Research Article

Nicole Bäuerle, Igor Gilitschenski and Uwe Hanebeck

Exact and approximate hidden Markov chain filters based on discrete observations

Abstract: We consider a *Hidden Markov Model* (HMM) where the integrated continuous-time Markov chain can be observed at discrete time points perturbed by a Brownian motion. The aim is to derive a filter for the underlying continuous-time Markov chain. The recursion formula for the discrete-time filter is easy to derive, however involves densities which are very hard to obtain. In this paper we derive exact formulas for the necessary densities in the case the state space of the HMM consists of two elements only. This is done by relating the underlying integrated continuous-time Markov chain to the so-called asymmetric telegraph process and by using recent results on this process. In case the state space consists of more than two elements we present three different ways to approximate the densities for the filter. The first approach is based on the continuous filter problem. The second approach is to derive a PDE for the densities and solve it numerically. The third approach is a crude discrete time approximation of the Markov chain. All three approaches are compared in a numerical study.

Keywords: Hidden Markov model, discrete Bayesian filter, Wonham filter, asymmetric telegraph process

1 Introduction

We consider a *Hidden Markov Model* (HMM) where the integrated continuous-time Markov chain can be observed at discrete time points perturbed by a Brownian motion. The aim is to derive a filter for the underlying continuous-time Markov chain. Thus, we have a continuous-time Hidden Markov Model with discrete obser-vations. Models of this type are widely used in finance, in telecommunication or in biology. For references, we refer to the examples in Section 2. The knowledge of the filter is for example necessary in case one has to control the system in a dynamic way. Usually the principle of estimation and control is valid and implies that the filter has to be computed first and then the control is applied, see, e.g., [4, 25, 28] for different kinds of applications. We assume that all parameters of the model are known or have been estimated before. In par-ticular, we assume that the number and values of the states of the hidden Markov chain are known. As far as model selection is concerned we refer, e.g., to [6, 12, 22]. Papers dealing with the estimation of state values and transition intensities are, e.g., [8, 13, 14, 26]. Since we assume a finite number of states for the Markov chain, the filter has a finite dimension and it is in principal clear from the existing theory how to construct it. However, the filter recursion involves densities which are difficult to compute explicitly.

We rely on results from [21] and [7] on the asymmetric telegraph process to obtain the exact filter in closed form in case the hidden Markov chain has only two states. We distinguish between the symmetric case, where the switching intensities are the same for both states, and the asymmetric case, where intensities are different. The first case is easier and related to the well-known *telegraph process* which has been introduced by Kac in 1956 (see [18]) and which we discuss in Section 3. The second problem is related to the asymmetric

telegraph process. Estimates for the switching intensity of a symmetric telegraph process have, e.g., been derived in [17, 32] and for an inhomogeneous telegraph process in [16].

In the general situation with more than two states, the required densities become rather complicated. So a procedure to compute these densities approximately is called for. We present three different ways to derive approximations. All three approaches are general and need no information about the size of the intensities, i.e. how fast the hidden Markov chain switches. Of course the faster the switching of the Markov chain, the harder it is to obtain a reasonable filter, given a fixed time lag between observations. If the switching intensity is too large compared to the observation time lag, the stationary distribution is a reasonable approximation. The first approach relies on results in [19, 23] where the authors derive an approximate continuous-time filter. When this continuous-time filter is discretized, it can also be used by someone who is only able to observe the data in discrete time. However, this approach is based on the assumption that two matrices commute which is not true in general and thus only provides an approximate solution. The second approach is to derive a PDE for the densities which are involved in the filter process and to solve this PDE numerically. This approach is exact up to the error of the numerical PDE solution. The third attempt relies on a naive approximation of the continuous-time Markov chain by a discrete-time Markov chain. The finer the approximation, i.e. the smaller the time step of the discrete-time Markov chain, the better the approximate filter but the higher the computational effort.

Our paper is organized as follows: Section 2 contains a precise mathematical formulation of the problem and derives the basic filter recursion. Section 3 is then devoted to the problem with two states. Results about the telegraph process are used to derive the filter explicitly in both the symmetric and the asymmetric case. Section 4 considers the general problem and three approximate filters are derived. Section 5 is devoted to numerical results. In particular, we compare our approximate filters in a setting with five states in situations with frequent or rare observations and with small or large variability of the error terms.

2 Hidden Markov chain filters

Suppose we have a filtered probability space $(\Omega, \mathcal{F}_T, (\mathcal{F}_t), \mathbb{P})$ with $T > 0$ and a hidden stationary continuous-time Markov chain $\varepsilon = (\varepsilon_t)_{t\in[0,T]}$ on it with finite state space $\mathcal{E} = \{1, \dots, d\}$ and $d \in \mathbb{N}$. The intensity matrix of this process is given by Q and the initial distribution by p_0. Let us denote by $p(t) = (p_1(t), \dots, p_d(t))^\top$ the distribution of ε_t for $t \in [0, T]$, then $p(t)$ satisfies the ODE

$$\frac{dp(t)}{dt} = Qp(t) \tag{2.1}$$

with $p(0) = p_0$. Note that for $i, j \in \mathcal{E}$, $i \neq j$ either $p_{ij}(t) \equiv 0$ or $p_{ij}(t) > 0$ for all $t > 0$. We assume w.l.o.g. that the latter case is valid for all $i \neq j$ otherwise we could exclude elements from the state space. The formal solution of (2.1) is given by

$$p(t) = p_0 e^{Qt}$$

where

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!}$$

is the matrix exponential. Moreover, the h-step transition probabilities are given by

$$p_{ij}(h) = \mathbb{P}(\varepsilon_{t+h} = j | \varepsilon_t = i) = (e^{Qh})_{i,j}. \tag{2.2}$$

Now assume further that $W = (W_t)_{t\in[0,T]}$ is a Brownian motion on our filtered probability space which is independent of the hidden continuous-time Markov chain ε and plays the role of a noise. Moreover, $\alpha_1, \dots, \alpha_d$ are different real numbers and $\sigma > 0$. We assume that we can observe the process

$$Z_t := \int_0^t \alpha_{\varepsilon_s} ds + \sigma W_t \tag{2.3}$$

at discrete time points $t_k = kh$ for $k = 0, 1, \dots, N$, $h > 0$ and $Nh = T$, i.e. the observation σ-algebra at time t_k is given by

$$\mathcal{Z}_k := \sigma(\{Z_{t_1} - Z_{t_0}, \dots, Z_{t_k} - Z_{t_{k-1}}\}), \quad k = 1, \dots, N$$

as opposed to the filtration $(\mathcal{F}_t)$ which contains both the information about ε and W. A simple interpretation of this model is as follows: Suppose a particle is moving on the real line, starting in zero. The velocity of the particle at time t is determined by α_{ε_t}. Thus, $\int_0^t \alpha_{\varepsilon(s)} ds$ gives the exact position of the particle at time t. However, we are only able to observe the position at discrete time points t_k together with a noise σW_{t_k}. The aim is now to filter from this observation the current velocity of the particle, i.e. to determine

$$\mathbb{P}(\varepsilon_{t_n} = j | \mathcal{Z}_n) =: \mu_n(j)$$

where $\mu_0 = p_0$ is given. It is well known that we can derive a recursive filter formula with the help of Bayesian analysis which is given by

$$\mu_n(j) = \frac{\sum_{i=1}^d p_{ij}(h) g_{ij}(z, h) \mu_{n-1}(i)}{\sum_{i=1}^d g_i(z, h) \mu_{n-1}(i)} \tag{2.4}$$

where $g_i(z, h)$ is the density of the conditional distribution of Z_h given $\varepsilon_0 = i$, and $g_{ij}(z, h)$ is the density of the conditional distribution of Z_h given $\varepsilon_0 = i$ and $\varepsilon_h = j$, see, e.g., [9, Chapter 2], [11, Chapter 3] and [2, Chapter 5]. Obviously by conditioning we have the relation

$$g_i(z, h) = \sum_{j \in \mathcal{E}} g_{ij}(z, h) p_{ij}(h). \tag{2.5}$$

Thus, it is in principle enough to derive the densities g_{ij} from which we obtain g_i. In case $\mathcal{E}$ consists of two elements only, say $\mathcal{E} = \{1, 2\}$, we will derive explicit expressions for g_i and g_{ij} in the next section. For general $\mathcal{E}$ we derive approximate expressions in Section 4.

Let us also shortly mention the situation when we have continuous observations. In this case the observation filtration is given by

$$\mathcal{Z}_t := \sigma(\{Z_s, s \in [0, t]\}), \quad t \in [0, T]$$

and we have to determine

$$\mathbb{P}(\varepsilon_t = j | \mathcal{Z}_t), \quad t \in [0, T].$$

This is the well-known *Wonham filter problem*. From the *Kallianpur–Striebel formula* we have the following representation:

$$\mathbb{P}(\varepsilon_t = j | \mathcal{Z}_t) = \frac{\xi_j(t)}{\sum_{i=1}^d \xi_i(t)} \tag{2.6}$$

where $(\xi_t)_{t \in [0,T]} := (\xi_1(t), \dots, \xi_d(t))_{t \in [0,T]}$ satisfies the *Zakai equations*

$$\xi_t = p_0 + \int_0^t Q^\top \xi_s ds + \int_0^t D \xi_s dZ_s \tag{2.7}$$

for $t \in [0, T]$ where D is a diagonal matrix with elements $\alpha_1/\sigma^2, \dots, \alpha_d/\sigma^2$ on the diagonal (see, e.g., [9, 23]). Obviously, this filter makes use of a continuous observation of (Z_t). However, we will later see that approximations of this filter at time t may only depend on Z_t which allows us to use this filter in case of discrete observations, too.

