Exponential stability for nonlinear filtering

Rami Atar
Department of Electrical Engineering
Technion- Israel Institute of Technology
Haifa 32000, Israel

Ofer Zeitouni*
Department of Electrical Engineering
Technion- Israel Institute of Technology
Haifa 32000, Israel

May 5, 1996

Abstract

We study the a.s. exponential stability of the optimal filter w.r.t. its initial conditions. A bound is provided on the exponential rate (equivalently, on the memory length of the filter) for a general setting both in discrete and in continuous time, in terms of Birkhoff's contraction coefficient. Criteria for exponential stability and explicit bounds on the rate are given in the specific cases of a diffusion process on a compact manifold, and discrete time Markov chains on both continuous and discrete-countable state spaces.

Résumé

Nous étudions la stabilité du filtre optimal par raport à ses conditions initiales. Le taux de décroissance exponentielle est calculé dans un cadre général, pour temps discret et temps continu, en terme du coefficient de contraction de Birkhoff. Des critères de stabilité exponentielle et des bornes explicites sur le taux sont calculées pour les cas particuliers d'une diffusion sur une variéte compacte, ainsi que pour des chaines de Markov sur un espace discret ou continu.

AMS SUBJECT CLASSIFICATION: 93E11, 60J57.
KEYWORDS: Nonlinear filtering. Exponential stability. Birkhoff contraction coefficient.

1 Introduction

To a Markov process x_t on a state space S, we attach an observation σ-field $\mathcal{F}_t$ generated by an observation corrupted by Gaussian white noise (precise definitions are postponed to Section 2). The optimal filtering problem consists then of computing the conditional law of x_t given $\mathcal{F}_t$. This

*This work was partially supported by the Israel Science Foundation, administered by the Israeli Academy of Sciences and Humanities.

work is concerned with the asymptotic dependence of the optimal filter on its initial conditions, and the relation between this dependence and the corrupting noise intensity.

It is known since the work of Kunita [12] that under mild conditions, the conditional law, viewed as a random process taking values in the space of probability measures, is stationary when appropriately initialized. Stettner [18] shows that whenever the state process is a Feller Markov process converging in law to its unique invariant measure, so is its conditional law. Actually, c.f. [19], the joint law of the state and its filtering process is Markovian even if the filter is wrongly initialized. It thus seems natural to investigate the rate of convergence and to study the sensitivity of the optimal filter to its initialization with the wrong initial measure. This issue is also highly relevant for numerical and practical computation of the optimal filter or its approximations, for almost never does one have access to the true initial distribution.

Several approaches exist to analyze this sensitivity. In a recent article [14], Ocone and Pardoux have studied L^p type of convergence, and showed that actually, the nonlinear filter initialized at the wrong initial condition converges (in an L^p sense) to the nonlinear filter initialized at the correct initial condition. In particular cases (most notably, the Kalman filter), this convergence is exponential. In general, however, no rates of convergence are given by this approach.

Another approach, first suggested in [7], is to use Lyapunov exponent techniques. Building on Bougerol [4], it is shown in [7] that under suitable algebraic conditions, the dependence on initial conditions of the optimal filter of finite state Markov processes observed in white noise decays exponentially. This approach was recently extended in [1], and the question is essentially resolved in the case where the state process is Markovian with finite state space. In particular, the decay rate is shown to be related to the Lyapunov exponents of the Zakai equation, and, with good observation, to increase as the perturbing noise becomes weaker.

Below we provide bounds on the convergence rate of the filter to its ergodic behaviour in a more general setting. The technique is based on considering the unnormalized filtering process as a positive flow in the space of measures, acting on the initial measure, and then bounding the amount of contraction generated by the flow, at a discrete skeleton of time instances. To this purpose, we have found the Hilbert metric and the related Birkhoff contraction coefficient, most appropriate. The idea to use this approach originated from a result of Peres [16, Prop. 5], where a related question is treated, concerning the contraction of finite dimensional random operators. Next, we treat specific filtering models, for which we prove exponential convergence rate, and, when possible, provide explicit bounds on the rate.

The rest of the paper is organized as follows. In Section 2, we define the setting of the optimal filtering problem we deal with, quote a general result of Birkhoff concerning positive operators, and use it to formulate a bound on the decay rate referred to above. This bound is used to deduce more specific results in the special cases, treated in the succeeding sections 3 and 4. In Section 3, the discrete time filtering problem is treated. First, a general uniform bound on the decay rate is obtained, depending only on the law of the state process itself. Then, in the case of a bounded domain one-dimensional state process, observed via a $1:1$ observation function, it is proved that the decay rate converges to ∞ as the perturbing noise intensity becomes smaller. In Section 4, bounds are derived for the filtering of a diffusion on a compact manifold, observed in white noise. Finally, in Section 5, a discrete time, countable state space process is treated, with an approach that is independent of the general results of Section 2. We provide a bound on the decay rate in the weak observation noise limit, which, in special cases, including the filtering of a nearest-neighbour process, is shown to provide the exact asymptotic behaviour.

2 General setting and a basic estimate

Let T denote either $\mathbb{R}_+$ or $\mathbb{Z}_+$ (cases which we refer to as continuous or discrete time, respectively), and consider a homogeneous Markov process $\{x_t\}_{t\in T}$ on the probability space $(\Omega_x, \mathcal{B}_x, \{\mathcal{B}_x(t)\}_{t\in T}, P_x)$ defined as follows. Let the state space S be a Polish space and let $\mathcal{S}$ be the σ-field of Borel subsets of S. In continuous time we set Ω_x to be the set of all mappings from $\mathbb{R}_+$ to S which are RCLL and equip it with Skorohod's metric relative to the metric defined on S. Let $\mathcal{B}_x$ be the Borel σ-field on Ω_x w.r.t. this metric. For $t \geq 0$ and $\omega \in \Omega_x$, let $x_t(\omega)$ denote the value of ω at time t. For $t \geq 0$, let $x|_{[0,t]}$ be the restriction of x to $[0, t]$. Let $\mathcal{B}_x(t) = \sigma\{x|_{[0,t]}\}$ be the smallest sub σ-field of $\mathcal{B}_x$ for which all $x|_{[0,t]}$ are measurable. We assume that the semigroup $G_t, t \geq 0$ associated with the transition probabilities $P_t(x, \cdot)$ is a Feller semigroup i.e., if f is any continuous function on S then so is $G_t f$, and $\lim_{t\to 0} G_t f(x) = f(x)$ uniformly on S. The infinitesimal generator of G_t is denoted by $\mathcal{L}$. In discrete time we set $\Omega_x = \prod_{i=0}^{\infty} S_i, \mathcal{B}_x(t) = \bigvee_{i=0}^{t} \mathcal{S}_i$ and $\mathcal{B}_x = \bigvee_{i=0}^{\infty} \mathcal{S}_i$ where $(S_i, \mathcal{S}_i)$ are all identical copies of $(S, \mathcal{S})$, and denote the transition kernel by $G(\cdot, \cdot)$. A Markovian family of measures $P_x^{(\alpha)}, \alpha \in S$ is defined on $(\Omega_x, \mathcal{B}_x)$ in continuous time by

$$P_x^{(\alpha)}(x_{t_1} \in E_1, \ldots, x_{t_n} \in E_n) = \int_{E_1\times\cdots\times E_n} G_{t_1}(\alpha, dx_1)\cdots G_{t_n - t_{n-1}}(x_{n-1}, dx_n)$$

