sicon 2004 **43** 2 pp.643–669

ASYMPTOTIC STABILITY OF THE WONHAM FILTER: ERGODIC AND NONERGODIC SIGNALS*

PETER BAXENDALE†, PAVEL CHIGANSKY‡, AND ROBERT LIPTSER‡

Abstract. The stability problem of the Wonham filter with respect to initial conditions is addressed. The case of ergodic signals is revisited in view of a gap in the classic work of H. Kunita (1971). We give new bounds for the exponential stability rates, which do not depend on the observations. In the nonergodic case, the stability is implied by identifiability conditions, formulated explicitly in terms of the transition intensities matrix and the observation structure.

Key words. nonlinear filtering, stability, Wonham filter

AMS subject classifications. 93E11, 60J57

1. Introduction.

The optimal filtering estimate of a signal from the record of noisy observations is usually generated by a nonlinear recursive equation subject to the signal a priori distribution. If the latter is unknown and the filtering equation is initialized by an arbitrary initial distribution, the obtained estimate is suboptimal in general. From an applications point of view, it is important to know whether such an estimate becomes close to the optimal one at least after enough time elapses. This property of filters to forget the initial conditions is far from being obvious and in fact generally remains an open and challenging problem.

In this paper, we consider the filtering setting for signals with a finite state space. Specifically, let $X = (X_t)_{t\geq 0}$ be a continuous time homogeneous Markov chain observed via

$$Y_t = \int_0^t h(X_s)ds + \sigma W_t \tag{1.1}$$

with the Wiener process $W = (W_t)_{t\geq 0}$, independent of X, some bounded function h, and $\sigma \neq 0$.

We assume that X_t takes values in the finite alphabet $\mathbb{S} = \{a_1, \dots, a_n\}$ and admits several ergodic classes. Namely,

$$\mathbb{S} = \Big\{ \underbrace{a_1^1, \dots, a_{n_1}^1}_{\mathbb{S}_1}, \dots, \underbrace{a_1^m, \dots, a_{n_m}^m}_{\mathbb{S}_m} \Big\},$$

where the subalphabets $\mathbb{S}_1, \dots, \mathbb{S}_m$ are noncommunicating in the sense that for any $i \neq j$ and $t \geq s$

$$P\big(X_t \in \mathbb{S}_j | X_s \in \mathbb{S}_i\big) = 0. \tag{1.2}$$

So, unless $m = 1$, X_t is a compound Markov chain with the transition intensities

*Received by the editors November 1, 2002; accepted for publication (in revised form) October 16, 2003; published electronically July 23, 2004.

http://www.siam.org/journals/sicon/43-2/41692.html

†Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113 (baxendal@math.usc.edu). The research of this author was supported by ONR grant N00014-96-1-0413.

‡Department of Electrical Engineering Systems, Tel Aviv University, 69978 Tel Aviv, Israel (pavelm@eng.tau.ac.il, liptser@eng.tau.ac.il).

matrix

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & 0 \\ 0 & \Lambda_2 & 0 \\ \dots & \dots & \dots \\ 0 & 0 & \Lambda_m \end{pmatrix} \tag{1.3}$$

of m ergodic classes and is not ergodic itself.

The filtering problem consists in computation of the conditional distribution,

$$\pi_t^\nu(1) = P(X_t^\nu = a_1 | \mathscr{Y}_{[0,t]}^\nu), \dots, \pi_t^\nu(n) = P(X_t^\nu = a_n | \mathscr{Y}_{[0,t]}^\nu),$$

where $\mathscr{Y}_{[0,t]}^\nu$ is the filtration, generated by $\{Y_s^\nu, 0 \le s \le t\}$ satisfying the usual conditions (henceforth, the superscript ν is used to emphasize that the distribution of X_0 is ν).

The vector-valued random process π_t^ν with entries $\pi_t^\nu(1), \dots, \pi_t^\nu(n)$ is generated by the Wonham filter [45] (see also [29, Chap. 9])

$$\begin{aligned} \pi_0^\nu &= \nu, \\ d\pi_t^\nu &= \Lambda^* \pi_t^\nu dt + \sigma^{-2} \big(\operatorname{diag}(\pi_t^\nu) - \pi_t^\nu (\pi_t^\nu)^* \big) h (dY_t^\nu - h^* \pi_t^\nu dt), \end{aligned} \tag{1.4}$$

where $\operatorname{diag}(x)$ is the scalar matrix with the diagonal $x \in \mathbb{R}^n$, h is the column vector with entries $h(a_1), \dots, h(a_n)$, and * is the transposition symbol. If ν is unknown and some other distribution β (on $\mathbb{S}$) is used to initialize the filter, the "wrong" conditional distribution $\pi_t^{\beta\nu}$ is obtained:

$$\begin{aligned} \pi_0^{\beta\nu} &= \beta, \\ d\pi_t^{\beta\nu} &= \Lambda^* \pi_t^{\beta\nu} dt + \sigma^{-2} \big(\operatorname{diag}(\pi_t^{\beta\nu}) - \pi_t^{\beta\nu} (\pi_t^{\beta\nu})^* \big) h (dY_t^\nu - h^* \pi_t^{\beta\nu} dt). \end{aligned} \tag{1.5}$$

According to the intuitive notion of stability, given at the beginning of this section, the filter defined in (1.5) is said to be asymptotically stable if

$$\lim_{t \to \infty} E \| \pi_t^\nu - \pi_t^{\beta\nu} \| = 0, \tag{1.6}$$

where $\| \cdot \|$ is the total variation norm.

If the state space of the Markov chain X consists of one ergodic class ($m = 1$), our setting is in the framework studied by Ocone and Pardoux [35]. In this case, there exists the unique invariant distribution μ, so that

$$\lim_{t \to \infty} \| S_t \gamma - \mu \| = 0, \tag{1.7}$$

where S_t is the semigroup corresponding to X and γ is an arbitrary probability distribution on $\mathbb{S}$. Moreover,

$$\lim_{t \to \infty} \int_{\mathbb{S}} | \mathcal{S}_t f(x) - \mu(f) | d\mu(x) = 0 \tag{1.8}$$

holds for any bounded $f : \mathbb{S} \mapsto \mathbb{R}$. So, it may seem that it remains only to assume

$$\nu \ll \beta \tag{1.9}$$

and allude to [35]. However, the proof of (1.6) given in [35] uses as its central argument the uniqueness theorem for the stationary measure of the filtering process π_t^ν which appeared in the work of H. Kunita [22]. Unfortunately, the proof of this theorem (Theorem 3.3 in [22]) contains a serious gap, as elaborated in the next section.

A different approach to the stability analysis of the filters for ergodic signals was initiated by Delyon and Zeitouni [19]. The authors studied the top Lyapunov exponent of the filtering equation

$$\gamma_\sigma(\beta', \beta'') = \overline{\lim_{t\to\infty}} \frac{1}{t} \log \left\| \pi_t^{\beta'\nu} - \pi_t^{\beta''\nu} \right\|, \quad \beta' \text{ and } \beta'' \text{ distributions on } \mathbb{S},$$

and showed that $\gamma_\sigma(\beta', \beta'') < 0$ too when Λ and h satisfy certain conditions. Moreover, the filter is found to be stable in the low signal-to-noise regime: $\overline{\lim}_{\sigma\to\infty} \gamma_\sigma(\beta', \beta'') \le \Re\big[\lambda^{\max}\big(\Lambda\big)\big]$ with $\lambda^{\max}(\Lambda)$ being the eigenvalue of Λ with the largest nonzero real part.

These results were further extended by Atar and Zeitouni [3], where it is shown that uniformly in $\sigma > 0$ and h

$$\gamma_\sigma(\beta', \beta'') \le -2 \min_{p\ne q} \sqrt{\lambda_{pq}\lambda_{qp}}, \text{ a.s.,} \tag{1.10}$$

and the high signal-to-noise asymptotics are obtained:

$$\overline{\lim_{\sigma\to 0}}\, \sigma^2\gamma_\sigma \le -\frac{1}{2}\sum_{i=1}^{d} \mu_i \min_{j\ne i} \big[h(a_i) - h(a_j)\big]^2,$$

$$\underline{\lim_{\sigma\to 0}}\, \sigma^2\gamma_\sigma \ge -\frac{1}{2}\sum_{i=1}^{d} \mu_i \sum_{j=1}^{d} \big[h(a_i) - h(a_j)\big]^2,$$

where μ is the ergodic measure of X.

The method in [3] (and its full development in [2]) does not rely on [22] and is based on the analysis of the Zakai equation, corresponding to (1.4) (see (5.2) below). The analysis is carried out by means of the Hilbert projective metric and the Birkhoff inequality, etc.; see section 5 for more details. This approach proved out its efficiency in several filtering scenarios (see [1], [9], [11]).

Other results and methods related to the filtering stability can be found in [4], [10], [12], [13], [14], [16], [17], [18], [15], [24], [25], [26], [27], [36], [37]. The linear Kalman–Bucy case, being the most understood, is extensively treated by several authors: [5], [32], [33], [19], [35], [28], [30] (sections 14.6 and 16.2).

In the present paper, we consider both ergodic and nonergodic signals. Applying the technique from Atar and Zeitouni [2], we show that in the ergodic case the asymptotic stability holds true without any additional assumptions. In other words, the conclusion of H. Kunita [22] is valid in the specific case under consideration.

In view of the counterexample given in section 3, it is clear that in general γ_σ may vanish at $\sigma = 0$. So, it is interesting to find out which ergodic properties of the signal are inherited by the filter regardless of the specific observation structure. In this connection we prove the inequality

$$\overline{\lim_{t\to\infty}} \frac{1}{t} \log \|\pi_t^{\beta\nu} - \pi_t^{\nu}\| \le -\sum_{r=1}^{n} \mu_r \min_{i\ne r} \lambda_{ri}.$$

Since μ is the positive measure on $\mathbb{S}$, unlike (1.10), this bound remains negative if at least one row of Λ has all nonzero entries.

Also we give the nonasymptotic bound (compare with (1.10))

$$\|\pi_t^{\nu} - \pi_t^{\beta\nu}\| \le C \exp\Big(-2t \min_{p\ne q} \sqrt{\lambda_{pq}\lambda_{qp}}\Big)$$

with some positive constant C depending on ν and β only.

For the discrete time case, related results can be found in Del Moral and Guionnet [18] and Le Gland and Mevel [24]. For example, in [24] the positiveness assumption for all transition probabilities is relaxed under certain constraints on the observation process noise density.

In the case of nonergodic signal, $m > 1$, we show that the filtering stability holds true if the ergodic classes can be identified via observations and the filter matched to each class is stable. We formulate explicit sufficient identifiability conditions in terms of Λ and h.

The paper is organized as follows. In section 2, we introduce the necessary notations and clarify the role of condition $\nu \ll \beta$ in the filtering stability (Proposition 2.1). This section also gives a link to the gap in Kunita's proof [22], while in section 3 the filtering setting is described for which the stability fails and the gap becomes evident.

The main results are formulated in section 4 and proved in sections 5 and 6.

2. Preliminaries and connection to the gap in [22].

2.1. Notations. Throughout, $\nu \ll \beta$ is assumed.

In order to explain our approach, let us consider a general setting when (X, Y) is a Markov process with paths from the Skorokhod space $\mathbb{D} = \mathbb{D}_{[0,\infty)}(\mathbb{R}^2)$ of right continuous functions having limits to the left functions. Moreover, the signal component X is a Markov process itself.

We introduce a measurable space $(\mathbb{D}, \mathscr{D})$, where $\mathscr{D} = \sigma\{(x_s, y_s), s \geq 0\}$ is the Borel σ-algebra on $\mathbb{D}$. Let $D = (\mathscr{D}_t)_{t\geq 0}$ be the filtration of $\mathscr{D}_t = \sigma\{(x_s, y_s), s \leq t\}$ and let $D^y = (\mathscr{D}^y_t)_{t\geq 0}$ be the filtration of $\mathscr{D}^y_t = \sigma\{y_s, s \leq t\}$.

As before, we write (X^ν_t, Y^ν_t) and (X^β_t, Y^β_t), when the distribution of X_0 is ν or β, respectively, meaning that both pairs are defined on the same probability space, have the same transition semigroup, but different initial distributions.

For a bounded measurable function f, we introduce $\pi^\nu_t(f) := E(f(X^\nu_t)|\mathscr{Y}^\nu_{[0,t]})$ and $\pi^\beta_t(f) := E(f(X^\beta_t)|\mathscr{Y}^\beta_{[0,t]})$. Since $\pi^\nu_t(f)$ and $\pi^\beta_t(f)$ are $\mathscr{Y}^\nu_{[0,t]}$- and $\mathscr{Y}^\beta_{[0,t]}$-measurable random variables, respectively, it is convenient to identify $\pi^\nu_t(f)$ and $\pi^\beta_t(f)$ with some $\mathscr{D}^y_t$-measurable functionals of trajectories $Y^\nu_{[0,t]} = \{Y^\nu_s, s \leq t\}$ and $Y^\beta_{[0,t]} = \{Y^\beta_s, s \leq t\}$.

For this purpose, let Q^ν and Q^β denote the distributions of (X^ν, Y^ν) and (X^β, Y^β) on $(\mathbb{D}, \mathscr{D})$, respectively, and Q^ν_t and Q^β_t be their restrictions on $[0, t]$, so that Q^ν_0, Q^β_0 are the distributions of $(X^\nu_0, Y^\nu_0), (X^\beta_0, Y^\beta_0)$. We also assume that

$$\frac{dQ^\nu_0}{dQ^\beta_0}(x, y) = \frac{d\nu}{d\beta}(x_0). \tag{2.1}$$

Since (X^ν_t, Y^ν_t) and (X^β_t, Y^β_t) have the same transition law, we have $Q^\nu \ll Q^\beta$ with

$$\frac{dQ^\nu}{dQ^\beta}(x, y) = \frac{d\nu}{d\beta}(x_0).$$

Without loss of generality, we assume that the filtrations D and D^y satisfy the general conditions with respect to $(Q^\nu + Q^\beta)/2$.

