS13] F. Lindsten, P. Bunch, S.J. Godsill, and T.B. Schon. Rao-Blackwellized particle smoothers for mixed linear/nonlinear state-space models. *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

- [LBS⁺16] F. Lindsten, P. Bunch, S. Sarkka, T.B. Schon, and S.J. Godsill. Rao-Blackwellized particle smoothers for conditionally linear Gaussian models. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):353–365, 2016.
- [RST65] A.E. Rauch, C.T. Striebel, and F. Tung. Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics journal*, 3(8):1445–1450, 1965.
- [Sar13] S. Sarkka. *Bayesian filtering and smoothing*. Cambridge University Press, 2013.
- [SBG12] S. Sarkka, P. Bunch, and S.J. Godsill. A backward-simulation based Rao-Blackwellized particle smoother for conditionally linear Gaussian models. *Proceedings of the 16th IFAC Symposium on System Identification (SYSID)*, 2012.
- [Sch97] E.S. Schwartz. The stochastic behaviour of commodity prices: Implications for pricing and hedging. *The Journal of Finance*, 3:923–973, 1997.
- [SGN05] T. Schon, F. Gustafsson, and P.-J. Nordlund. Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7):2279–2289, 2005.