where $0 < t_1 < \cdots < t_n$ and $E_1, \ldots, E_n \in \mathcal{S}$. In discrete time $P_x^{(\alpha)}$ is induced by $G(\cdot, \cdot)$ on $(\Omega_x, \mathcal{B}_x)$ in a similar manner. With any probability measure q on $(S, \mathcal{S})$ there will be associated a Markovian measure P_x^q defined as

$$P_x^q = \int_S P_x^{(\alpha)} q(d\alpha).$$

Further, let $\{\nu_t\}_{t\in T}$ be a process on $(\Omega_\nu, \mathcal{B}_\nu, \{\mathcal{B}_\nu(t)\}_{t\in T}, P_\nu)$ defined as follows. In continuous time, Ω_ν is the set of continuous mappings from $\mathbb{R}_+$ to $\mathbb{R}^d$, $\mathcal{B}_\nu$ is the Borel σ-field w.r.t. the sup-norm, and $\mathcal{B}_\nu(t) = \sigma\{\nu|_{[0,t]}\}$, where $\nu|_{[0,t]}$ is defined similarly to the above. Let P_ν be the d-dimensional standard Wiener measure on $(\Omega_\nu, \mathcal{B}_\nu)$. In discrete time $\{\nu_t\}_{t\in\mathbb{Z}_+}$ is a standard normal random walk in $\mathbb{Z}^d$ starting at the origin. Now let $\Omega = \Omega_x \times \Omega_\nu$, $\mathcal{B} = \mathcal{B}_x \times \mathcal{B}_\nu$ and $P^q = P_x^q \times P_\nu$. We fix a specific measure p_0 on $(S, \mathcal{S})$ that will be called the "exact" initial distribution of $\{x_t\}_{t\in T}$, and denote $P_x = P_x^{p_0}, P = P^{p_0}$. Expectation w.r.t. P^q $(P, P^{(\alpha)})$ will be denoted by E^q $(E, E^{(\alpha)}$, respectively). Let $g : S \to \mathbb{R}^d$ be a measurable function, and let the observation process on $\mathbb{R}^d$ be defined by

$$y_t = \int_0^t g(x_s)ds + \nu_t \quad t \in \mathbb{R}_+,$$

$$y_t = \sum_1^t g(x_i) + \nu_t \quad t \in \mathbb{Z}_+$$

where by convention $\sum_1^0 = 0$. We shall denote the observation filtration of sub σ-fields by

$$\mathcal{F}_t = \sigma\{y|_{[0,t]}\} \quad t \in T.$$

The filtering of x_t based on $\mathcal{F}_t$ is defined as the conditional distribution of x_t conditioned on $\mathcal{F}_t$ i.e., for all measurable bounded f,

$$p_t^q(f) = E^q[f(x_t)|\mathcal{F}_t] \quad t \in T. \tag{1}$$

We denote the "exact" filtering process by $p_t(f) = p_t^{p_0}(f)$.

In the continuous time case, under further assumptions of compactness of S and continuity of g, it follows from Kunita [12, Theorem 1.1] that $\{p_t\}, t \geq 0$ satisfies the Kushner equation namely,

$$p_t(f) = p_0(f) + \int_0^t p_s(\mathcal{L}f)ds + \int_0^t (p_s(fg) - p_s(f)p_s(g),\, dy_s - p_s(g)ds)\,, \quad f \in \mathcal{D}(\mathcal{L}).$$

Zakai's approach of the reference measure provides another expression for p_t, in terms of a bilinear SPDE [20]. Motivated by this approach, we define the reference measure P_0 on $(\Omega, \mathcal{B}(t), \mathcal{B})$ as follows: $P_0^{(\alpha)}(A \cap B) = P_x^{(\alpha)}(A)P'(B)$ for all $A \in \mathcal{B}_x$ and $B \in \mathcal{B}'$, where $\mathcal{B}'$ is the smallest sub σ-field of $\mathcal{B}$ for which all $y_t, 0 \leq t < \infty$ are measurable. P' is a standard Wiener measure in the continuous case, and a standard normal random walk in the discrete case. This defines the joint law of x and y, and since ν is determined by x and y, also the joint law of x and ν. Define similarly P_0^q and P_0, and denote by $E_0^{(\alpha)}, E_0^q$ and E_0 expectations under $P_0^{(\alpha)}, P_0^q$ and P_0, respectively.

Define now on $(\Omega, \mathcal{B}(t), \mathcal{B}, P_0^q)$ the process

$$\Lambda_t = \exp\left(\int_0^t (g(x_s),\, dy_s) - \frac{1}{2}\int_0^t \|g(x_s)\|^2 ds\right) \quad t \in \mathbb{R}_+, \tag{2}$$

$$\Lambda_t = \exp\left(\sum_1^t (g(x_i),\, \Delta y_i) - \frac{1}{2}\sum_1^t \|g(x_i)\|^2\right) \quad t \in \mathbb{Z}_+ \setminus \{0\}, \quad \Lambda_0 = 1, \tag{3}$$

where $\Delta y_i = y_i - y_{i-1}$. Assume that $E\int_0^t \|g(x_s)\|^2 ds < \infty$ for any $t \in \mathbb{R}_+$. Then Λ_t of (2) is a martingale. It follows from Girsanov's theorem [13, chapter 6.3], [10, chapter 3.5] in the continuous case, and from Bayes' rule in the discrete case, that for every $t \in T \setminus \{0\}$, the restriction $P^{q,t}$ of P^q to $\mathcal{B}(t)$ is absolutely continuous w.r.t. the restriction $P_0^{q,t}$ of P_0^q to $\mathcal{B}(t)$, and that

$$\frac{dP^{q,t}}{dP_0^{q,t}} = \Lambda_t.$$

It follows that for any measurable bounded function f, P^q-a.s.,

$$p_t^q(f) = \frac{E_0^q(f(x_t)\Lambda_t|\mathcal{F}_t)}{E_0^q(\Lambda_t|\mathcal{F}_t)}. \tag{4}$$

The unnormalized conditional measure $\rho_t^q(f),\ t \in T$, defined by

$$\rho_t^q(f) = E_0^q(f(x_t)\Lambda_t|\mathcal{F}_t), \tag{5}$$

therefore has the property that $p_t^q(f) = \rho_t^q(f)/\rho_t^q(1)$. Since S is assumed to be a Polish space, Ω with Skorohod's metric is itself a Polish space. The map $x_0^t \mapsto x_t$ is measurable, and thus there exists a regular conditional probability distribution [15, pp. 146–7]

$$P_0^{(\alpha)}(\cdot\,|\mathcal{F}_t, x_t = \beta),$$

satisfying for $B \in \mathcal{B}(t)$ and $A \in \mathcal{S}$

$$P_0^{(\alpha)}(B \cap \{x_t \in A\} | \mathcal{F}_t) = \int_A P_0^{(\alpha)}(B | \mathcal{F}_t, x_t = \beta) P_0^{(\alpha)}(x_t \in d\beta).$$

We thus obtain from (5) by conditioning first on $\{\mathcal{F}_t, x_t\}$, the following representation for the unnormalized conditional distribution

$$\rho_t^q(f) = \int_S \int_S f(\beta) G_t(\alpha, d\beta) I_t(\alpha, \beta; y_0^t) q(d\alpha), \tag{6}$$

where we define

$$I_t(\alpha, \beta; y_0^t) = E_0^{(\alpha)}(\Lambda_t | \mathcal{F}_t, x_t = \beta). \tag{7}$$

We shall also use the following notation for any finite measures μ, λ and a measurable function f on $(S, \mathcal{S})$:

$$I_t^\lambda(\mu; y_0^t)(f) = \int_S \int_S I_t(\alpha, \beta; y_0^t) f(\beta) \mu(d\alpha) \lambda(d\beta).$$

In order to be able to compare different filtering processes (namely, $p_t^q(f)$, for different values of q), we use (4) and (5) to define $p_t^q(\cdot)$ and $\rho_t^q(\cdot)$, respectively, as stochastic processes on the common probability space $(\Omega, \mathcal{B}, P)$, as follows. We first regard, for each q, equation (4) as the definition of a mapping $y|_{[0,t]} \mapsto p_t^q$. Then, the law of p_t^q on $(\Omega, \mathcal{B}, P)$ is induced from that of y_t on the same probability space by this mapping. The definition of ρ_t^q is similar.

