Research article

Filtering of hidden Markov renewal processes by continuous and counting observations

Andrey Borisov*

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44/2 Vavilova Str., 119333 Moscow, Russia

* **Correspondence:** Email: aborisov@frccsc.ru; Tel: +74991356260; Fax: +74959304505.

Abstract: This paper introduces a subclass of Markov renewal processes (MRPs) and presents a solution to the optimal filtering problem in a stochastic observation system, where the state is modeled by an MRP and observed indirectly through noisy measurements. The MRPs considered here can be interpreted as continuous-time Markov chains (CTMCs) with a finite set of abstract states representing distributions of random vectors. The paper outlines the probabilistic properties of MRPs, emphasizing the ability to express any arbitrary function of the MRP as the solution to a linear stochastic differential system (SDS) with a martingale on the right-hand side (RHS). Using these properties, an optimal filtering problem is formulated in stochastic observation systems, where the hidden state belongs to the class of MRPs, and the observations consist of both diffusion and counting components. The drift terms in all observations depend on the system state. An optimal filtering estimate for a scalar function of the MRP is provided through the solution of an SDS with innovation processes on the RHS. Additionally, the paper presents a version of the Kushner-Stratonovich equation, describing the evolution of the conditional probability density function (PDF). To demonstrate the practical application of the estimation method, the paper presents a communications-related example, focusing on monitoring the qualitative state and numerical characteristics of a network channel using noisy observations of round-trip time (RTT) and packet loss flow. The paper also highlights the robustness of the filtering algorithm in scenarios where the MRP distribution is uncertain.

Keywords: hidden Markov model; Markov renewal process; martingale representation; optimal filtering problem; Kushner-Stratonovich equation
Mathematics Subject Classification: 60G35, 62M05, 94A40

1. Introduction

Markov processes [1, 2], and in particular Markov jump processes (MJPs), are widely used both as models of real-world phenomena and as tools for solving a broad range of analysis, estimation/identification, and control/optimization problems [3]. Their popularity stems from the simplicity of their probabilistic description, the efficiency of their mathematical framework for system analysis, and the potential for stochastic analysis under additional but non-restrictive assumptions. However, the need for more precise modeling of various jump-like phenomena leads to increasingly complex mathematical models, resulting in the emergence of multivariate point processes (MPPs) [4]. Markov renewal processes [5, 6] offer a middle ground between MJPs and MPPs. Any jump-like process can be expressed as a sequence of the pair "inter-arrival time – process value". For MRPs, this sequence forms a Markov chain. This allows for the relaxation of the Markov property in continuous time while still describing the process distribution via the transition kernel. However, this simplification comes at the cost of significantly complicating optimal estimation and control of MRPs. Specifically, the dynamics of a general MRP are not governed by a single finite-dimensional stochastic system, but rather by a sequence of such systems glued at random MRP transition times.

The goal of this paper is to introduce a subclass of MRPs of practical interest. This subclass is broader than finite-state CTMCs and shows promise as a model for various jump-like processes arising in fields such as tracking and navigation, telecommunications, financial mathematics, etc. The paper does not aim to describe the broadest class of MRPs for subsequent theoretical inferences. Instead, it focuses on MRPs that allow a straightforward "kinematic" description via the solution to a linear SDS with a martingale on the RHS. The advantage of this approach lies in the relative simplicity of filtering estimates, which, in turn, facilitates the design of efficient numerical filtering algorithms.

The paper is structured as follows: Section 2 introduces the subclass of MRPs under investigation, detailing their key probabilistic properties and showing how any arbitrary function of the MRP can be represented as the solution to a closed linear SDS with a martingale on the RHS.

Section 3 is dedicated to formulating the optimal filtering problem. The hidden system state is modeled by the studied MRP, while the available observations consist of both continuous diffusion and counting components. The drift coefficients of all observations are functions of the system state. The optimality criterion is the traditional mean squared error, meaning the filtering problem involves finding the conditional expectation (CE) of a signal process based on the available observations.

Section 4 forms the core of the paper and contains the solution to the filtering problem. The optimal filtering estimate of the signal process is expressed through the solution of a potentially non-closed SDS with innovation processes on the RHS. To address this non-closedness, we derive an analog of the Kushner-Stratonovich equation, which describes the evolution of the conditional PDF and aids in computing the CE for arbitrary functions of the MRP state.

Section 5 presents the results of an extensive numerical analysis of the proposed filter. A monitoring problem is considered, in which the qualitative state and numerical characteristics of communication channels are inferred from noisy observations of RTT and packet loss flows. The robustness of the filter against uncertainty in the MRP distribution is also demonstrated.

Section 6 concludes the paper with final remarks.

2. A sub-class of Markov renewal processes

First, we present an informal description of a subclass of Markov renewal processes (MRPs), which we later use as a hidden state in stochastic observation systems.

We introduce the process $Z_t \triangleq \mathrm{col}(\theta_t, Y_t) \in \mathbb{R}^{N+M}$, $t \in [0, T]$. The first N-dimensional component θ_t represents a CTMC with the state set $\mathbb{S}^N \triangleq \{e_1, \ldots, e_N\}$ (here, $\mathbb{S}^N$ stands for the set of the unit coordinate vectors in $\mathbb{R}^N$), initial distribution $p_0 \triangleq \mathrm{col}(p_0^1, \ldots, p_0^N)$, and the transition rate matrix (TRM) $\Lambda(t)$. The second compound component $Y_t \in \mathbb{R}^M$ also has piece-wise constant paths changing synchronously with θ_t. If $\{\tau_n\}_{n \in \mathbb{N}}$ is a sequence of θ_t transition instants and $\{\theta_{\tau_n}\}_{n \in \mathbb{N}}$ is known, then $\{Y_{\tau_n}\}_{n \in \mathbb{N}}$ is a sequence of mutually independent random vectors with the conditional distribution $\Pi(B) \triangleq \mathrm{col}(\Pi^1(B), \ldots, \Pi^N(B))$:

$$\mathcal{P}\{Y_{\tau_n} \in B | \theta_{[0,T]}\} = \int_B \theta_{\tau_n}^\top \Pi(dy) \quad \mathcal{P} - \text{a.s.} \quad \forall\, B \in \mathcal{B}(\mathbb{R}^M).$$

The distribution of the initial value Y_0 is defined via $\Phi(A) \triangleq \mathrm{col}(\Phi^1(A), \ldots, \Phi^N(A))$ quite analogously:

$$\mathcal{P}\{Y_0 \in B | \theta_{[0,T]}\} = \int_B \theta_0^\top \Phi(dy) \quad \mathcal{P} - \text{a.s.} \quad \forall\, B \in \mathcal{B}(\mathbb{R}^M).$$

One can interpret the sub-vector Y_t as a CTMC with the state set constituted by the random vectors with the distribution $\Pi(\cdot)$. There exists a Wiener–Poisson canonical space with filtration $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \in [0,T]})$ [7], such that the introduced process Z_t is properly defined on it and $\mathcal{F}_t$-adapted.

Further in the presentation we use the following notations:

- $\mathbf{1}$ is a row vector of an appropriate dimensionality formed by units.
- $\mathbf{0}$ is a zero row vector of an appropriate dimensionality.
- I is a unit matrix of an appropriate dimensionality.
- $\mathbf{I}_B(x)$ is an indicator function of the set B.
- N_t is the number of process θ_t transitions occurring on the interval $[0, t]$.
- $\mathcal{T}_n(s, t) \triangleq \mathcal{P}\{N_t - N_s = 0 | \theta_s = e_n\} = \exp\left(\int_s^t \Lambda_{nn}(u)du\right)$ $(0 \leqslant s < t)$ is a survival function of θ_t inter-arrival time under the condition $\theta_s = e_n$; $\mathcal{T}(s, t) \triangleq \mathrm{row}(\mathcal{T}_1(s, t), \ldots, \mathcal{T}_N(s, t))$.
- $\lambda(t) \triangleq \mathrm{row}(\Lambda_{11}(t), \ldots, \Lambda_{NN}(t))$ is a row-vector collecting the diagonal elements of TRM $\Lambda(t)$; $\widetilde{\Lambda}(t) \triangleq \Lambda(t) - \mathrm{diag}\, \lambda(t)$.
- $\mathcal{P}(s, t) = \|\mathcal{P}_{ij}(s, t)\|_{i,j=\overline{1,N}}$ is a transition matrix of the CTMC θ_t on the interval $[s, t]$ $(s \leqslant t)$: $\mathcal{P}_{ij}(s, t) \triangleq \mathcal{P}\{\theta_t = e_j \mid \theta_s = e_i\}$; $\mathcal{P}(s, t)$ is a solution to the Kolmogorov system: $\mathcal{P}'_t(s, t) = \mathcal{P}(s, t)\Lambda(t)$, $0 \leqslant s < t$, $\mathcal{P}(s, s) \equiv I$.
- Any function $f(e, y)$: $\mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}$ can be rewritten in the form $f(e, y) = \overline{f}(y)e$, where $\overline{f}(y) \triangleq \mathrm{row}(f(e_1, y), \ldots, f(e_N, y))$.
- $\mathsf{E}_f^n \triangleq \int_{\mathbb{R}^M} f(y, e_n)\Pi_n(dy)$, $\mathsf{E}_f = \mathrm{col}(\mathsf{E}_f^1, \ldots, \mathsf{E}_f^N) = \int_{\mathbb{R}^M} \mathrm{diag}\, \overline{f}(y)\Pi(dy)$.
- Any probability measure $\mathcal{Q}(\cdot)$ on $(\mathbb{S}^N \times \mathbb{R}^M, 2^{\mathbb{S}^N} \times \mathcal{B}(\mathbb{R}^M))$ can be expressed via the distribution $m^Q(B) = \mathrm{col}(m_1^Q(B), \ldots, m_N^Q(B))$, where $m_n(B) \triangleq \mathcal{Q}\{\theta = e_n, Y \in B\}$.

By definition, Z_t is Markovian on $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \in [0,T]})$ and has the following properties [2]:

1) The matrix-valued transition probability function $\mathsf{P}(y, s, B, t) = \|\mathsf{P}_{ij}(y, s, B, t)\|_{i,j=\overline{1,N}}$: $\mathsf{P}_{ij}(y, s, B, t) \triangleq \mathcal{P}\{\theta_t = e_j, Y_t \in B \mid \theta_s = e_i, Y_s = y\}$, $0 \leqslant s < t \leqslant T$ has the form

$$\mathsf{P}(y, s, B, t) = \mathbf{I}_B(y)\, \mathrm{diag}\, \mathcal{T}(s, t) + \mathrm{diag}\, \Pi(B)\, (\mathcal{P}(s, t) - \mathrm{diag}\, \mathcal{T}(s, t)). \tag{2.1}$$

The function $P(y, s, B, t)$ is a solution to the system

$$
\begin{cases}
\mathsf{P}'_t(y, s, B, t) = \mathsf{P}(y, s, B, t) \operatorname{diag} \lambda(t) + \operatorname{diag}(\Pi(B)) \, \mathcal{P}(s, t) \widetilde{\Lambda}(t), \\
\mathcal{P}'_t(s, t) = \mathcal{P}(s, t) \Lambda(t), \qquad 0 \leqslant s < t \leqslant T, \\
\mathsf{P}(y, s, B, s) = \mathbf{I}_B(y) I, \qquad \mathcal{P}(s, s) = I.
\end{cases} \tag{2.2}
$$

2) For an arbitrary bounded Borel function $f(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}$, the infinitesimal generator $\mathcal{A}_t$ of the process Z_t has the form

$$
\mathcal{A}_t f(e, y) \triangleq \lim_{h \downarrow 0} \frac{\mathsf{E}\{f(\theta_{t+h}, Y_{t+h}) | \theta_t = e, Y_t = y\} - f(e, y)}{h} = [\overline{f}(y) \operatorname{diag} \lambda(t) + \mathsf{E}_f^\top \widetilde{\Lambda}^\top(t)] e. \tag{2.3}
$$

3) For an arbitrary probability distribution $m(B) = \operatorname{col}(m_1(B), \dots, m_N(B))$ on $(\mathbb{S}^N \times \mathbb{R}^M, 2^{\mathbb{S}^N} \times \mathcal{B}(\mathbb{R}^M))$ the operator $\mathcal{A}_t^*$, adjoint to the generator $\mathcal{A}_t$, has the form

$$
\mathcal{A}_t^* m(B) = \operatorname{diag}(\lambda(t)) m(B) + \operatorname{diag}(\Pi(B)) \widetilde{\Lambda}^\top(t) \, m(\mathbb{R}^M). \tag{2.4}
$$

4) The distribution $P(t, B) = \operatorname{col}(P^1(t, B), \dots, P^N(t, B))$ of the process Z_t ($P^n(t, B) \triangleq \mathcal{P}\{\theta_t = e_n, Y_t \in B\}$) is a solution to the system

$$
P'_t(t, B) = \operatorname{diag}(\lambda(t)) P(t, B) + \operatorname{diag}(\Pi(B)) \widetilde{\Lambda}^\top(t) \, P(t, \mathbb{R}^M), \quad P(0, B) = \operatorname{diag}(\phi(B)) p_0. \tag{2.5}
$$

Formally, the compound component Y_t of the process Z_t represents an MRP. The MJP θ_t serves as a subordinator: it defines the transition times $\{\tau_n\}$ for Y_t and selects the distributions $\Pi^n(\cdot)$ for the values of Y_{τ_n}. In general, the MRP Y_t is non-Markovian. It exhibits the Markov property only when the support sets of the distributions $\Pi^n(\cdot)$ are disjointed [8]. Below in the presentation, the components θ_t and Y_t act as the hidden states in stochastic observation systems aimed at estimating these states. Therefore, we combine them into Z_t and refer to it as the MRP.