Applications of this model are given next.

Example 2.1. Financial data suggest that parameters of asset prices depend on external macroeconomic factors which may be described by a continuous-time Markov chain (cf. [27]). A popular model for example is to take the classical Black–Scholes model with parameters driven by a factor process which is represented by a continuous-time Markov chain. Thus, the log-return of an asset price S_t at time t would be

$$\log S_t = \int_0^t \Big(\alpha_{\varepsilon_s} - \frac{1}{2} \sigma^2 \Big) ds + \sigma W_t$$

where (W_t) is a Brownian motion independent of (ε_t), α is the drift or appreciation rate and σ is a fixed volatility. The parameter σ may also depend on (ε_t). Models where the factor process is assumed to be known are among others treated in [1, 5, 7, 33, 34]. If (ε_t) is not observable and σ constant, we have exactly the generic situation which we consider in this paper. The model is then reasonably described by an HMM or Markov switching models for price parameters. The aim is to filter the underlying economic factor. For applications of this model see [3, 10, 14, 25, 28, 29]. In this case it is often only possible to observe the log-return, i.e. the integrated HMM at discrete time points. For example some asset prices are only quoted on a daily basis.

Example 2.2. In communication networks, information arrives to a multiplexer, switch or information processor at a rate which changes randomly and often shows a high degree of correlation in time. For example Heffes and Lucantoni [15] modeled the input stream of a statistical multiplexer consisting of a mix of data and packetized voice sources as a Markov-modulated Poisson process. Stern and Elwalid [31] used a Markov-modulated continuous flow model for the information stream at a multiplexer and analyzed the system performance. These and other findings fostered the investigation of so-called fluid queues, see, e.g., [20]. A more recent reference is [24] where Rabehasaina and Sericola are interested in the stationary queueing level of a fluid queue whose buffer content evolution process (Q_t) is given by the SDE

$$dQ_t = \alpha_{\varepsilon_t} dt - \mu_{\varepsilon_t} Q_t dt + \sigma_{\varepsilon_t} Q_t dW_t.$$

Here again, if one is able to observe, perturbed by an error, the amount of fluid which came in, then the filter computes the distribution of the current inflow rate. Such fluid queues can also be used to model production processes.

3 HMM problem with two states

3.1 Symmetric *Q*-matrix

In case the state space consists of two elements only, i.e. $\mathcal{E} = \{1, 2\}$, and the intensities are symmetric, i.e.

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}$$

with $\lambda > 0$, the model can be reduced to the so-called *telegraph process*. The telegraph process $(I_t)_{t\in[0,T]}$ is given by

$$I_t := \int_0^t (-1)^{N_s} ds \tag{3.1}$$

where $(N_t)_{t\in[0,T]}$ is a Poisson process with intensity λ. The telegraph process has been introduced by Kac in lecture notes from 1956, see also [18]. It has been generalized in different ways and studied in various applications. Its density function satisfies a hyperbolic second order differential equation which is known as the telegraph equation. For a recent paper with many references see [21]. One usually distinguishes the *symmetric* case, where the particle moves with symmetric velocities +1 and −1 and switches the direction with same intensity $\lambda > 0$, and the *asymmetric* case, where we have different velocities and different switching intensities.

Now let us denote

$$J_t := \int_0^t \alpha_{\varepsilon_s} ds \tag{3.2}$$

and define $a := \frac{1}{2}(\alpha_1 + \alpha_2)$, $b_1 = \frac{1}{2}(\alpha_1 - \alpha_2)$ and $b_2 = \frac{1}{2}(\alpha_2 - \alpha_1)$. Then, given $\varepsilon_0 = i$ for $i = 1, 2$ we have

$$J_t = at + b_i I_t, \quad t \in [0, T],$$

i.e. J_t is a linear transformation of I_t. The transition probabilities in (2.2) in this case simplify to

$$p_{12}(t) = p_{21}(t) = \frac{1}{2} - \frac{1}{2}e^{-2\lambda t}, \quad t \in [0, T].$$

It is now possible to derive explicit formulas for the densities $g_i, g_{ij}, i, j \in \{1, 2\}$, which appear in the filter formula (2.4). The density for the telegraph process I_h at time h given the initial state can already be found in the appendix of [7]. In order to derive the densities for our application, let us introduce the following notations. The modified Bessel functions are given by

$$B_0(z) = \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!^2}, \quad B_1(z) = \frac{z}{2}\sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!(k+1)!}, \quad z \geq 0 \tag{3.3}$$

and we denote the density of the normal distribution with zero expectation and standard deviation σ by

$$\phi_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}, \quad x \in \mathbb{R}.$$

Then we obtain the following theorem, the proof of which can be found in Appendix A.

Theorem 3.1. *The density of Z_h given $\varepsilon_0 = i$ for $i = 1, 2$ is*

$$g_i(z, h) = e^{-\lambda h}\Big[\phi_{\sigma\sqrt{h}}(z - h(b_i + a)) + \frac{1}{|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)} \phi_{\sigma\sqrt{h}}(z - x)\tilde{q}_{h,\lambda}\Big(\frac{x - ah}{b_i}\Big)dx\Big], \quad z \in \mathbb{R},$$

with

$$\tilde{q}_{h,\lambda}(x) = \frac{\lambda}{2}\Big[B_0\Big(\lambda\sqrt{(h^2 - x^2)}\Big) + \Big(\frac{h + x}{h - x}\Big)^{\frac{1}{2}} B_1\Big(\lambda\sqrt{(h^2 - x^2)}\Big)\Big], \quad x \in [-h, h].$$

The density of Z_h given $\varepsilon_0 = \varepsilon_h = i$ is

$$g_{ii}(z, h) = \frac{1}{\cosh(\lambda h)}\Big[\phi_{\sigma\sqrt{h}}(z - h(b_i + a)) + \frac{1}{|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)} \phi_{\sigma\sqrt{h}}(z - x)\tilde{q}_{h,+}\Big(\frac{x - ah}{b_i}\Big)dx\Big], \quad z \in \mathbb{R},$$

with

$$\tilde{q}_{h,+}(x) = \frac{\lambda}{2}\Big(\frac{h + x}{h - x}\Big)^{\frac{1}{2}} B_1\Big(\sqrt{(h^2 - x^2)}\lambda\Big), \quad x \in [-h, h].$$

The density of Z_h given $\varepsilon_0 = i, \varepsilon_h = j$ with $i, j \in \{1, 2\}$ and $i \neq j$ is

$$g_{ij}(z, h) = \frac{1}{\sinh(\lambda h)|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)} \phi_{\sigma\sqrt{h}}(z - x)\tilde{q}_{h,-}\Big(\frac{x - ah}{b_i}\Big)dx, \quad z \in \mathbb{R},$$

with

$$\tilde{q}_{h,-}(x) = \frac{\lambda}{2} B_0\Big(\sqrt{(h^2 - x^2)}\lambda\Big), \quad x \in [-h, h].$$

3.2 Non-symmetric *Q* matrix

In this section we again assume that the state space consists of two elements only, i.e. $\mathcal{E} = \{1, 2\}$, but the transition intensities are arbitrary, i.e. $q_{12} = \lambda_1 > 0$, $q_{21} = \lambda_2 > 0$. The transition probabilities in (2.2) are now given by

$$p_{12}(t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} - \frac{\lambda_1}{\lambda_1 + \lambda_2}e^{-(\lambda_1 + \lambda_2)t}, \quad p_{21}(t) = \frac{\lambda_2}{\lambda_1 + \lambda_2} - \frac{\lambda_2}{\lambda_1 + \lambda_2}e^{-(\lambda_1 + \lambda_2)t}. \tag{3.4}$$

In this case we interpret (I_t) in (3.1) as the asymmetric telegraph process. This means that

$$I_t := \int_0^t (-1)^{N_s} ds$$

where now

$$N_t = |\{0 \le s \le t : \varepsilon_s \neq \varepsilon_{s-}\}|$$

is the number of state transitions in $[0, t]$ of the Markov chain (ε_t). Again explicit formulas for the densities $g_i, g_{ij}, i, j \in \{1, 2\}$ can be derived. The density of I_h given $\varepsilon_0 = i$ has been obtained in [7] by using the Girsanov theorem for point processes. In [21] López and Ratanov use a differential equations approach. For our situation we get the following theorem, where we assume w.l.o.g. that the initial state is $\varepsilon_0 = 1$. The proof can be found in Appendix B.