For fixed t, let $H^\beta_t(y)$ be a $\mathscr{D}^y_t$-measurable functional so that $H^\beta_t(Y^\beta) = \pi^\beta_t(f)$ a.s. Moreover, due to $Q^\nu \ll Q^\beta$, a version of $H^\beta_t(y)$ can be chosen such that the random variable $H^\beta_t(Y^\nu)$ is well defined. Then we identify $\pi^{\beta\nu}_t(f)$ with $H^\beta_t(Y^\nu)$.

We do not assume that $\beta \ll \nu$ (and thus $Q^\beta \not\ll Q^\nu$), so this construction fails for $\pi_t^{\nu\beta}(f)$. Nevertheless, a version of $H_t^\nu(y)$ can be chosen such that $H_t^\nu(Y^\nu) = \pi_t^\nu(f)$ a.s. and used for the definition of $\pi_t^{\nu\beta}(f)$. Indeed, let $\overline{Q}^\beta$ and $\overline{Q}^\nu$ be the distributions of Y^ν and Y^β, respectively, i.e., the marginal distributions of Q^β and Q^ν, obviously, $\overline{Q}^\nu \ll \overline{Q}^\beta$ as well as $\overline{Q}_t^\nu \ll \overline{Q}_t^\beta$; the restrictions of $\overline{Q}^\nu$ and $\overline{Q}^\beta$ on the interval $[0,t]$. Moreover, $\frac{d\overline{Q}_t^\nu}{d\overline{Q}_t^\beta}(Y^\beta) = E(\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}_{[0,t]}^\beta)$. Now define

$$\pi_t^{\nu\beta}(f) := H_t^\nu(Y^\beta) I\left(\frac{d\overline{Q}_t^\nu}{d\overline{Q}_t^\beta}(Y^\beta) > 0\right).$$

We introduce the decreasing filtration $\mathscr{X}_{[t,\infty)}^\beta = \sigma\{X_s^\beta, s \geq t\}$, the tail σ-algebra

$$\mathscr{T}(X^\beta) = \bigcap_{t\geq 0} \mathscr{X}_{[t,\infty)}^\beta, \tag{2.2}$$

and σ-algebras $\mathscr{X}_t^\beta = \sigma\{X_t^\beta\}$, $\quad \mathscr{Y}_{[0,\infty)}^\beta = \bigvee_{t\geq 0} \mathscr{Y}_{[0,t]}^\beta$.

Set

$$\pi_t^{\beta_0}(f) = E\big(f(X_t^\beta)|\mathscr{Y}_{[0,t]}^\beta \vee \mathscr{X}_0^\beta\big). \tag{2.3}$$

2.2. Filter stability. For bounded and measurable f, the estimate $\pi_t^\nu(f)$ is asymptotically stable with respect to β if

$$\lim_{t\to\infty} E\big|\pi_t^\nu(f) - \pi_t^{\beta\nu}(f)\big| = 0. \tag{2.4}$$

Note that, when the signal process takes values in a finite alphabet and (2.4) holds for any bounded f, then (2.4) and (1.6) are equivalent.

We establish below that (2.4) holds if for large values of t the additional measurement X_0^β is useless for estimation of $f(X_t^\beta)$ via $Y_{[0,t]}^\beta$ or, analogously, if the additional measurement X_t^β is useless for estimation of $\frac{d\nu}{d\beta}(X_0^\beta)$ via $Y_{[0,\infty)}^\beta$.

Proposition 2.1. *Assume $\nu \ll \beta$. Then, any of the conditions*

1.

$$\lim_{t\to\infty} E\big|\pi_t^\beta(f) - \pi_t^{\beta_0}(f)\big| = 0, \tag{2.5}$$

2.

$$E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,\infty)}^\beta\Big) = \lim_{t\to\infty} E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,\infty)}^\beta \vee \mathscr{X}_{[t,\infty)}^\beta\Big) \tag{2.6}$$

provides (2.4).

Proof. Let us first show that, under $\nu \ll \beta$, for any bounded f

$$\begin{aligned} &E\big|\pi_t^{\beta\nu}(f) - \pi_t^\nu(f)\big| \\ &= E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big)E\Big(f(X_t^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big) - E\Big(\frac{d\nu}{d\beta}(X_0^\beta)f(X_t^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big)\Big|. \end{aligned} \tag{2.7}$$

Write

$$\begin{aligned} E\big|\pi_t^{\beta\nu}(f) - \pi_t^\nu(f)\big| &= E\frac{d\nu}{d\beta}(X_0^\beta)\big|\pi_t^\beta(f) - \pi_t^{\nu\beta}(f)\big| \\ &= EE\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big)\big|\pi_t^\beta(f) - \pi_t^{\nu\beta}(f)\big| = E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big)\big(\pi_t^\beta(f) - \pi_t^{\nu\beta}(f)\big)\Big| \\ &= E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big)E\Big(f(X_t^\beta)\big|\mathscr{Y}_{[0,t]}^\beta\Big) - E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\pi_t^{\nu\beta}(f)\big|\mathscr{Y}_{[0,t]}^\beta\Big)\Big|. \end{aligned}$$

So, it remains to show

$$E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\pi_t^{\nu\beta}(f)\big|\mathscr{Y}^\beta_{[0,t]}\Big)=E\Big(\frac{d\nu}{d\beta}(X_0^\beta)f(X_t^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big). \tag{2.8}$$

With $\mathscr{D}_t^y$-measurable and bounded function $\Psi_t(y)$ we get

$$\begin{aligned}&E\Big\{\Psi_t(Y^\beta)E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\pi_t^{\nu\beta}(f)\big|\mathscr{Y}^\beta_{[0,t]}\Big)\Big\}=E\Big(\Psi_t(Y^\beta)\frac{d\nu}{d\beta}(X_0^\beta)\pi_t^{\nu\beta}(f)\Big)\\&=E\Big(\Psi_t(Y^\nu)\pi_t^\nu(f)\Big)=E\Big(\Psi_t(Y^\nu)f(X_t^\nu)\Big)=E\Big(\Psi_t(Y^\beta)\frac{d\nu}{d\beta}(X_0^\beta)f(X_t^\beta)\Big)\end{aligned}$$

and notice that (2.8) is valid by the arbitrariness of Ψ_t.

The proof of (2.5)⇒(2.4). Using (2.7) and

$$E\Big(\frac{d\nu}{d\beta}(X_0^\beta)f(X_t^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big)=E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\pi_t^{\beta_0}(f)\big|\mathscr{Y}^\beta_{[0,t]}\Big),$$

we derive

$$\begin{aligned}&E\big|\pi_t^{\beta\nu}(f)-\pi_t^\nu(f)\big|=E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big)\pi_t^\beta(f)-E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\pi^{\beta_0}(f)\big|\mathscr{Y}^\beta_{[0,t]}\Big)\Big|\\&=E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big(\pi_t^\beta(f)-\pi^{\beta_0}(f)\big)\big|\mathscr{Y}^\beta_{[0,t]}\Big)\Big|\le E\frac{d\nu}{d\beta}(X_0^\beta)\big|\pi_t^\beta(f)-\pi_t^{\beta_0}(f)\big|,\end{aligned}$$

where the Jensen inequality has been used. Let for definiteness $|f|\le K$ with some constant K. Then $\pi_t^\beta(f)$, $\pi_t^{\beta_0}(f)$ can also be chosen such that $|\pi_t^\beta(f)|$ and $|\pi_t^{\beta_0}(f)|$ are bounded by K. Hence, for any $C>0$, we have

$$E\big|\pi_t^{\beta\nu}(f)-\pi_t^\nu(f)\big|\le CE\big|\pi_t^\beta(f)-\pi_t^{\beta_0}(f)\big|+2KP\Big(\frac{d\nu}{d\beta}(X_0^\beta)>C\Big).$$

Therefore, $\overline{\lim}_{t\to\infty}E\big|\pi_t^{\beta\nu}(f)-\pi_t^\nu(f)\big|\le 2KP\big(\frac{d\nu}{d\beta}(X_0^\beta)>C\big)$ and by the Chebyshev inequality $P\big(\frac{d\nu}{d\beta}(X_0^\beta)>C\big)\le C^{-1}\to 0,\ C\to\infty$.

The proof of (2.6)⇒(2.4). By (2.7)

$$\begin{aligned}&E\big|\pi_t^{\beta\nu}(f)-\pi_t^\nu(f)\big|\\&=E\left|E\Big(f(X_t^\beta)E\Big[\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}^\beta_{[0,t]}\Big]\Big|\mathscr{Y}^\beta_{[0,t]}\Big)-E\Big(f(X_t^\beta)\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}^\beta_{[0,t]}\Big)\right|.\end{aligned}$$

Notice also

$$E\Big(f(X_t^\beta)\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}^\beta_{[0,t]}\Big)=E\Big(f(X_t^\beta)E\Big[\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big]\big|\mathscr{Y}^\beta_{[0,t]}\Big).$$

Since $|f|\le K$, by the Jensen inequality we have

$$\begin{aligned}&E\big|\pi_t^{\beta\nu}(f)-\pi_t^\nu(f)\big|\\&\qquad\le KE\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}^\beta_{[0,t]}\Big)-E\Big(\frac{d\nu}{d\beta}(X_0^\beta)|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)\Big|.\end{aligned}\tag{2.9}$$

Both random processes $E(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,t]})$ and $E(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)})$ are uniformly integrable forward and backward martingales with respect to the filtrations $(\mathscr{Y}^\beta_{[0,t]})_{t\ge0}$ and $(\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)})_{t\ge0}$. Therefore, they admit limits a.s. in $t\to\infty$: $E(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)})$ and $\lim\limits_{t\to\infty}E(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)})$, respectively. By (2.6)

$$\lim_{t\to\infty}\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big)-E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)\Big|=0.$$

We show also that

$$\lim_{t\to\infty}E\Big|E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big)-E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)\Big|=0. \tag{2.10}$$

Denote by α_t any of $E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,t]}\Big)$ and $E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)$ and

$$\alpha_\infty=\lim_{t\to\infty}\alpha_t.$$

It is clear that (2.10) holds true if $\lim_{t\to\infty}E|\alpha_t-\alpha_\infty|=0$. Since $\lim_{t\to\infty}\alpha_t=\alpha_\infty$, $\alpha_t\ge0$, and $E\alpha_t\equiv E\alpha_\infty=1$, by the Scheffe theorem we get the desired property.

Thus the right-hand side of (2.9) converges to zero and the result follows. $\square$

2.3. Connection to the gap in [22]. In [22], H. Kunita studies[1] ergodic properties of the filtering process π_t^ν. He considers π_t^ν as a Markov process with values in the space of probability measures and claims (in Theorem 3.3) that there exists the unique invariant measure being "limit point" of marginal distributions of π_t^ν, $t\nearrow\infty$. As was later shown in [35], this result is the key to the stability analysis under (1.8).

Below we demonstrate that the main argument, used in the proof of Theorem 3.3 of [22], cannot be taken for granted. We discuss this issue in the context of Proposition 2.1. Suppose the Markov process X is ergodic in the sense of (1.7) and (1.8). It is well known that its tail σ-algebra $\mathscr{T}(X^\beta)$ (see (2.2) for definition) is empty a.s. It is very tempting in this case to change the order of intersection and supremum as follows:

$$\bigcap_{t\ge0}\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}=\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{T}(X^\beta)\quad\text{a.s.} \tag{2.11}$$

Then, the right-hand side of (2.6) is transformed to

$$\begin{aligned}\lim_{t\to\infty}E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)&=E\left(\frac{d\nu}{d\beta}(X_0^\beta)\big|\bigcap_{t\ge0}\Big\{\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big\}\right)\\&=E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{T}(X^\beta)\Big)=E\Big(\frac{d\nu}{d\beta}(X_0^\beta)\big|\mathscr{Y}^\beta_{[0,\infty)}\Big)\end{aligned}$$

and (2.6) would be correct, regardless (!) of any other ingredients of the problem (e.g., with $\sigma=0$ in (1.1)).

In [22], the relation of (2.11) type plays the key role in verification of the uniqueness for the invariant measure corresponding to $\pi_t^\nu, t\ge0$. However, the validity of (2.11) is far from being obvious. According to Williams [44], it "...tripped up even

[1]The notations of this paper are used here.

Kolmogorov and Wiener" (see Sinai [39, p. 837] for some details). The reader can find a discussion concerning (2.11) in von Weizsäcker [43]; unfortunately, the counterexample there is incorrect. A proper counterexample to (2.11) is given in Exercise 4.12 in Williams [44], which, however, seems somewhat artificial in the filtering context. It turns out that the example, considered by Delyon and Zeitouni in [19] (see [21] by Kaijser for its earlier discrete time version), is nothing but another case when (2.11) fails.

For the reader's convenience, we give below a detailed analysis of this example.

It is important to note that the counterexamples mentioned above do not fit exactly into the setup considered by Kunita. They merely indicate that (2.11) is not evident and so the claim of Theorem 3.3 in [22] remains a conjecture.

Generally, the stability of nonlinear filters for ergodic Markov processes remains an open problem, and some results [23], [40], [41], [6], [8], [7], [35] based on [22] have to be revised.

3. Counterexample. Below we give a detailed discussion of one counterexample to (2.11). Consider Markov process X with values in $\mathbb{S} = \{1, 2, 3, 4\}$, with the initial distribution ν and the transition intensities matrix

$$\Lambda = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \tag{3.1}$$

All states of Λ communicate, and so X is an ergodic Markov process (see, e.g., [34]) with the unique invariant measure $\mu = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$. Let $h(x) = I(x = 1) + I(x = 3)$, that is,

$$Y_t = \int_0^t \big[I(X_s = 2) + I(X_s = 3)\big] ds + \sigma W_t.$$

By Theorem 4.1 below, the filter is stable in this case for any $\sigma > 0$.