The proposed class of MRPs is a promising tool for modeling real-world phenomena characterized by a combination of "qualitative state" and "numerical characteristics". Let us illustrate this concept.

First, consider the motion of a maneuvering target [9, 10]. As is well-known, maneuvers are performed with various accelerations, which we assume to follow a piecewise constant process [11]. Typically, there are only a few types of maneuvers, such as:

– Nearly uniform rectilinear motion.
– Uniformly accelerated or decelerated rectilinear motion.
– Turning.

Thus, the process θ_t of maneuver changes has three possible states and can be modeled as an MJP [12]. Note, that the numerical parameters Y_t of a particular maneuver can vary in each instance. These parameters are described by a three-dimensional vector, including tangential acceleration and two components of normal acceleration.

Second, consider the price evolution of a risky asset. It is assumed that the financial market is governed by a hidden regime-switching process θ_t [13, 14]. There are relatively few possible market regimes [15, 16], making it natural to describe them using a finite-state MJP [17]. The local numerical characteristics of the market Y_t, including the instantaneous interest rate and volatility, are random and follow specific probability distributions for each regime.

The third example of applying the proposed class of MRPs to mathematical modeling relates to telecommunications and is detailed in Section 5.

Let us introduce a process $f(Z_t)$ that is a scalar function of Z_t: $f = f(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}$. It is easy to verify that $f(Z_t)$ can be expressed as a linear function of the process $\mathsf{F}_t \triangleq \mathrm{col}(\theta_t, f(Z_t)\theta_t) \in \mathbb{R}^{2N}$: $f(Z_t) = \mathrm{row}(\mathbf{0}, \mathbf{1})\mathsf{F}_t$. The process F_t is called the process associated with $f(Z_t)$. Below in the presentation we reserve the notation f_t for the second N-dimensional sub-vector of F_t: $\mathsf{f}_t \triangleq f(Z_t)\theta_t$.

From the Dynkin formula [2], it follows the martingale representation of an arbitrary function of Z_t. If $f = f(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}$ is the function, such that

$$\int_{\mathbb{R}^M} \|\bar{f}(y)\|^2 (\Pi(dy) + \Phi(dy)) < \infty, \tag{2.6}$$

then the process F_t has a finite second moment and represents the unique strong solution to the linear SDS

$$\mathsf{F}_t = \mathsf{F}_0 + \int_0^t D^f(s)\mathsf{F}_{s-}ds + \mu_t^f, \tag{2.7}$$

where $D^f(t)$ is $2N \times 2N$-dimensional matrix-valued non-random function

$$D^f(t) \triangleq \begin{bmatrix} \Lambda^\top(t) & 0 \\ \mathrm{diag}\, \mathsf{E}_f \widetilde{\Lambda}^\top(t) & \mathrm{diag}\, \lambda(t) \end{bmatrix}$$

and $\mu_t^f \in \mathbb{R}^{2N}$ is an $\mathcal{F}_t$-adapted square integrable martingale.

If the MRP Z_t satisfies the condition $\int_{\mathbb{R}^M} \|y\|^2 (\Pi(dy) + \Phi(dy)) < \infty$, then the process $\mathsf{Z}_t \triangleq \mathrm{col}(\theta_t, Y_t^1 \theta_t, \ldots, Y_t^M \theta_t) \in \mathbb{R}^{(M+1)N}$, associated with Z_t, is a solution to the linear SDS

$$\mathsf{Z}_t = \mathsf{Z}_0 + \int_0^t D^Z(s)\mathsf{Z}_s ds + \mu_t^Z, \tag{2.8}$$

where $D^Z(t)$ is $(M+1)N \times (M+1)N$-dimensional matrix-valued non-random function

$$D^Z(t) \triangleq \begin{bmatrix} \Lambda^\top(t) & 0 & 0 & \ldots & 0 \\ \mathrm{diag}\, \mathsf{E}_{Y^1}\widetilde{\Lambda}^\top(t) & \mathrm{diag}\, \lambda(t) & 0 & \ldots & 0 \\ \mathrm{diag}\, \mathsf{E}_{Y^2}\widetilde{\Lambda}^\top(t) & 0 & \mathrm{diag}\, \lambda(t) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathrm{diag}\, \mathsf{E}_{Y^M}\widetilde{\Lambda}^\top(t) & 0 & 0 & \ldots & \mathrm{diag}\, \lambda(t) \end{bmatrix}$$

and $\mu_t^Z \in \mathbb{R}^{(M+1)N}$ is an $\mathcal{F}_t$-adapted square integrable martingale.

The generalized Itô rule allows us to derive a formula for the mutual quadratic characteristic of the functions of the MPRs. If $g = g(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}$ is another function, satisfying (2.6) and the process $\mathsf{G}_t \triangleq \mathrm{col}(\theta_t, g(Z_t)\theta_t) \in \mathbb{R}^{2N}$, associated with $g(Z_t)$, admits the martingale representation similar to (2.7) with a martingale μ_t^g, then the mutual quadratic characteristic of f_t and g_t components has the form

$$\langle \mathsf{f}, \mathsf{g} \rangle_t = \int_0^t \Big[\mathrm{diag}\, \mathsf{E}_{fg}\, \mathrm{diag}(\widetilde{\Lambda}^\top(s)\theta_s) - \mathrm{diag}(\mathsf{f}_s)\, \mathrm{diag}\, \lambda(s)\, \mathrm{diag}(\mathsf{g}_s) \tag{2.9}$$
$$- \mathrm{diag}(\mathsf{f}_s)\widetilde{\Lambda}(s)\, \mathrm{diag}\, \mathsf{E}_g - \mathrm{diag}\, \mathsf{E}_f \widetilde{\Lambda}^\top(s)\, \mathrm{diag}(\mathsf{g}_s) \Big] ds.$$

The class of MRPs under study is, of course, not as general as the one proposed in [18]. Nevertheless, the fact that an arbitrary function of the studied MRP can be expressed through the solution of a linear SDS has the potential to simplify the form of the optimal filtering equations. Moreover, it could lead to a successful solution of conditionally-optimal filtering problems, such as optimal polynomial filtering. Additionally, the adjoint operator $\mathcal{A}^*$ is sufficiently simple, allowing the temporal evolution of the process distribution to be described by the functions $\mathcal{T}(s, t)$ and $\mathcal{P}(s, t)$. This provides a relatively simple analog to the Kushner-Stratonovich equation, which describes the evolution of the conditional distribution of the state process. In the case of a diffusion observation system, the Kushner-Stratonovich equation falls within the class of stochastic partial integro-differential equations. However, in the present case, it is expected that the Kushner-Stratonovich equation will take the form of a stochastic integral equation. This simplified form suggests the potential for various efficient algorithms to solve it numerically.

3. Problem statement

On the Wiener–Poisson basis $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}\}_{t \in [0,T]})$, we consider the observation system

$$
\begin{aligned}
\mathsf{Z}_t &= \mathsf{Z}_0 + \int_0^t D^Z(s)\mathsf{Z}_s ds + \mu_t^Z, & q_t &= q(\mathsf{Z}_t) = \overline{q}(Y_t)\theta_t, \\
\xi_t &= \int_0^t g(\mathsf{Z}_s)ds + \int_0^t R^{1/2}dw_s, & \eta_t &= \int_0^t h(\mathsf{Z}_s)ds + \mu_t^\eta.
\end{aligned}
\tag{3.1}
$$

Here,

- $\mathsf{Z}_t \in \mathbb{R}^{(M+1)N}$ is a system state, which is the process associated with the MRP $Z_t \in \mathbb{R}^{N+M}$; it is described by the martingale representation (2.8).
- $q_t \in \mathbb{R}$ is an estimated scalar process, which is a function of the MRP Z_t.
- $\xi_t \in \mathbb{R}^K$ is a continuous observation process.
- $\eta_t \in \mathbb{R}^L$ is an observable process with counting components.

In the observation system (3.1)

- $\mu_t^Z \in \mathbb{R}^{(M+1)N}$ and $\mu_t^\eta \in \mathbb{R}^L$ are $\mathcal{F}_t$-adapted martingales from the representations of the MRP Z_t and counting observations η_t.
- $g(\mathsf{Z}_t) = g(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}^K$ and $h(\mathsf{Z}_t) = h(e, y) : \mathbb{S}^N \times \mathbb{R}^M \to \mathbb{R}^L$ constitute the "useful signals", governed by the estimated state, in the observations.
- $w_t \in \mathbb{R}^K$ is an $\mathcal{F}_t$-adapted standard Wiener process, characterizing the noise in the continuous observations ξ_t; the matrix-valued function R denotes the noise intensity in the observations ξ_t (here, $R^{1/2}$ stands for the symmetric square root of the non-negative square matrix R).

For the proper formulation of the optimal filtering problem, we should make the following assumptions concerning the observation system (3.1):

A1. In the Wiener-Poisson basis $\mathcal{F}_t \equiv \sigma\{Z_s, w_s, \eta_s : 0 \leqslant s \leqslant t\}$.
A2. All trajectories of Z_t and η_t are càdlàg functions.
A3. The TRM $\Lambda(t)$ consists of the piecewise continuous components.

A4. For any $n = \overline{1, N}$ the functions $g(e_n, y)$, $h(e_n, y)$, and $q(e_n, y)$ are continuous in y. There exists a constant C_1 such that $\sum_{n=1}^{N} \int_{\mathbb{R}^M} \left(\|g(e_n, y)\|^2 + \|h(e_n, y)\|^2 + q^2(e_n, y) \right) (\Pi(dy) + \Phi(dy)) \leqslant C_1 < \infty$.

A5. There exists a constant $C_2 > 0$ such that $\min_{e_n \in \mathbb{S}^N} h(e_n, y) > C_2$ and $R \geqslant C_2 I$.

A6. The components of the martingale μ_t^η are orthogonal to each other, i.e.,

$$\langle \mu^\eta, \mu^\eta \rangle_t \equiv \int_0^t \operatorname{diag} h(s, Z_s) ds. \tag{3.2}$$

A7. The purely discontinuous martingales μ^Z and μ^η are strongly orthogonal, i.e., $\langle \mu^Z, \mu^\eta \rangle_t \equiv 0$.

Let $O_t \triangleq \{\xi_s, \eta_s : 0 \leqslant s \leqslant t\}$ be a natural filtration generated by the observations available up to time t. The optimal filtering problem for the signal process q_t is to find $\widehat{q}_t \triangleq \mathsf{E}\{q_t | O_t\}$.

Conditions A1 and A2 ensure the correct application of the stochastic analysis framework [19], and Eq (3.1) represents a properly defined SDS. Condition A1 also implies that all randomness in the canonical space is generated solely by the random processes in (3.1).

Condition A3 guarantees that the transition matrix $\mathcal{P}(s, t)$ of the CTMC θ_t is the solution to the Kolmogorov differential system.

Condition A4 assures that the state Z_t, observations ξ_t, η_t, and the estimated process q_t have finite moments up to the second order; hence, the CE $\widehat{q}_t$ is optimal in the mean square sense.

Condition A5 is standard for the filtering problem. It means the uniform non-degeneracy of both the observation noise in the continuous observations ξ_t and intensity of the counting observations η_t. The condition also guarantees the legitimacy of the Girsanov change of measure.

Any observable counting process can be factorized into the process with the orthogonal components, satisfying Condition A6; the mutual jumps in various components of the "genuine" counting observations could be separated into the new counting observable processes.

Finally, Conditions A1–A6 look mainly technical but are non-restrictive in practice.

4. Solution to filtering problem

Below in the presentation we use the notation $\widehat{c}_t \triangleq \mathsf{E}\{c(Z_t) | O_t\}$ for any function $c(e, y)$, for which $\sum_{n=1}^{N} \int_{\mathbb{R}^m} \|c(e_n, y)\|^2 (\Pi(dy) + \Phi(dy)) < +\infty$, and $\widehat{c}_{t-} \triangleq \lim_{s \uparrow t} \widehat{c}_s$.

Theorem 4.1. *The optimal filtering estimate $\widehat{q}_t$ of the signal process $q(Z_t)$ has the form $\widehat{q}_t = \operatorname{row}(0, 1)\widehat{\mathsf{Q}}_t$, where $\widehat{\mathsf{Q}}_t = \mathsf{E}\{\operatorname{col}(\theta_t, q(Z_t)\theta_t) | O_t\}$ is a filtering estimate of the process Q_t associated with $q(Z_t)$; $\widehat{\mathsf{Q}}_t$ is a solution to the SDS*

$$\widehat{\mathsf{Q}}_t = \mathsf{E}\{\mathsf{Q}_0\} + \int_0^t D^Q(s)\widehat{\mathsf{Q}}_{s-} ds + \int_0^t \left(\widehat{\mathsf{Q}g^\top}_{s-} - \widehat{\mathsf{Q}}_{s-}\widehat{g^\top}_{s-} \right) R^{-1/2} d\nu_s + \int_0^t \left(\widehat{\mathsf{Q}h^\top}_{s-} - \widehat{\mathsf{Q}}_{s-}\widehat{h^\top}_{s-} \right) \operatorname{diag}^{-1}(\widehat{h}_{s-}) d\zeta_s, \tag{4.1}$$

where

$$\nu_t = R^{-1/2} \int_0^t (d\xi_s - \widehat{g}_{s-} ds) \tag{4.2}$$

and

$$\zeta_t = \int_0^t (d\eta_s - \widehat{h}_{s-} ds) \tag{4.3}$$

are the innovation processes.