Theorem 3.2. *The density of Z_h given $\varepsilon_0 = 1$ is*

$$g_1(z,h) = e^{-\lambda_1 h}\Big[\phi_{\sigma\sqrt{h}}(z - h(b_1 + a)) + \frac{1}{|b_1|}\int_{h(a-|b_1|)}^{h(a+|b_1|)} \phi_{\sigma\sqrt{h}}(z-x)\tilde{p}_h\Big(\frac{x-ah}{b_1}\Big)dx\Big], \quad z \in \mathbb{R}.$$

with

$$\tilde{p}_h(x) = \frac{\lambda_1}{2}\exp\Big(\frac{\lambda_1-\lambda_2}{2}(h-x)\Big)\Big[B_0\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big) + \Big(\frac{(h+x)\lambda_2}{(h-x)\lambda_1}\Big)^{\frac{1}{2}} B_1\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big)\Big], \quad x \in [-h, h].$$

The density of Z_h given $\varepsilon_0 = \varepsilon_h = 1$ is, for $z \in \mathbb{R}$,

$$g_{11}(z,h) = \frac{(\lambda_1+\lambda_2)}{\lambda_2 e^{\lambda_1 h} + \lambda_1 e^{-\lambda_2 h}}\Big[\phi_{\sigma\sqrt{h}}(z - h(b_1+a)) + \frac{1}{|b_1|}\int_{h(a-|b_1|)}^{h(a+|b_1|)} \phi_{\sigma\sqrt{h}}(z-x)\tilde{q}_{h,+,1}\Big(\frac{x-ah}{b_1}\Big)dx\Big],$$

with

$$\tilde{q}_{h,+,1}(x) = \frac{\lambda_1}{2}\exp\Big(\frac{\lambda_1-\lambda_2}{2}(h-x)\Big)\Big(\frac{(h+x)\lambda_2}{(h-x)\lambda_1}\Big)^{\frac{1}{2}} B_1\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big), \quad x \in [-h, h].$$

The density of Z_h given $\varepsilon_0 = 1, \varepsilon_h = 2$ is

$$g_{12}(z,h) = \frac{(\lambda_1+\lambda_2)}{e^{\lambda_1 h} - e^{-\lambda_2 h}} \cdot \frac{1}{|b_1|}\int_{h(a-|b_1|)}^{h(a+|b_1|)} \phi_{\sigma\sqrt{h}}(z-x)\tilde{q}_{h,-,1}\Big(\frac{x-ah}{b_1}\Big)dx, \quad z \in \mathbb{R},$$

with

$$\tilde{q}_{h,-,1}(x) = \frac{1}{2}\exp\Big(\frac{\lambda_1-\lambda_2}{2}(h-x)\Big)B_0\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big), \quad x \in [-h, h].$$

4 Approximate filter for more states

With the use of analytic solutions for partial differential equations (see Section 4.2 below and [21]), it is in principle possible to derive filters if $\mathcal{E}$ consists of more than two states. However, formulas get quite complicated, in particular when the state space is large. Instead we try to derive approximate filters in this case. The first approach is based on the continuous filter obtained by (2.6) and (2.7).

4.1 Approach via continuous filter (quasi-exact)

The solution of the Zakai equation (2.7) which is a homogeneous linear Itô SDE is discussed in [19, 23]. If the matrices Q and D (recall that D is the diagonal matrix with elements $\alpha_1/\sigma^2, \dots \alpha_d/\sigma^2$ on the diagonal) commute, i.e. if $QD = DQ$ then an explicit solution of (2.7) would be

$$\xi_t = \exp\Big(Q^\top t - \frac{1}{2}D^2 t + DZ_t\Big). \tag{4.1}$$

However this is not true if the matrices do not commute, but still the expression can serve as an approximate filter, see [19, 23]. Note that this solution only involves Z_t and not the path of the process and is thus also feasible for a filter with discrete observations. We will use (2.6) and (4.1) as one approximation for our discrete filter in Section 5. Since this approach relies on an assumption which is not valid in general, it is not possible

to show any kind of convergence of this approximation scheme. Alternatively, (ξ_t) can be approximated by a discrete process using the Milstein scheme in order to solve the SDE in (2.7) numerically:

$$\hat{\xi}_{n+1} = \Big[I + \Big(Q^\top - \frac{1}{2}D^2\Big)h + D\Big(I + \frac{1}{2}D\Big)(Z_{t_{n+1}} - Z_{t_n})\Big]\hat{\xi}_n. \tag{4.2}$$

Here $\hat{\xi}_n := \hat{\xi}_{t_n}$ where $t_n = nh$ as before. Again we see that this approximation only involves $(Z_{t_{n+1}} - Z_{t_n})$ which can be observed in our situation, too. Thus the approximate filter can also be used in the discrete setting. For (4.2) it is known that the continuous approximation $\hat{\xi}$ which is obtained from $\hat{\xi}_n$ converges with strong order 1 to the true solution of (2.7) if the time step size of the discretization tends to zero. Moreover, this convergence is also true for the corresponding filter (see [23, Theorem 4.1]). Hence for frequent observations, the approximation in (4.2) is supposed to work well in our setting. Note however, that a practical implementation of (4.2) requires some smoothing in order to avoid affects like $\hat{\xi}_n$ getting negative.

4.2 Approach via PDEs

Here we derive a system of PDEs for the densities g_{ij} which appear in the filter (2.4). From g_{ij} we obtain the densities g_i by (2.5). As soon as continuous-time Markov chains are involved it is a common tool to work with differential equations, see, e.g., the system of ODEs for the probability distribution of a Markov chain (2.1) or the partial differential equation for the position of a particle given by the telegraph process, see [18]. This equation is a hyperbolic second order differential equation and is known as the telegraph (or damped wave) equation. Partial differential equations for the densities of the telegraph process in the asymmetric case have been derived in [21].

In what follows define for $i, j \in \mathcal{E}$, $x \in \mathbb{R}$ and $t \geq 0$ the distribution

$$F_{ij}(x, t) = \mathbb{P}(J_t \leq x, \varepsilon_t = j | \varepsilon_0 = i)$$

and the corresponding density by $\tilde{f}_{ij}(x, t)$. By $f_{ij}(x, t)$ we denote the density of $\mathbb{P}(J_t \leq x | \varepsilon_0 = i, \varepsilon_t = j)$. Thus, we obviously have the relation

$$\tilde{f}_{ij}(x, t) = f_{ij}(x, t)p_{ij}(t). \tag{4.3}$$

From f_{ij} we can compute immediately by convolution the required density g_{ij} for our filter using (2.3). Note that the density $\tilde{f}_{ii}(x, t)$ has an atom at $x = \alpha_i t$ of size $e^{-q_i t}$ where $q_i = -q_{ii}$, since with probability $e^{-q_i t}$ no switch occurs in the time interval $[0, t]$ in which case $J_t = \alpha_i t$.

We will now derive a PDE for $\tilde{f}_{ij}$ by conditioning on what happens in the first time interval $[0, h]$. We obtain

$$F_{ij}(x, t) = (1 - q_i h + o(h))F_{ij}(x - \alpha_i h, t - h) + \int_0^h \sum_{k \neq j} q_i e^{-q_i s} p_{ik} F_{ik}(x - \alpha_i s, t - s)ds + o(h),$$

where p_{ik} is the transition probability of the embedded Markov chain given by

$$p_{ik} = \frac{q_{ik}}{q_i} \quad \text{for } i \neq k.$$

Rearranging terms and dividing by h yields

$$\begin{aligned}\frac{1}{h}(F_{ij}(x, t) - F_{ij}(x - \alpha_i h, t)) + \frac{1}{h}(F_{ij}(x - \alpha_i h, t) - F_{ij}(x - \alpha_i h, t - h)) \\ = -q_i F_{ij}(x, t) + \frac{1}{h}\int_0^h \sum_{k \neq i} q_i e^{-q_i s} p_{ik} F_{ik}(x - \alpha_i s, t - s)ds + \frac{o(h)}{h}.\end{aligned}$$

Now letting $h \downarrow 0$, we obtain

$$\alpha_i \frac{\partial}{\partial x} F_{ij}(x, t) + \frac{\partial}{\partial t} F_{ij}(x, t) = \sum_k q_{ik} F_{kj}(x, t)$$

(note that the limit on the right-hand side exists, hence also on the left-hand side). Applying the operator $\frac{\partial}{\partial x}$ on both sides, we derive the following equation for the densities:

$$\alpha_i \frac{\partial}{\partial x}\tilde{f}_{ij}(x,t) + \frac{\partial}{\partial t}\tilde{f}_{ij}(x,t) = \sum_k q_{ik}\tilde{f}_{kj}(x,t).$$

In matrix form this equations can be written as

$$D^\alpha \frac{\partial}{\partial x} G(x,t) + \frac{\partial}{\partial t} G(x,t) = QG(x,t) \tag{4.4}$$

where $G = (\tilde{f}_{ij})_{i,j\in\mathcal{E}}$ and D^α is a diagonal matrix with $\alpha_1, \dots, \alpha_n$ on the diagonal. Moreover, we have the boundary conditions $\tilde{f}_{ij}(x,0) = \delta_{ij}$. This system of PDEs can then be solved numerically. The domain of $\tilde{f}_{ij}(x,t)$ is given by $x \in [\min_i \alpha_i t, \max_i \alpha_i t]$ and $t > 0$. Outside the densities vanish. This approach is exact, however involves a numerical computation. One difficulty is the point mass which appears. However, the point mass is known and can be separated. More precisely, we know that $\tilde{f}_{ii}(x,t) = \delta_{\alpha_i t}(x)e^{-q_i t} + \hat{f}_{ii}(t,x)$ where $\hat{f}$ is smooth.