3.1. Noiseless observation. Consider the case $\sigma = 0$.

It will be convenient to redefine the observation process as follows:

$$Y_t = [I(X_t = 1) + I(X_t = 3)].$$

We assume $\nu \ll \beta$ and notice that (2.1) holds true. We omit the superscripts ν and β when the initial condition does not play a significant role. Since X is an ergodic Markov process, satisfying (1.8), $\mathscr{T}(X) = (\Omega, \varnothing)$ a.s.

PROPOSITION 3.1.

$$\bigcap_{t \geq 0} \Big(\mathscr{Y}_{[0,\infty)} \vee \mathscr{X}_{[t,\infty)}\Big) \supsetneq \mathscr{Y}_{[0,\infty)} \ \ a.s. \tag{3.2}$$

Proof. It suffices to show that X_0 is a $\bigcap_{t\geq 0}(\mathscr{Y}_{[0,\infty)} \vee \mathscr{X}_{[t,\infty)})$-measurable random variable and at the same time $X_0 \notin \mathscr{Y}_{[0,\infty)}$.

The structure of matrix Λ admits only cyclic transitions in the following order:

$$\cdots \to \{3\} \to \{4\} \to \{1\} \to \{2\} \to \{3\} \to \cdots.$$

TABLE 3.1
Typical trajectory of π_t for $Y_0 = 1$.

t	$[0,\tau_1)$	$[\tau_1,\tau_2)$	$[\tau_2,\tau_3)$	$[\tau_3,\tau_4)$	$[\tau_4,\tau_5)$	...
Y_t	1	0	1	0	1	...
$\pi_t(1)$	$\frac{\nu_1}{\nu_1+\nu_3}$	0	$\frac{\nu_3}{\nu_1+\nu_3}$	0	$\frac{\nu_1}{\nu_1+\nu_3}$	...
$\pi_t(2)$	0	$\frac{\nu_1}{\nu_1+\nu_3}$	0	$\frac{\nu_3}{\nu_1+\nu_3}$	0	...

So, since Y and X jump simultaneously, X_0 can be recovered exactly from the trajectory $Y_s, s \le t$, and X_t for any $t > 0$, i.e., X_0 is $\mathscr{X}_t \vee \mathscr{Y}_{[0,t]}$-measurable. Owing to $\mathscr{X}_t \vee \mathscr{Y}_{[0,t]} \subset \mathscr{X}_{[t,\infty)} \vee \mathscr{Y}_{[0,\infty)}$, X_0 is measurable with respect to

$$\bigcap_{t\ge 0}\Big(\mathscr{Y}_{[0,\infty)} \vee \mathscr{X}_{[t,\infty)}\Big).$$

Denote by $(\tau_i)_{i\ge 1}$ the time moments where Y jumps. It is not hard to check that $(\tau_i)_{i\ge 0}$ is independent of (X_0, Y_0) and, moreover,

$$\mathscr{Y}_{[0,t]} = \bigvee_{i\ge 0}\sigma\{\tau_i \le t\} \vee \sigma\{Y_0\}.$$

Thus for any $t \ge 0$

$$\begin{aligned} P\big(X_0 = 1|\mathscr{Y}_{[0,t]}\big) &= P\Bigg(X_0 = 1|\bigvee_{i\ge 0}\sigma\{\tau_i \le t\} \vee \sigma\{Y_0\}\Bigg) \\ &= P\big(X_0 = 1|Y_0\big) = \frac{\nu_1}{\nu_1+\nu_3}Y_0. \end{aligned} \tag{3.3}$$

Since (3.3) is valid for any $t \ge 0$, we conclude that

$$P\big(X_0 = 1|\mathscr{Y}_{[0,\infty)}\big) = \frac{\nu_1}{\nu_1+\nu_3}Y_0.$$

Obviously $I(X_0 = 1) \ne \frac{\nu_1}{\nu_1+\nu_3}Y_0$ and thus X_0 is not $\mathscr{Y}_{[0,\infty)}$-measurable. $\square$

3.2. Invariant measures of π_t and the filter instability. Since $I_t(2) + I_t(4) = 1 - Y_t$ and $I_t(1) + I_t(3) = Y_t$, only $I_t(1)$ and $I_t(2)$ have to be filtered while $\pi_t(3) = Y_t - \pi_t(1)$ and $\pi_t(4) = (1-Y_t) - \pi_t(2)$. The derivation of the filtering equations is sketched in the appendix.

PROPOSITION 3.2. *The optimal filtering estimate satisfies*

$$\begin{aligned} d\pi_t(1) &= \big(1 - \pi_{t-}(2)\big)(1 - Y_{t-})dY_t + \pi_{t-}(1)Y_{t-}dY_t, \\ d\pi_t(2) &= -\pi_{t-}(2)(1 - Y_{t-})dY_t - \pi_{t-}(1)Y_{t-}dY_t \end{aligned}$$

subject to $\pi_0(1) = \frac{\nu_1}{\nu_1+\nu_3}Y_0$, $\pi_0(2) = \frac{\nu_2}{\nu_2+\nu_4}(1 - Y_0)$.

Let us examine the behavior of the filter from Proposition 3.2. A pair of typical trajectories are given in Table 3.1 (for $Y_0 = 1$) and Table 3.2 (for $Y_0 = 0$).

It is not hard to see that Y is itself a Markov chain with values in $\{0, 1\}$ and the transition intensities matrix $\left(\begin{smallmatrix}-1 & 1\\ 1 & -1\end{smallmatrix}\right)$, and thus its invariant measure is $\mu' = \big(1/2 \quad 1/2\big)$. Hence, the invariant measure Φ of the filtering process $(\pi_t(1), \pi_t(2))$ is

TABLE 3.2
Typical trajectory of π_t for $Y_0 = 0$.

t	$[0,\tau_1)$	$[\tau_1,\tau_2)$	$[\tau_2,\tau_3)$	$[\tau_3,\tau_4)$	$[\tau_4,\tau_5)$	...
Y_t	0	1	0	1	0	...
$\pi_t(1)$	0	$\frac{\nu_2}{\nu_2+\nu_4}$	0	$\frac{\nu_4}{\nu_2+\nu_4}$	0	...
$\pi_t(2)$	$\frac{\nu_2}{\nu_2+\nu_4}$	0	$\frac{\nu_4}{\nu_2+\nu_4}$	0	$\frac{\nu_2}{\nu_2+\nu_4}$	...

concentrated on eight vectors

$$\phi_1 = \begin{pmatrix} \frac{\nu_1}{\nu_1+\nu_3} \\ 0 \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} 0 \\ \frac{\nu_1}{\nu_1+\nu_3} \end{pmatrix}, \quad \phi_3 = \begin{pmatrix} \frac{\nu_3}{\nu_1+\nu_3} \\ 0 \end{pmatrix}, \quad \phi_4 = \begin{pmatrix} 0 \\ \frac{\nu_3}{\nu_1+\nu_3} \end{pmatrix},$$
$$\phi_5 = \begin{pmatrix} \frac{\nu_2}{\nu_2+\nu_4} \\ 0 \end{pmatrix}, \quad \phi_6 = \begin{pmatrix} 0 \\ \frac{\nu_2}{\nu_2+\nu_4} \end{pmatrix}, \quad \phi_7 = \begin{pmatrix} \frac{\nu_4}{\nu_2+\nu_4} \\ 0 \end{pmatrix}, \quad \phi_8 = \begin{pmatrix} 0 \\ \frac{\nu_4}{\nu_2+\nu_4} \end{pmatrix}$$

with

$$\Phi(\phi_i) = (\nu_1+\nu_3)/4, \quad i = 1,2,3,4,$$
$$\Phi(\phi_i) = (\nu_2+\nu_4)/4, \quad i = 5,6,7,8,$$

and, consequently, Φ is not unique. Moreover, the optimal filter is not stable in the sense of (1.6). In fact, for different initial conditions, the filtering distribution $\pi_t, t > 0$, can "sit" on different vectors!

4. Main results.

4.1. Ergodic case. Markov chain X is ergodic if and only if all entries of its transition intensities matrix Λ *communicate*, i.e., for any pair of indices i and j, a string of indices $\{\ell_1, \dots, \ell_m\}$ can be found so that $\lambda_{i\ell_1}\lambda_{\ell_1\ell_2}\dots\lambda_{\ell_m j} \neq 0$ (see, e.g., [34]). In this case, the distribution of X_t converges to the positive invariant distribution μ being the unique solution of $\Lambda^*\mu = 0$ in the class of vectors with positive entries the sum of which is equal to one.

THEOREM 4.1. *If all states of Λ communicate, then there exists a positive constant c such for any ν and β*

$$\overline{\lim_{t\to\infty}} \frac{1}{t} \log \|\pi_t^{\beta\nu} - \pi_t^{\nu}\| < -c \ a.s.$$

Remark 1. Clearly, Theorem 4.1 provides (1.6). Also it allows us to conclude that $\lim_{t\to\infty} \|\pi_t^{\beta\nu} - \pi_t^{\nu}\| = 0$ a.s. for β concentrated in a single state of $\mathbb{S}$. Then, in particular, we have

$$\lim_{t\to\infty} \|\pi_t^{\mu_0} - \pi_t^{\mu}\| = 0$$

which is the main argument in the proof of existence of the unique invariant measure for the process $(\pi_t)_{t\geq 0}$. This fact corroborates Kunita's result from [22] in the finite state space setup of Theorem 4.1.

Actually, Theorem 4.1 verifies the logarithmic rate in $t \to \infty$ which is in general a function of Λ, h and σ. However, stronger assumptions on Λ guarantee exponential or logarithmic rates, regardless of h and σ (σ is only required to be nonzero).

THEOREM 4.2. *Assume all states of* Λ *communicate. Then*

$$\varlimsup_{t\to\infty} \frac{1}{t}\log\|\pi_t^{\beta\nu}-\pi_t^{\nu}\| \le -\sum_{r=1}^{n}\mu_r \min_{i\neq r}\lambda_{ri}. \tag{4.1}$$

Remark 2. The bound (4.1) is negative if at least one row of Λ has all nonzero entries.

THEOREM 4.3. *Assume all entries of* Λ *are nonzero.*

1. *If* $\nu \ll \beta$, *then*

$$E\|\pi_t^{\beta\nu}-\pi_t^{\nu}\| \le n\sum_{j=1}^{n}\frac{d\nu}{d\beta}(a_j)\exp\Big(-2t\min_{p\neq q}\sqrt{\lambda_{pq}\lambda_{qp}}\Big),\ t>0. \tag{4.2}$$

2. *If* $\nu\sim\beta$, *then*

$$\|\pi_t^{\beta\nu}-\pi_t^{\nu}\| \le n^2\max_j\frac{d\nu}{d\beta}(a_j)\max_j\frac{d\beta}{d\nu}(a_j)\exp\Big(-2t\min_{p\neq q}\sqrt{\lambda_{pq}\lambda_{qp}}\Big),\ t>0. \tag{4.3}$$

4.2. Nonergodic case. Let $m\ge 2$ and Λ be given in (1.3). If $X_0\in\mathbb{S}_j$, then X is a Markov process with values in $\mathbb{S}_j$ with transition intensities matrix Λ_j. We denote this process by X^j. In addition to h, introduce column vectors h_j, $j=1,\dots,m$, with entries $h(a_1^j),\dots,h(a_{n_j}^j)$, respectively.

THEOREM 4.4. *Assume the following.*

A-1. For any j, all states of Λ_j communicate.

A-2. For each j,k with $j\neq k$, either

$$h_j^*\mu^j \neq h_k^*\mu^k$$

or

$$h_j^*\operatorname{diag}(\mu^j)\Lambda_j^q h_j \neq h_k^*\operatorname{diag}(\mu^k)\Lambda_k^q h_k,\ \text{for some}\ \ 0\le q\le n_j+n_k-1.$$

Then the asymptotic stability (1.6) *holds true.*

The condition A-1 is inherited from Theorem 4.1 to ensure the stability within each ergodic class, while under A-2, $\mathscr{Y}_{[0,\infty)}$ completely identifies the class in which X actually resides.

5. Proofs for the ergodic case. Recall that under $m=1$, X is a homogeneous ergodic Markov chain with values in the finite alphabet $\mathbb{S}=\{a_1,\dots,a_n\}$ with the transition intensities matrix Λ. The unique invariant measure $\mu=(\mu_1,\dots,\mu_n)$ is the positive distribution on $\mathbb{S}$. Let ν be the distribution of X_0 and β a probability measure on $\mathbb{S}$. The observation process Y is defined in (1.1). Recall that the entries of π_t^ν and $\pi_t^{\beta\nu}$ are the true and "wrong" conditional probabilities, respectively, as defined in the introduction.

5.1. The proof of Theorem 4.1. We use the method proposed by Atar and Zeitouni in [2], which is elaborated for the considered filtering setup for the reader's convenience.