Proof of Theorem 4.1 is given in Appendix A.

Remark 4.1. *In the case of pure counting observations, the possibility of describing the optimal filtering estimate using Eq (4.1) has been proven in a recent and novel paper [18] for the general class of MRPs.*

One can see that Eq (4.1), which describes the evolution of the estimate $\widehat{Q}_t$, is not a closed-form equation. Besides the estimate itself, the RHS includes the CEs of the observation drifts and their products with the estimated process. The question of whether or not an optimal filtering estimate is a solution to a finite-dimensional system has its history [20–22]. The general answer sounds rather pessimistic: the optimal filters have a finite-dimensional form only for some narrow class of the observation systems [23–25].

An alternative approach to solving the optimal filtering problem involves finding the conditional distribution of the state Z_t. Let $\widehat{\psi}(t, y) \triangleq \mathrm{col}(\widehat{\psi}^1(t, y), \ldots, \widehat{\psi}^N(t, y))$ denote the conditional PDF, defined such that the identity $\mathcal{P}\{\theta_t = e_n, Y_t \in B \mid O_t\} \equiv \int_B \widehat{\psi}^n(t, y))dy$ holds $\mathcal{P}$-a.s. for all $t \in [0; T]$, $n = \overline{1, N}$ and $B \in \mathcal{B}(\mathbb{R}^M)$. The optimal filtering estimate $\widehat{q}_t = \mathsf{E}\{q(Z_t)|O_t\}$ takes the form

$$\widehat{q}_t = \sum_{n=1}^{N} \int_{\mathbb{R}^M} q(e_n, y)\widehat{\psi}^n(t, y)dy = \int_{\mathbb{R}^M} \overline{q}(y)\widehat{\psi}(t, y)dy.$$

If the observation system is described by the diffusion processes, then under certain additional assumptions, the conditional PDF exists and satisfies the Kushner-Stratonovich equation [26]. We impose an additional assumption to derive an analogous equation for the observation systems in (3.1).

A8. The distributions $\Pi(\cdot)$ and $\Phi(\cdot)$ are absolutely continuous with respect to the Lebesgue measure. The corresponding vector-valued PDFs are $\pi(y) = \mathrm{col}(\pi^1(y), \ldots, \pi^N(y))$ and $\phi(y) = \mathrm{col}(\phi^1(y), \ldots, \phi^N(y))$.

Condition A8 ensures that the distribution $P(t, B) = \mathrm{col}(P^1(t, B), \ldots, P^N(t, B))$ of the process Z_t has a PDF $\psi(t, y) = \mathrm{col}(\psi^1(t, y), \ldots, \psi^N(t, y))$. This PDF satisfies the following Kolmogorov system

$$\psi'_t(t, y) = \mathrm{diag}(\lambda(t))\psi(t, y) + \mathrm{diag}(\pi(y))\,\widetilde{\Lambda}^\top(t) \int_{\mathbb{R}^M} \psi(t, u)du, \quad \psi(0, y) = \mathrm{diag}(p_0)\,\phi(y).$$

Theorem 4.2. *Under conditions A1–A8, the evolution of $\widehat{\psi}(t, y)$ is described by the system*

$$\widehat{\psi}(t, y) = \mathrm{diag}(p_0)\phi(y) + \int_0^t \left[\mathrm{diag}\,\lambda(s)\widehat{\psi}(s-, y) + \mathrm{diag}\,\pi(y)\widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-}\right] ds \qquad (4.4)$$

$$+ \int_0^t \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)\left(\overline{g}(y) - \widehat{g}_{s-}\mathbf{1}\right)^\top R^{-1/2}d\nu_s$$

$$+ \int_0^t \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)\left(\overline{h}(y) - \widehat{h}_{s-}\mathbf{1}\right)^\top \mathrm{diag}^{-1}(\widehat{h}_{s-})d\zeta_s,$$

where $\widehat{\theta}_s$, $\widehat{g}_s$ and $\widehat{h}_{s-}$ are expressed via $\widehat{\psi}(s, y)$:

$$\widehat{\theta}_s = \int_{\mathbb{R}^M} \widehat{\psi}(s, y)dy, \qquad \widehat{g}_s = \int_{\mathbb{R}^M} \overline{g}(y)\widehat{\psi}(s, y)dy, \qquad \widehat{h}_{s-} = \int_{\mathbb{R}^M} \overline{h}(y)\widehat{\psi}(s-, y)dy, \qquad (4.5)$$

and $\overline{g}(y)$ and $\overline{h}(y)$ are the matrix-valued functions

$$\overline{g}(y) \triangleq \begin{bmatrix} g^1(e_1, y) & \dots & g^1(e_N, y) \\ \dots & \dots & \dots \\ g^K(e_1, y) & \dots & g^K(e_N, y) \end{bmatrix}, \quad \overline{h}(y) \triangleq \begin{bmatrix} h^1(e_1, y) & \dots & h^1(e_N, y) \\ \dots & \dots & \dots \\ h^L(e_1, y) & \dots & h^L(e_N, y) \end{bmatrix}.$$

Proof of Theorem 4.2 is given in Appendix B.

5. Numerical study

5.1. Nominal filtering performance

To demonstrate the performance of the proposed filtering estimate, we present an applied problem in the field of communications. The problem involves monitoring the qualitative state and numerical characteristics of heterogeneous (wired/wireless) communication channels, which are hidden from direct observation, with only indirect noisy observations available.

The "channel health" is fully described by the MRP $Z_t \triangleq \operatorname{col}(\theta_t, Y_t)$. The first component, θ_t, which represents the qualitative state, is a CTMC with values in $\mathbb{S}^4$:

- $\theta_t = e_1$: moderate load, where the "bottleneck" buffer of the channel is empty.
- $\theta_t = e_2$: pre-congestion state, indicating that the "bottleneck" buffer is non-empty.
- $\theta_t = e_3$: congestion phase, where the "bottleneck" buffer is full.
- $\theta_t = e_4$: signal loss in the wireless channel hop.

The TRM Λ and the initial distribution p_0 of the CTMC are

$$\Lambda = \begin{bmatrix} -3.825 \times 10^{-2} & 3.75 \times 10^{-2} & 0 & 7.5 \times 10^{-4} \\ 1.5 \times 10^{-1} & -2.01 \times 10^{-1} & 5.025 \times 10^{-2} & 7.5 \times 10^{-4} \\ 0 & 2.4975 \times 10^{-1} & -2.505 \times 10^{-1} & 7.5 \times 10^{-4} \\ 2.4975 \times 10^{-1} & 0 & 0 & -2.4975 \times 10^{-1} \end{bmatrix}, \quad p_0 = \begin{bmatrix} 0.748 \\ 0.186 \\ 0.037 \\ 0.029 \end{bmatrix}.$$

The second component, $Y_t \in \mathbb{R}^2$, represents the current numerical characteristics of the channel:

- Y_t^1 is a current RTT value.
- Y_t^2 indicates the fraction of lost packets in the packet flow.

Given the trajectory of the CTMC $\{\theta_t\}$, the components Y_t^1 and Y_t^2 are mutually independent and have the uniform conditional PDFs $\pi^1(y^1 | \theta_t)$ and $\pi^2(y^2 | \theta_t)$:

$$\pi^1(y^1 | \theta_t = e_1) = 100 \times \mathbf{I}_{[0.01; 0.02]}(y^1), \quad \pi^1(y^1 | \theta_t = e_2) = 142.857 \times \mathbf{I}_{[0.018; 0.025]}(y^1),$$

$$\pi^1(y^1 | \theta_t = e_3) = 200 \times \mathbf{I}_{[0.022; 0.027]}(y^1), \quad \pi^1(y^1 | \theta_t = e_4) = 250 \times \mathbf{I}_{[0.026; 0.03]}(y^1),$$

$$\pi^2(y^2 | \theta_t = e_1) = 1000 \times \mathbf{I}_{[0.0005; 0.0015]}(y^2), \quad \pi^2(y^2 | \theta_t = e_2) = 1000 \times \mathbf{I}_{[0.0005; 0.0015]}(y^2),$$

$$\pi^2(y^2 | \theta_t = e_3) = 11.111 \times \mathbf{I}_{[0.01; 0.1]}(y^2), \quad \pi^2(y^2 | \theta_t = e_4) = 4 \times \mathbf{I}_{[0.05; 0.3]}(y^2).$$

The observations include the noisy RTT measurements

$$\xi_t = \int_0^t Y_s^1 ds + 0.0002 \, w_t, \tag{5.1}$$

and the counting process of the packet losses

$$\eta_t = \int_0^t \frac{Y_s^2}{Y_s^1} ds + \mu_t^\eta. \tag{5.2}$$

The channel state filtering problem involves the online estimation of the process $Z_t = \text{col}(\theta_t, Y_t)$ given the available observations ξ_t and η_t.

The nonlinearity in the drift term of (5.2) arises because the instantaneous packet loss intensity is proportional to the packet loss fraction, while the total packet sending rate is inversely proportional to the RTT.

To avoid errors introduced by the numerical discretization of continuous-time processes, we use different time scales for simulating the state, observations, and the state filtering procedure: $\delta_t = 10^{-4}$ for the simulation and $\Delta_t = 10^{-2}$ for the estimation procedure.

Since the estimation problem requires solving the Kushner-Stratonovich equation, we apply a grid method with space step sizes $\delta_{y^1} = 10^{-3}$ and $\delta_{y^2} = 6 \times 10^{-3}$. For the filtering process, we utilize a hybrid algorithm combining the MJP filtering method for time-discretized observations [27, 28] and the Euler–Maruyama algorithm adapted for jumps [29]. The idea of the algorithm is to approximate the continuous component Y_t of the MRP on a finite grid with N_m nodes. The proposed algorithm is stable, ensuring that the resulting numerical solution satisfies the natural conditions of non-negativity and normalization, which are expected for the actual solution $\widehat{\psi}_n(t, y)$. Analyzing the algorithm [28], we can conclude that the computation of each time layer $\widehat{\psi}_n(t_i, \cdot)$ at time instant t_i requires $O(N^3 \times N_m^3)$ operations.

Although the example is artificial, it exhibits several noteworthy features with practical significance. First, the chosen numerical characteristics of the channel, representing various qualitative states, closely approximate real-world values. Second, the noise intensity in the RTT observations is relatively high. Third, the average duration of states e_3 and e_4 is rather short, making these states difficult to identify. Fourth, the support sets $\mathcal{D}_n$ of the component Y distributions overlap across different channel states θ, complicating the recovery of θ from observations of Y. Lastly, the drift in the counting observations is the nonlinear function of the state. Together, these factors make the example particularly challenging.

Figure 1 presents the results of the observation system simulation, with the following elements:

– Color filling represents the current system state θ_t: from e_1 until e_4.
– The true value of the current RTT Y_1^1.
– The true value of the current loss fraction Y_t^2 (displayed on the auxiliary ordinate axis).
– The ratio $\frac{\Delta \xi_t}{\Delta_t}$, representing the continuous-time observation.
– The observable process η_t, reflecting the packet losses.

The observations exhibit minor fluctuations, indicating transitions in the state Z_t. However, visually identifying the exact current state θ_t and especially estimating its numerical characteristics Y_t, remains challenging. These subtle differences in the observations do not provide sufficient clarity for direct visual interpretation of the channel behavior.

Figure 2 shows the filtering performance for the CTMC θ_t:

– The true state θ_t.
– The filtering estimate $\widehat{\theta}_t^c$ calculated by the continuous observations ξ_t.

– The filtering estimate $\widehat{\theta_t}$ calculated by the continuous and counting observations ξ_t and η_t.

The figure highlights that incorporating counting observations significantly enhances the accuracy of the state estimate θ_t, particularly for the more dynamic "fast" states e_3 and e_4. These states likely involve more abrupt transitions, where counting observations provide valuable additional information.

Figure 3 presents the filtering results for the components of Y_t:

– Color filling indicates the current state θ_t: from e_1 until e_4.
– The true values of Y_t^1 and Y_t^2.
– The filtering estimate $\widehat{Y_t^c}$ calculated by the continuous observations ξ_t.
– The filtering estimate $\widehat{Y_t}$ calculated by the continuous and counting observations ξ_t and η_t.