Our goal is to study the behaviour under P of the asymptotic rate

$$\gamma(q, q') = \limsup_{t \to \infty} \frac{1}{t} \log d_{TV}(p_t^q, p_t^{q'}). \tag{8}$$

We shall establish exponential stability of the filtering process by providing conditions for $\gamma(q, q')$ to be negatively bounded, uniformly in q and q'. Note that whenever either the Kushner or the Zakai equations apply, this provides exponential stability of these equations w.r.t. their initial conditions (see [8] for a study of the related topic of Lyapunov exponents for a general bilinear SPDE).

We now deviate to some classical definitions and results regarding Hilbert's metric and contraction of positive operators, which we adapt to our setting. Thus we focus on positive operators on finite measures. Let $\mathcal{V}$ be the vector space of finite signed measures on $(S, \mathcal{S})$. Define on it the partial order relation $\leq$ by $\lambda \leq \mu$ for $\lambda, \mu \in \mathcal{V}$ iff $\lambda(A) \leq \mu(A)$ for every $A \in \mathcal{S}$. Let $\mathcal{C} \subset \mathcal{V}$ be the resulting positive cone. Then $\lambda \in \mathcal{C}$ iff λ is a finite non-zero measure on $(S, \mathcal{S})$. Moreover, $\mathcal{V}$ is an integrally closed, directed vector space (see [3, pp. 289, 290] for definitions). Two elements $\lambda, \mu \in \mathcal{C}$ are called *comparable* if $\alpha\lambda \leq \mu \leq \beta\lambda$ for suitable positive scalars α, β. The *Hilbert metric,* also called the *projective distance,* on $\mathcal{C}$ is defined as

$$h(\lambda, \mu) = \log \frac{\sup_{A \in \mathcal{S} \ \mu(A) > 0} \lambda(A)/\mu(A)}{\inf_{A \in \mathcal{S} \ \mu(A) > 0} \lambda(A)/\mu(A)} \qquad \text{if } \lambda, \mu \in \mathcal{C} \text{ are comparable} \tag{9}$$

and $h(\lambda, \mu) = \infty$ otherwise. The function h defined above is a pseudo-metric on $\mathcal{C}$ (see e.g., [3, chapter 16] or, in a discrete setup [17]) and hence is a metric on the space of all probability measures on $(S, \mathcal{S})$ that are comparable to any fixed measure $\lambda_0 \in \mathcal{C}$.

Let now K be a linear *positive* operator on $\mathcal{V}$ i.e., which maps $\mathcal{C}$ into itself. Then K is a contraction under the Hilbert metric and as has been shown by Birkhoff [2] and Hopf [9] (see also [3, chapter 16], [17])

$$\tau(K) \equiv \sup_{0<h(\lambda,\mu)<\infty} \frac{h(K\lambda, K\mu)}{h(\lambda,\mu)} = \tanh \frac{H(K)}{4} \tag{10}$$

where $H(K)$ is the h-diameter of the transform $K\mathcal{C}$ of $\mathcal{C}$ under K i.e.,

$$H(K) = \sup_{\lambda,\mu\in\mathcal{C}} h(K\lambda, K\mu).$$

The function τ defined in (10) is called *Birkhoff's contraction coefficient.* We shall denote by $\mathcal{P}$ the subset of $\mathcal{C}$ of probability measures on $(S, \mathcal{S})$.

Note that if for a specific $\lambda \in \mathcal{C}$, the operator K has the kernel representation

$$K\mu(f) = \int_S \int_S K(\alpha,\beta) f(\beta) \mu(d\alpha) \lambda(d\beta), \tag{11}$$

where $K(\alpha, \beta)$ is non-negative, then

$$H(K) = \log \operatorname{ess\,sup} \frac{K(\alpha,\beta)K(\alpha',\beta')}{K(\alpha,\beta')K(\alpha',\beta)} \tag{12}$$

with the convention $\frac{0}{0} = 1$ and $\frac{1}{0} = \infty$. The supremum above is strict over $\alpha, \alpha' \in S$, and is essential over $\beta, \beta' \in S$ w.r.t. λ.

Back to the filtering problem, let now $q \in \mathcal{V}$ be any signed measure, and for $t \in T$ let $J_{0,t} : \mathcal{V} \to \mathcal{V}$ denote the ($\mathcal{F}_t$-measurable) mapping $q \mapsto \rho_t^q$ defined by equation (6). Hence, for $q \in \mathcal{P}$, $\rho_t^q = J_{0,t}q$. Next, for $s, t \in T$ with $s \leq t$ define $J_{s,t} = J_{0,t-s} \circ \theta_s$ where θ_s is the shift operator on Ω. Thus $J_{s,t}$ is a positive linear operator on $\mathcal{V}$. Moreover, the smoothing property for the conditional expectation implies that for $q \in \mathcal{P}$ and $0 \leq s \leq t$

$$\rho_t^q = J_{s,t} J_{0,s} q. \tag{13}$$

We assume that there exists a unique invariant measure, p_s, of the Markovian family $P_x^{(\alpha)}$ (associated with the state process), and let the stationary measure be denoted by $P_s = P_x^{p_s}$ (and $E_s = E_x^{p_s}$). We assume the following ergodicity assumptions of the state process.

Assumption A1 P_s is ergodic.

Assumption A2 The measures P_x and P_s are absolutely continuous on the tail σ-field.

Our basic estimate will be the following

Theorem 1 *Let the assumptions of section 2 hold for either $T = \mathbb{Z}^+$ or $T = \mathbb{R}^+$. Let $\delta \in T \setminus \{0\}$. Then the following hold.*

(a) For any comparable $q, q' \in \mathcal{P}$, P-a.s.,

$$\gamma(q, q') \leq \frac{1}{\delta} E_s \log \tau(J_{0,\delta}). \tag{14}$$

(b) *If for every $\alpha \in S$, $G_\delta(\alpha,\cdot)$ is absolutely continuous w.r.t. a specific measure $\lambda \in \mathcal{C}$, then for any comparable $q, q' \in \mathcal{P}$, P-a.s.,*

$$\gamma(q,q') \leq \frac{1}{\delta} E_s \log \tanh \frac{H(G_\delta) + H(I_\delta^\lambda(\cdot\,; y_0^\delta))}{4}. \tag{15}$$

(c) *If for every $\alpha \in S$, $G_\delta(\alpha,\cdot)$ is comparable to a specific measure $\lambda \in \mathcal{C}$, and $I_\delta(\alpha,\beta;y_0^\delta)$ can be bounded above and below by positive bounds that do not depend on α and β, then (14) and (15) hold P-a.s. for any $q, q' \in \mathcal{P}$.*