Recall the following facts from the theory of nonnegative matrices. For a pair (p,q) of nonnegative measures on $\mathbb{S}$ (i.e., vectors with nonnegative entries), the Hilbert projective metric $H(p,q)$ is defined as the following (see, e.g., [38]):

$$H(p,q)=\begin{cases}\log\dfrac{\max_{j:q_j>0}(p_j/q_j)}{\min_{i:q_i>0}(p_i/q_i)}, & p\sim q,\\ \infty, & p\not\sim q.\end{cases} \tag{5.1}$$

The Hilbert metric is known to satisfy the following properties:

1. $H(c_1p, c_2q) = H(p, q)$ for any positive constants c_1 and c_2.
2. For matrix A with nonnegative entries (A_{ij}),
$$H\big(Ap, Aq\big) \le \tau(A)H\big(p, q\big) \quad \text{(see, e.g., [38])},$$
where $\tau(A) = \frac{1-\sqrt{\psi(A)}}{1+\sqrt{\psi(A)}}$ is the Birkhoff contraction coefficient with
$$\psi(A) = \min_{i,j,k,\ell} \frac{A_{ik}A_{j\ell}}{A_{i\ell}A_{jk}}.$$
3. $\|p - q\| \le \frac{2}{\log 3} H(p, q)$ (Lemma 1 in [2]).

Returning to the filtering problem, let us first consider the special case when $\nu = \mu$, and thus the signal X^μ is the stationary Markov chain. It is well known that $\pi^\mu_t = \eta^\mu_t / \langle \mathbf{1}, \eta^\mu_t \rangle$, where $\mathbf{1}$ denotes the vector with unit entries, $\langle \cdot, \cdot \rangle$ is the usual inner product, and η^μ_t solves the Zakai equation

$$d\eta^\mu_t = \Lambda^* \eta^\mu_t dt + \sigma^{-2} \operatorname{diag}(h) \eta^\mu_t dY^\mu_t \tag{5.2}$$

subject to $\eta^\mu_0 = \mu$. Similarly, $\pi^{\beta\mu}_t = \eta^{\beta\mu}_t / \langle \mathbf{1}, \eta^{\beta\mu}_t \rangle$, where $\eta^{\beta\mu}_t$ is the solution of (5.2) subject to $\eta^{\beta\mu}_0 = \beta$.

The Zakai equation possesses the unique strong solution which is linear with respect to the initial condition. Hence, $\eta^\mu_t = J_{[0,t]}\mu$ and $\eta^{\beta\mu}_t = J_{[0,t]}\beta$, $t > 0$, where $J_{[0,t]}$ is the random Cauchy matrix corresponding to (5.2).

The matrix $J_{[0,t]}$ can be factored (here $\lfloor t \rfloor$ is the integer part of t):

$$J_{[0,t]} = J_{[\lfloor t \rfloor, t]} \left(\prod_{n=2}^{\lfloor t \rfloor} J_{[n-1,n]} \right) J_{[0,1]}.$$

The properties of the Hilbert metric, listed above, provide

$$\begin{aligned} \big\|\pi^\mu_t - \pi^{\beta\mu}_t\big\| &\le \frac{2}{\log 3} H\big(\pi^\mu_t, \pi^{\beta\mu}_t\big) = \frac{2}{\log 3} H\big(J_{[0,t]}\mu, J_{[0,t]}\beta\big) \\ &\le \frac{2}{\log 3} \tau\big(J_{[\lfloor t \rfloor, t]}\big) \prod_{n=2}^{\lfloor t \rfloor} \tau\big(J_{[n-1,n]}\big) H\big(J_{[0,1]}\mu, J_{[0,1]}\beta\big). \end{aligned}$$

Assume for a moment that $H\big(J_{[0,1]}\mu, J_{[0,1]}\beta\big) < \infty$ a.s. Then

$$\begin{aligned} \varlimsup_{t\to\infty} \frac{1}{t} \log \big\|\pi^\mu_t - \pi^{\beta\mu}_t\big\| &\le \varlimsup_{t\to\infty} \frac{1}{\lfloor t \rfloor} \sum_{n=2}^{\lfloor t \rfloor} \log \tau\big(J_{[n-1,n]}\big) \\ &\le \varlimsup_{t\to\infty} \frac{1}{\lfloor t \rfloor} \sum_{n=2}^{\lfloor t \rfloor} \big\{ -1 \vee \log \tau\big(J_{[n-1,n]}\big)\big\} = E\big[-1 \vee \log \tau\big(J_{[0,1]}\big)\big] \le 0. \end{aligned} \tag{5.3}$$

The equality is implied by the law of large numbers, which is valid since $-1 \le \big\{-1 \vee \log \tau\big(J_{[n-1,n]}\big)\big\} \le 0$ and $\log \tau\big(J_{[n-1,n]}\big)$ is generated by

$$\{X^\mu_s - X^\mu_{n-1},\ W_s - W_{n-1}\},\ n-1 \le s < n,$$

where the processes X^μ and W are independent and X^μ is an ergodic Markov chain.

Let $J^{\nu}_{[n-1,n]}$ be the matrices defined similarly to $J_{[n-1,n]}$ with Y^{μ} replaced by Y^{ν}. Recall that μ is the positive measure on $\mathbb{S}$, so that $\nu \ll \mu$ and, in turn, $\overline{Q}^{\nu} \ll \overline{Q}^{\mu}$ (here $\overline{Q}^{\mu}$ is the distribution of Y^{μ}).

Since (5.3) holds $\overline{Q}^{\mu}$-a.s., it also holds $\overline{Q}^{\nu}$-a.s., i.e., with $J_{[n-1,n]}$ replaced by $J^{\nu}_{[n-1,n]}$ which gives the following theorem.

THEOREM 5.1. (version of Theorem 1(a) in Atar and Zeitouni [2]). *Assume that all states of* Λ *communicate, i.e.,* X *is an ergodic Markov chain. Assume* $J_{[0,1]}\beta$ *and* $J_{[0,1]}\nu$ *have positive entries a.s. Then,*

$$\varlimsup_{t\to\infty} \frac{1}{t} \log \left\|\pi^{\nu}_t - \pi^{\beta\nu}_t\right\| \le E\big[-1 \vee \log \tau\big(J_{[0,1]}\big)\big] \ a.s. \tag{5.4}$$

Now the statement of Theorem 4.1 follows from the lemma below.

LEMMA 5.2. *The right-hand side of* (5.4) *is strictly negative.*

Proof. It suffices to show that all entries of $J_{[0,1]}$ are positive a.s. For fixed i, j, we have

$$J_{[0,t]}(i,j) = \delta_{ij} + \int_0^t J_{[0,s]}(i,j)\big[\lambda_{ii}ds + \sigma^{-2}h(a_i)dY^{\mu}_s\big] + \int_0^t \sum_{r\ne i} \lambda_{ri} J_{[0,s]}(r,j)ds.$$

With the help of the Itô formula and with

$$\phi_t(i) = \exp\big\{\lambda_{ii}t + \sigma^{-2}h(a_i)Y^{\mu}_t - (1/2)\sigma^{-2}h^2(a_i)t\big\}$$

we derive

$$\begin{aligned} J_{[0,t]}(j,j) &= \phi_t(j)\Big(1 + \int_0^t \phi^{-1}_s(j) \sum_{r\ne j} \lambda_{rj} J_{[0,s]}(r,j)ds\Big),\\ J_{[0,t]}(i,j) &= \phi_t(i) \int_0^t \phi^{-1}_s(i) \sum_{r\ne i} \lambda_{ri} J_{[0,s]}(r,j)ds, \quad i \ne j. \end{aligned} \tag{5.5}$$

Also notice that the entries of $J_{[0,t]}$ are unnormalized conditional probabilities and so nonnegative a.s. Since all states of Λ communicate, for a pair of indices (i,j) there is a string of indexes $j = i_\ell, \dots, i_1 = i$ such that $\lambda_{i_\ell i_{\ell-1}}, \dots, \lambda_{i_2 i_1} > 0$. So from (5.5), it follows that a.s.

$$\begin{aligned} &J_{[0,t]}(i_\ell, i_\ell) \ge \phi_t(i_\ell) > 0,\\ &J_{[0,t]}(i_{\ell-1}, i_\ell) \ge \phi_t(i_{\ell-1}) \int_0^t \phi^{-1}_s(i_{\ell-1})\lambda_{i_\ell i_{\ell-1}} J_{[0,s]}(i_\ell, i_\ell)ds > 0,\\ &J_{[0,t]}(i_{\ell-2}, i_\ell) \ge \phi_t(i_{\ell-2}) \int_0^t \phi^{-1}_s(i_{\ell-2})\lambda_{i_{\ell-1} i_{\ell-2}} J_{[0,s]}(i_{\ell-1}, i_\ell)ds > 0 \end{aligned}$$

for any $t > 0$, and so on until we get $J_{[0,t]}(i_1, i_\ell) > 0$, $t > 0$. □

5.2. The proof of Theorem 4.2. Denote $\rho_{ji}(t) = P\big(X^{\beta}_0 = a_j | \mathscr{Y}^{\beta}_{[0,t]}, X^{\beta}_t = a_i\big)$. If β is a positive distribution, then by Lemma 9.5 in [29, Chap. 9] we have

$$\begin{aligned} \rho_{ji}(0) &= \begin{cases} 1, & j = i,\\ 0, & j \ne i, \end{cases}\\ \frac{d\rho_{ji}(t)}{dt} &= \sum_{r\ne i} \frac{\lambda_{ri}\pi^{\beta}_t(r)}{\pi^{\beta}_t(i)}\big(\rho_{jr}(t) - \rho_{ji}(t)\big), \quad i = 1, \dots, n. \end{aligned} \tag{5.6}$$

Remark 3. By the arguments used in the proof of Lemma 5.2, it can be readily shown that $\pi_t^\beta(i) > 0$ a.s., $i = 1, \dots, n$, for any $t > 0$. Then (5.6) remain valid for $t > t_0$ for any $t_0 > 0$ initialized by

$$\rho_{ji}(t_0) = P\big(X_0^\beta = a_j | \mathscr{Y}_{[0,t_0]}^\beta, X_{t_0}^\beta = a_i\big).$$

Set $i^\diamond(t) = \operatorname{argmax}_{i\in\mathbb{S}} \rho_{ji}(t)$ and $i_\diamond(t) = \operatorname{argmin}_{i\in\mathbb{S}} \rho_{ji}(t)$ (if the maximum or the minimum is attained at several indices, the lowest one is taken by convention). Set

$$\rho^\diamond(t) := \rho_{ji^\diamond(t)}(t) \quad \text{and} \quad \rho_\diamond(t) := \rho_{ji_\diamond(t)}(t). \tag{5.7}$$

LEMMA 5.3. *The processes $\rho^\diamond(t)$ and $\rho_\diamond(t)$ have absolutely continuous paths with*

$$\begin{aligned} d\rho^\diamond(t) &= \sum_{i=1}^n I(i^\diamond(t) = i)\dot\rho_{ji}(t)dt, \\ d\rho_\diamond(t) &= \sum_{i=1}^n I(i_\diamond(t) = i)\dot\rho_{ji}(t)dt. \end{aligned} \tag{5.8}$$

The proof of this lemma uses two results formulated in Propositions 5.4 and 5.5 below.

PROPOSITION 5.4 (Theorem A.6.3 in Dupuis and Ellis [20]). *Let $g = g(t)$ be an absolutely continuous function mapping of $[0,1]$ into $\mathbb{R}$. Then for each real number a the set $\{t : g(t) = a, \dot g(t) \neq 0\}$ has Lebesgue measure* 0.

PROPOSITION 5.5. *Let $X(t,\omega)$ be a random process with absolutely continuous paths with respect to dt in the sense that there exists a measurable random process $x(t,\omega)$ such that $\int_0^t |x(s,\omega)|ds < \infty$ a.s., $t > 0$, and*

$$X(t,\omega) = X(0,\omega) + \int_0^t x(s,\omega)ds. \tag{5.9}$$

Then

$$|X(t,\omega)| = |X(0,\omega)| + \int_0^t \operatorname{sign}(X(s,\omega))x(s,\omega)ds,$$

where $\operatorname{sign}(0) = 0$.

Proof. Set $V_t(\omega) = \int_0^t |x(s,\omega)|ds$ and notice that for any $t' \le t''$ it holds that

$$\big||X(t'',\omega)| - |X(t',\omega)|\big| \le |X(t'',\omega) - X(t',\omega)| \le (V_{t''}(\omega) - V_{t'}(\omega)).$$

Hence, for fixed ω, the function $|X(t,\omega)|$ possesses bounded total variation for any finite time interval. Denote by $U_t(\omega)$ this total variation corresponding to $[0,t]$. Obviously, $dU_t(\omega) \ll dV_t(\omega) \ll dt$. Recall that $U_t(\omega) = U_t'(\omega) + U_t''(\omega)$, where $U_t'(\omega)$, $U_t''(\omega)$ are increasing continuous in t functions such that for any $t > 0$ and measurable set A from $\mathbb{R}_+$, $\int_{A\cap[0,t]} dU_s''(\omega) = 0$ and $\int_{(\mathbb{R}_+\setminus A)\cap[0,t]} dU_s'(\omega) = 0$, and at the same time $|X(t,\omega)| = U_t''(\omega) - U_t'(\omega)$. Since $dU_t' \ll dU_t(\omega)$, $dU_t'' \ll dU_t(\omega)$, it follows that

$d|X(t,\omega)| \ll dU_t(\omega) \ll dV_t(\omega) \ll dt$ and so that

$$|X(t,\omega)| = |X(0,\omega)| + \int_0^t g(s,\omega)ds \tag{5.10}$$

though we may not claim that $g(t,\omega)$ is measurable in (t,ω).