The observation process η_t plays a crucial role in estimating Y_t^2, the lost packet fraction, as this component directly reflects the packet loss events. In the filtering process, incorporating η_t notably improves the accuracy of the filtering estimate for Y_t^2. This enhancement is especially evident in the time interval $[0; 12.0]$, where the system state remains $\theta_t \equiv e_4$ (signal loss state). During this interval, the continuous observations alone allow the estimate $\widehat{\theta_t^c}$ to track the true state θ_t quite well. However, without access to η_t, the filtering estimate $\widehat{Y_t^{2,c}}$ for Y_t^2 merely aligns with the expectation of Y_t^2 under the condition $\theta_t = e_4$, yielding an expected value of approximately 0.175, given the uniform distribution $\mathcal{R}[0.05; 0.3]$. This causes the estimate to oscillate near this mean value, regardless of the actual variations in Y_t^2.
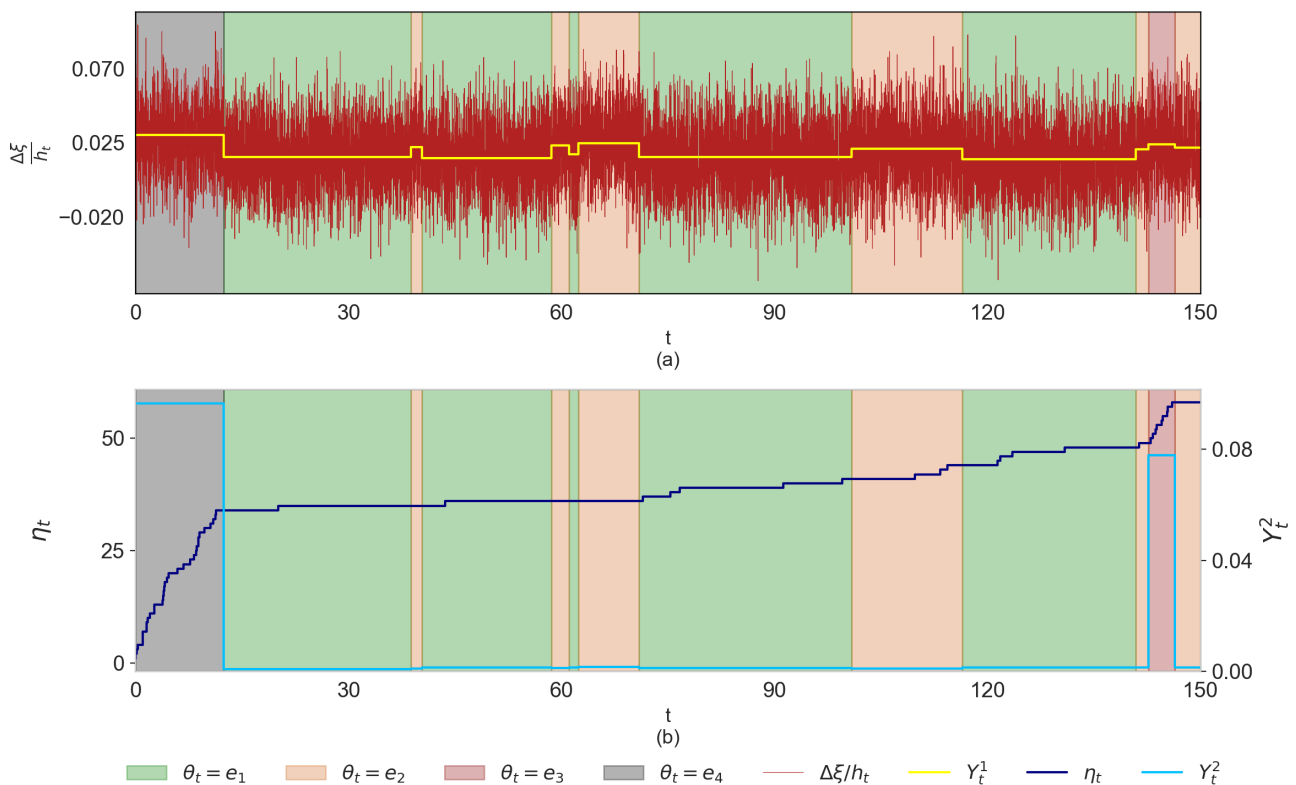

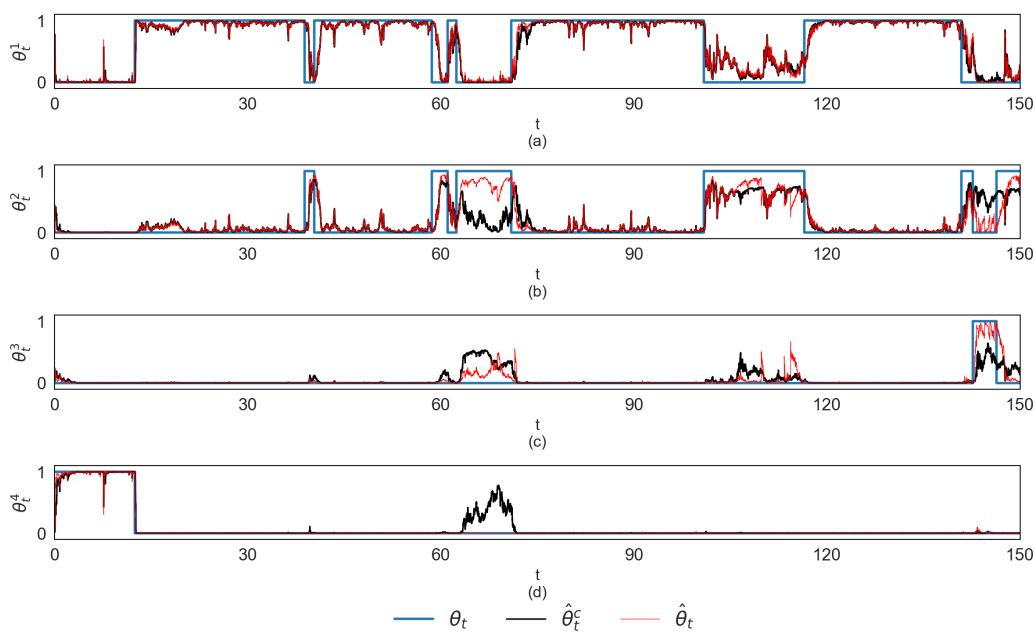**Figure 1.** The channel state, numerical characteristics, and available observations.

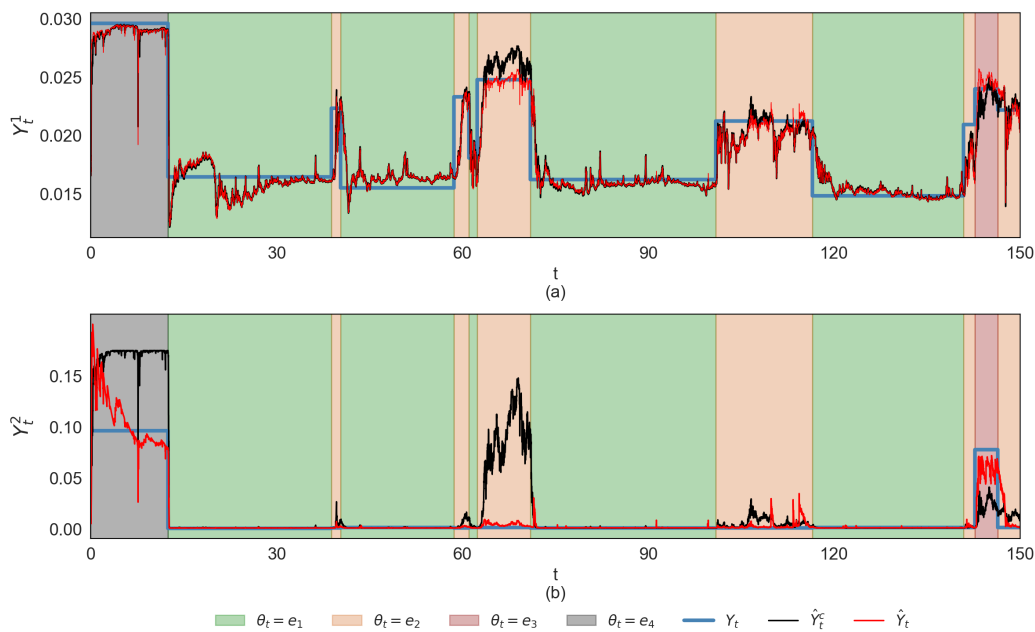**Figure 2.** The filtering results of the channel state θ_t.

Figure 3. The filtering results of the channel characteristics Y_t.

By contrast, when η_t is utilized in the filtering process, the estimate $\widehat{Y_t^2}$ becomes much more accurate, as η_t provides direct information about packet losses. This highlights that the lost packet fraction Y_t^2 can only be accurately estimated by observing the flow of packet losses through the counting process η_t. Hence, combining both continuous and counting observations significantly enhances the filtering performance for Y_t^2.

The filtering errors in both $\widehat{\theta_t}$ and $\widehat{Y_t}$ often exhibit peaks due to mismatches between the local random

behavior of the observations and the true state of the system. This phenomenon is clearly seen in Figures 2 and 3, particularly in the time interval $[7; 10]$, where the filtering estimates temporarily deviate from the true values. During this period, both $\widehat{Y}_t^1$ and $\widehat{Y}_t^2$ tend to underestimate their actual values. Moreover, a noticeable peak appears in the conditional probability $\widehat{\theta}_t^1$, suggesting that the channel is moderately loaded (i.e., $\theta_t = e_1$) even though the true state is $\theta_t = e_4$ (signal loss phase). This discrepancy is explained by examining the observations in Figure 1, which show that during this interval, the continuous observation process $\frac{\delta \xi_t}{h_t}$ is significantly lower than the actual RTT value. Additionally, no packet losses are recorded, making the local observations resemble those of a channel in a moderately loaded state with high throughput and minimal packet loss fraction. Despite this mismatch, the filtering error is short-lived. The estimation is rapidly corrected as new, more relevant observations become available, which help the filter realign with the true state of the system. This highlights the adaptive nature of the filtering algorithm, which compensates for temporary observation anomalies over time.

To illustrate the joint evolution of the conditional PDF $\widehat{\psi}(t, y)$ and the filtering estimates $\widehat{Y}_t$, we provide Figures 4 and 5. These figures focus on a short time interval $[55.0; 75.0]$, during which the process Z_t undergoes several jumps.

Figure 4 and 5 contain similar plots:

  – The true values of Y_t^1 and Y_t^2.
  – The filtering estimates $\widehat{Y}_t^1$ and $\widehat{Y}_t^2$.
  – Evolution of marginal conditional PDFs $\widehat{\psi}^1(t, y_1)$ and $\widehat{\psi}^2(t, y_2)$.

One can observe that after each transition of the state Y_t, the corresponding filtering estimate is also recalibrated through modifications in the conditional PDF.
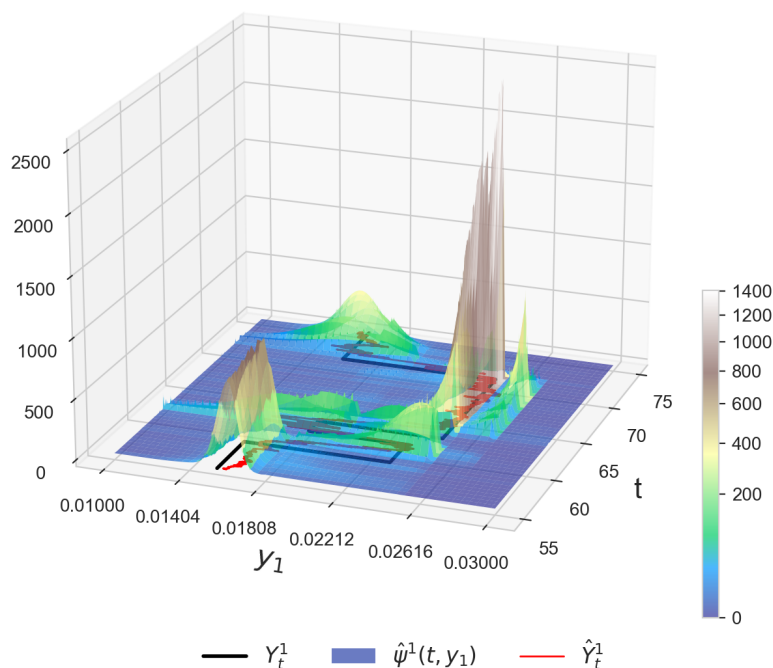

Figure 4. The evolution of the marginal conditional PDF $\widehat{\psi}^1(t, y_1)$ and filtering estimate $\widehat{Y}_t^1$.
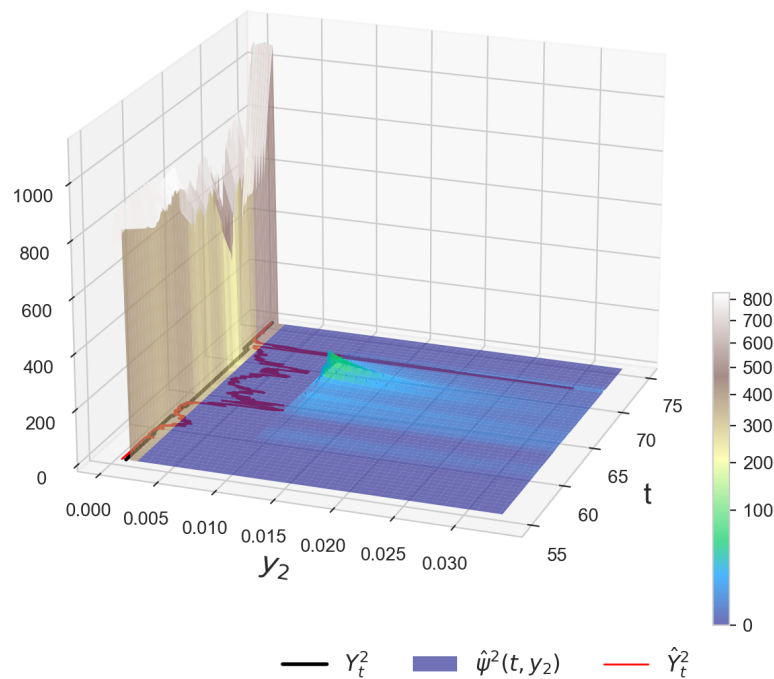
Figure 5. The evolution of the marginal conditional PDF $\widehat{\psi}^2(t, y_2)$ and filtering estimate $\widehat{Y}_t^2$.