To prove Theorem 1, we need the

Lemma 1 *Let $\lambda, \mu \in \mathcal{P}$. Then*

$$d_{TV}(\lambda,\mu) \equiv 2 \sup_{A \in \mathcal{S}} |\lambda(A) - \mu(A)| \leq 2 \wedge (e^{h(\lambda,\mu)} - 1) \leq \frac{2}{\log 3} h(\lambda,\mu). \tag{16}$$

Proof of Lemma 1 In case that λ and μ are not comparable (16) holds trivially. Otherwise set $\mathcal{A} = \{A \in \mathcal{S} : \lambda(A) \geq \mu(A) > 0\}$ and $\mathcal{A}' = \{A \in \mathcal{S} : \mu(A) > \lambda(A) > 0\}$. Note that $\mathcal{A}$ cannot be empty. Moreover, since if $\mathcal{A}'$ is empty it follows that $\lambda \equiv \mu$ with which (16) again holds trivially, we shall assume the contrary. Thus we have

$$1 \leq \frac{\lambda(A)}{\mu(A)} \leq \frac{\lambda(A)}{\mu(A)} \frac{\mu(B)}{\lambda(B)} \leq e^{h(\lambda,\mu)} \quad A \in \mathcal{A}, B \in \mathcal{A}'$$

which implies

$$\begin{aligned} 0 \leq \lambda(A) - \mu(A) &\leq \mu(A)(e^{h(\lambda,\mu)} - 1) \quad A \in \mathcal{A} \\ 0 \leq \mu(B) - \lambda(B) &\leq \lambda(B)(e^{h(\lambda,\mu)} - 1) \quad B \in \mathcal{A}' \\ d_{TV}(\lambda,\mu) = \sup_{A \in \mathcal{A}} \lambda(A) - \mu(A) - \lambda(A^c) + \mu(A^c) &\leq e^{h(\lambda,\mu)} - 1 \end{aligned}$$

and (16) follows. □

Proof of Theorem 1:
For either $t \in \mathbb{Z}_+$ or $t \in \mathbb{R}_+$ let $n = [\frac{t}{\delta}]$. Use Lemma 1, the insensitivity of h to multiplication by a positive scalar, the property (13), the definition (10) of τ and the fact that $\tau(\cdot) \leq 1$ to get

$$\begin{aligned} \frac{1}{t} \log d_{TV}(p_t^q, p_t^{q'}) - \frac{1}{t} \log \frac{2}{\log 3} &\leq \frac{1}{t} \log h(p_t^q, p_t^{q'}) = \frac{1}{t} \log h(\rho_t^q, \rho_t^{q'}) \\ &\leq \frac{1}{t} \log h(\rho_{n\delta}^q, \rho_{n\delta}^{q'}) \\ &= \frac{1}{t} \log \left[\frac{h(\rho_{n\delta}^q, \rho_{n\delta}^{q'})}{h(\rho_{(n-1)\delta}^q, \rho_{(n-1)\delta}^{q'})} \cdots \frac{h(\rho_\delta^q, \rho_\delta^{q'})}{h(q,q')} h(q,q') \right] \\ &\leq \frac{n}{t} \frac{1}{n} \sum_{i=0}^{n-1} \log \tau(J_{i\delta,(i+1)\delta}) + \frac{1}{t} \log h(q,q'). \end{aligned} \tag{17}$$

Since $\log \tau(J_{i\delta,(i+1)\delta})$ depends only on $(x|_{[i\delta,(i+1)\delta]}, \nu|_{[i\delta,(i+1)\delta]} - \nu_{i\delta})$, the result (14) follows from Assumption A1 and Birkhoff's ergodic theorem, for $P_x = P_s$. Since (14) is a tail event, Assumption A2 implies part (a) of the theorem.

Next, under the absolute continuity assumption, let $G_\delta(\alpha, d\beta) = \hat{G}_\delta(\alpha,\beta)\lambda(d\beta)$ and define $J_{0,\delta}(\alpha,\beta) = \hat{G}_\delta(\alpha,\beta) I_\delta(\alpha,\beta; y_0^\delta)$. Then by (6) we get

$$J_{0,\delta}(q)(f) = \rho_\delta^q(f) = \int_S \int_S J_{0,\delta}(\alpha,\beta) f(\beta) q(d\alpha)\lambda(d\beta).$$

Thus $J_{0,\delta}$ has the kernel representation (11), and (12) can be used. We have

$$\begin{aligned} H(J_{0,\delta}) &= \log \operatorname{ess\,sup} \frac{J_{0,\delta}(\alpha,\beta) J_{0,\delta}(\alpha',\beta')}{J_{0,\delta}(\alpha,\beta') J_{0,\delta}(\alpha',\beta)} \\ &\leq \log \operatorname{ess\,sup} \frac{\hat{G}_\delta(\alpha,\beta)\hat{G}_\delta(\alpha',\beta')}{\hat{G}_\delta(\alpha,\beta')\hat{G}_\delta(\alpha',\beta)} + \log \operatorname{ess\,sup} \frac{I_\delta(\alpha,\beta;y_0^\delta) I_\delta(\alpha',\beta';y_0^\delta)}{I_\delta(\alpha,\beta';y_0^\delta) I_\delta(\alpha',\beta;y_0^\delta)} \\ &= H(G_\delta) + H(I_\delta^\lambda(\cdot\,; y_0^\delta)). \end{aligned}$$

Now (15) is deduced from (14) using (10).

As for part (c), note that

$$\frac{d(J_{0,\delta} q)}{d\lambda}(\beta) = \int_S \hat{G}(\alpha,\beta) I_\delta(\alpha,\beta;y_0^\delta) q(d\alpha).$$

Therefore, since by assumption both $\hat{G}$ and I_δ are bounded, the R.N. derivative above is bounded above and below (by bounds that do not depend on β), and ρ_δ^q and $\rho_\delta^{q'}$ must be comparable. Thus the argument expressed in (17) holds where $h(q,q')$ is replaced by $h(\rho_\delta^q, \rho_\delta^{q'})$. □

Remark: Note that ergodicity of the process $g(x_t)$ would suffice instead of that of x_t. In that case P_s is to be understood as the measure P^q with the appropriate q to make $g(x_t)$ a stationary process.

3 The discrete time case

In this section we are concerned with the case where the state process is a discrete time Markov chain. The filtering of a continuous time process observed in discrete time is a special case of this setting, if we understand the transition kernel $G(\cdot,\cdot)$ as that of the propagator of the continuous process, at time equal one observation-period. We keep all the assumptions and definitions of section 2 in the case $T = \mathbb{Z}_+$. We denote by G the unconditional discrete time propagator at time 1 i.e., $G\mu(A) = \int_S G(x,A)\mu(dx)$ for any $A \in \mathcal{S}$. Note that in this case I_1 defined in (7) is given by

$$I_1(\alpha,\beta;y_0^1) = \exp\left((g(\beta), \Delta y_1) - \frac{1}{2}\|g(\beta)\|^2\right). \tag{18}$$

The latter is independent of α and hence if g is bounded, $H(I_1(\cdot\,; y_0^1)) = 0$. Moreover, if the unconditional kernel satisfies the condition

$$c_1\mu_0(\cdot) \leq G(x,\cdot) \leq c_2\mu_0(\cdot) \quad x \in S \tag{19}$$

for positive constants c_1, c_2 and a $\mu_0 \in \mathcal{C}$, then by definition (9), $h(G\lambda, G\mu)$ is uniformly bounded for $\lambda, \mu \in \mathcal{P}$. Hence by the right hand side of (10), $\tau(G) < 1$. In the latter case, note also that for every $q, q' \in \mathcal{P}$ (not necessarily comparable) ρ_1^q and $\rho_1^{q'}$ are comparable. Therefore as a corollary of Theorem 1 we get