Now, we show that $\text{sign}(X(s,\omega))x(s,\omega)$ is a measurable version of $g(s,\omega)$. By (5.9), we have $X^2(t,\omega) = X^2(0,\omega) + 2\int_0^t X(s,\omega)x(s,\omega)ds$. At the same time, by (5.10) it holds that $|X(t,\omega)|^2 = |X(0,\omega)|^2 + 2\int_0^t |X(s,\omega)|g(s,\omega)ds$. Hence, the following identity is valid: For any $t \geq 0$

$$\int_0^t |X(s,\omega)|g(s,\omega)ds \equiv \int_0^t X(s,\omega)x(s,\omega)ds.$$

Therefore, $|X(s,\omega)|g(s,\omega) = X(s,\omega)x(s,\omega)$ for almost all s with respect to Lebesgue measure. Consequently, we have $I(|X(s,\omega)| \neq 0)g(s,\omega) = \text{sign}(X(s,\omega))x(s,\omega)$ for almost all s with respect to Lebesgue measure. It remains to show that

$$I(X(s,\omega) = 0)g(s,\omega) = 0$$

for almost all s with respect to Lebesgue measure. Taking into account (5.10), it suffices to prove that $\int_0^\infty I(X(s,\omega) = 0)d|X(s,\omega)| = 0$ a.s. On the other hand, whereas $d|X(t,\omega)| \ll dV_t(\omega)$, it suffices to show that $\int_0^\infty I(X(s,\omega) = 0)dV_s(\omega) = 0$ a.s. The latter holds by Proposition 5.4. □

Now we give the proof for Lemma 5.3.

Proof. Let us introduce $\rho^{\diamond,i}(t) = \rho_{j1} \vee \rho_{j2} \vee \cdots \vee \rho_{ji}$ and $\rho_{\diamond,i}(t) = \rho_{j1} \wedge \rho_{j2} \wedge \cdots \wedge \rho_{ji}$ and notice that $\rho^{\diamond,n}(t) = \rho^\diamond(t)$, $\rho_{\diamond,n}(t) = \rho_\diamond(t)$.

The use of obvious identities

$$\begin{aligned} \rho^{\diamond,2}(t) + \rho_{\diamond,2}(t) &= \rho_{j1}(t) + \rho_{j2}(t), \\ \rho^{\diamond,2}(t) - \rho_{\diamond,2}(t) &= |\rho_{j1}(t) - \rho_{j2}(t)| \end{aligned}$$

and the fact, provided by Proposition 5.5, that $d|\rho_{j1}(t) - \rho_{j2}(t)| = p(t,\omega)dt$ with measurable derivative $p(\omega,t)$, allow us to claim that $\rho^{\diamond,2}(t)$ and $\rho_{\diamond,2}(t)$ are absolutely continuous with respect to dt with measurable derivatives.

Further, taking into account $\rho^{\diamond,i}(t) = \rho^{\diamond,i-1}(t) \vee \rho_{ji}$ and $\rho_{\diamond,i}(t) = \rho_{\diamond,i-1}(t) \wedge \rho_{ji}(t)$ and consequent identities

$$\begin{aligned} \rho^{\diamond,i}(t) + \rho^{\diamond,i-1}(t) \wedge \rho_{ji}(t) &= \rho^{\diamond,i-1}(t) + \rho_{ji}(t), \\ \rho^{\diamond,i}(t) - \rho^{\diamond,i-1}(t) \wedge \rho_{ji}(t) &= |\rho^{\diamond,i-1}(t) - \rho_{ji}(t)|, \\ \rho_{\diamond,i-1}(t) \vee \rho_{ji}(t) + \rho_{\diamond,i}(t) &= \rho_{\diamond,i-1}(t) + \rho_{ji}(t), \\ \rho_{\diamond,i-1}(t) \vee \rho_{ji}(t) - \rho_{\diamond,i}(t) &= |\rho_{\diamond,i-1}(t) - \rho_{ji}(t)|, \end{aligned}$$

absolute continuity for $\rho^\diamond(t)$ and $\rho_\diamond(t)$ is verified by the induction method.

Thus, $d\rho^\diamond(t) = u(t)dt$ with some density $u(t)$ such that $\int_0^t |u(s)|ds < \infty$ a.s., $t > 0$. On the other hand, since $\sum_{i=1}^n I(i^\diamond(t) = i) = 1$, we have

$$\rho^\diamond(t) = \rho^\diamond(0) + \int_0^t \sum_{i=1}^n I(i^\diamond(s) = i)u(s)ds.$$

So, it suffices to show that for any $t > 0$ and any $i = 1, 2, \dots, n$

$$\int_0^t I(i^\diamond(s) = i)|u(s) - \dot{\rho}_{ji}(s)|ds = 0 \text{ a.s.}$$

The latter holds true by Proposition 5.4, since

$$\begin{aligned}
&\int_0^t I(i^\diamond(s) = i)|u(s) - \dot{\rho}_{ji}(s)|ds \\
&= \int_0^t I(\rho^\diamond(s) - \rho_{ji}(s) = 0)|u(s) - \dot{\rho}_{ji}(s)|ds \\
&= \int_0^t I\big(\rho^\diamond(s) - \rho_{ji}(s) = 0, u(s) - \dot{\rho}_{ji}(s) \neq 0\big)|u(s) - \dot{\rho}_{ji}(s)|ds = 0. \qquad \square
\end{aligned}$$

LEMMA 5.6. *Under the assumptions of Theorem* 4.2*,*

$$\varlimsup_{t\to\infty} \frac{1}{t} \log \max_{1\le j,k,\ell\le n} \big|\rho_{jk}(t) - \rho_{j\ell}(t)\big| \le -\sum_{r=1}^n \mu_r \min_{i\neq r} \lambda_{ri}. \tag{5.11}$$

Proof. By (5.6) and (5.8), we have[2]

$$\begin{aligned}
\frac{d\rho_\diamond(t)}{dt} &= \sum_{r\neq i_\diamond(t)} \frac{\lambda_{ri_\diamond(t)}\pi_t^\beta(r)}{\pi_t^\beta(i_\diamond(t))}\big(\rho_{jr}(t) - \rho_\diamond(t)\big), \\
\frac{d\rho^\diamond(t)}{dt} &= \sum_{r\neq i^\diamond(t)} \frac{\lambda_{ri^\diamond(t)}\pi_t^\beta(r)}{\pi_t^\beta(i^\diamond(t))}\big(\rho_{jr}(t) - \rho^\diamond(t)\big).
\end{aligned} \tag{5.12}$$

In what follows, we will omit the time variable in $i_\diamond(t)$ and $i^\diamond(t)$ for brevity. Set $\triangle_t = \rho^\diamond(t) - \rho_\diamond(t)$. By (5.12) we have

$$\begin{aligned}
\frac{d\triangle_t}{dt} &= -\sum_{r\neq i^\diamond} \frac{\lambda_{ri^\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i^\diamond)}\big(\rho^\diamond(t) - \rho_{jr}(t)\big) - \sum_{r\neq i_\diamond} \frac{\lambda_{ri_\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i_\diamond)}\big(\rho_{jr}(t) - \rho_\diamond(t)\big) \\
&= -\triangle_t\left(\frac{\lambda_{i_\diamond i^\diamond}\pi_t^\beta(i_\diamond)}{\pi_t^\beta(i^\diamond)} + \frac{\lambda_{i^\diamond i_\diamond}\pi_t^\beta(i^\diamond)}{\pi_t^\beta(i_\diamond)}\right) \\
&\quad -\triangle_t\left(\sum_{\substack{r\neq i^\diamond(t)\\ r\neq i_\diamond(t)}} \left[\frac{\lambda_{ri^\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i^\diamond)}\left(\frac{\rho^\diamond(t) - \rho_{jr}(t)}{\triangle_t}\right) + \frac{\lambda_{ri_\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i_\diamond)}\left(\frac{\rho_{jr}(t) - \rho_\diamond(t)}{\triangle_t}\right)\right]\right).
\end{aligned} \tag{5.13}$$

Letting $0/0 = 1/2$, set $\alpha_r(t) = \frac{\rho^\diamond(t)-\rho_{jr}(t)}{\triangle_t}$. Then, we get $1 - \alpha_r(t) = \frac{\rho_{jr}(t)-\rho_\diamond(t)}{\triangle_t}$

[2]In (5.12)–(5.14) we use for brevity a form of differential equalities (inequalities) which are valid for any ω and almost all t with respect to Lebesgue measure.

and $0 \le \alpha_r(t) \le 1$ and (5.13) implies

$$
\begin{aligned}
\frac{d\triangle_t}{dt} = &- \triangle_t\Bigg(\frac{\lambda_{i_\diamond i^\diamond}\pi_t^\beta(i_\diamond)}{\pi_t^\beta(i^\diamond)} + \frac{\lambda_{i^\diamond i_\diamond}\pi_t^\beta(i^\diamond)}{\pi_t^\beta(i_\diamond)}\Bigg) \\
&- \triangle_t\left(\sum_{\substack{r\ne i^\diamond(t)\\ r\ne i_\diamond(t)}}\left[\alpha_r(t)\frac{\lambda_{ri^\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i^\diamond)} + \big(1-\alpha_r(t)\big)\frac{\lambda_{ri_\diamond}\pi_t^\beta(r)}{\pi_t^\beta(i_\diamond)}\right]\right) \\
\le &- \triangle_t\Big(\lambda_{i_\diamond i^\diamond}\pi_t^\beta(i_\diamond) + \lambda_{i^\diamond i_\diamond}\pi_t^\beta(i^\diamond)\Big) \\
&- \triangle_t\left(\sum_{\substack{r\ne i^\diamond(t)\\ r\ne i_\diamond(t)}}\Big[\alpha_r(t)\lambda_{ri^\diamond} + \big(1-\alpha_r(t)\big)\lambda_{ri_\diamond}\Big]\pi_t^\beta(r)\right) \\
\le &- \triangle_t\left(\lambda_{i_\diamond i^\diamond}\pi_t^\beta(i_\diamond) + \lambda_{i^\diamond i_\diamond}\pi_t^\beta(i^\diamond) + \sum_{\substack{r\ne i^\diamond(t)\\ r\ne i_\diamond(t)}}\Big[\lambda_{ri^\diamond}\wedge\lambda_{ri_\diamond}\Big]\pi_t^\beta(r)\right).
\end{aligned} \tag{5.14}
$$

Recall that all offdiagonal entries of Λ are nonnegative and $\sum_{r=1}^n \lambda_{ir} = 0$ for any i. Then, $\big|\lambda_{i_\diamond i^\diamond}| \wedge |\lambda_{i_\diamond i_\diamond}\big| \ge \lambda_{i_\diamond i^\diamond}$, $\big|\lambda_{i^\diamond i^\diamond}| \wedge |\lambda_{i^\diamond i_\diamond}\big| \ge \lambda_{i^\diamond i_\diamond}$, and (5.14) provides

$$
\begin{aligned}
\frac{d\triangle_t}{dt} &\le -\triangle_t\sum_{r=1}^n\Big(|\lambda_{ri^\diamond}|\wedge|\lambda_{ri_\diamond}|\Big)\pi_t^\beta(r) \le -\triangle_t\sum_{r=1}^n\min_{1\le i\le n}|\lambda_{ri}|\pi_t^\beta(r) \\
&= -\triangle_t\sum_{r=1}^n\pi_t^\beta(r)\min_{i\ne r}\lambda_{ri}.
\end{aligned}
$$

Since the derivative $\frac{d\triangle_t}{dt}$ is defined for each ω and almost everywhere (a.e.) in t with respect to dt, the above inequality $\frac{d\triangle_t}{dt} \le -\triangle_t\sum_{r=1}^n \pi_t^\beta(r)\min_{i\ne r}\lambda_{ri}$ is also valid a.e. So, it allows us to define a.e. the function

$$
H(t) = -\triangle_t\sum_{r=1}^n\pi_t^\beta(r)\min_{i\ne r}\lambda_{ri} - \frac{d\triangle_t}{dt}.
$$

Moreover, for definiteness, we may redefine $H(t)$ everywhere so as $H(t) \ge 0$. Then we have

$$
d\triangle_t = -\left[\triangle_t\sum_{r=1}^n\pi_t^\beta(r)\min_{i\ne r}\lambda_{ri} + H(t)\right]dt.
$$

Notice also that $\int_0^t |H(s)|ds < \infty$ a.s. for any $t > 0$ and recall that $\triangle_0 = 1$. Then, we get

$$
\triangle_t = \exp\left(-\int_0^t\sum_{r=1}^n\pi_s^\beta(r)\min_{i\ne r}\lambda_{ri}ds\right) - \int_0^t\exp\left(-\int_s^t\sum_{r=1}^n\pi_u^\beta(r)\min_{i\ne r}\lambda_{ri}du\right)H(s)ds
$$

and in turn

$$
\frac{1}{t}\log\triangle_t \le -\sum_{r=1}^n\Big(\min_{i\ne r}\lambda_{ri}\Big)\frac{1}{t}\int_0^t\pi_s^\beta(r)ds.
$$

So, it is left to verify that

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \pi_s^\beta(r)ds=\mu_r \quad \text{a.s.} \tag{5.15}$$

Similarly to (1.4), π_t^β satisfies

$$\begin{aligned}\pi_0^\beta =&\beta,\\ d\pi_t^\beta =&\Lambda^*\pi_t^\beta dt+\sigma^{-2}\big(\operatorname{diag}(\pi_t^\beta)-\pi_t^\beta(\pi_t^\beta)^*\big)h(dY_t^\beta-h^*\pi_t^\beta dt).\end{aligned}$$

Recall that $\sigma^{-1}\big(Y_t^\beta-\int_0^t h^*\pi_s^\beta ds\big)$ is the innovation Wiener process (see, e.g., Theorem 9.1 in Chapter 10 in [30]). Hence $M_t=\int_0^t\big(\operatorname{diag}(\pi_s^\beta)-\pi_s^\beta(\pi_s^\beta)^*\big)h(dY_s^\beta-h^*\pi_s^\beta ds)$ is a vector-valued continuous martingale. Its entries $M_t(i)$, $i=1,\dots,n$, have predictable quadratic variation processes $\langle M(i)\rangle_t$ with the following property: For some positive constant c, $d\langle M(i)\rangle_t\le cdt$. Then by Theorem 10, Chapter 3 in [31], $\lim_{t\to\infty}\frac{1}{t}M_t(i)=0$ a.s. This fact and the boundedness of π_t^β provide $\Lambda^*\lim_{t\to\infty}\frac{1}{t}\int_0^t\pi_s^\beta ds=0$. The vector $Z_t=\frac{1}{t}\int_0^t\pi_s^\beta ds$ has nonnegative entries, whose sum equals 1. Therefore the limit vector Z_∞, obeying the same property, is the unique solution of the linear algebraic equation $\Lambda^*Z_\infty=0$, i.e., $Z_\infty=\mu$. $\square$