In practice, the variance of the filtering error serves not only as a performance index for the estimation but also as a basis for synthesizing control with incomplete information. In only a limited number of fortunate scenarios within stochastic observation systems can one derive the error variance analytically. The observation systems investigated in this article do not fall into this category, leading us to employ the Monte Carlo method (MCM) to compute the sample variance of the filtering error.

For our analysis, we utilize a sample size of $Q = 10\,000$. We compare the variance of the filtering error to that of the estimated process itself. This comparison is significant because the variance of the estimated process can be viewed as a performance metric for the unconditional mathematical expectation, which serves as a trivial estimator. By examining these variances, we gain valuable insights into the effectiveness of our filtering approach.

Figure 6 presents the performance characteristics of the CTMC θ_t estimates:

- The value $\mathsf{D}\theta_t \triangleq \mathsf{E}\left\{\|\theta_t - \mathsf{E}\{\theta_t\}\|^2\right\}$.
- The second moment $\mathsf{D}\widehat{\theta}_t^c \triangleq \mathsf{E}\left\{\|\widehat{\theta}_t^c - \theta_t\|^2\right\}$ of the estimate $\widehat{\theta}_t^c$ error (calculated by the MCM).
- The second moment $\mathsf{D}\widehat{\theta}_t \triangleq \mathsf{E}\left\{\|\widehat{\theta}_t - \theta_t\|^2\right\}$ of the estimate $\widehat{\theta}_t$ error (calculated by the MCM).

Figure 7 is analogous to Figure 6 and presents the performance characteristics of Y_t estimates:

- The variances of the components Y_t^i, $i = 1, 2$: $\mathsf{D}Y_t^i \triangleq \mathsf{E}\left\{\left(Y_t^i - \mathsf{E}\left\{Y_t^i\right\}\right)^2\right\}$, $i = 1, 2$.
- The second moment $\mathsf{D}\widehat{Y}_t^{i,c} \triangleq \mathsf{E}\left\{(\widehat{Y}_t^{i,c} - Y_t^i)^2\right\}$ of $\widehat{Y}_t^{i,c}$ error, $i = 1, 2$ (calculated by the MCM).
- The second moment $\mathsf{D}\widehat{Y}_t^i \triangleq \mathsf{E}\left\{(\widehat{Y}_t^i - Y_t^i)^2\right\}$ of $\widehat{Y}_t^i$ error, $i = 1, 2$ (calculated by the MCM).
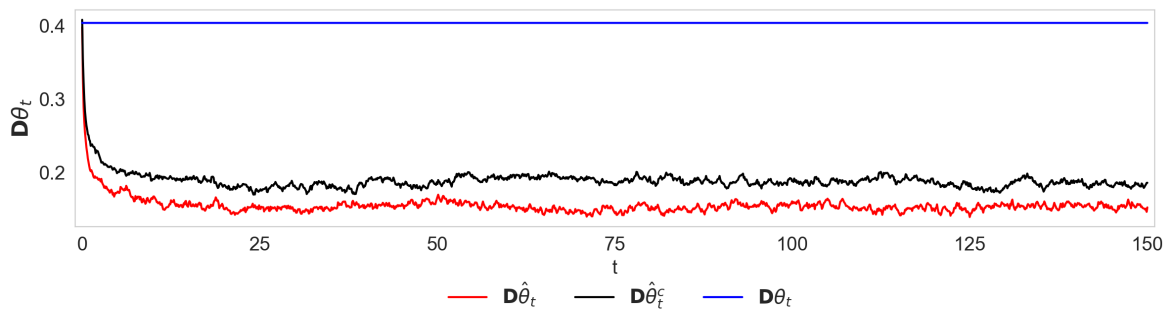
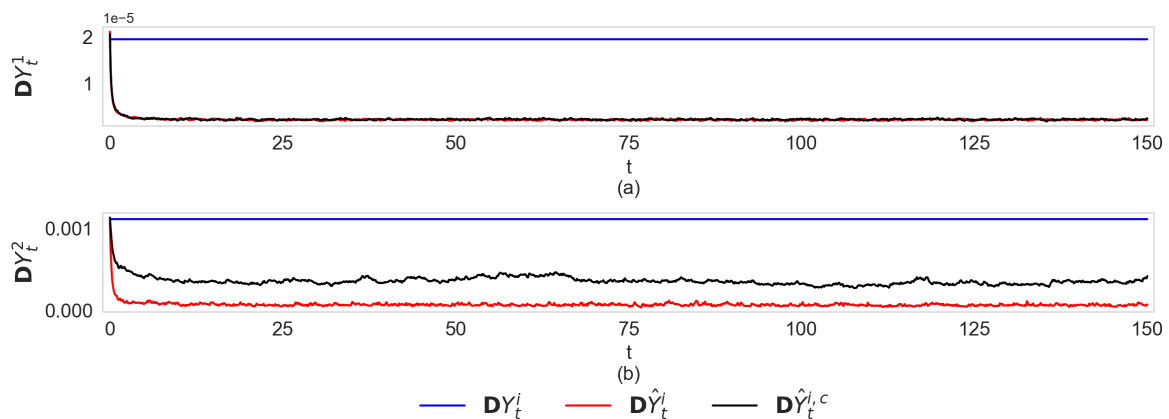**Figure 6.** The estimation quality of the CTMC θ_t.

Figure 7. The estimation quality of the sub-vector Y_t.

In addition to Figures 2 and 3, which illustrate specific trajectories of the system state and its estimates, Figures 6 and 7 support formal conclusions. Moreover, for each filtering estimate of the system state Z_t or its sub-vectors, it is possible to calculate the residual variance ratio using the formula

$$\rho_{\widehat{Z}} \triangleq \frac{\int_0^T \mathsf{E}\left\{\|Z_t - \widehat{Z}_t^q\|^2\right\} dt}{\int_0^T \mathsf{E}\left\{\|Z_s - \mathsf{E}\left\{Z_s\right\}\|^2\right\} ds}. \tag{5.3}$$

From the physical point of view, this integral characteristic represents the ratio of the error "power" to the "power" of the estimated signal itself. A value close to 0 indicates a highly accurate estimate. If the value is slightly lower than 1, the estimate is only marginally better than a trivial one. When the ratio exceeds 1, the proposed estimate is ineffective, as it performs worse than the trivial estimate.

The residual variance ratios of the estimates $\widehat{\theta}^c$, $\widehat{Y}^{1,c}$, and $\widehat{Y}^{2,c}$ calculated using only the continuous observations ξ_t are as follows: $\rho_{\widehat{\theta}^c} = 0.468$, $\rho_{\widehat{Y}^{1,c}} = 0.122$, and $\rho_{\widehat{Y}^{2,c}} = 0.322$. In contrast, the corresponding values for the estimates using the entire observation set are $\rho_{\widehat{\theta}} = 0.384$, $\rho_{\widehat{Y}^1} = 0.117$, and $\rho_{\widehat{Y}^2} = 0.064$. Thus, the use of counting observations of packet losses significantly improves the estimation quality of the channel state θ, has a slight impact on the performance of the current RTT Y^1 estimate, and greatly enhances the estimate of the lost packet fraction Y^2.

The performance of $\widehat{Z}_t$, derived from the entire set of available observations, may be regarded as unimpressive. However, the proposed estimate is optimal in the mean square sense, meaning

its precision cannot be improved with the current observations. To enhance estimation precision, additional observers should be employed, and their observations incorporated into the filtering process.

5.2. Robustness of filtering algorithm

In applied problems, the conditional distribution of the sub-vector Y_t in the MRP Z_t is partially or completely unknown. To design filtering algorithms that are robust to potential deviations of $\pi(\cdot)$, various approaches can be employed, particularly the minimax approach [30]. This method involves searching for the least favorable distribution $\pi^*(\cdot)$ and subsequently constructing the filter in the form of either (4.1) or (4.4). However, this approach has two significant drawbacks. First, the minimax filtering algorithm is computationally intensive. Second, the resulting estimates are typically overly conservative because they are designed to accommodate the least favorable choice of $\pi(\cdot)$, which often results in inappropriately low precision. The distribution $\pi^*(\cdot)$ can become impractical and unlikely to occur unless there is intentional counteraction.

To develop a more robust version of the filtering algorithm, we propose selecting a specific variant of the distribution $\pi(\cdot)$ and using it in either (4.1) or (4.4). In the case of the bounded support sets $\mathcal{D}_\ell$ of the distributions $\pi_\ell(\cdot)$, we propose the uniform ones over $\mathcal{D}_\ell$.

Let us illustrate the performance of the proposed robust filtering estimate using the example discussed above. We will consider three variants of distributions with the common support sets $\mathcal{D}_\ell$:

  – The continuous uniform distribution over the sets $\mathcal{D}_\ell$ (see previous subsection).
  – The continuous symmetric triangular distributions over the sets $\mathcal{D}_\ell$.
  – Three-point uniform distributions concentrated at the ends of the intervals and their midpoints.

We will compare the filtering estimates using the MCM with a sample size of $Q = 5\,000$.

Figure 8 presents the results of the numerical study of the CTMC θ_t estimates for the triangular distribution $\pi(\cdot)$:

  – The unconditional mean square $\mathsf{D}\theta_t = \mathsf{E}_{tr}\left\{\|\theta_t - \mathsf{E}_{tr}\{\theta_t\}\|^2\right\}$.
  – The mean square error (MSE) of the optimal estimate $\widehat{\theta}_t^{tr,tr}$ in the case of the triangular distribution: $\mathsf{D}\widehat{\theta}_t^{tr,tr} = \mathsf{E}_{tr}\left\{\|\widehat{\theta}_t^{tr,tr} - \theta_t\|^2\right\}$ (calculated by the MCM).
  – The MSE of the robust estimate $\widehat{\theta}_t^{tr,u}$ in the case of the triangular distribution $\mathsf{D}\widehat{\theta}_t^{tr,u} = \mathsf{E}_{tr}\left\{\|\widehat{\theta}_t^{tr,u} - \theta_t\|^2\right\}$: the real $\pi(\cdot)$ is triangular, and the filtering algorithm uses the uniform one (calculated by the MCM).
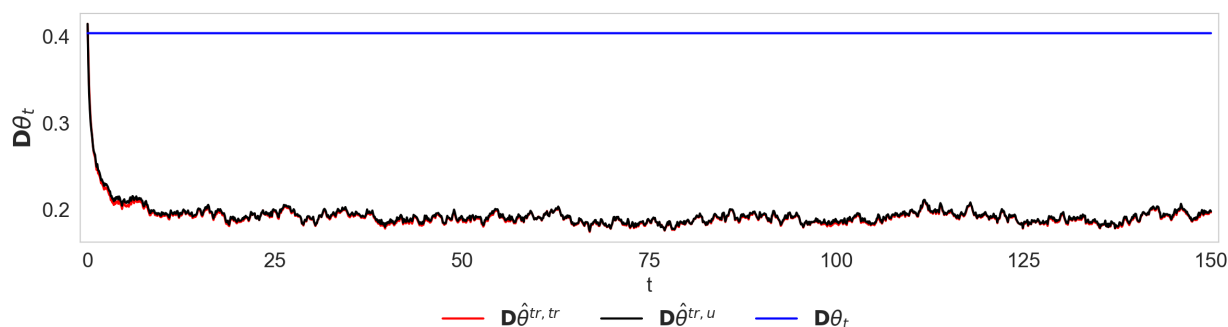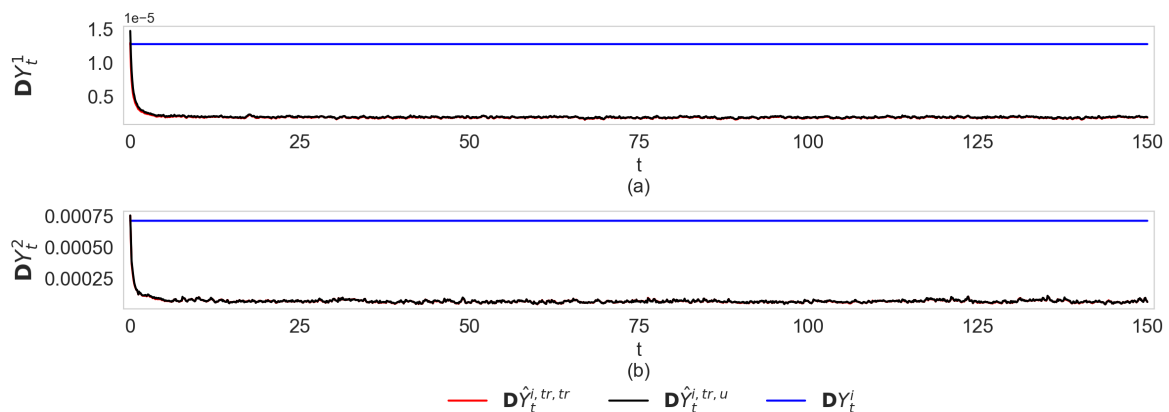

Figure 8. Performance of the robust filtering estimate of θ_t: the case of triangular $\pi(\cdot)$.