4.3 Approach via discretization

Here we use a very simple approximation of the densities involved in the filter by discretizing the continuous-time Markov chain. More precisely, when we consider $(\varepsilon_{m\cdot\Delta t})_m$ with $\Delta t = h/N$, then we obtain a discrete-time Markov chain with transition matrix $p(\Delta t) = e^{Q\Delta t}$. We can now approximate the random variable J_h in (3.2) by

$$\hat{J}_h := \Delta t \sum_{m=1}^{N} \alpha_{\varepsilon_{m\cdot\Delta t}}.$$

Since $\mathcal{E}$ is finite, the random variable $\hat{J}_h$ is obviously discrete and can only take a finite number of values. We denote by $\mathcal{D}$ the finite set of all possible values. Furthermore, we denote a multi-index $\mathbf{j}$ as a vector $\mathbf{j} = (j_1, \dots, j_N)$, where each $j_k \in \mathcal{E}$, i.e. $\mathbf{j} \in \mathcal{E}^N$. For $d \in \mathcal{D}$ denote

$$\Lambda(d) := \{\mathbf{j} \in \mathcal{E}^N : \alpha_{j_1} + \cdots + \alpha_{j_N} = d/\Delta t\}.$$

In what follows let $\mathbb{P}_i(\cdot) := \mathbb{P}(\cdot|\varepsilon_0 = i)$ and $\mathbb{P}_{ij}(\cdot) := \mathbb{P}(\cdot|\varepsilon_0 = i, \varepsilon_h = j)$ be the two conditional probability measures. Then we obtain

$$\mathbb{P}_i(\hat{J}_h = d) = \mathbb{P}_i\Big(\sum_{m=1}^{N} \alpha_{\varepsilon_{m\cdot\Delta t}} = d/\Delta t\Big) = \sum_{\mathbf{j}\in\Lambda(d)} \mathbb{P}_i(\varepsilon_{\Delta t} = j_1) \prod_{k=1}^{N-1} \mathbb{P}(\varepsilon_{(k+1)\Delta t} = j_{k+1}|\varepsilon_{k\Delta t} = j_k).$$

Thus, we can approximate the density $g_i(z,h)$ of the distribution of Z_h given $\varepsilon_0 = i$ by

$$\hat{g}_i(z,h) = \sum_{d\in\mathcal{D}} \mathbb{P}_i(\hat{J}_h = d)\phi_{\sigma\sqrt{h}}(z-d). \tag{4.5}$$

Analogously, for fixed $i, j \in \mathcal{E}$ and $d \in \mathcal{D}$ denote

$$\Lambda(j,d) := \{\mathbf{j} \in \mathcal{E}^{N-1} : \alpha_{j_1} + \cdots + \alpha_{j_{N-1}} = d/\Delta t - \alpha_j\}.$$

Then we obtain

$$\begin{aligned}\mathbb{P}_i(\hat{J}_h = d|\varepsilon_h = j) &= \mathbb{P}_i\Big(\sum_{m=1}^{N-1} \alpha_{\varepsilon_{m\cdot\Delta t}} = d/\Delta t - \alpha_j \;\Big|\; \varepsilon_h = j\Big) \\ &= \sum_{\mathbf{j}\in\Lambda(j,d)} \frac{\mathbb{P}_i(\varepsilon_{\Delta t} = j_1)}{\mathbb{P}_i(\varepsilon_h = j)} \prod_{k=1}^{N-2} \mathbb{P}(\varepsilon_{(k+1)\Delta t} = j_{k+1}|\varepsilon_{k\Delta t} = j_k) \cdot \mathbb{P}(\varepsilon_h = j|\varepsilon_{(N-1)\Delta t} = j_{N-1}).\end{aligned}$$

Thus, we can approximate the density $g_{ij}(z,h)$ of the distribution of Z_h given $\varepsilon_0 = i, \varepsilon_h = j$ by

$$\hat{g}_{ij}(z,h) = \sum_{d\in\mathcal{D}} \mathbb{P}_{i,j}(\hat{J}_h = d)\phi_{\sigma\sqrt{h}}(z-d). \tag{4.6}$$

Indeed for $\Delta t \to 0$ this density converges to the exact one, since on one hand by construction $\hat{J}_h \to J_h$, $\mathbb{P}$-a.s. for $\Delta t \to 0$, and thus it also converges weakly. On the other hand, the expression, e.g., in (4.5) can be written as $\mathbb{E}_i[\phi_{\sigma\sqrt{h}}(z - \hat{J}_h)]$ and since ϕ is continuous and bounded, the convergence follows immediately. The same is true for the expression in (4.6).

5 Numerical examples

This section illustrates the results presented in this work using some numerical examples. Basically, we consider two cases. First, we evaluate a scenario in which the underlying continuous time Markov chain has two states in order to provide an illustration comparing the exact filter to the approximations. Second, we consider a five state case. For application of the continuous filter (4.1), it is important to keep in mind, that the induced approximation error not only stems from the fact that we have discrete measurements, but also from the fact that the considered matrices D and Q do not commute.

All examples were implemented using Matlab R2014a on a laptop with an Intel i7-2620M CPU and 8GB RAM. The system of PDEs (4.4) was solved using the algorithm proposed by Skeel and Berzins [30], which is directly implemented in Matlab. From the solution $\tilde{f}_{ij}$, we derived $f_{ij}(x, t)$ using (4.3). Then, a numerical convolution procedure was performed for obtaining the values of the density $g_{ij}(z, h)$ on the grid points and $g_i(z, h)$ was obtained using (2.5). Cubic interpolation was used for evaluating these densities at other points.

5.1 Two-states example

We consider the scenario $\alpha_1 = -3$, $\alpha_2 = 1$, $\sigma = 1$, initial distribution of the continuous-time Markov chain $p_0 = (0.1,\ 0.9)$, and intensity matrix

$$Q = \begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}.$$

In Figure 1, we show a contour plot of the density $g_i(z, h)$ of Z_h conditioned on the value of $\varepsilon_0 = i$. In both cases time $h \in [0, 1]$ is plotted on the y-axis, where $z \in [-5, 5]$ is on the x-axis. In Figure 2, we show the contour plot of the density $g_{ij}(z, h)$ of Z_h conditioned on both, ε_0 and ε_h. It can be seen that the additional restriction of ε_h has quite some impact on the shape of the density. For evaluation of the filters, the observed process was generated using 100 discretization points and 50 observations in each time step, i.e. $h = \frac{1}{50}$. Both, the process and its observations are shown in Figure 3. More precisely, one curve shows the integrated continuous-time Markov chain $(\int_0^t \alpha_{\varepsilon_s} ds)$ (Int. CTMC), one curve the Brownian motion (W_t) and the third one (True Process) the combination $(Z_t) = (\int_0^t \alpha_{\varepsilon_s} ds + \sigma W_t)$.

In Figure 4, we show filter results in terms of the expectation value $\mu_n(1)\alpha_1 + \mu_n(2)\alpha_2$ at time nh of the true state. This is done for two scenarios. In the first case, every measurement is used. In the second case, only every fifth measurement is used. As expected, more frequent measurements result in a faster convergence of the estimate after a change of the underlying state. On the other hand, the filter with fewer observations is less sensitive to observed errors. The exact filter is almost indistinguishable from the PDE based approach. The time discretization based filter assumed the Markov process to perform at most one jump between each measurement, i.e., we used $N = 1$.

The computation times of the filters are given in Table 1. The high computation time for the exact filter is due to the need for performing numerical integration in each filter step. The numbers of the PDE based filter do not include the time required for numerically solving the PDE, because the resulting solution is reused in every filter step, and thus, the impact on the average computation time depends on the number of performed filter steps. The PDE was solved on an equidistant grid on $[-5, 5]$ consisting of 3000 discretization points for 2000 equidistantly distributed time-points on $[0, 1]$.