Corollary 1 *Assume that the state process $\{x_n\}_{n\in\mathbb{Z}^+}$ is ergodic, the transition kernel $G(\cdot,\cdot)$ satisfies (19) and that g is bounded. Then for any $q, q' \in \mathcal{P}$, P-a.s.,*

$$\gamma(q, q') \leq \log \tau(G) < 0.$$

Next we treat a specific example in which the state space is the unit interval and the observation function is real-valued and $1:1$. We assume that the observation process is given by

$$y_n = \sum_{i=1}^{n} \sigma^{-1} g(x_i) + \nu_n.$$

The positive scalar σ is used to parameterize the observation- (or equivalently, the noise-) intensity. Below we use the notion $\gamma_\sigma(\cdot,\cdot)$ for $\gamma(\cdot,\cdot)$ defined in (8), to denote the dependence on σ. In what follows we assume that the measures $G(x,\cdot)$ and ρ_n have densities w.r.t. the Lebesgue measure on S. Thus we let $G(x,\cdot)$ and ρ_n denote these densities rather than the measures, and for a measurable f on S, $x \in S$ we write $Gf(y) = \int_S f(x)G(x,y)dy$. The specific assumptions and result are summarized below.

Theorem 2 *Let $S = [0,1]$ and assume that $G(\cdot,\cdot) \in C^3(S^2)$. Let $q, q' \in \mathcal{P}$ and assume $g \in C^4(S)$. Assume moreover, that both $G(\cdot,\cdot)$ and the derivative $\dot{g}$ are positively bounded from below (on S^2 and S, respectively). Then*

$$\limsup_{\sigma\to 0} \frac{\gamma_\sigma(q,q')}{\log\frac{1}{\sigma}} \leq -1 \quad a.s. \tag{20}$$

Remark: A previous result [1, Theorems 3,7] in finite state space provides a bound for γ which is of order $\Omega(\sigma^{-2})$ as $\sigma \to 0$. Moreover, direct computations show that both in the Gaussian and in the Beneš cases (on $\mathbb{R}$), the behaviour is in $\Omega(\sigma^{-1})$. In view of these facts, it seems that the above result is not the optimal possible (see also the remark after this proof).

Proof: Denote

$$\Delta_n(\beta) = \exp\left(\sigma^{-1} g(\beta)\Delta y_n - \frac{1}{2}\,\sigma^{-2} g^2(\beta)\right) \tag{21}$$

and let the linear operator T_n be defined by

$$T_n f(\beta) = \Delta_n(\beta) G f(\beta). \tag{22}$$

Then it follows from (6) and (18) that $\rho_n(\beta) = T_n \rho_{n-1}(\beta)$. Since by assumption on $G(\cdot,\cdot)$, x_t must be ergodic, Theorem 1 can be used. We take $\delta = 2$ in (14) and apply equality (10). Our goal is to show that

$$\limsup_{\sigma\to 0} \frac{E_s \log \tau(J_{0,2})}{2\log\frac{1}{\sigma}} \leq -1. \tag{23}$$

Let f be measurable and bounded, and write the explicit expression for T_2T_1f as

$$T_2T_1f(w) = \Delta_2(w)\int_S G(v,w)\Delta_1(v)\int_S G(u,v)f(u)dudv.$$

Let $\psi(T_2T_1) = \exp(-H(J_{0,2}))$. Then, noting that the factor Δ_2 cancels out, it follows that

$$\psi(T_2T_1) = \inf_{u,u',w,w'\in S} \frac{\int_S \Delta_1(v)G(u,v)G(v,w)dv}{\int_S \Delta_1(v)G(u',v)G(v,w)dv}\frac{\int_S \Delta_1(v)G(u',v)G(v,w')dv}{\int_S \Delta_1(v)G(u,v)G(v,w')dv}. \tag{24}$$

We show below that the integrands in (24) are controlled by the behaviour of Δ_1, which tends to concentrate on one point as $\sigma \to 0$, causing the cross-ratio to converge to 1. Substituting $\tilde{v} = g(v)$, using the differentiability and monotonicity of g and the definition of Δ_n (21), the first integral in the numerator of (24) reads

$$\int_{g(S)} \frac{1}{\dot{g}(g^{-1}(\tilde{v}))}\exp(\frac{\tilde{v}y_1}{\sigma} - \frac{\tilde{v}^2}{2\sigma^2})G(u,g^{-1}(\tilde{v}))G(g^{-1}(\tilde{v}),w)d\tilde{v} \equiv \int_{g(S)} \tilde{h}(u,\tilde{v},w)d\tilde{v}\ ,$$

where by assumptions on g we have $g(S) \equiv \{g(s)|s\in S\} = [g(0),g(1)]$. Dividing $\tilde{h}(u,\tilde{v},w)$ by $\sqrt{2\pi}\sigma\exp(y_1^2/2)$, which does not depend on u,w, it follows that

$$\psi(T_2T_1) = \inf_{u,u',w,w'\in S} \frac{\int_{g(S)} h(u,\tilde{v},w)d\tilde{v}}{\int_{g(S)} h(u',\tilde{v},w)d\tilde{v}}\frac{\int_{g(S)} h(u',\tilde{v},w')d\tilde{v}}{\int_{g(S)} h(u,\tilde{v},w')d\tilde{v}} \tag{25}$$

where

$$h(u,\tilde{v},w) = \frac{\tilde{h}(u,\tilde{v},w)}{\sqrt{2\pi}\sigma\exp(y_1^2/2)} \equiv \frac{1}{\dot{g}(g^{-1}(\tilde{v}))}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\tilde{v}-\sigma y_1)^2}{2\sigma^2}\right)G(u,g^{-1}(\tilde{v}))G(g^{-1}(\tilde{v}),w).$$

For $\epsilon > 0$ define now $A = \{y_1|\sigma y_1 \in [g(0)+\epsilon, g(1)-\epsilon]\}$, and recall that $\sigma y_1 = g(x_\delta) + \sigma\nu_1$. Then

$$\begin{aligned}P_s(y_1 \notin A) &\leq P_s\left(g(x_\delta) \in [g(0)+2\epsilon, g(1)-2\epsilon] \cap |\sigma\nu_1| \geq \epsilon\right) + P_s\left(g(x_\delta) \notin [g(0)+2\epsilon, g(1)-2\epsilon]\right)\\ &\leq c(e^{-\epsilon^2/2\sigma^2} + \epsilon),\end{aligned}$$

for some constant c independent of ϵ and σ, provided that σ is sufficiently small. Since τ is always ≤ 1 it follows that

$$\begin{aligned}E_s \log\tau(J_{0,2}) &\leq E_s[\log\tau(J_{0,2})|y_1 \in A]P_s(y_1\in A)\\ &\leq E_s[\log\tau(J_{0,2})|y_1\in A](1 - c(e^{-\epsilon^2/2\sigma^2}+\epsilon)).\end{aligned} \tag{26}$$