Figure 9 presents similar estimation results for Y_t under the triangular distribution $\pi(\cdot)$:

- The unconditional variance $\mathsf{D}Y_t^\ell = \mathsf{E}_{tr}\left\{\left(Y_t^\ell - \mathsf{E}_{tr}\left\{Y_t^\ell\right\}\right)^2\right\}$, $\ell = 1, 2$.
- The MSE of the optimal estimate $\widehat{Y}_t^{\ell,tr,tr}$ in the case of the triangular distribution: $\mathsf{D}\widehat{Y}_t^{\ell,tr,tr} = \mathsf{E}_{tr}\left\{(\widehat{Y}_t^{\ell,tr,tr} - Y_t^\ell)^2\right\}$ (calculated by the MCM).
- The MSE of the robust estimate $\widehat{Y}_t^{\ell,tr,u}$ in the case of the triangular distribution $\mathsf{D}\widehat{Y}_r^{\ell,tr,u} = \mathsf{E}_{tr}\left\{(\widehat{Y}_t^{\ell,tr,u} - Y_t^\ell)^2\right\}$: the true distribution $\pi(\cdot)$ is triangular, and the filtering algorithm uses the uniform one (calculated by the MCM).

Figure 9. Performance of the robust filtering estimate of Y_t: the case of triangular $\pi(\cdot)$.

Figure 10 presents the results of the comparative numerical study of the CTMC θ_t estimates for the three-point discrete uniform distribution $\Pi(\cdot)$:

- The unconditional mean square $\mathsf{D}\theta_t = \mathsf{E}_{3p}\left\{\|\theta_t - \mathsf{E}_{3p}\{\theta_t\}\|^2\right\}$.
- The MSE of the optimal estimate $\widehat{\theta}_t^{3p,3p}$ in the case of the three-point distribution $\mathsf{D}\widehat{\theta}_t^{3p,3p} = \mathsf{E}_{3p}\left\{\|\widehat{\theta}_t^{3p,3p} - \theta_t\|^2\right\}$ (calculated by the MCM).
- The MSE of the robust estimate $\widehat{\theta}_t^{3p,u}$ in the case of the three-point discrete uniform distribution $\mathsf{D}\widehat{\theta}_t^{3p,u} = \mathsf{E}_{3p}\left\{\|\widehat{\theta}_t^{3p,u} - \theta_t\|^2\right\}$: the real $\Pi(\cdot)$ is three-point discrete uniform, and the filtering algorithm uses the continuous uniform one (calculated by the MCM).
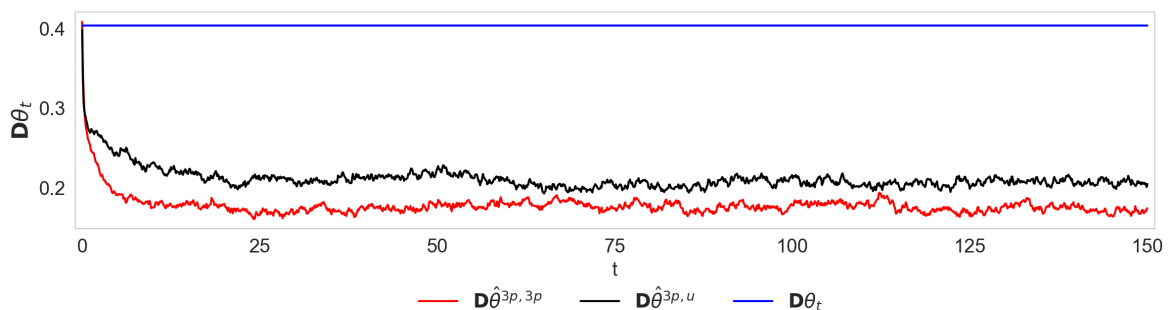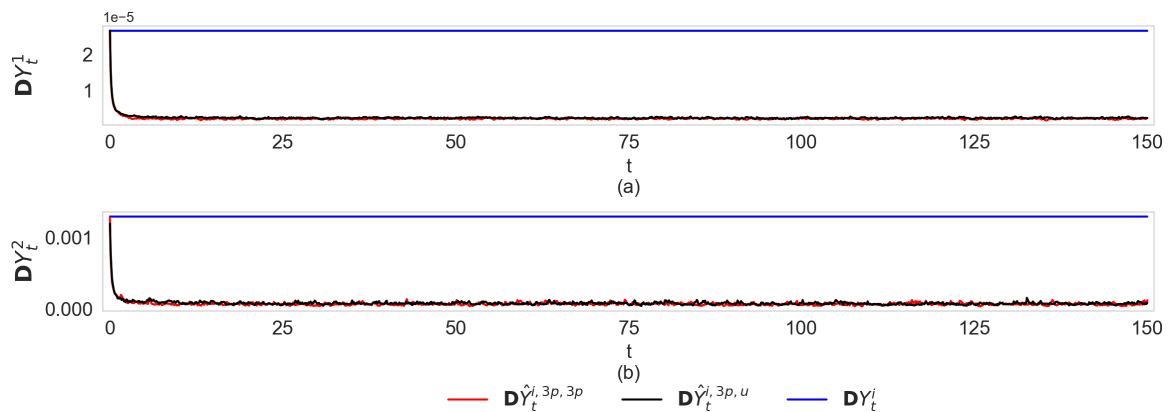

Figure 10. Performance of the robust filtering estimate of θ_t: the case of three-point discrete uniform $\Pi(\cdot)$.

Figure 11 presents similar estimation results for Y_t under the three-point discrete uniform distribution $\Pi(\cdot)$:

– The unconditional variance $\mathsf{D}Y_t^\ell = \mathsf{E}_{3p}\left\{\left(Y_t^\ell - \mathsf{E}_{3p}\left\{Y_t^\ell\right\}\right)^2\right\}$, $\ell = 1, 2$.

– The MSE of the optimal estimate $\widehat{Y}_t^{\ell,3p,3p}$ error in the case of the three-point discrete uniform distribution: $\mathsf{D}\widehat{Y}_t^{\ell,3p,3p} = \mathsf{E}_{3p}\left\{(\widehat{Y}_r^{\ell,3p,3p} - Y_t^\ell)^2\right\}$ (calculated by the MCM).

– The MSE of the robust estimate $\widehat{Y}_t^{\ell,tr,u}$ error in the case of the three-point discrete uniform distribution: $\mathsf{D}\widehat{Y}_r^{\ell,3p,u} = \mathsf{E}_{3p}\left\{(\widehat{Y}_r^{\ell,3p,u,q} - Y_t^{\ell,q})^2\right\}$: the real $\Pi(\cdot)$ is three-point discrete uniform, and the filtering algorithm uses the continuous uniform one (calculated by the MCM).

Figure 11. Performance of the robust filtering estimate of Y_t: the case of three-point discrete uniform $\Pi(\cdot)$.

To evaluate the performance loss of the robust filtering estimate $\widehat{Z}^{i,u}$ (where i = "tr" for triangular distribution or "3p" for three-point distribution) in comparison with the optimal estimate, we use the following integral index. Let $\rho_{\widehat{Z}^{i,u}}$ denote the performance index (5.3) of the robust filtering estimate when the real distribution corresponds to i. Additionally, let $\rho_{\widehat{Z}^{i,i}}$ represent the performance index (5.3) calculated for the optimal filtering estimate $\widehat{Z}^{i,i}$. We propose to consider the value $\varkappa_Z^i \triangleq \left(\frac{\rho_{\widehat{Z}^{i,u}}}{\rho_{\widehat{Z}^{i,i}}} - 1\right) \times 100\%$, which is the percentage increase in the index ρ when replacing the optimal estimator with the robust one.

Table 1 contains the characteristics $\varkappa$ calculated for the robust filtering estimates of CTMC θ and for the components Y^1 and Y^2 separately.

Table 1. Performance of robust filtering estimate.

Estimated component	Triangular distribution	Three-point distribution
θ	0.4 %	18.7 %
Y^1	3.6 %	8.9 %
Y^2	1.6 %	11.6 %

Note that in the considered case, the performance losses do not exceed 19%. In the adjacent estimation problem involving a priori uncertainty, one can select the least favorable distribution from

the discrete options [30]. Therefore, it is not surprising that the performance loss is greater for the three-point uniform distribution Π: the actual least favorable distribution is concentrated at distant points within the support sets $\mathcal{D}_\ell$, and three-point distribution is close to it. By analyzing the presented figures and the table, we can conclude that the proposed robust filtering algorithm demonstrates acceptable accuracy and could be beneficial in situations where there is uncertainty in the probability distribution $\pi(\cdot)$ of the sub-vector Y_t.

6. Conclusions

In summary, we can present the results of the paper as follows:

1) The paper introduces a subclass of MRPs that has practical value for the mathematical modeling of real-world objects and phenomena.
2) The optimal filtering problem for the MRP, considering both continuous and counting observations, is properly formulated and solved. The optimal filtering estimate of a scalar signal process is defined through the solution of a potentially non-closed SDS. Additionally, a variant of the Kushner-Stratonovich equation, which describes the evolution of the conditional PDF of the system state, is derived.
3) The high quality of the derived estimate is illustrated through an applied example related to telecommunications, where the filter enables monitoring of the network channel qualitative state and numerical characteristics based on observations of the RTT and packet loss flow.
4) The presented filtering algorithm demonstrates robustness to a priori uncertainty in the probability distribution of the estimated MRP.

These results can serve as a foundation for future studies.

First, the class of observation systems can be expanded to include the considered MRPs as states for subsequent solutions to the optimal filtering problem. A promising direction involves utilizing continuous-time observations with multiplicative noise [31] and estimating the CTMC θ_t using the noiseless observations of the component Y_t. The filtering problems in this context are challenging because they do not allow for a Girsanov change of measure, which would reduce the original filtering problem to one involving Wiener and Poisson processes as observations [32].

Second, the Kushner-Stratonovich equation (4.4) is the non-linear stochastic partial integro-differential equation. It is simpler than the original Kushner-Stratonovich equation derived for diffusion observation systems because it does not include partial derivatives with respect to the state variable y. However, to address various applied estimation problems, efficient numerical algorithms for solving (4.4) and an analysis of their accuracy are required.

Third, the proposed filtering algorithm can be adapted to account for a priori uncertainty in the parameters of the MRP distribution. The current version already demonstrates some robustness to imprecise knowledge of the Y_t distribution. However, the algorithm can be further developed in several directions, including uncertain parameter identification [33], the design of a corresponding guaranteed filter [30], and fuzzy logic adaptations [34, 35].

Fourth, the paper highlights telecommunications as an application area for the mathematical modeling of real phenomena using the studied MRPs and the subsequent solutions to estimation problems. The applicability of MRPs and the corresponding estimation framework could be

broader, encompassing areas such as navigation and maneuvering target tracking [9, 10], financial mathematics [36], biology [37], medicine [38], etc.

Acknowledgments

Conflict of interest

The author declares no conflict of interest.

References

1. E. B. Dynkin, *Markov processes*, Heidelberg: Springer, **1** (2012). https://doi.org/10.1007/978-3-662-00031-1

2. S. Ethier, T. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, 2009.

3. D. White, A survey of applications of Markov decision processes, *J. Oper. Res. Soc.*, **44** (2014), 1073–1096. https://doi.org/10.1057/jors.1993.181

4. J. Jacod, Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, **31** (1975), 235–253. https://doi.org/10.1007/BF00536010

5. N. Limnios, G. Oprişan, *Semi-Markov processes and reliability*, Boston: Birkhäuser, 2001. https://doi.org/10.1007/978-1-4612-0161-8

6. C. Cocozza-Thivent, *Markov renewal and piecewise deterministic processes*, Springer Cham, 2021. https://doi.org/10.1007/978-3-030-70447-6

7. Y. Ishikawa, H. Kunita, Malliavin calculus on the Wiener-Poisson space and its application to canonical SDE with jumps, *Stoch. Process. Appl.*, **116** (2006), 1743–1769. https://doi.org/10.1016/j.spa.2006.04.013

8. A. Borisov, Analysis and estimation of the states of special Markov jump processes. I. Martingale Representation, *Autom. Remote Control*, **65** (2004), 44–57. https://doi.org/10.1023/B:AURC.0000011689.11915.24

9. Y. Bar-Shalom, X. R. Li, T. Kirubarajan, *Estimation with applications to tracking and navigation: Theory algorithms and software*, John Wiley & Sons, 2002. https://doi.org/10.1002/0471221279

10. X. R. Li, V. P. Jilkov, Survey of maneuvering target tracking. Part V. Multiple-model methods, *IEEE Trans. Aerosp. Electron. Syst.*, **41** (2005), 1255–1321. https://doi.org/10.1109/TAES.2005.1561886

11. D. Delahaye, S. Puechmorel, P. Tsiotras, E. Feron, Mathematical models for aircraft trajectory design: A survey, In: *Air traffic management and systems. Lecture notes in electrical engineering*, Tokyo: Springer, **290** (2014), 205–247. https://doi.org/10.1007/978-4-431-54475-3_12

12. J. Lan, X. R. Li, V. P. Jilkov, C. Mu, Second-order Markov chain based multiple-model algorithm for maneuvering target tracking, *IEEE Trans. Aerosp. Electron. Syst.*, **49** (2013), 3–19. https://doi.org/10.1109/TAES.2013.6404088

13. Y. Shen, T. K. Siu, Asset allocation under stochastic interest rate with regime switching, *Econom. Model.*, **29** (2012), 1126–1136. https://doi.org/10.1016/j.econmod.2012.03.024

14. S. Goutte, Pricing and hedging in stochastic volatility regime switching models, *J. Math. Finance* **3** (2013), 70–80. http://dx.doi.org/10.4236/jmf.2013.31006

15. J. D. Hamilton, Macroeconomic regimes and regime shifts, In: *Handbook of macroeconomics*, Elsevier, **2** (2016), 163–201. https://doi.org/10.1016/bs.hesmac.2016.03.004

16. R. Sueppel, Classifying market regimes, 2021. Available from: `https://research.macrosynergy.com/classifying-market-regimes/#classifying-market-regimes`.