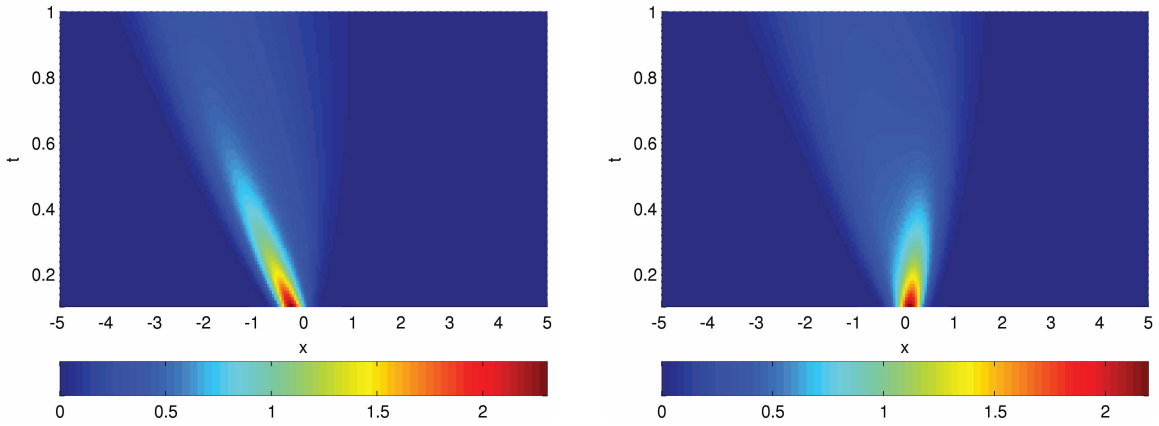
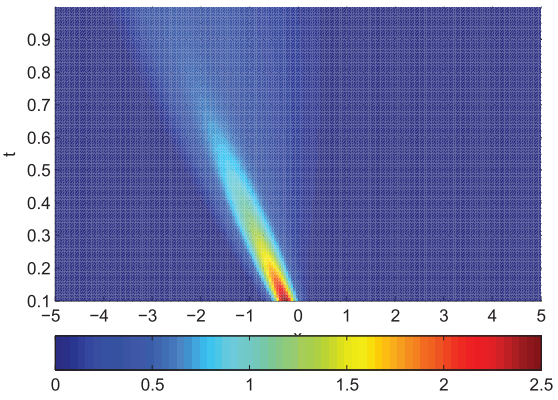
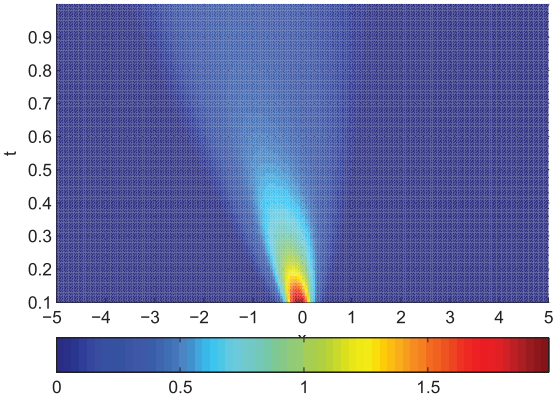


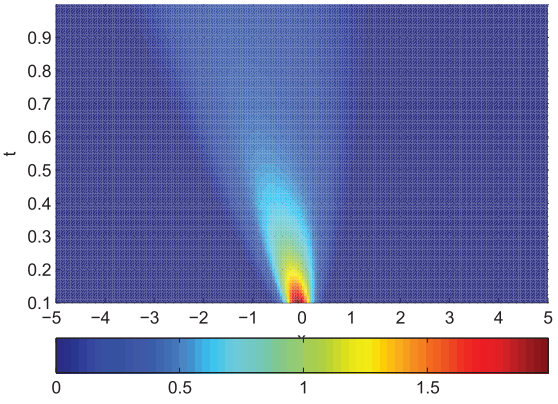
Figure 1. Contour plot of density $g_i(z, h)$ of Z_h conditioned on $\varepsilon_0 = i = 1$ (left) and $\varepsilon_0 = i = 2$ (right). Time is on the y-axis from 0 to 1.



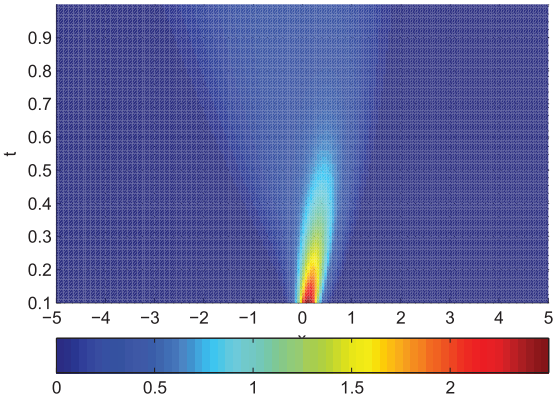
(a) $\varepsilon_0 = 1, \varepsilon_h = 1$



(b) $\varepsilon_0 = 1, \varepsilon_h = 2$



(c) $\varepsilon_0 = 2, \varepsilon_h = 1$



(d) $\varepsilon_0 = 2, \varepsilon_h = 2$

Figure 2. Contour plot of density $g_{ij}(z, h)$ of Z_t conditioned on $\varepsilon_0 = i$ and $\varepsilon_h = j$. Time is on the y-axis from 0 to 1.

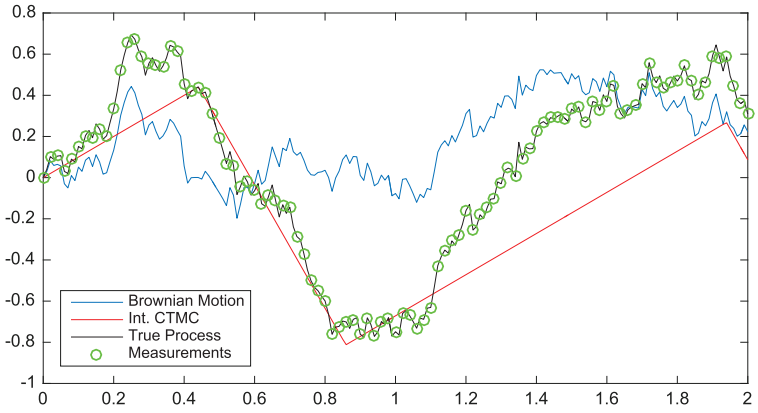
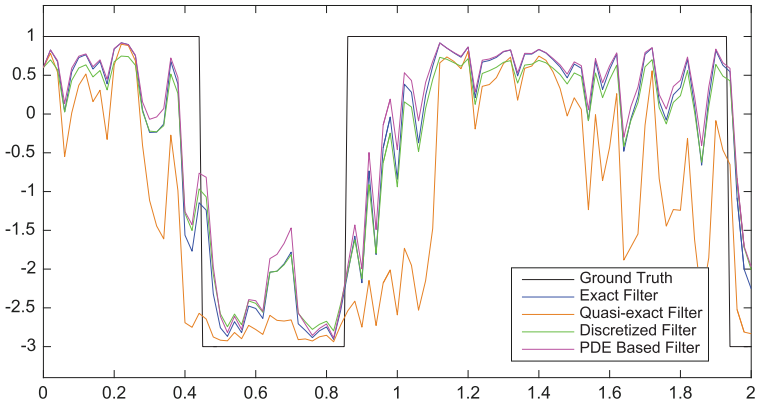
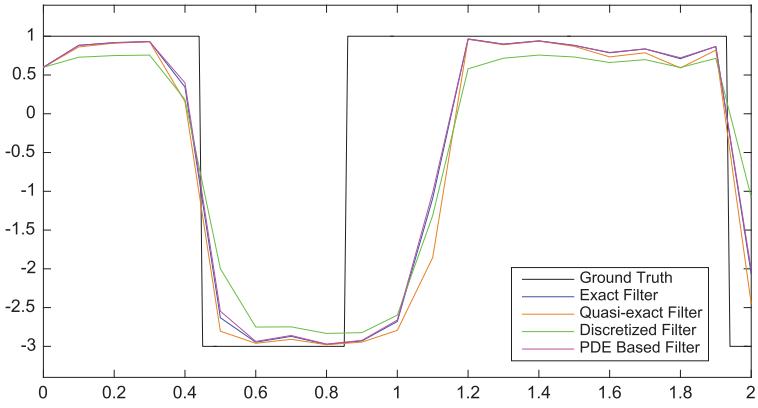


Figure 3. Observed process and its components (2-state case).



(a) Every measurement used



(b) Every fifth measurement used

Figure 4. True state of CTMC and its estimates (2-state case).

Filter	Time
Exact	248.53
Quasi-Exact	0.18
PDE Based	3.32
Discretized	4.73

(a) Two-states example

Filter	Time
Exact	—
Quasi-Exact	0.23
PDE Based	4.99
Discretized	171.48

(b) Five-states example

Table 1. Computation time (in ms) per filter step.

5.2 Five-states example

Now, we take a look at a more complex example where the α_i (for $i = 1, \dots, 5$) are given by the values -3, -1, 0, 1, 2 respectively. Furthermore, we used $p_0 = (0.1,\ 0.3,\ 0.3,\ 0.2,\ 0.1)$ and

$$Q = \begin{pmatrix} -1 & 0.5 & 0.3 & 0.1 & 0.1 \\ 0.4 & -1 & 0.3 & 0.1 & 0.2 \\ 0.1 & 0.1 & -1 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.3 & -1 & 0.5 \\ 0.1 & 0.1 & 0.3 & 0.5 & -1 \end{pmatrix}$$

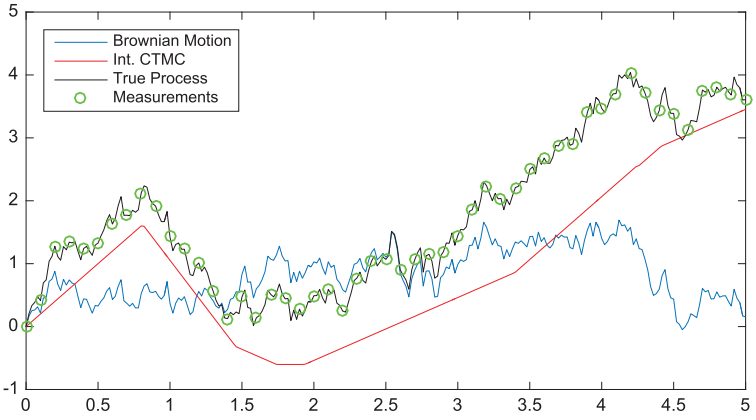
as initial distribution and intensity matrix. Two different diffusion parameters, $\sigma = 1$ and $\sigma = 2$, were considered in this example in order to observe the impact of σ on the quality of the filters. This time, we simulated five time steps with 50 discretization points and 10 observations per time step. The true process is shown in Figure 5. We use the same notation as in Figure 3.