To prove Theorem 4.2, without loss generality, due to Remark 3, we may assume that $\nu\sim\beta$. Then, we show that for any $t\ge 0$ and $i=1,\dots,n$

$$\big|\pi_t^\nu(i)-\pi_t^{\beta\nu}(i)\big|\le n\max_j\frac{d\nu}{d\beta}(a_j)\max_j\frac{d\beta}{d\nu}(a_j)\max_{1\le i,j,k\le d}\big|\rho_{ji}(t)-\rho_{jk}(t)\big|. \tag{5.16}$$

Recall that Q^ν and Q^β are distributions of (X^ν,Y^ν) and (X^β,Y^β), respectively, which are equivalent, by virtue of $\nu\sim\beta$, with

$$\frac{dQ^\beta}{dQ^\nu}(X^\nu,Y^\nu)\equiv\frac{d\beta}{d\nu}(X_0^\nu)\quad\text{and}\quad\frac{dQ^\nu}{dQ^\beta}(X^\beta,Y^\beta)\equiv\frac{d\nu}{d\beta}(X_0^\beta).$$

Now, we show that for any $i=1,\dots,d$ and $t>0$, Q^ν- and Q^β-a.s.

$$\pi_t^{\beta\nu}(i)=\frac{\sum_{j=1}^n\Big(\frac{d\beta}{d\nu}(a_j)P\big(X_0^\nu=a_j),X_t^\nu=a_i|\mathscr{Y}_{[0,t]}^\nu\Big)}{E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}_{[0,t]}^\nu\Big)}. \tag{5.17}$$

To this end, with any bounded $\mathscr{D}_t^y$-measurable function $\psi_t(y)$, write

$$\begin{aligned}&E\psi_t(Y^\nu)\pi_t^{\beta\nu}(i)E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}_{[0,t]}^\nu\Big)=E\psi_t(Y^\nu)\pi_t^{\beta\nu}(i)\frac{d\beta}{d\nu}(X_0^\nu)\\ &=E\psi_t(Y^\nu)\pi_t^{\beta\nu}(i)\frac{dQ^\beta}{dQ^\nu}(X^\nu,Y^\nu)=E\psi_t(Y^\beta)\pi_t^\beta(i)\\ &=E\psi_t(Y^\beta)I(X_t^\beta=a_i)=E\psi_t(Y^\nu)I(X_t^\nu=a_i)\frac{dQ^\beta}{dQ^\nu}(X^\nu,Y^\nu)\\ &=E\psi_t(Y^\nu)I(X_t^\nu=a_i)\frac{d\beta}{d\nu}(X_0^\nu)=E\psi_t(Y^\nu)E\Big(I(X_t^\nu=a_i)\frac{d\beta}{d\nu}(X_0^\nu)\big|\mathscr{Y}_{[0,t]}^\nu\Big).\end{aligned}$$

Hence, by the arbitrariness of $\psi_t(y)$,

$$\pi_t^{\beta\nu}(i)E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}_{[0,t]}^\nu\Big)=E\left(I(X_t^\nu=a_i)\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}_{[0,t]}^\nu\right).$$

Further, $Q^\nu \sim Q^\beta$ provides $E\big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\big) > 0$, Q^ν- and Q^β-a.s., so that

$$\pi_t^{\beta\nu}(i) = \frac{E\big(I(X_t^\nu = a_i)\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\big)}{E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\Big)}$$

and it remains to notice that

$$E\left(I(X_t^\nu = a_i)\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\right) = \sum_{j=1}^n \frac{d\beta}{d\nu}(a_j)P\big(X_t^\nu = a_i, X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big).$$

Taking into consideration (5.17), we find

$$\begin{aligned}\big|\pi_t^\nu(i) - \pi_t^{\beta\nu}(i)\big| &= \left|\pi_t^\nu(i) - \frac{\sum_{j=1}^n\Big(\frac{d\beta}{d\nu}(a_j)P\big(X_0^\nu = a_j, X_t^\nu = a_i|\mathscr{Y}^\nu_{[0,t]}\big)\Big)}{E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\Big)}\right| \\ = \frac{\Big|\sum_{j=1}^n \frac{d\beta}{d\nu}(a_j)\Big(\pi_t^\nu(i)P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) - P\big(X_0^\nu = a_j, X_t^\nu = a_i|\mathscr{Y}^\nu_{[0,t]}\big)\Big)\Big|}{E\Big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\Big)}.\end{aligned}$$

Then, since by the Jensen inequality $1/E\big(\frac{d\beta}{d\nu}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\big) \le E\big(\frac{d\nu}{d\beta}(X_0^\nu)|\mathscr{Y}^\nu_{[0,t]}\big)$, we get the chain of estimates

$$\begin{aligned}\big|\pi_t^\nu(i) - \pi_t^{\beta\nu}(i)\big| &\le \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{a_j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j) \\ &\quad\times\left|\sum_{j=1}^n \pi_t^\nu(i)\Big(P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) - P\big(X_0^\nu = a_j\big|X_t^\nu = a_i, \mathscr{Y}^\nu_{[0,t]}\big)\Big)\right| \\ &\le \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{a_j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j) \\ &\quad\times\sum_{j=1}^n \pi_t^\nu(i)\Big|P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) - P\big(X_0^\nu = a_j|X_t^\nu = a_i, \mathscr{Y}^\nu_{[0,t]}\big)\Big| \qquad (5.18)\\ &\le \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j) \\ &\quad\times\sum_{j=1}^n \Big|P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) - P\big(X_0^\nu = a_j|X_t^\nu = a_i, \mathscr{Y}^\nu_{[0,t]}\big)\Big| \\ &= \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{a_j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j)\sum_{j=1}^n\Big|P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) - \rho_{ji}(t)\Big|.\end{aligned}$$

The obvious formula $P\big(X_0^\nu = a_j|\mathscr{Y}^\nu_{[0,t]}\big) = \sum_{k=1}^n \pi_t^\nu(k)\rho_{jk}(t)$, and (5.18) provide

$$\begin{aligned}\big|\pi_t^\nu(i) - \pi_t^{\beta\nu}(i)\big| &\le \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{a_j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j)\sum_{j=1}^n\Big|\sum_{k=1}^n \pi_t^\nu(k)\rho_{jk}(t) - \rho_{ji}(t)\Big| \\ &\le \max_{a_j\in\mathbb{S}}\frac{d\beta}{d\nu}(a_j)\max_{a_j\in\mathbb{S}}\frac{d\nu}{d\beta}(a_j)\sum_{j=1}^n\sum_{k=1}^n \pi_t^\nu(k)\big|\rho_{jk}(t) - \rho_{ji}(t)\big| \qquad (5.19)\end{aligned}$$

and (5.16). Thus, by Lemma 5.6, the desired statement (4.1) holds true.

5.3. The proof of Theorem 4.3. We start with the following lemma.

LEMMA 5.7. *Under the assumptions of Theorem* 4.3, *for any* $t > 0$

$$\max_{1\le j,k,\ell\le n} \big|\rho_{jk}(t) - \rho_{j\ell}(t)\big| \le \exp\Big(-2t\min_{p\ne q}\sqrt{\lambda_{pq}\lambda_{qp}}\Big). \tag{5.20}$$

Proof. Here we follow the notations from Lemma 5.6. From (5.14), it follows that

$$\frac{d\triangle_t}{dt} \le -\triangle_t\left(\frac{\lambda_{i_\diamond i^\diamond}\pi^\beta_t(i_\diamond)}{\pi^\beta_t(i^\diamond)} + \frac{\lambda_{i^\diamond i_\diamond}\pi^\beta_t(i^\diamond)}{\pi^\beta_t(i_\diamond)}\right) \tag{5.21}$$

subject to $\triangle_0 = 1$. Set $\tau = \inf\{t : i^\diamond(t) = i_\diamond(t)\}$. Since $\triangle_t$ is a nonincreasing function, $\triangle_t \equiv 0$ for $t \ge \tau$, and (5.20) holds trivially. For $t < \tau$, as previously we find

$$\begin{aligned}
\triangle_t &\le \exp\left\{-\int_0^t\left(\frac{\lambda_{i_\diamond i^\diamond}\pi^\beta_s(i_\diamond)}{\pi^\beta_s(i^\diamond)} + \frac{\lambda_{i^\diamond i_\diamond}\pi^\beta_s(i^\diamond)}{\pi^\beta_s(i_\diamond)}\right)ds\right\}\\
&\le \exp\left\{-\int_0^t\min_{x\ge 0}\left(\lambda_{i_\diamond i^\diamond}x + \lambda_{i^\diamond i_\diamond}\frac{1}{x}\right)ds\right\}\\
&= \exp\left\{-\int_0^t 2\sqrt{\lambda_{i_\diamond i^\diamond}\lambda_{i^\diamond i_\diamond}}ds\right\} \le \exp\Big(-2t\min_{p\ne q}\sqrt{\lambda_{pq}\lambda_{qp}}\Big),
\end{aligned}$$

and (5.20) follows. ∎

To prove the first statement of the theorem, taking into account $\nu \ll \beta$ we replicate a fragment from the proof of Proposition 2.1.

Using the notations introduced in section 2.1, write $\pi^\nu_t(i) := \pi^\nu_t(f)$ and $\pi^{\beta\nu}_t(i) := \pi^{\beta\nu}_t(f)$ for $f(x) = I(x = a_i)$. Then,

$$E\big|\pi^{\beta\nu}_t(i) - \pi^\nu_t(i)\big| \le E\Big|E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,t]}\Big) - E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)\Big| \tag{5.22}$$

and, since (X^β, Y^β) is a Markov process,

$$E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big) = E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,t]}\vee\mathscr{X}^\beta_t\Big).$$

Then,

$$\begin{aligned}
&E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,t]}\Big) - E\Big(\frac{d\nu}{d\beta}(X^\beta_0)\big|\mathscr{Y}^\beta_{[0,\infty)}\vee\mathscr{X}^\beta_{[t,\infty)}\Big)\\
&= \sum_{j=1}^n\frac{d\nu}{d\beta}(a_j)\Big(P\big(X^\beta_0 = a_j|\mathscr{Y}^\beta_{[0,t]}\big) - P\big(X^\beta_0 = a_j|\mathscr{Y}^\beta_{[0,t]}\vee\mathscr{X}^\beta_t\big)\Big)\\
&= \sum_{j=1}^n\sum_{\ell=1}^n I(X^\beta_t = a_\ell)\frac{d\nu}{d\beta}(a_j)\Big(P\big(X^\beta_0 = a_j|\mathscr{Y}^\beta_{[0,t]}\big) - \rho_{j\ell}(t)\Big) \qquad (5.23)\\
&= \sum_{j=1}^n\sum_{\ell=1}^n\sum_{k=1}^n\pi^\beta_t(k)I(X^\beta_t = a_\ell)\frac{d\nu}{d\beta}(a_j)\big(\rho_{jk}(t) - \rho_{j\ell}(t)\big)\\
&\le \max_{1\le j,k,\ell\le n}\big|\rho_{jk}(t) - \rho_{j\ell}(t)\big|\sum_{j=1}^n\frac{d\nu}{d\beta}(a_j).
\end{aligned}$$

The first statement of Theorem 4.3 follows from (5.22), (5.23), and Lemma 5.7. The second statement follows from (5.16) and Lemma 5.7.

6. Proofs for the nonergodic case. Recall that in the nonergodic setting under consideration

$$\mathbb{S} = \Big\{ \underbrace{a_1^1, \dots, a_{n_1}^1}_{\mathbb{S}_1}, \dots, \underbrace{a_1^m, \dots, a_{n_m}^m}_{\mathbb{S}_m} \Big\}, \quad m \ge 2,$$

with subalphabets $\mathbb{S}_1, \dots, \mathbb{S}_m$ noncommunicating in the sense of (1.2).

6.1. Auxiliary lemmas. In this subsection, $\widetilde{X}_t^j$ is an independent copy of X_t^j with the initial distribution μ^j, defined on some auxiliary probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{P})$, and $\widetilde{E}$ is the expectation with respect to $\widetilde{P}$. Recall that μ^j is the invariant measure, so that $\widetilde{X}_t^j$ is a stationary process.

LEMMA 6.1. *Fix $r > 0$ and define $Z_n = \sum_{i=1}^n \big(Y_{ir}^\beta - Y_{(i-1)r}^\beta\big)^2$. Then with $n \to \infty$*

$$\frac{1}{n} Z_n \to r + \sum_{j=1}^m I(X_0^\beta \in \mathbb{S}_j) \widetilde{E} \left(\int_0^r h(\widetilde{X}_s^j)\, ds \right)^2.$$

Proof. Define

$$F(i) = E\left[\left(\int_0^r h(X_s^\beta)\, ds \right)^2 \Big| X_0^\beta = a_i \right]$$

and $\mathscr{G}_n = \sigma\{Y_{[0,nr]}\} \vee \sigma\{X_{[0,nr]}\}$. Then $E[(Y_{(n+1)r}^\beta - Y_{nr}^\beta)^2 | \mathscr{G}_n] = r + F(X_{nr}^\beta)$ so that the sequence $M_n = Z_n - nr - \sum_{i=0}^{n-1} F(X_{ir}^\beta)$ is a martingale with respect to the filtration $(\mathscr{G}_n)_{n \ge 1}$. It is easy to verify that there exists $K < \infty$ such that for all n we have $E(M_{n+1} - M_n)^2 \le K$. It follows that $(1/n) M_n \to 0$ a.s. as $n \to \infty$ (see, e.g., Chapter VII, Section 5, Theorem 4 in [42]).