17. X. Zhang, R. J. Elliott, T. K. Siu, J. Guo, Markovian regime-switching market completion using additional Markov jump assets, *IMA J. Manag. Math.*, **23** (2012), 283–305. https://doi.org/10.1093/imaman/dpr018

18. M. A. Kouritzin, Sampling and filtering with Markov chains, *Signal Process.*, **225** (2024), 109613. https://doi.org/10.1016/j.sigpro.2024.109613

19. R. Sh. Liptser, A. N. Shiryayev, *Theory of martingales*, Dordrecht: Springer, 1989. https://doi.org/10.1007/978-94-009-2438-3

20. R. W. Brockett, Nonlinear systems and nonlinear estimation theory, In: *Stochastic systems: The mathematics of filtering and identification and applications*, **78** (1981), 441–477. https://doi.org/10.1007/978-94-009-8546-9_23

21. V. E. Beneš, Exact finite-dimensional filters for certain diffusions with nonlinear drift, *Stochastics*, **5** (1981), 65–92. https://doi.org/10.1080/17442508108833174

22. M. Hazewinkel, S. I. Marcus, H. J. Sussmann, Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem, *Syst. Control Lett.*, **3** (1983), 331–340. https://doi.org/10.1016/0167-6911(83)90074-9

23. T. Björk, Finite optimal filters for a class of nonlinear diffusions with jumping parameters, *Stochastics*, **6** (1982), 121–138. https://doi.org/10.1080/17442508208833198

24. S. Tang, Brockett's problem of classification of finite-dimensional estimation algebras for nonlinear filtering systems, *SIAM J. Control Optim.*, **39** (2000), 900–916. https://doi.org/10.1137/S036301299833464X

25. W. Dong, J. Shi, A survey of estimation algebras in application of nonlinear filtering problems, *Commun. Inf. Syst.*, **19** (2019), 193–217. https://dx.doi.org/10.4310/CIS.2019.v19.n2.a4

26. H. J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications, *J. Soc. Indust. Appl. Math. Ser. A*, **2** (1962), 106–119. https://doi.org/10.1137/0302009

27. A. V. Borisov, Numerical schemes of Markov jump process filtering given discretized observations I: Accuracy characteristics, *Inform. Appl.*, **13** (2019), 68–75. https://doi.org/10.14357/19922264190411

28. A. V. Borisov, Numerical schemes of Markov jump process filtering given discretized observations II: Additive noises case, *Inform. Appl.*, **14** (2020), 17–23. https://doi.org/10.14357/19922264200103

29. E. Platen, N. Bruti-Liberati, *Numerical solution of stochastic differential equations with jumps in finance*, Berlin Heidelberg: Springer-Verlag, 2010. https://doi.org/10.1007/978-3-642-13694-8

30. A. V. Borisov, The Wonham filter under uncertainty: A game-theoretic approach, *Automatica*, **47** (2011), 1015–1019. https://doi.org/10.1016/j.automatica.2011.01.056

31. A. Borisov, I. Sokolov, Optimal filtering of Markov jump processes given observations with state-dependent noises: Exact solution and stable numerical schemes, *Mathematics*, **8** (2020), 506. https://doi.org/10.3390/math8040506

32. E. Wong, B. Hajek, *Stochastic processes in engineering systems*, New York: Springer, 1985. https://doi.org/10.1007/978-1-4612-5060-9

33. A. Borisov, A. Gorshenin, Identification of continuous-discrete hidden Markov models with multiplicative observation noise, *Mathematics*, **10** (2022), 3062. https://doi.org/10.3390/math10173062

34. P. Cheng, S. He, V, Stojanovic, X. Luan, F. Liu, Fuzzy fault detection for Markov jump systems with partly accessible hidden information: An event-triggered approach, *IEEE Trans. Cybernet.*, **52** (2022), 7352–7361. https://doi.org/10.1109/TCYB.2021.3050209

35. X. Zhang, H. Wang, V. Stojanovic, P. Cheng, S. He, X. Luan, Asynchronous fault detection for interval type-2 fuzzy nonhomogeneous higher level Markov jump systems with uncertain transition probabilities, *IEEE Trans. Fuzzy Syst.*, **30** (2022), 2487–2499. https://doi.org/10.1109/TFUZZ.2021.3086224

36. R. S. Mamon, R. J. Elliott, *Hidden Markov models in finance*, New York: Springer, 2014. https://doi.org/10.1007/978-1-4899-7442-6

37. D. J. Wilkinson, *Stochastic modelling for systems biology*, 2nd Eds., Boca Raton: CRC Press, 2011. https://doi.org/10.1201/b11812

38. A. Bureau, S. Shiboski, J. P. Hughes, Applications of continuous time hidden Markov models to the study of misclassified disease outcomes, *Stat. Med.*, **22** (2003), 441–462. https://doi.org/10.1002/sim.1270

39. S. N. Cohen, R. J. Elliott, *Stochastic calculus and applications*, New York: Birkhäuser, 2015. https://doi.org/10.1007/978-1-4939-2867-5

40. F. Klebaner, R. Liptser, When a stochastic exponential is a true Martingale. Extension of the Beneš method, *Theory Probab. Appl.*, **58** (2014), 38–62. https://doi.org/10.1137/S0040585X97986382

41. R. S. Liptser, A. N. Shiryayev, *Statistics of random processes: I. General theory*, Heidelberg: Springer-Verlag Berlin, 2001. https://doi.org/10.1007/978-3-662-13043-8

A. Proof of Theorem 4.1

We anticipate the proof of the theorem itself by

Lemma A.1. *The innovation process v_t is an O_t-adapted Wiener process. The innovation process ζ_t is an O_t-adapted purely discontinuous martingale with the quadratic characteristic*

$$\langle \zeta, \zeta \rangle_t = \int_0^t \operatorname{diag} \widehat{h}_{s-} ds.$$

Proof. First, we verify the martingale property of v_t. To do this, we consider two time instants $0 \leqslant u < t \leqslant T$:

$$\mathsf{E}\{v_t - v_u \mid O_u\} = \mathsf{E}\left\{R^{-1/2} \int_u^t (g(Z_s) - \widehat{g}_{s-}) \, ds + w_t - w_u \mid O_u\right\}$$

$$= R^{-1/2} \int_u^t \mathsf{E}\{g(Z_s) - \widehat{g}_{s-} \mid O_u\} \, ds + \mathsf{E}\{\mathsf{E}\{w_t - w_u \mid \mathcal{F}_u\} \mid O_u\} = 0,$$

because $\widehat{g}_s \neq \widehat{g}_{s-}$ at a finite subset of $[0, T]$ $\mathcal{P}$ – a.s. So, v_t is a martingale.

Further, to define $\langle v, v \rangle_t$, we apply the Itô rule to the product $v_t v_t^{\top}$, keeping in mind that $\langle w, w \rangle_t = tI$:

$$v_t v_t^{\top} = \int_0^t v_{s-} dv_s^{\top} + \int_0^t dv_s v_{s-}^{\top} + \langle v, v \rangle_t$$

$$= \int_0^t v_s \left(R^{-1/2}(g(Z_s) - \widehat{g}_{s-}) ds + dw_s\right)^{\top} + \int_0^t \left(R^{-1/2}(g(Z_s) - \widehat{g}_{s-}) ds + dw_s\right) v_s^{\top} + \langle w, w \rangle_t.$$

We have the equality $\langle v, v \rangle_t = \langle w, w \rangle_t = tI$, and since the quadratic characteristic of w_t is the non-random function with respect to the filtration $\{\mathcal{F}_t\}$, it also is with respect to $\{O_t\}$. The martingale v_t has the quadratic characteristic tI and $\mathcal{P}$-a.s. continuous trajectories, hence this is a K-dimensional Wiener process [39, Thm 14.4.1].

Now, we verify the martingale property of ζ_t. We consider two time instants $0 \leqslant u < t \leqslant T$ and use the fact that $\widehat{h}_{s-} \neq \widehat{h}_s$ at most at a finite set of instants on $[0, T]$ with probability 1:

$$\mathsf{E}\{\zeta_t - \zeta_u \mid O_u\} = \mathsf{E}\left\{\int_u^t \left(h(Z_s-) - \widehat{h}_{s-}\right) ds + \mu_t^{\eta} - \mu_u^{\eta} \mid O_u\right\}$$

$$= \int_u^t \mathsf{E}\left\{h(Z_s) - \widehat{h}_s \mid O_u\right\} ds + \mathsf{E}\{\mathsf{E}\{\mu_t^{\eta} - \mu_u^{\eta} \mid \mathcal{F}_u\} \mid O_u\} = 0. \quad \mathcal{P} - \text{a.s.}$$

So, ζ_t is a martingale.

Further, to define $\langle \zeta, \zeta \rangle_t$, we apply the Itô rule to the product $\zeta_t \zeta_t^{\top}$ keeping in mind the equality $\langle \eta, \eta \rangle_t = \int_0^t \operatorname{diag} h(Z_{s-}) ds$:

$$\zeta_t \zeta_t^{\top} = \int_0^t \zeta_{s-} d\zeta_s^{\top} + \int_0^t d\zeta_s \zeta_{s-}^{\top} + \sum_{0 \leqslant \tau \leqslant t} \Delta\zeta_\tau \Delta\zeta_\tau^{\top} = \int_0^t \zeta_{s-} d\zeta_s^{\top} + \int_0^t d\zeta_s \zeta_{s-}^{\top} + \sum_{0 \leqslant \tau \leqslant t} \Delta\eta_\tau \Delta\eta_\tau^{\top}$$

$$= \int_0^t \zeta_{s-}d\zeta_s^\top + \int_0^t d\zeta_s\zeta_{s-}^\top + [\eta,\eta]_t = \int_0^t \text{diag}\,\widehat{h}_{s-}ds + \mu_t^4 + \underbrace{\int_0^t \zeta_{s-}d\zeta_s^\top + \int_0^t d\zeta_s\zeta_{s-}^\top}_{o_t\text{–adapted martingale}}.$$

Hence, $\langle \zeta, \zeta \rangle_t = \int_0^t \text{diag}\,\widehat{h}_{s-}ds.$ □

Now, we go back to the proof of Theorem 4.1. First, if $\widehat{Q}_t$ is an optimal estimate of the process Q_t associated with $q(Z_t)$, and $q(Z_t) = \text{row}(\mathbf{0},\mathbf{1})Q_t$ is the linear transformation of Q_t, then $\widehat{q}_t = \text{row}(\mathbf{0},\mathbf{1})\widehat{Q}_t$. To derive an equation describing the evolution of $\widehat{Q}_t$, we use an approach suggested in [32, Chap. 7] based on the uniqueness of the special semimartingale decomposition [19].

The associated process Q_t can be described by an SDS, which is an appropriate version of (2.7)

$$Q_t = Q_0 + \int_0^t D^Q(s)Q_{s-}ds + d\mu_t^Q, \tag{A.1}$$

where $\mu_t^Q \in \mathbb{R}^{2N}$ is an $\mathcal{F}_t$-adapted square integrable martingale. Conditioning both sides of (A.1) by O_t, one can obtain, that

$$\widehat{Q}_t = \mathsf{E}\{Q_0\} + \int_0^t D^Q(s)\widehat{Q}_{s-}ds + \int_0^t \gamma_s d\nu_s + \int_0^t \Gamma_s d\zeta_s, \tag{A.2}$$

where $\gamma_s = \gamma(s,\omega_s) \in \mathbb{R}^{2N \times K}$ and $\Gamma_s = \Gamma(s,\omega_s) \in \mathbb{R}^{2N \times L}$ are O_t-predictable random matrix-valued processes that should be determined [32, Lemmas 7.4.1 and 7.4.2]. However, decomposition (A.2) is true when there exists a suitable Girsanov measure transform for which ξ_t is the Wiener process, and η_t is the Poisson one. To meet this condition, the equality

$$\mathsf{E}\{\Lambda_T\} = 1 \tag{A.3}$$

should be true [39] for

$$\Lambda_t = \exp\left(-\int_0^t g^\top(Z_s)R^{-1/2}dw_s - \frac{1}{2}\int_0^t g^\top(Z_s)R^{-1}g(Z_s)ds + \sum_{\ell=1}^L \int_0^t \left(\left(h^\ell(Z_s)-1\right)ds - \ln h^\ell(Z_{s-})d\eta_s^\ell\right)\right).$$

The processes $g(\cdot)$ and $h(\cdot)$ in (3.1) can be expressed via the solution to some linear SDS, and this is sufficient condition for the fulfillment of (A.3) [40, Theorem 4.1].