The filter results are shown in Figure 6 (again by showing the expectation value $\sum_{i=1}^{5} \alpha_i \mu_n(i)$ at time nh of the obtained distributions). For a high number of measurements, the PDE based filter is almost indistinguishable from the discretized filter (which uses four discretization cells here, i.e., $N = 4$). This is due to the fact that the PDE based approach yields the exact density which is only corrupted by numerical errors induced from solving the PDE and by interpolation errors arising when evaluating the densities. Obviously, the expected number of jumps between two consecutive measurements decreases as the frequency of measurements goes up. Thus, in the considered scenario, the densities obtained from the discretization approach yield a good approximation of the exact density when every measurement is used. This approximation quality does not depend on the diffusion parameter σ. However, a larger σ results in less clear differences between all filtering approaches, because less information can be obtained about the underlying true state of the continuous time Markov chain.

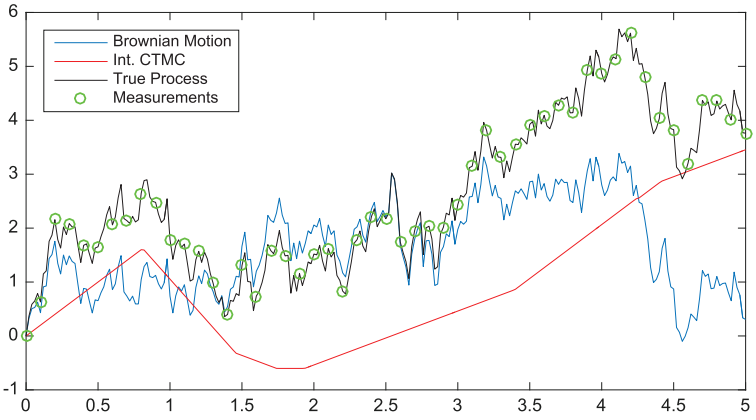
Looking at the computation time in Table 1 gives a somewhat different picture this time compared to the two-states case. Once again, we only show the computation times of the run using every measurement and $\sigma = 1$ (it does not differ significantly for other runs). This time, the discretized filter requires a significantly higher amount of computation time. This is due to its recursive implementation and use of a higher number of discretization cells. However, this could also be optimized by obtaining the set $\mathcal{D}$ and the densities $\hat{g}_i(z, h)$ and $\hat{g}_{ij}(z, h)$ (according to (4.5) and (4.6), respectively) before the actual filter run. In that case, evaluation of these densities comes down to evaluating a precomputed Gaussian mixture density.

5.3 Impact of parameters

Of course our filters depend on the model parameters. In particular, we assume that the different velocities α_i and the intensity matrix Q are known because of some physical constraints or because they have been estimated before. All three approximate filters use these data and depend on it in a complicated way. This can already been seen in the simple two-state case where we have analytic formulas and where the values of Q and α_i enter, e.g., the Bessel functions. The parameters which one may be able to control are the volatility of the Brownian motion σ and the distance h between observation time points. As mentioned before from Figure 6 we see that σ has a significant influence on all three filters since with higher σ it is more difficult to distinguish between states, but all approximate filters still behave in the same way. The size of h has to be seen relative to the size of the intensities q_{ii}. In order to be able to detect changes in the states we should have $h \ll (\sup_i |q_{ii}|)^{-1}$. We can see again from Figure 6 that a filter with smaller h is able to detect state changes faster whereas the filter with fewer observations is less sensitive to observed errors. But again this effect is similar for all three filters. We did one case study where we assumed that the parameters α, Q and p_0 have not been estimated correctly. Indeed we assume that the values of α, Q and p_0 in Section 5.1 are the true ones and that the observations belong to these model parameters. However, we estimated the parameters to

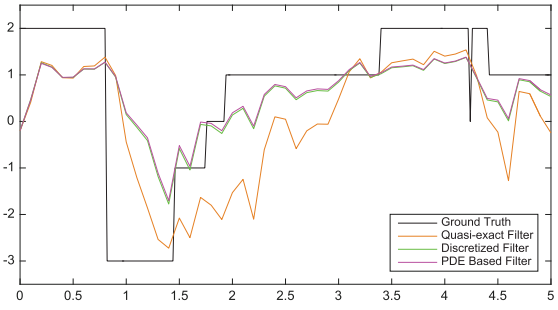


(a) $\sigma = 1$

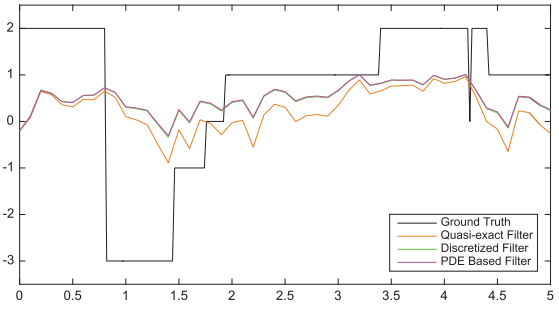


(b) $\sigma = 2$

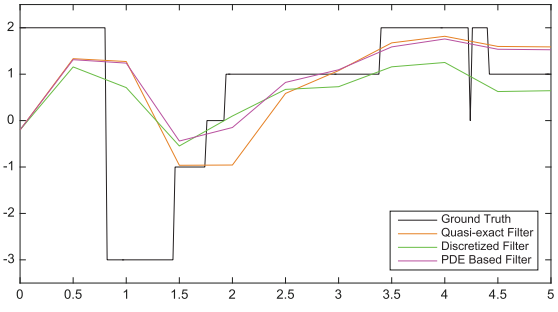
Figure 5. Observed process and its components (5-state case).



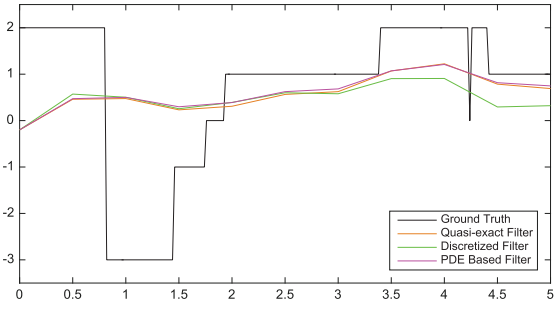
(a) $\sigma = 1$, every measurement used



(b) $\sigma = 2$, every measurement used

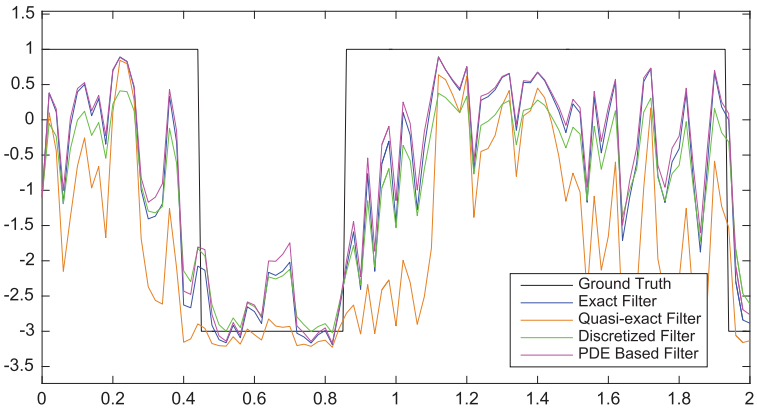


(c) $\sigma = 1$, every fifth measurement used

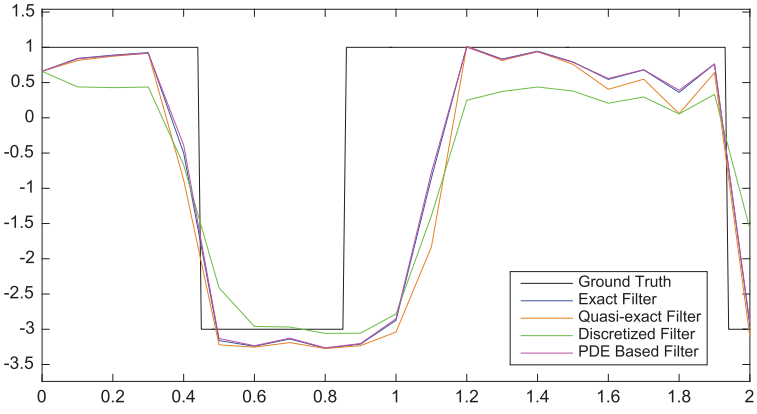


(d) $\sigma = 2$, every fifth measurement used

Figure 6. True state of CTMC and its estimates (5-state case).