We thus assume $y_1 \in A$ from now on. For $M > 0$ let us divide the integration region in (25) according to $|\tilde{v} - \sigma y_1| > \sigma M$ and its complement. Note that

$$0 \leq \int_{g(S)\cap\{\tilde{v}:|\tilde{v}-\sigma y_1|>\sigma M\}} h(u,\tilde{v},w)d\tilde{v} \leq \frac{c_1}{\sigma}\exp(-M^2/2) \tag{27}$$

for some constant $c_1 > 0$, and that if we keep $M\sigma < \epsilon$ then $g(S)\cap\{\tilde{v} : |\tilde{v}-\sigma y_1| \leq \sigma M\} = \{\tilde{v} : |\tilde{v}-\sigma y_1| \leq \sigma M\}$. Thus it follows from (25) and (27) that

$$\psi(T_2T_1) \geq \tag{28}$$
$$\inf_{u,u',w,w'\in S} \frac{\int_{|\tilde{v}-\sigma y_1|\leq\sigma M} h(u,\tilde{v},w)d\tilde{v}}{\int_{|\tilde{v}-\sigma y_1|\leq\sigma M} h(u',\tilde{v},w)d\tilde{v} + \frac{c_1}{\sigma}\exp(-M^2/2)}\ \frac{\int_{|\tilde{v}-\sigma y_1|\leq\sigma M} h(u',\tilde{v},w')d\tilde{v}}{\int_{|\tilde{v}-\sigma y_1|\leq\sigma M} h(u,\tilde{v},w')d\tilde{v} + \frac{c_1}{\sigma}\exp(-M^2/2)}.$$

Expand now the function

$$\varphi(u,\tilde{v},w) \equiv \frac{1}{\dot{g}(g^{-1}(\tilde{v}))} G(u,g^{-1}(\tilde{v}))G(g^{-1}(\tilde{v}),w)$$

with respect to $\tilde{v}$ around σy_1 (note that $y_1 \in A$) to get

$$\varphi(u,\tilde{v},w) = \varphi(u,\sigma y_1,w)+\varphi_2(u,\sigma y_1,w)(\tilde{v}-\sigma y_1)+\frac{1}{2}\varphi_{22}(u,\sigma y_1,w)(\tilde{v}-\sigma y_1)^2+\frac{1}{6}\varphi_{222}(u,\xi,w)(\tilde{v}-\sigma y_1)^3.$$

Here $|\xi - \sigma y_1| \le |\tilde{v} - \sigma y_1|$; the derivatives with respect to the second variable indeed exist and are continuous due to the differentiability assumptions on $G(\cdot,\cdot)$ and $g(\cdot)$. The integrals in (28) may thus be estimated by

$$\int_{|\tilde{v}-\sigma y_1|\le \sigma M} h(u,\tilde{v},w)d\tilde{v} = \varphi(u,\sigma y_1,w) + \frac{1}{2}\,\sigma^2\varphi_{22}(u,\sigma y_1,w) + \text{res} \tag{29}$$

where

$$|\text{res}| \le \int_{|\tilde{v}-\sigma y_1|\le \sigma M} \frac{1}{\sqrt{2\pi}\sigma} \max|\varphi_{222}|(\sigma M)^3 d\tilde{v} + c_2 e^{-M^2} \le c_3(\sigma^3 M^4 + e^{-M^2}) \tag{30}$$

for some $c_2, c_3 > 0$, independent of ϵ, σ and M, provided that M is sufficiently large. Setting $M = \sigma^{-1/8}$ so that $\sigma^3 M^4 = \sigma^{2.5}$, and combining (28), (29) and (30), it follows that for $y_1 \in A$

$$\psi(T_2T_1) \ge \inf_{u,u',w,w' \in S} A_{u,u'}^{w,w'} \tag{31}$$

where

$$A_{u,u'}^{w,w'} = \frac{(\varphi(u,\sigma y_1,w) + \frac{1}{2}\,\varphi_{22}(u,\sigma y_1,w)\sigma^2 - c_4\sigma^{2.5})(\varphi(u',\sigma y_1,w') + \frac{1}{2}\,\varphi_{22}(u',\sigma y_1,w')\sigma^2 - c_4\sigma^{2.5})}{(\varphi(u',\sigma y_1,w) + \frac{1}{2}\,\varphi_{22}(u',\sigma y_1,w)\sigma^2 + c_4\sigma^{2.5})(\varphi(u,\sigma y_1,w') + \frac{1}{2}\,\varphi_{22}(u,\sigma y_1,w')\sigma^2 + c_4\sigma^{2.5})}.$$

Noting that $\varphi(u,\tilde{v},w)\varphi(u',\tilde{v},w')/\left[\varphi(u',\tilde{v},w)\varphi(u,\tilde{v},w')\right] \equiv 1$, and moreover that φ and φ_{22} are bounded while φ is bounded away from zero (all bounds are uniform), it follows from (31) that for $y_1 \in A$,

$$\psi(T_2T_1) \ge 1 - c_5\sigma^2 \tag{32}$$

for some $c_5 \ge 0$ provided that σ is sufficiently small. Thus by (10) it holds that for $y_1 \in A$,

$$\tau(J_{0,2}) = \frac{1-\sqrt{\psi(T_2T_1)}}{1+\sqrt{\psi(T_2T_1)}} \le 1 - \psi(T_2T_1) \le c_5\sigma^2$$

and

$$\frac{\log\tau(J_{0,2})}{2\log\frac{1}{\sigma}} \le -1 + \frac{\log c_5}{2\log\frac{1}{\sigma}}. \tag{33}$$

By (26), (33) and Fatou's lemma, one gets

$$\begin{aligned}
&\limsup_{\sigma\to 0} \frac{E_s \log\tau(J_{0,2})}{2\log\frac{1}{\sigma}} \le \limsup_{\sigma\to 0} \frac{E_s[\log\tau(J_{0,2})|y_1 \in A]}{2\log\frac{1}{\sigma}}(1 - c(e^{-\epsilon^2/2\sigma^2} + \epsilon)) \\
&\le E_s\left[\limsup_{\sigma\to 0}(-1 + \frac{\log c_5}{2\log\frac{1}{\sigma}})(1 - c(e^{-\epsilon^2/2\sigma^2} + \epsilon))\middle|\, y_1 \in A\right] = -1 + c\epsilon,
\end{aligned}$$

and taking $\epsilon \to 0$ one obtains (23). The result now follows from Theorem 1. $\square$

Remark: One checks that the coefficient corresponding to σ^2 that results from expanding the cross ratio in (31) with respect to σ is *not* identically zero, and especially *does* assume negative values for some $u, u', w, w' \in S$ so the term σ^2 in (32) cannot be replaced by a higher power of σ, and therefore it seems that the above bound is the asymptotically best bound one can get by the technique used.

4 The case of a diffusion on a compact manifold

In this section we treat the filtering of a diffusion process x_t, on a compact Riemannian manifold M of dimension m. All assumptions and definitions of section 2 are kept with $T = \mathbb{R}_+$. Embedding the manifold in $\mathbb{R}^d$, some large enough d, one may write the Itô equation

$$dx_t = b(x_t)dt + \tilde{\sigma}(x_t)dW_t, \quad x_0 = x\,, \tag{34}$$

with a one dimensional observation process:

$$dy_t = \sigma^{-1}g(x_t)dt + d\nu_t, \quad y_0 = 0. \tag{35}$$

Here W and ν are independent standard Brownian motions on $\mathbb{R}^d$ and $\mathbb{R}$, respectively, $b = \{b^{(i)}\}_{1\le i\le d}$, $\tilde{\sigma} = \{\tilde{\sigma}^{(ij)}\}_{\substack{1\le i\le d\\ 1\le j\le d}}$ and g are Lipschitz-continuous on M. The function g is also twice continuously differentiable, and we denote by g_i and g_{ij} its partial derivatives of first and second order, respectively, in the directions of the coordinate vectors in $\mathbb{R}^d$. Furthermore, we assume that the coefficients b and $\tilde{\sigma}$ lead to a strictly elliptic generator on M, with heat kernel $G_t(x,y)$. By comparing to standard bounds on heat kernels of strictly elliptic operators on $\mathbb{R}^d$ (e.g., [6, chapter 3]), the ellipticity assumption implies that there exist positive constants c_0, c_1 and c_2, such that

$$c_0 e^{-c_1/t} \le G_t(x,y) \le c_2 t^{-m/2} \quad x, y \in M. \tag{36}$$

As in section 3, σ is used to parameterize the intensity of observation, and we use the notion $\gamma_\sigma(\cdot,\cdot)$ instead of $\gamma(\cdot,\cdot)$ to denote the dependence on σ.