Let us consider the product $\widehat{Q}_t\xi_t^\top$. Due to the Itô rule and the fact $\langle Q, \xi \rangle_t \equiv 0$, we have

$$Q_t\xi_t^\top = \int_0^t Q_s d\xi_s^\top + \int_0^t dQ_s\xi_s^\top = \int_0^t \left(Q_s g(Z_s)^\top R^{-1/2} + D^Q(s)Q_s\xi_s^\top\right)ds + \underbrace{\int_0^t \left(Q_s dw_s^\top + d\mu_s^Q\xi_s^\top\right)}_{\mathcal{F}_t\text{–measurable martingale}}.$$

Conditioning both sides of the last expression with respect to O_t, we obtain the first variant of decomposition of $\widehat{Q}_t\xi_t^\top$:

$$\mathsf{E}\{Q_t\xi_t^\top|O_t\} = \widehat{Q}_t\xi_t^\top = \int_0^t \left(\widehat{Qg^\top}_{s-}R^{-1/2} + D^Q(s)\widehat{Q}_{s-}\xi_s^\top\right)ds + \mu_t^5, \tag{A.4}$$

where μ_t^5 is an O_t-adapted martingale.

Now, we use (4.1) and the expression $\xi_t = \int_0^t \widehat{g}_{s-} ds + R^{1/2} \nu_t$:

$$\widehat{Q}_t \xi_t^\top = \int_0^t \widehat{Q}_{s-} d\xi_s^\top + \int_0^t d\widehat{Q}_s \xi_s^\top + \langle \widehat{Q}, \xi \rangle_t \tag{A.5}$$

$$= \int_0^t \left(\widehat{Q}_{s-} \widehat{g}_{s-}^\top R^{-1/2} + D^Q(s) \widehat{Q}_{s-} \xi_s^\top + \gamma_s \right) ds + \underbrace{\int_0^t \left(\widehat{Q}_{s-} d\nu_s^\top R^{1/2} + (\gamma_s d\nu_s + \Gamma_s d\zeta_s) \xi_s^\top \right)}_{\triangleq \mu_t^6},$$

where μ_t^6 is an O_t-adapted martingale.

Formulae (A.4) and (A.5) represent the same semimartingale, and this presumes $\mathcal{P}$-a.s. fulfillment of the equality $\widehat{Qg^\top}_{t-} R^{-1/2} + D^Q(t) \widehat{Q}_{t-} \xi_t^\top = \widehat{Q}_{t-} \widehat{g}_{t-}^\top R^{-1/2} + D^Q(t) \widehat{Q}_{t-} \xi_t^\top + \gamma_t$ for almost any $t \in [0, T]$. It is easy to verify that the O_t-predictable process

$$\gamma_t = \left(\widehat{Qg^\top}_{t-} - \widehat{Q}_{t-} \widehat{g}_{t-}^\top \right) R^{-1/2} \tag{A.6}$$

satisfies the last equality.

We define the integrand Γ_t analogously. First, we obtain the decomposition of the product $\widehat{Q}_t \eta_t^\top$:

$$Q_t \eta_t^\top = \int_0^t Q_{s-} d\eta_s^\top + \int_0^t dQ_s \eta_{s-}^\top + \sum_{0 \leqslant \tau \leqslant t} \Delta Q_\tau \Delta \eta_\tau^\top$$

$$= \int_0^t \left(Q_s h(Z_s)^\top + D^Q(s) Q_s \eta_s^\top \right) ds + \underbrace{\int_0^t \left(Q_{s-} d(\mu_s^\eta)^\top + d\mu_s^Q \eta_{s-}^\top \right)}_{\mathcal{F}_t\text{--measurable martingale}}.$$

Conditioning both sides of the last expression with respect to O_t, we obtain the first variant of decomposition of $\widehat{Q}_t \eta_t^\top$:

$$\mathsf{E}\{Q_t \eta_t^\top | O_t\} = \widehat{Q}_t \eta_t^\top = \int_0^t \left(\widehat{Qh^\top}_{s-} + D^Q(s) \widehat{Q}_{s-} \eta_{s-}^\top \right) ds + \mu_t^7, \tag{A.7}$$

where μ_t^7 is an O_t-adapted martingale.

Now, we apply the Itô rule to $\widehat{Q}_t \eta_t^\top$ and use the expression $\eta_t = \int_0^t \widehat{h}_{s-} ds + \zeta_t$:

$$\widehat{Q}_t \eta_t^\top = \int_0^t \widehat{Q}_{s-} d\eta_s^\top + \int_0^t d\widehat{Q}_s \eta_{s-}^\top + \sum_{0 \leqslant \tau \leqslant t} \Delta \widehat{Q}_\tau \Delta \eta_\tau^\top = \int_0^t \left(\widehat{Q}_{s-} \widehat{h}_{s-}^\top + D^Q(s) \widehat{Q}_s \eta_s^\top + \Gamma_s \operatorname{diag} \widehat{h}_{s-} \right) ds + \mu_t^8, \tag{A.8}$$

where μ_t^8 is an O_t-adapted martingale. Formulae (A.7) and (A.8) represent the same semimartingale, and this presumes $\mathcal{P}$-a.s. fulfillment of the equality $\widehat{Qh^\top}_t + D^Q(t) \widehat{Q}_t \eta_t^\top = \widehat{Q}_t \widehat{h}_{t-}^\top + D^Q(t) \widehat{Q}_t \eta_t^\top + \Gamma_t \operatorname{diag} \widehat{h}_{t-}$ for almost any $t \in [0, T]$. It is easy to verify that the O_t-predictable process

$$\Gamma_t = \left(\widehat{Qh^\top}_{t-} - \widehat{Q}_{t-} \widehat{h}_{t-}^\top \right) \operatorname{diag}^{-1} \widehat{h}_{t-} \tag{A.9}$$

satisfies the last equality. The substitution of (A.6) and (A.9) into (A.2) completes the proof.

B. Proof of Theorem 4.2

The existence of the conditional PDF $\widehat{\psi}(t, y)$ follows from an abstract variant of the Bayes rule given both the continuous and counting observations by analogy with [41, Thm 7.23]. Conditions A1–A8 guarantee the legitimacy of this existence. If $\widehat{\psi}(t, y)$ does exist, then expressions (4.5) for $\widehat{\theta}_s$, $\widehat{g}_s$, and $\widehat{h}_{s-}$ are obvious.

Using (4.1), one can obtain the following representation for the estimate $\widehat{q}_t = \mathsf{E}\{q(Z_t)|O_t\}$:

$$\widehat{q}_t = \mathsf{E}\{q(Z_0)\} + \int_0^t \left[\lambda(s)\widehat{\mathsf{q}}_{s-} + \mathsf{E}_q^\top \widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-}\right] ds \tag{B.1}$$

$$+ \int_0^t \left(\widehat{qg^\top}_{s-} - \widehat{q}_{s-}\widehat{g^\top}_{s-}\right) R^{-1/2} dv_s + \int_0^t \left(\widehat{qh^\top}_{s-} - \widehat{q}_{s-}\widehat{h^\top}_{s-}\right) \mathrm{diag}^{-1}(\widehat{h}_{s-}) d\zeta_s.$$

We choose the function $q(\cdot)$, defining the estimated process, in the form $q(z) = q(e, y) = e_n^\top e \mathbf{I}_B(y)$ for some $n \in \{1, \dots, N\}$ and $B \in \mathcal{B}(\mathbb{R}^M)$, hence $q(Z_t) = e_n^\top \theta_t \mathbf{I}_B(Y_t)$. The terms in both sides of (B.1) can be written as follows:

$$\widehat{q}_t = \int_B \widehat{\psi}^n(t, y) dy = e_n^\top \int_B \widehat{\psi}(t, y) dy, \tag{B.2}$$

$$\mathsf{E}\{q(Z_0)\} = p_0^n \int_B \phi^n(y) dy = e_n^\top \int_B \mathrm{diag}(p_0)\phi(y) dy, \tag{B.3}$$

$$\int_0^t \lambda(s)\widehat{\mathsf{q}}_{s-} ds = \int_0^t \lambda(s)\mathsf{E}\{\theta_{s-}e_n^\top \theta_{s-}\mathbf{I}_B(Y_{s-})|O_{s-}\} ds \tag{B.4}$$

$$= \int_0^t \lambda(s)\,\mathrm{diag}(e_n) \int_B \widehat{\psi}(s-, y) dy\, ds = e_n^\top \int_B \left[\int_0^t \mathrm{diag}(\lambda(s))\widehat{\psi}(s-, y) ds\right] dy,$$

$$\int_0^t \mathsf{E}_q^\top \widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-} ds = \int_0^t e_n^\top \,\mathrm{diag}\left(\int_B \pi(y) dy\right)\widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-} ds = e_n^\top \,\mathrm{diag}\left(\int_B \pi(y) dy\right)\int_0^t \widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-} ds, \tag{B.5}$$

$$\int_0^t \widehat{qg^\top}_{s-} R^{-1/2} dv_s = \int_0^t \mathsf{E}\{e_n^\top \theta_{s-}\mathbf{I}_B(Y_{s-})g^\top(\theta_{s-}, Y_{s-})|O_{s-}\} R^{-1/2} dv_s \tag{B.6}$$

$$= e_n^\top \int_0^t \left[\int_B \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)\overline{g}^\top(y) dy\right] R^{-1/2} dv_s$$

$$= e_n^\top \int_B \left[\int_0^t \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)\overline{g}^\top(y) R^{-1/2} dv_s\right] dy,$$

$$\int_0^t \widehat{q}_{s-}\widehat{g}_{s-}^\top R^{-1/2} dv_s = \int_0^t e_n^\top \left[\int_B \widehat{\psi}(s-, y) dy\right]\widehat{g}_{s-}^\top R^{-1/2} dv_s \tag{B.7}$$

$$= e_n^\top \int_B \left[\int_0^t \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)(\widehat{g}_{s-}\mathbf{1})^\top R^{-1/2} dv_s\right] dy,$$

$$\int_0^t \widehat{qh^\top}_{s-}\,\mathrm{diag}^{-1}(\widehat{h}_{s-}) d\zeta_s = e_n^\top \int_B \left[\int_0^t \mathrm{diag}\left(\widehat{\psi}(s-, y)\right)\overline{h}^\top(y)\,\mathrm{diag}^{-1}(\widehat{h}_{s-}) d\zeta_s\right] dy, \tag{B.8}$$

$$\int_0^t \widehat{q}_{s-} \widehat{h}_{s-}^\top \operatorname{diag}^{-1}(\widehat{h}_{s-})d\zeta_s = e_n^\top \int_B \left[\int_0^t \operatorname{diag}\left(\widehat{\psi}(s-,y)\right)\left(\widehat{h}_{s-}\mathbf{1}\right)^\top \operatorname{diag}^{-1}(\widehat{h}_{s-})d\zeta_s \right]dy. \qquad \text{(B.9)}$$

The change of the integration order in (B.4)–(B.9) is proper due to the Fubini theorem. From the expressions above and (B.1), it follows that

$$e_n^\top \int_B \Big[\widehat{\psi}(t,y) - \operatorname{diag}(p_0)\phi(y) - \int_0^t \left(\operatorname{diag}(\lambda(s))\widehat{\psi}(s-,y) + \operatorname{diag}\left(\int_B \pi(y)dy\right)\widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-}\right)ds \qquad \text{(B.10)}$$

$$- \int_0^t \operatorname{diag}\left(\widehat{\psi}(s-,y)\right)\left(\overline{g}^\top(y) - \widehat{g}_{s-}\mathbf{1}\right)^\top R^{-1/2}dv_s$$

$$- \int_0^t \operatorname{diag}\left(\widehat{\psi}(s-,y)\right)\left(\overline{h}^\top(y) - \widehat{h}_{s-}\mathbf{1}\right)^\top \operatorname{diag}^{-1}\left(\widehat{h}_{s-}\right)d\zeta_s\Big]dy = 0.$$

From the arbitrariness of $n \in \{1, \dots, N\}$ and $B \in \mathcal{B}(\mathbb{R}^M)$, it follows that (B.10) holds when the equality

$$\widehat{\psi}(t,y) = \operatorname{diag}(p_0)\phi(y) + \int_0^t \left[\operatorname{diag}\lambda(s)\widehat{\psi}(s-,y) + \operatorname{diag}\pi(y)\widetilde{\Lambda}^\top(s)\widehat{\theta}_{s-}\right]ds$$

$$+ \int_0^t \operatorname{diag}\left(\widehat{\psi}(s-,y)\right)(\overline{g}(y) - \widehat{g}_{s-}\mathbf{1})^\top R^{-1/2}dv_s + \int_0^t \operatorname{diag}\left(\widehat{\psi}(s-,y)\right)\left(\overline{h}(y) - \widehat{h}_{s-}\mathbf{1}\right)^\top \operatorname{diag}^{-1}(\widehat{h}_{s-})d\zeta_s,$$

is true $\mathcal{P}$-a.s. and almost everywhere with respect to the Lebesgue measure. The theorem is proved.