(a) Every measurement used



(b) Every fifth measurement used

Figure 7. True state of CTMC and its estimates with *wrong* parameters (2-state case).

be $\alpha_1 = -3.3$, $\alpha_2 = 1.1$, $p_0 = (0.5,\ 0.5)$ and

$$Q = \begin{pmatrix} -2.5 & 2.5 \\ 7.8 & -7.8 \end{pmatrix}.$$

Using these values in our filters, we obtain the results in Figure 7 where we compare the filters with the true values. We can see that all filters are worse but the Quasi-Exact filter shows the worst performance. As discussed before, the Quasi-Exact filter contains an intrinsic error. Its performance depends on the distance of QD to DQ. When estimation errors occur, this difference might be larger than expected and contributes to a worse performance. In this case we do not recommend to use the Quasi-Exact filter.

6 Conclusion

In this paper we considered a Hidden Markov Model where the integrated continuous-time Markov chain can be observed at discrete time points perturbed by a Brownian motion. Using recent results on the asymmetric telegraph process, we derived exact formulas for a filter for the underlying continuous-time Markov chain, given this chain has only two different states. In case of more states we propose three approximate algorithms to compute the filter numerically. We investigated the performance of these filters with the help of two examples: one with two states and the other with five states. Problems like this arise for example when we have to filter the underlying economic state from observed asset prices.

A Proof of Theorem 3.1

The density of the integrated telegraph process I_h is given in [7]:

$$\mathbb{P}(I_h \in dx)/dx = e^{-\lambda h}\delta_h(x) + q_{h,\lambda}(x)$$

where $\delta_h(x)$ is the Dirac-measure in point h and

$$q_{h,\lambda}(x) = \frac{\lambda}{2}e^{-\lambda h}\Big[B_0\Big(\lambda\sqrt{(h^2-x^2)}\Big) + \Big(\frac{h+x}{h-x}\Big)^{\frac{1}{2}}B_1\Big(\lambda\sqrt{(h^2-x^2)}\Big)\Big], \quad x \in [-h,h]$$

where the modified Bessel functions have been defined in (3.3). The density of J_h conditioned on $\varepsilon_0 = i$ is then by the density transformation formula given by

$$\mathbb{P}(J_h \in dx|\varepsilon = i)/dx = e^{-\lambda h}\delta_{h(a+b_i)}(x) + \frac{1}{|b_i|}q_{h,\lambda}\Big(\frac{x-ah}{b_i}\Big).$$

Finally, the density of $Z_h = J_h + \sigma W_h$ given $\varepsilon_0 = i$ is obtained by the convolution formula

$$g_i(z,h) = e^{-\lambda h}\phi_{\sigma\sqrt{h}}(z - h(b_i+a)) + \frac{1}{|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)}\phi_{\sigma\sqrt{h}}(z-x)q_{h,\lambda}\Big(\frac{x-ah}{b_i}\Big)dx.$$

Thus, the first part of Theorem 3.1 is shown. Next we determine the density of I_h given N_h is even. From this we can compute g_{ii} as in the previous calculation. First note that

$$\begin{aligned}\mathbb{P}(I_h \in dx|N_h \in 2\mathbb{N}_0)/dx &= \frac{\mathbb{P}(I_h \in dx, N_h \in 2\mathbb{N}_0)/dx}{\mathbb{P}(N_h \in 2\mathbb{N}_0)}\\ &= \frac{\sum_{k=0}^{\infty}\mathbb{P}(I_h \in dx, N_h = 2k)/dx}{\mathbb{P}(N_h \in 2\mathbb{N}_0)}\\ &= \sum_{k=0}^{\infty}\mathbb{P}(I_h \in dx|N_h = 2k)/dx\frac{\mathbb{P}(N_h = 2k)}{\mathbb{P}(N_h \in 2\mathbb{N}_0)}.\end{aligned}$$

For $k \geq 1$ the density of I_h given $N_h = 2k$ is

$$\mathbb{P}(I_h \in dx|N_h = 2k)/dx = \frac{(2k)!}{(k-1)!k!}(h^2-x^2)^{k-1}\frac{h+x}{(2h)^{2k}}, \quad x \in [-h,h],$$

see [7, Lemma A.1]. Moreover, we have

$$\begin{aligned}\mathbb{P}(N_h = 2k) &= e^{-\lambda h}\frac{(\lambda h)^{2k}}{(2k)!},\\ \mathbb{P}(N_h \in 2\mathbb{N}_0) &= \sum_{k=0}^{\infty}e^{-\lambda h}\frac{(\lambda h)^{2k}}{(2k)!} = e^{-\lambda h}\cosh(\lambda h).\end{aligned}$$

For $k = 0$ the density of I_h given $N_h = 0$ has a point mass of one on h, i.e. $\mathbb{P}(I_h \in dx|N_h = 0)/dx = \delta_h(x)$ and $\mathbb{P}(N_h = 0) = e^{-\lambda h}$. Altogether we obtain

$$\begin{aligned}\mathbb{P}(I_h \in dx|N_h \in 2\mathbb{N}_0)/dx &= \delta_h(x)\cosh(\lambda h)^{-1} + \sum_{k=1}^{\infty}\mathbb{P}(I_h \in dx|N_h = 2k)/dx\frac{(\lambda h)^{2k}}{(2k)!\ \cosh(\lambda h)}\\ &= \cosh(\lambda h)^{-1}\delta_h(x) + q_{h,+}(x)\end{aligned}$$

where

$$q_{h,+}(x) = \cosh(\lambda h)^{-1}\frac{\lambda}{2}\Big(\frac{h+x}{h-x}\Big)^{\frac{1}{2}}B_1\Big(\sqrt{(h^2-x^2)}\lambda\Big), \quad x \in [-h,h].$$

Next we determine the density $q_{h,-}(x)$ of I_h given N_h is odd. For $k \geq 1$ the density of I_h given $N_h = 2k+1$ is

$$\mathbb{P}(I_h \in dx|N_h = 2k+1)/dx = \frac{(2k+1)!}{(k!)^2}\frac{(h^2-x^2)^k}{(2h)^{2k+1}}, \quad x \in [-h,h],$$

see [7, Lemma A.1]. Moreover, we have

$$\mathbb{P}(N_h = 2k+1) = e^{-\lambda h}\frac{(\lambda h)^{2k+1}}{(2k+1)!},$$
$$\mathbb{P}(N_h \in 2\mathbb{N}_0 + 1) = \sum_{k=0}^{\infty} e^{-\lambda h}\frac{(\lambda h)^{2k+1}}{(2k+1)!} = e^{-\lambda h}\sinh(\lambda h).$$

Altogether we obtain

$$\begin{aligned} q_{h,-}(x) &= \sum_{k=0}^{\infty} \mathbb{P}(I_h \in dx | N_h = 2k+1)/dx \frac{(\lambda h)^{2k+1}}{(2k+1)!\sinh(\lambda h)} \\ &= \sinh(\lambda h)^{-1}\frac{\lambda}{2} B_0\Big(\sqrt{(h^2 - x^2)}\lambda\Big), \quad x \in [-h, h]. \end{aligned}$$

Hence the density of J_h given $\varepsilon_0 = \varepsilon_h = i$ is again obtained by the density transformation formula

$$\mathbb{P}(J_h \in dx | \varepsilon_0 = \varepsilon_h = i)/dx = \cosh(\lambda h)^{-1}\delta_{h(a+b_i)}(x) + \frac{1}{|b_i|} q_{h,+}\Big(\frac{x - ah}{b_i}\Big),$$

and finally the density of $Z_h = J_h + \sigma W_h$ given $\varepsilon_0 = \varepsilon_h = i$ is

$$g_{ii}(z,h) = \phi_{\sigma\sqrt{h}}(z - h(b_i + a))\cosh(\lambda h)^{-1} + \frac{1}{|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)} \phi_{\sigma\sqrt{h}}(z-x) q_{h,+}\Big(\frac{x-ah}{b_i}\Big)dx.$$

Analogously we obtain the density of J_h given $\varepsilon_0 = i, \varepsilon_h = j$ with $i \neq j$ by

$$\mathbb{P}(J_h \in dx | \varepsilon_0 = i, \varepsilon_h = j)/dx = \frac{1}{|b_i|} q_{h,-}\Big(\frac{x - ah}{b_i}\Big)$$

and the density of Z_h given $\varepsilon_0 = i, \varepsilon_h = j$ is

$$g_{ij}(z,h) = \frac{1}{|b_i|}\int_{h(a-|b_i|)}^{h(a+|b_i|)} \phi_{\sigma\sqrt{h}}(z-x) q_{h,-}\Big(\frac{x-ah}{b_i}\Big)dx, \quad z \in \mathbb{R}.$$

B Proof of Theorem 3.2

According to [21], the density of the asymmetric telegraph process I_h given $\varepsilon_0 = 1$ is

$$\mathbb{P}(I_h \in dx | \varepsilon_0 = 1)/dx = e^{-\lambda_1 h}\delta_h(x) + p_h(x)$$

where $\delta_h(x)$ is the Dirac-measure in point h and

$$\begin{aligned} p_h(x) = \frac{\lambda_1}{2}\exp\Big(-\lambda_1\Big(\frac{x+h}{2}\Big) - \lambda_2\Big(\frac{h-x}{2}\Big)\Big) \\ \cdot\Big[B_0\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big) + \Big(\frac{(h+x)\lambda_2}{(h-x)\lambda_1}\Big)^{\frac{1}{2}} B_1\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big)\Big], \quad x \in [-h,h]. \end{aligned}$$

From this, the density of Z_h given $\varepsilon_0 = 1$ follows as in the previous proof.