Now consider $(1/n) \sum_{i=0}^{n-1} F(X_{ir}^\beta)$. If $X_0 \in \mathbb{S}_j$, then $X_t \in \mathbb{S}_j$ for all $t \ge 0$ and the process is ergodic in $\mathbb{S}_j$ with stationary distribution μ^j. Applying the ergodic theorem for each class $\mathbb{S}_j$ we obtain

$$\frac{1}{n} \sum_{i=0}^{n-1} F(X_{ir}^\beta) \to \sum_{j=1}^m \widetilde{E}(F(\widetilde{X}_0)) I(X_0 \in \mathbb{S}_j) = \sum_{j=1}^m \widetilde{E} \left(\int_0^r h(\widetilde{X}_s^j)\, ds \right)^2 I(X_0^\beta \in \mathbb{S}_j)$$

as $n \to \infty$ a.s. Finally

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} Z_n &= \lim_{n\to\infty} \frac{1}{n} M_n + r + \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} F(X_{ir}^\beta) \\ &= r + \sum_{j=1}^m \widetilde{E} \left(\int_0^r h(\widetilde{X}_s^j)\, ds \right)^2 I(X_0^\beta \in \mathbb{S}_j) \end{aligned}$$

and we are done. $\square$

With $\widetilde{X}_t^j$ defined as in Lemma 6.1 and $r \ge 0$ let $d_j(r) = \widetilde{E}(\int_0^r h(\widetilde{X}_s^j)\, ds)^2$.

LEMMA 6.2. *For any $k \ne j$ the following are equivalent:*

i. *$d_k(r) = d_j(r)$ for all $r \ge 0$;*

ii. *$h_k^* \operatorname{diag}(\mu_k) \Lambda_k^q h_k = h_j^* \operatorname{diag}(\mu_j) \Lambda_j^q h_j$ for all $0 \le q \le n_i + n_j - 1$.*

Proof. Notice first that

$$\begin{aligned} d_j(r) &= 2\widetilde{E}\int_0^r\int_0^s h(\widetilde{X}_u^j)h(\widetilde{X}_s^j)\,du\,ds = 2\int_0^r\int_0^s \widetilde{E}h(\widetilde{X}_u^j)h(\widetilde{X}_s^j)\,du\,ds \\ &= 2\int_0^r\int_0^s \widetilde{E}h(\widetilde{X}_0^j)h(\widetilde{X}_{s-u}^j)\,du\,ds = 2\int_0^r\int_0^s \widetilde{E}h(\widetilde{X}_0^j)h(\widetilde{X}_v^j)\,dv\,ds. \end{aligned}$$

Now, introduce the vector $\widetilde{I}_t^j$ with entries $I(\widetilde{X}_t^j = a_1^j), \dots, I(\widetilde{X}_t^j = a_{n_j}^j)$ and notice also that

$$\begin{aligned} \widetilde{E}h(\widetilde{X}_0^j)h(\widetilde{X}_v^j) &= \widetilde{E}h_j^*\widetilde{I}_0^j(\widetilde{I}_v^j)^*h_j = \widetilde{E}h_j^*\widetilde{I}_0^j(\widetilde{I}_0^j)^*e^{\Lambda_j v}h_j \\ &= h_j^*\widetilde{E}\,\mathrm{diag}(\widetilde{I}_0^j)e^{\Lambda_j v}h_j = h_j^*\,\mathrm{diag}(\mu^j)e^{\Lambda_j v}h_j. \end{aligned}$$

Therefore $d_j(r) = 2\int_0^r\int_0^s h_j^*\,\mathrm{diag}(\mu^j)e^{\Lambda_j v}h_j dv ds$, so $d_j(0) = {d_j}'(0) = 0$ and

$$ {d_j}''(r) = 2h_j^*\,\mathrm{diag}(\mu^j)e^{\Lambda_j r}h_j. $$

Differentiating with respect to r a further q times and then putting $r = 0$ we get

$$ d_j^{(2+q)}(0) = 2h_j^*\,\mathrm{diag}(\mu^j)\Lambda_j^q h_j. $$

It follows immediately that if $d_k(r) = d_j(r)$ for all $r \ge 0$, then

$$ h_k^*\,\mathrm{diag}(\mu^k)\Lambda_k^q h_k = h_j^*\,\mathrm{diag}(\mu^j)\Lambda_j^q h_j $$

for all $q \ge 0$ and so in particular for all $0 \le q \le n_k + n_j - 1$.

Suppose conversely that $h_j^*\,\mathrm{diag}(\mu^j)\Lambda_j^q h_j = h_k^*\,\mathrm{diag}(\mu^k)\Lambda_k^q h_k$ for all $0 \le q \le n_k + n_j - 1$. The Cayley–Hamilton theorem applied to the $(n_k+n_j)\times(n_k+n_j)$ block diagonal matrix $\left(\begin{smallmatrix}\Lambda_k & 0\\ 0 & \Lambda_j\end{smallmatrix}\right)$ gives constants $c_0, c_1, \dots, c_{n_k+n_j-1}$ so that

$$ \Lambda_k^{n_k+n_j} = \sum_{q=0}^{n_k+n_j-1} c_q\Lambda_k^q \quad\text{and}\quad \Lambda_j^{n_i+n_j} = \sum_{q=0}^{n_k+n_j-1} c_q\Lambda_j^q. $$

Therefore we have $h_k^*\,\mathrm{diag}(\mu^k)\Lambda_k^q h_k = h_j^*\,\mathrm{diag}(\mu^j)\Lambda_j^q h_j$ for all $q > n_j + n_k - 1$ as well. Using the fact that $e^{\Lambda_j r} = \sum_{q=0}^\infty \frac{r^q\Lambda_j^q}{q!}$, we see that ${d_k}''(r) = {d_j}''(r)$ for all $r \ge 0$, and hence $d_k(r) = d_j(r)$ for all $r \ge 0$. $\square$

LEMMA 6.3. *Assume* A-2. *For any* β

$$ \lim_{t\to\infty} E\Big|P\big(X_0^\beta \in \mathbb{S}_j|\mathscr{Y}_{[0,t]}^\beta\big) - I(X_0^\beta \in \mathbb{S}_j)\Big| = 0,\ j \ge 1. $$

Proof. We use the notation $Z_n^{(r)}$ to express the dependence on r of the function Z_n in Lemma 6.1. We have $\frac{1}{n}Y_n^\beta \to \sum_{j=1}^m h_j^*\mu^j I(X_0^\beta \in \mathbb{S}_j)$ and

$$ \frac{1}{n}Z_n^{(r)} \to r + \sum_{j=1}^m d_j(r)I(X_0^\beta \in \mathbb{S}_j) $$

as $n \to \infty$ a.s. Using assumption A-2 and Lemma 6.2 we can find an integer ℓ and numbers $r_i > 0, i = 1, \dots, \ell$, and construct a random variable of the form $V_n = (Y_n^\beta, Z_n^{(r_1)} - nr_1, \dots, Z_n^{(r_\ell)} - nr_\ell)$ so that $\frac{1}{n}V_n \to \sum_{j=1}^m v_j I(X_0^\beta \in \mathbb{S}_j)$ as $n \to \infty$, P-a.s., where the $v_1, \dots, v_m$ are distinct vectors in $\mathbb{R}^{\ell+1}$. Therefore $\{X_0^\beta \in \mathbb{S}_j\}$ is $Y_{[0,\infty)}^\beta$-measurable a.s. and the result follows immediately. $\square$

6.2. The proof of Theorem 4.4. By Proposition 2.1, it suffices to show that

$$\lim_{t\to\infty} E\big\|\pi_t^\beta - \pi_t^{\beta_0}\big\| = 0.$$

We introduce a new filter, intermediate between π_t^β and $\pi_t^{\beta_0}$. Define the random variable U by $U = j$ on the set $\{X_0^\beta \in \mathbb{S}_j\}$, and then define

$$\pi_t^{\beta,U}(i) = P(X_t^\beta = a_i | \mathscr{Y}_{[0,t]}^\beta, U).$$

Then

$$\begin{aligned}
\big\|\pi_t^\beta - \pi_t^{\beta,U}\big\| &= \sum_{i=1}^n \Big| P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta) - P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, U)\Big| \\
&= \sum_{i=1}^n \left| \sum_{j=1}^m P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, U=j) \Big(P(U=j|\mathscr{Y}_{[0,t]}^\beta) - I(U=j)\Big)\right| \\
&\le \sum_{j=1}^m \Big| P(U=j|\mathscr{Y}_{[0,t]}^\beta) - I(U=j)\Big|
\end{aligned}$$

and

$$\begin{aligned}
\big\|\pi_t^{\beta,U} - \pi_t^{\beta_0}\big\| &= \sum_{i=1}^n \Big| P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, U) - P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, X_0^\beta)\Big| \\
= \sum_{i=1}^n \sum_{j=1}^m & I(U=j) \Big| P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, U=j) - P(X_t^\beta = a_i|\mathscr{Y}_{[0,t]}^\beta, U=j, X_0^\beta)\Big| \\
= \sum_{j=1}^m & I(U=j)\big\|\pi_t^{\beta^j} - \pi_t^{\beta_0^j}\big\|,
\end{aligned}$$

where β^j denotes the conditional distribution of β restricted to the subalphabet $\mathbb{S}_j$. By Lemma 6.3,

$$\sum_{j=1}^m \Big| P(U=j|\mathscr{Y}_{[0,t]}^\beta) - I(U=j)\Big| \xrightarrow[t\to\infty]{P} 0$$

while $\sum_{j=1}^m I(U=j)\big\|\pi_t^{\beta^j} - \pi_t^{\beta_0^j}\big\| \xrightarrow[t\to\infty]{\mathbb{L}_1} 0$ by applying Theorem 4.1 to each $\mathbb{S}_j$.

Appendix. Proof of Proposition 3.2.

Proof (sketch). We use the following construction for X. Let X_0 be a random variable with values in $\mathbb{S} = \{1,2,3,4\}$ and $P(X_0 = j) = \nu_j$, $j = 1,\dots,4$. Introduce independent of X_0 the matrix-valued process

$$\mathcal{N}_t = \begin{pmatrix} -N_{12}(t) & N_{12}(t) & 0 & 0 \\ 0 & -N_{23}(t) & N_{23}(t) & 0 \\ 0 & 0 & -N_{34}(t) & N_{34}(t) \\ N_{41}(t) & 0 & 0 & -N_{41}(t) \end{pmatrix}, \tag{A.1}$$

where $N_{ij}(t)$ are independent copies of the Poisson process with the unit rate. Let us consider the Itô equation

$$I_t = I_0 + \int_0^t d\mathcal{N}_s^* I_{s-} \tag{A.2}$$

with I_0 the vector with entries $I_0(j) = I(X_0 = j)$, $j = 1, \dots, 4$. Since the jumps of Poisson processes $N_{ij}(t)$'s are disjoint, for any $t > 0$ the vector I_t has only one nonzero entry. Moreover, whereas the increments of $\mathcal{N}_t$ are independent for nonoverlapping intervals, I_t is a Markov process. It is readily checked that, with the row vector $g = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$, $X_t = gI_t$ is a Markov process with values in $\mathbb{S}$ and the transition intensities matrix Λ and $I_t(j) = I(X_t = j)$, $j = 1, \dots, 4$.

We will follow Theorem 4.10.1 from [31]. The random process Y has piecewise constant paths with jumps of two magnitudes, $+1$ and -1. Due to (A.2), its saltus measure $p(dt, dy)$ is completely described by

$$\begin{aligned} p(dt, \{1\}) &= \big\{I_{t-}(4)dN_{41}(t) + I_{t-}(2)dN_{23}(t)\big\}, \\ p(dt, \{-1\}) &= \big\{I_{t-}(1)dN_{12}(t) + I_{t-}(3)dN_{34}(t)\big\}. \end{aligned}$$

So, the compensator $\overline{q}(dt, dy)$ of $p(dt, dy)$ with respect to the filtration $(\mathscr{Y}_{[0,t]})_{t\ge 0}$ is defined as

$$\begin{aligned} \overline{q}(dt, \{1\}) &= \big(\pi_{t-}(4) + \pi_{t-}(2)\big)dt = (1 - Y_{t-})dt, \\ \overline{q}(dt, \{-1\}) &= \big(\pi_{t-}(1) + \pi_{t-}(3)\big)dt = Y_{t-}dt. \end{aligned} \tag{A.3}$$

Notice also that

$$p(dt, \{1\}) = (1 - Y_{t-})dY_t \quad \text{and} \quad p(dt, \{-1\}) = -Y_{t-}dY_t. \tag{A.4}$$

Equation (A.2) also gives "drift+martingale" presentation for $I_1(t)$, $I_2(t)$:

$$\begin{aligned} dI_t(1) &= \big(-I_t(1) + I_t(4)\big)dt + dM_1(t), \\ dI_t(2) &= \big(I_t(1) - I_t(2)\big)dt + dM_2(t) \end{aligned} \tag{A.5}$$

with martingales

$$\begin{aligned} M_1(t) &= \int_0^t \Big(-I_{s-}(1)d(N_{12}(s) - s) + I_{s-}(4)d(N_{41}(s) - s)\Big), \\ M_2(t) &= \int_0^t \Big(I_{s-}(1)d(N_{12} - s) - I_{s-}(2)d(N_{23}(s) - s)\Big). \end{aligned}$$

Then, by Theorem 4.10.1 in [31], adapted to the case considered, we have

$$\begin{aligned} d\pi_1(t) &= \big(-\pi_t(1) + \pi_t(4)\big)dt + \int H_1(\omega, t, y)\big[p(dt, dy) - \overline{q}(dt, dy)\big], \\ d\pi_2(t) &= \big(\pi_t(1) - \pi_t(2)\big)dt + \int H_2(\omega, t, y)\big[p(dt, dy) - \overline{q}(dt, dy)\big], \end{aligned} \tag{A.6}$$

where $H_i(\omega, t, y)$, $i = 1, 2$, are $\mathscr{P}(Y) \otimes \mathscr{B}(\mathbb{R})$-measurable functions (here $\mathscr{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$ and $\mathscr{P}(Y)$ is the predictable σ-algebra on $\Omega \times \mathbb{R}_+$ with respect to the filtration $(\mathscr{Y}_{[0,t]})_{t\ge 0}$). Moreover

$$H_i(\omega, t, y) = \mathsf{M}\big(\triangle M_i + I_-(i) | \mathscr{P}(\mathcal{Y}) \otimes \mathscr{B}(\mathbb{R})\big)(\omega, t, y) - \pi_{t-}(i),$$

where $\triangle M_i$ and $I_-(i)$ are the processes $M_i(t) - M_i(t-)$ and $I_{t-}(i)$, respectively, and $\mathsf{M}\big(\,\cdot\,|\mathscr{P}(Y) \otimes \mathscr{B}(\mathbb{R})\big)$ is the conditional expectation with respect to the measure $\mathsf{M}(d\omega, dt, dy) = P(d\omega)p(dt, dy)$ given $\mathscr{P}(Y) \otimes \mathscr{B}(\mathbb{R})$.