Theorem 3 *Let the state and observation processes be as in (34) and (35), and assume that the ellipticity condition holds. Then there exist a non-random function $c(\sigma) > 0$ and constants $c' > 0, c'' > 0$ s.t. for any $q, q' \in \mathcal{P}$, the following hold P-a.s.*

(a) $\gamma_\sigma(q,q') \le -c(\sigma)\ , \quad \sigma > 0\ ,$

(b) $\limsup_{\sigma\to\infty} \gamma_\sigma(q,q') < -c'\ ,$

(c) $\limsup_{\sigma\to 0} \gamma_\sigma(q,q') e^{c''/\sigma} \le -1\ .$

Remarks: 1) (c) above provides an upper bound on γ_σ which converge to 0 as $\sigma \to 0$. This bound is certainly not believed to be tight (compare with Theorems 2 and 4 in the discrete time setting). The significance of this bound is in those cases for which γ_σ indeed seems to converges to 0 as $\sigma \to 0$ (see [7] for such an example).
2) Under additional assumptions, with $\mathcal{L}^*$ denoting the forward operator of x_t on M, the Zakai equation for the filtering of x_t given $\mathcal{F}_t$ is written

$$du_t = \mathcal{L}^* u_t dt + \sigma^{-1} g u_t dy_t, \tag{37}$$

where u_0 equals the density of q. Though below we do not make use of the representation (37), Theorem 3 yields exponential stability for its normalized solution.

Proof: Throughout the proof c denotes a deterministic constant whose value may change from line to line. The proof is based on (15). By ellipticity, x_t is ergodic, and in view of (12), $H(G_\delta) < \infty$. Next, we fix $\delta > 0$, and let λ be the Riemannian measure on M. Note that in view of (15) and part (c) of Theorem 1, to prove (a), it suffices to bound $H(I_\delta^\lambda(\cdot\,; y_0^\delta))$ by a finite, possibly random, bound. We do so by bounding $I_\delta(\alpha, \beta; y_0^\delta)$ defined in (7) by bounds that do not depend on α and β, and using the representation (12). To this end, note that (see (2), (7) and (35)),

$$I_\delta(\alpha,\beta;y_0^\delta) = E_0^{(\alpha)}\left[\exp(\int_0^\delta \frac{g(x_s)}{\sigma} dy_s - \frac{1}{2}\int_0^\delta \frac{g(x_s)^2}{\sigma^2} ds)\middle| x_\delta = \beta, \mathcal{F}_\delta\right] \quad \alpha,\beta \in M. \tag{38}$$

We begin with the upper bound. Integrating by parts in (38), we have that

$$I_\delta(\alpha,\beta;y_0^\delta) = E_0^{(\alpha)}\left[\exp(A_\delta + \int_0^\delta (M_s, dx_s) + \int_0^\delta N_s ds)\middle| x_\delta = \beta, \mathcal{F}_\delta\right],$$

where

$$A_\delta = \sigma^{-1} g(x_\delta) y_\delta,$$
$$M_s = \{M_s^{(i)}\}_{1\le i\le d}, \quad M_s^{(i)} = -\sigma^{-1} y_s g_i(x_s),$$
$$N_s = -\frac{1}{2}\left(\sigma^{-2} g^2(x_s) + \sigma^{-1} y_s \sum_{i,j,k} g_{ij}(x_s)\tilde\sigma^{(ik)}(x_s)\tilde\sigma^{(jk)}(x_s)\right).$$

Note that $b^{(i)}, \tilde\sigma^{(ij)}, g, g_i$ and g_{ij} are all bounded, hence

$$I_\delta(\alpha,\beta;y_0^\delta) \le \exp\left(|A_\delta| + \bar N_\delta\right) E_0^{(\alpha)}\left(\exp\int_0^\delta (M_s, dx_s)\middle| x_\delta = \beta, \mathcal{F}_\delta\right) \tag{39}$$

where $y_m = \sup_{0\le t\le\delta} |y_t|$, and

$$\bar N_\delta = c\delta(\sigma^{-2} + \sigma^{-1} y_m).$$

Denote

$$F_t = \exp\int_0^t (M_s, dx_s).$$

We bound the conditional expectation $E_0^{(\alpha)}(F_\delta|x_\delta=\beta,\mathcal{F}_\delta)$ in (39). First, by Fatou's lemma and continuity of F_t with respect to t,

$$E_0^{(\alpha)}(F_\delta|x_\delta=\beta,\mathcal{F}_\delta)\le\lim_{n\to\infty}E_0^{(\alpha)}(F_{\delta_n}|x_\delta=\beta,\mathcal{F}_\delta), \tag{40}$$

where δ_n is a sequence increasing to δ. By the Markov property and the boundedness from below of $G_\delta(\cdot,\cdot)$,

$$\begin{aligned} E_0^{(\alpha)}(F_{\delta_n}|x_\delta=\beta,\mathcal{F}_\delta) &= E_0^{(\alpha)}\left[F_{\delta_n}G_{\delta-\delta_n}(x_{\delta_n},\beta)|\mathcal{F}_\delta\right]/G_\delta(\alpha,\beta) \\ &\equiv b_n(\alpha,\beta,y_0^\delta)/G_\delta(\alpha,\beta) \end{aligned} \tag{41}$$

where we have defined b_n as the conditional expectation in (41). Note that since $F_{\delta_{n-1}}$ is measurable w.r.t. $\sigma\{x_0^{\delta_n},y_0^\delta\}$, it follows by the same argument that

$$b_{n-1}(\alpha,\beta,y_0^\delta)=E_0^{(\alpha)}\left[F_{\delta_{n-1}}G_{\delta-\delta_n}(x_{\delta_n},\beta)|\mathcal{F}_\delta\right].$$

In view of (39) and (40), in order to show the existence of a uniform upper bound on $J(\alpha,\beta)$, it suffices to bound b_n uniformly in α,β and n. Now,

$$\begin{aligned} &|b_n(\alpha,\beta,y_0^\delta)-b_{n-1}(\alpha,\beta,y_0^\delta)| \\ &=\left|E_0^{(\alpha)}\left[(F_{\delta_n}-F_{\delta_{n-1}})G_{\delta-\delta_n}(x_{\delta_n},\beta)|\mathcal{F}_\delta\right]\right|\le Q_1Q_2 \end{aligned} \tag{42}$$

where we have used Hölder's inequality and denoted

$$\begin{aligned} Q_1 &= E_0^{(\alpha)\,1/(m+1)}\left[|F_{\delta_n}-F_{\delta_{n-1}}|^{m+1}|\mathcal{F}_\delta\right], \\ Q_2 &= \left(\int G_{\delta-\delta_n}^{1+1/m}(\xi,\beta)G_{\delta_n}(\alpha,\xi)d\xi\right)^{m/(m+1)}. \end{aligned}$$