Next we have to determine the density of Z_h given $\varepsilon_0 = \varepsilon_h = 1$. Note that we first have

$$g_{11}(z,h) = \mathbb{P}(Z_h \in dx | \varepsilon_0 = \varepsilon_h = 1)/dx = \frac{\mathbb{P}(Z_h \in dx, \varepsilon_h = 1 | \varepsilon_0 = 1)/dx}{\mathbb{P}(\varepsilon_h = 1 | \varepsilon_0 = 1)}.$$

From the first equation of (3.4) we obtain that

$$\mathbb{P}(\varepsilon_h = 1 | \varepsilon_0 = 1) = p_{11}(h) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)h}.$$

From the results in [21] we can derive $\mathbb{P}(I_h \in dx, \varepsilon_h = 1|\varepsilon_0 = 1)/dx$. We have that

$$\mathbb{P}(I_h \in dx, \varepsilon_h = 1|\varepsilon_0 = 1)/dx = e^{-\lambda_1 h}\delta_h(x) + q_{h,+,1}(x)$$

where

$$q_{h,+,1}(x) = \frac{1}{2}\exp\Big(-\lambda_1\Big(\frac{x+h}{2}\Big) - \lambda_2\Big(\frac{h-x}{2}\Big)\Big)\Big[\sqrt{\lambda_1\lambda_2}\Big(\frac{h+x}{h-x}\Big)^{\frac{1}{2}} B_1\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big)\Big], \quad x \in [-h, h].$$

The density $g_{11}(z, h)$ follows by joining the results.

We proceed similar for the density of Z_h given $\varepsilon_0 = 1, \varepsilon_h = 2$. Here we have

$$g_{12}(z, h) = \mathbb{P}(Z_h \in dx|\varepsilon_0 = 1, \varepsilon_h = 2)/dx = \frac{\mathbb{P}(Z_h \in dx, \varepsilon_h = 2|\varepsilon_0 = 1)/dx}{\mathbb{P}(\varepsilon_h = 2|\varepsilon_0 = 1)}.$$

From the first equation of (3.4) we obtain

$$\mathbb{P}(\varepsilon_h = 2|\varepsilon_0 = 1) = p_{12}(h) = \frac{\lambda_1}{\lambda_1+\lambda_2} - \frac{\lambda_1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)t}$$

and for the density of I_h and $\varepsilon_h = 2$ given $\varepsilon_0 = 1$ in the asymmetric case note that

$$\mathbb{P}(I_h \in dx, \varepsilon_h = 2|\varepsilon_0 = 1)/dx = \frac{\lambda_1}{2}\exp\Big(-\lambda_1\Big(\frac{x+h}{2}\Big) - \lambda_2\Big(\frac{h-x}{2}\Big)\Big) B_0\Big(\sqrt{\lambda_1\lambda_2(h^2-x^2)}\Big), \quad x \in [-h, h].$$

Using the usual transformation yields the statement.

Acknowledgment: We would like to thank our student Fehmi Mabrouk for some initial work on this topic during his Diploma Thesis which encouraged us to further pursue this topic.

References

[1] N. Bäuerle and U. Rieder, Portfolio optimization with Markov-modulated stock prices and interest rates, *IEEE Trans. Automat. Control* **49** (2004), 442–447.
[2] N. Bäuerle and U. Rieder, *Markov Decision Processes with Applications to Finance*, Springer, Berlin, 2011.
[3] N. Bäuerle, S. Urban and L. Veraart, The relaxed investor with partial information, *SIAM J. Financial Math.* **3** (2012), 304–327.
[4] E. Bayraktar and M. Ludkovski, Sequential tracking of a hidden Markov chain using point process observations, *Stochastic Process. Appl.* **119** (2009), 1792–1822.
[5] O. L. V. Costa and M. V. Araujo, A generalized multi-period mean-variance portfolio optimization with Markov switching parameters, *Automatica* **44** (2008), 2487–2497.
[6] J. Dannemann and H. Holzmann, Testing for two states in a hidden Markov model, *Canad. J. Statist.* **36** (2008), 505–520.
[7] G. B. Di Masi, Y. Kabanov and W. Runggaldier, Mean-square hedging of options on a stock with Markov volatilities, *Teor. Veroyatnost. i Primenen.* **39** (1994), 211–222.
[8] R. J. Elliott, V. Krishnamurthy and J. Sass, Moment based regression algorithm for drift and volatility estimation in continuous time Markov switching models, *Econometrics J.* **11** (2008), 244–270.
[9] R. J. Elliott, L. Aggoun and J. B. Moore, *Hidden Markov models*, Springer, New York, 1995.
[10] R. Frey, A. Gabih and R. Wunderlich, Portfolio optimization under partial information with expert opinions, *Int. J. Theor. Appl. Finance* **15** (2012), DOI 10.1142/S0219024911006486.
[11] B. Fristedt, N. Jain and N. Krylov, *Filtering and Prediction: A primer*, American Mathematical Society, Providence, 2007.
[12] S. Frühwirth-Schnatter, Fully Bayesian analysis of switching Gaussian state space models, *Ann. Inst. Statist. Math.* **53** (2001), 31–49.
[13] M. Hahn, S. Frühwirth-Schnatter and J. Sass, Markov chain Monte Carlo methods for parameter estimation in multidimensional continuous-time Markov switching models, *J. Financ. Econometrics* **8** (2010), 88–121.
[14] M. Hahn and J. Sass, Parameter estimation in continuous time Markov switching models: A semi-continuous Markov chain Monte Carlo approach, *Bayesian Anal.* **4** (2009), 63–84.
[15] H. Heffes and D. Lucantoni, A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, *IEEE J. Select. Areas Commun.* **4** (1986), 856–867.
[16] S. M. Iacus, Statistical analysis of the inhomogeneous telegrapher's process, *Statist. Probab. Lett.* **55** (2001), 83–88.

[17] S. M. Iacus and N. Yoshida, Estimation for the discretely observed telegraph process, *Teor. Ĭmovīr. Mat. Stat.* **78** (2008), 32–42.
[18] M. Kac, A stochastic model related to the telegrapher's equation, *Rocky Mountain J. Math.* **4** (1974), 497–509.
[19] P. E. Kloeden, E. Platen and H. Schurz, Higher order approximate Markov chain filters, in: *Stochastic Processes*, Springer, New York (1993), 181–190.
[20] V. G. Kulkarni, Fluid models for single buffer systems, in: *Frontiers in Queueing*, CRC, Boca Raton (1997), 321–338.
[21] O. López and N. Ratanov, On the asymmetric telegraph processes, *J. Appl. Probab.* **51** (2014), 569–589.
[22] E. Otranto and G. M. Gallo, A nonparametric Bayesian approach to detect the number of regimes in Markov switching models, *Econometric Rev.* **21** (2002), 477–496.
[23] E. Platen and R. Rendek, Quasi-exact approximation of hidden Markov chain filters, *Commun. Stoch. Anal.* **4** (2010), 129–142.
[24] L. Rabehasaina and B. Sericola, A second-order Markov-modulated fluid queue with linear service rate, *J. Appl. Probab.* **41** (2004), 758–777.
[25] U. Rieder and N. Bäuerle, Portfolio optimization with unobservable Markov-modulated drift process, *J. Appl. Probab.* **42** (2005), 362–378.
[26] T. Rydén, An EM algorithm for estimation in Markov-modulated Poisson processes, *Comput. Statist. Data Anal.* **21** (1996), 431–447.
[27] T. Rydén, T. Teräsvirta and S. Åsbrink, Stylized facts of daily return series and the hidden Markov models, *J. Appl. Econometrics* **13** (1998), 217–244.
[28] J. Sass and U. G. Haussmann, Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain, *Finance Stoch.* **8** (2004), 553–577.
[29] J. Sass and R. Wunderlich, Optimal portfolio policies under bounded expected loss and partial information, *Math. Methods Oper. Res.* **72** (2010), 25–61.
[30] R. D. Skeel and M. Berzins, A method for the spatial discretization of parabolic equations in one space variable, *SIAM J. Sci. Stat. Comput.* **11** (1990), 1–32.
[31] T. Stern and A. I. Elwalid, Analysis of separable Markov-modulated rate models for information-handling systems, *Adv. Appl. Probab.* **1** (1991), 105–139.
[32] Y.-C. Yao, Estimation of noisy telegraph processes: Nonlinear filtering versus nonlinear smoothing, *IEEE Trans. Inform. Theory* **31** (1985), 444–446.
[33] G. Yin and X. Y. Zhou, Markowitz's mean-variance portfolio selection with regime switching: From discrete-time models to their continuous-time limits, *IEEE Trans. Automat. Control* **49** (2004), 349–360.
[34] X. Zhang, T. Siu and Q. Meng, Portfolio selection in the enlarged Markovian regime-switching market, *SIAM J. Control Optim.* **48** (2010), 3368–3388.