By (A.5), $\triangle M_i(t) + I_{t-}(i) = I_t(i)$ and the structure of compensator $\overline{q}$ provides (here $\triangle I_t(i) = I_t(i) - I_{t-}(i)$)

$$\mathsf{M}\big(I(i)|\mathscr{P}(Y)\otimes\mathscr{B}(\mathbb{R})\big) - \pi_{t-}(i) = \mathsf{M}\big(\triangle I(i)|\mathscr{P}(Y)\otimes\mathscr{B}(\mathbb{R})\big).$$

The desired conditional expectation is determined uniquely from the following identity: For any bounded, compactly supported in t and $\mathscr{P}(\mathcal{Y})\otimes\mathscr{B}(\mathbb{R})$-measurable function $\phi(\omega,t,y)$

$$\begin{aligned} &E\int_0^\infty\int \phi(\omega,t,y)\triangle I_t(i)p(dt,dy)\\ &\qquad = E\int_0^\infty\int \phi(\omega,t,y)\mathsf{M}\big(\triangle I(i)|\mathscr{P}(Y)\otimes\mathscr{B}(\mathbb{R})\big)(\omega,t,y)\overline{q}(dt,dy). \end{aligned}$$

By (A.2)

$$\begin{aligned} \triangle I_t(1) &= -I_{t-}(1)\triangle N_{12}(t) + I_{t-}(4)\triangle N_{41}(t),\\ \triangle I_t(2) &= I_{t-}(1)\triangle N_{12}(t) - I_{t-}(2)\triangle N_{23}(t), \end{aligned}$$

and so

$$\begin{aligned} \triangle I_t(1)p(dt,\{1\}) &= I_{t-}(4)dN_{41}(t),\\ \triangle I_t(1)p(dt,\{-1\}) &= -I_{t-}(1)dN_{12}(t),\\ \triangle I_t(2)p(dt,\{1\}) &= -I_{t-}(2)dN_{23}(t),\\ \triangle I_t(2)p(dt,\{-1\}) &= I_{t-}(1)dN_{12}(t). \end{aligned}$$

Owing to the obvious relations

$$\begin{gathered} I_4(t)\equiv I_4(t)(1-Y_t),\quad I_2(t)\equiv I_2(t)(1-Y_t),\\ I_1(t)\equiv I_1(t)Y_t,\quad I_3(t)\equiv I_3(t)Y_t \end{gathered}$$

we have

$$\begin{aligned} \pi_{t-}(2)dt = \pi_{t-}(2)(1-Y_{t-})dt,\quad \pi_{t-}(2)dt = \pi_{t-}(2)(1-Y_{t-})dt,\\ \pi_{t-}(1)dt = \pi_{t-}(1)Y_{t-}dt,\quad \pi_{t-}(3)dt = \pi_{t-}(3)Y_{t-}dt. \end{aligned} \tag{A.7}$$

Taking into account (A.3), we find

$$\begin{aligned} H_1(\omega,t,y) &= \begin{cases} \pi_{t-}(4), & y=1,\\ -\pi_{t-}(1), & y=-1, \end{cases}\\ H_2(\omega,t,y) &= \begin{cases} -\pi_{t-}(2), & y=1,\\ \pi_{t-}(1), & y=-1. \end{cases} \end{aligned}$$

In accordance with (A.3), (A.4), the formulae for H_1, H_2, and (A.7), we transform (A.6) to

$$\begin{aligned} d\pi_1(t) &= \big(-\pi_t(1)+\pi_t(4)\big)dt + \pi_{t-}(4)(1-Y_{t-})(dY_t - dt) + \pi_{t-}(1)Y_{t-}(dY_t+dt)\\ &= \pi_{t-}(4)(1-Y_{t-})dY_t + \pi_{t-}(1)Y_{t-}dY_t\\ &= \big(1-\pi_{t-}(2)\big)(1-Y_{t-})dY_t + \pi_{t-}(1)Y_{t-}dY_t,\\ d\pi_2(t) &= \big(\pi_t(1)-\pi_t(2)\big)dt - \pi_{t-}(2)(1-Y_{t-})(dY_t-dt) - \pi_{t-}(1)Y_{t-}(dY_t+dt)\\ &= -\pi_{t-}(2)(1-Y_{t-})dY_t - \pi_{t-}(1)Y_{t-}dY_t. \qquad \square \end{aligned}$$

Acknowledgments. The authors gratefully acknowledge Boris Tsirelson for bringing [43] and the example in [44] to their attention, Rami Atar for suggesting use of Theorem 1 in [2] for the proof of Theorem 4.1, and the anonymous referees whose comments and advice allowed us to improve the paper significantly.

REFERENCES

[1] R. Atar, *Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain*, Ann. Probab., 26 (1998), pp. 1552–1574.

[2] R. Atar and O. Zeitouni, *Exponential stability for nonlinear filtering*, Ann. Inst. H. Poincaré Probab. Statist., 33 (1997), pp. 697–725.

[3] R. Atar and O. Zeitouni, *Lyapunov exponents for finite state nonlinear filtering*, SIAM J. Control Optim., 35 (1997), pp. 36–55.

[4] R. Atar, F. Viens, and O. Zeitouni, *Robustness of Zakai's equation via Feynman-Kac representations*, in Stochastic Analysis, Control, Optimization, and Applications, W. M. McEneaney, G. Yin, and Q. Zhang, eds., Birkhäuser Boston, Boston, 1998, pp. 339–352.

[5] V. E. Beneš and I. Karatzas, *Estimation and control for linear, partially observable systems with non-Gaussian initial distribution*, Stochastic Process Appl., 14 (1983), pp. 233–248.

[6] A. G. Bhatt, A. Budhiraja, and R. L. Karandikar, *Markov property and ergodicity of the nonlinear filter*, SIAM J. Control Optim., 39 (2000), pp. 928–949.

[7] A. Budhiraja, *Ergodic properties of the nonlinear filter*, Stochastic Process Appl., 95 (2001), pp. 1–24.

[8] A. Budhiraja, *On invariant measures of discrete time filters in the correlated signal-noise case*, Ann. Appl. Probab., 12 (2002), pp. 1096–1113.

[9] A. Budhiraja and D. Ocone, *Exponential stability of discrete-time filters for bounded observation noise*, Systems Control Lett., 30 (1997), pp. 185–193.

[10] A. Budhiraja and D. Ocone, *Exponential stability in discrete-time filtering for nonergodic signals*, Stochastic Process. Appl., 82 (1999), pp. 245–257.

[11] A. Budhiraja and H. J. Kushner, *Robustness of nonlinear filters over the infinite time interval*, SIAM J. Control Optim., 36 (1998), pp. 1618–1637.

[12] A. Budhiraja and H. J. Kushner, *Approximation and limit results for nonlinear filters over an infinite time interval*, SIAM J. Control Optim., 37 (1999), pp. 1946–1979.

[13] A. Budhiraja and H. J Kushner, *Approximation and limit results for nonlinear filters over an infinite time interval.* II. *Random sampling algorithms*, SIAM J. Control Optim., 38 (2000), pp. 1874–1908.

[14] F. Cerou, *Long time behavior for some dynamical noise free nonlinear filtering problems*, SIAM J. Control Optim., 38 (2000), pp. 1086–1101.

[15] J. M. C. Clark, D. L. Ocone, and C. Coumarbatch, *Relative entropy and error bounds for filtering of Markov process*, Math. Control Signals Systems, 12 (1999), pp. 346–360.

[16] G. Da Prato, M. Fuhrman, and P. Malliavin, *Asymptotic ergodicity of the process of conditional law in some problem of non-linear filtering*, J. Funct. Anal., 164 (1999), pp. 356–377.

[17] P. Del Moral and A. Guionnet, *On the stability of measure valued processes with applications to filtering*, C. R. Acad. Sci. Paris Ser. I Math., 329 (1999), pp. 429–434.

[18] P. Del Moral and A. Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Ann. Inst. H. Poincaré Probab. Statist., 37 (2001), pp. 155–194.

[19] B. Delyon and O. Zeitouni, *Lyapunov exponents for filtering problem*, in Applied Stochastic Analysis, M. H. A. Davis and R. J.Elliot, eds., Gordon & Breach, New York, 1991, pp. 511–521.

[20] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.

[21] T. Kaijser, *A limit theorem for partially observed Markov chains*, Ann. Probab., 3 (1975), pp. 677–696.

[22] H. Kunita, *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal., 1 (1971), pp. 365–393.

[23] H. Kunita, *Ergodic properties of nonlinear filtering processes*, in Spatial Stochastic Processes, Progr. Probab. 19, Birkhäuser Boston, Boston, 1991, pp. 233–256.

[24] F. Le Gland and L. Mevel, *Exponential forgetting and geometric ergodicity in hidden Markov models*, Math. Control Signals Systems, 13 (2000), pp. 63–93.

[25] P. Del Moral and L. Miclo, *On the stability of nonlinear Feynman-Kac semigroups*, Ann.

Fac. Sci. Toulouse Math.6 , 11 (2002), pp. 135–175.

[26] F. Le Gland and N. Oudjane, *Stability and Uniform Approximation of Nonlinear Filters Using the Hilbert Metric, and Application to Particle Filters*, RR-4215, INRIA, LeChesnay, France, 2001. Available online at http://www.inria.fr/rrrt/rr-4215.html.

[27] F. Le Gland and N. Oudjane, *A Robustification Approach to Stability and to Uniform Particle Approximation of Nonlinear Filters: The Example of Pseudo-mixing Signals*, 4431, INRIA, 2002.

[28] A. Le Breton and M. Roubaud, *Asymptotic optimality of approximate filters in stochastic systems with colored noises*, SIAM J. Control Optim., 39 (2000), pp. 917–927.

[29] R. Sh. Liptser and A. N. Shiryaev, *Statistics of Random Processes* I, 2nd ed., Springer-Verlag, Berlin, 2001.

[30] R. Sh. Liptser and A. N. Shiryaev, *Statistics of Random Processes* II, 2nd ed., Springer-Verlag, Berlin, 2001.

[31] R. Sh. Liptser and A. N. Shiryaev, *Theory of Martingales*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.

[32] A. M. Makowski, *Filtering formula for partially observed linear systems with non-Gaussian initial conditions*, Stochastics, 16 (1986), pp. 1–24.

[33] A. M. Makowski and R. B. Sowers, *Discrete-time filtering for linear systems with non-Gaussian initial conditions: Asymptotic behaviors of the difference between the MMSE and the LMSE estimates*, IEEE Trans. Automat. Control, 37 (1992), pp. 114–120.

[34] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, UK, 1997.

[35] D. Ocone and E. Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM J. Control Optim., 34 (1996), pp. 226–243.

[36] D. Ocone, *Asymptotic stability of Beneš filters*, Stochastic Anal. Appl., 17 (1999), pp. 1053–1074.

[37] D. Ocone, *Entropy inequalities and entropy dynamics in nonlinear filtering of diffusion processes*, in Stochastic Analysis, Control, Optimization and Applications, Systems Control Found. Appl., Birkhäuser Boston, Boston, 1999, pp. 477–496.

[38] E. Seneta, *Nonnegative Matrices and Markov Chains*, 2nd ed., Springer-Verlag, New York, 1981.

[39] Ya. D. Sinai, *Kolmogorov's work on ergodic theory*, Ann. Probab., 17 (1989), pp. 833–839.

[40] L. Stettner, *On invariant measure of filtering processes*, in Stochastic Differential Systems, Lecture Notes in Control and Inform. Sci. 126, Springer-Verlag, Berlin, 1989, pp. 279–292.

[41] L. Stettner, *Invariant measures of the pair: State, approximate filtering process*, Colloq. Math., 62 (1991), pp. 347–351.

[42] A. N. Shiryaev, *Probability*, 2nd ed., Springer-Verlag, New York, 1996.

[43] H. von Weizsäcker, *Exchanging the order of taking suprema and countable intersections of σ-algebras*, Ann. Inst. H. Poincaré Sect. B, 19 (1983), pp. 91–100.

[44] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, UK, 1991.

[45] W. M. Wonham, *Some applications of stochastic differential equations to optimal nonlinear filtering*, J. Soc. Indust. Appl. Math. Ser. A Control, 2 (1965), pp. 347–369.