To bound Q_1 we first use the inequality $|e^x-1|\le|x|(e^x+e^{-x})$ and the Cauchy-Schwartz inequality. Denoting $m'=2m+1$ one has

$$\begin{aligned} Q_1 &= E_0^{(\alpha)\,1/(m+1)}\left[|(\exp\int_0^{\delta_{n-1}}(M_s,dx_s))(\exp\int_{\delta_{n-1}}^{\delta_n}(M_s,dx_s)-1)|^{m+1}\middle|\,\mathcal{F}_\delta\right] \\ &\le E_0^{(\alpha)\,1/(m+1)}\left[|(\int_{\delta_{n-1}}^{\delta_n}(M_s,dx_s))[\exp\int_0^{\delta_n}(M_s,dx_s)+\exp-\int_0^{\delta_n}(M_s,dx_s)]|^{m+1}\middle|\,\mathcal{F}_\delta\right] \\ &\le E_0^{(\alpha)\,1/m'}\left[|\int_{\delta_{n-1}}^{\delta_n}(M_s,dx_s)|^{m'}\middle|\,\mathcal{F}_\delta\right] \\ &\quad\cdot E_0^{(\alpha)\,1/m'}\left[2^{m'}[\exp m'\int_0^{\delta_n}(M_s,dx_s)+\exp -m'\int_0^{\delta_n}(M_s,dx_s)]\middle|\,\mathcal{F}_\delta\right] \end{aligned}$$

Next, using Minkowski's inequality, and the Burkholder-Davis-Gundy inequality [10, p. 166] for the martingales $\int^t M_s^{(i)}dW_s^{(j)}$, one obtains

$$E_0^{(\alpha)\,1/m'}\left[\left|\int_{\delta_{n-1}}^{\delta_n}(M_s,dx_s)\right|^{m'}\middle|\,\mathcal{F}_\delta\right]$$

$$
\begin{aligned}
&\le E_0^{(\alpha)\,1/m'}\left[\left|\int_{\delta_{n-1}}^{\delta_n}(M_s,b(x_s)ds)\right|^{m'}\middle|\mathcal{F}_\delta\right]+E_0^{(\alpha)\,1/m'}\left[\left|\int_{\delta_{n-1}}^{\delta_n}(M_s,\tilde\sigma(x_s)dW_s)\right|^{m'}\middle|\mathcal{F}_\delta\right]\\
&\le c\sigma^{-1}y_m(\delta_n-\delta_{n-1})^{1/2}.
\end{aligned}
$$

While

$$
\begin{aligned}
E_0^{(\alpha)\,1/m'}\left[e^{m'\int_0^{\delta_n}(M_s,dx_s)}\middle|\mathcal{F}_\delta\right] &= E_0^{(\alpha)\,1/m'}\left[e^{m'\int_0^{\delta_n}(M_s,b(x_s)ds)+m'\int_0^{\delta_n}(M_s,\tilde\sigma(x_s)dW_s)}\middle|\mathcal{F}_\delta\right]\\
&\le e^{c\delta_n\sigma^{-1}y_m}\prod_{j=1}^{d}E_0^{(\alpha)\,c}\left[e^{cA_j}\middle|\mathcal{F}_\delta\right]
\end{aligned}
$$

where

$$
A_j=\int_0^{\delta_n}\sum_i\sigma^{-1}y_sg_i(x_s)\tilde\sigma^{(ij)}(x_s)dW_s^{(j)}.
$$

By time change for the stochastic integral, one has that, conditioned on $\mathcal{F}_\delta$, $A_j\overset{d}{=}B_{\eta_j}$, where B_t, conditioned on $\mathcal{F}_\delta$, is a standard BM and

$$
\eta_j=\int_0^{\delta_n}\left(\sum_i\sigma^{-1}y_sg_i(x_s)\tilde\sigma^{(ij)}(x_s)\right)^2ds\le c\delta\sigma^{-2}y_m^2\equiv\eta^*
$$

and hence

$$
E_0^{(\alpha)}\left[e^{cA_j}\middle|\mathcal{F}_\delta\right]\le E\left[e^{c\sup_{0\le t\le\eta^*}B_t}\middle|\mathcal{F}_\delta\right]\le 2e^{c\eta^*}.
$$

One therefore concludes that

$$
Q_1\le c\sigma^{-1}y_m(\delta_n-\delta_{n-1})^{1/2}e^{c\sigma^{-1}y_m\delta+c\sigma^{-2}y_m^2\delta}. \tag{43}
$$

On the other hand, using (36) one obtains

$$
Q_2\le c(\delta-\delta_n)^{-m/2(m+1)}\delta^{-m^2/2(m+1)}. \tag{44}
$$

Now, if we chose $\delta_n=\delta(1-2^{-n})$ and combine (42), (43) and (44), it follows that

$$
|b_n(\alpha,\beta,y_0^\delta)-b_{n-1}(\alpha,\beta,y_0^\delta)|\le c\sigma^{-1}y_m\delta^{(1-m)/2}2^{-n/2(m+1)}e^{c\sigma^{-1}y_m\delta+c\sigma^{-2}y_m^2\delta}.
$$

Since $b_0=G_\delta(\alpha,\beta)$ it follows that

$$
b_n\le c\delta^{-m/2}+c\sigma^{-1}y_m\delta^{(1-m)/2}e^{c\sigma^{-1}y_m\delta+c\sigma^{-2}y_m^2\delta},
$$

which is bounded uniformly in n,α,β. Thus from (39) and (41) and using (36), we obtain the upper bound

$$
I_\delta(\alpha,\beta;y_0^\delta)\le ce^{c/\delta}e^{c\sigma^{-1}y_m+c\delta\sigma^{-2}}\left(\delta^{-m/2}+\sigma^{-1}y_m\delta^{(1-m)/2}e^{c\sigma^{-1}y_m\delta+c\sigma^{-2}y_m^2\delta}\right). \tag{45}
$$

The lower bound follows easily from the upper bound, since by Cauchy-Schwartz $Ee^X\ge 1/Ee^{-X}$. Hence

$$
I_\delta(\alpha,\beta;y_0^\delta)\ge ce^{-c/\delta}e^{-c\sigma^{-1}y_m-c\delta\sigma^{-2}}\left(\delta^{-m/2}+\sigma^{-1}y_m\delta^{(1-m)/2}e^{c\sigma^{-1}y_m\delta+c\sigma^{-2}y_m^2\delta}\right)^{-1}. \tag{46}
$$

Actually, the above suffices for parts (a) and (b) in view of (12) and (15), since the bounds (45) and (46) are both positive and finite for fixed σ as well as in the limit $\sigma \to \infty$.

As for part (c), we shall need the concrete dependence on σ and δ of the bound on γ_σ derived from (15) using (45) and (46). Since $y_m \le c\delta\sigma^{-1} + \max_{s\le\delta} \nu_s$ for c large enough, it follows that $P^{(\alpha)}(y_m \le c\delta\sigma^{-1} + 4\sqrt{\delta}) \ge 1/2$. Denote $a_\sigma^\delta = c\delta\sigma^{-1} + 4\sqrt{\delta}$. Thus, combining (15), (45), (46) and the inequality $\tanh(\frac{1}{2}\log z) \le 1 - z^{-1}$, $z > 0$ we get

$$\gamma_\sigma(q, q') \le$$
$$\frac{1}{2\delta}\log\left[1 - c\delta^{m/2}e^{-c/\delta}e^{-c\sigma^{-1}a_\sigma^\delta - c\delta\sigma^{-2}}\left(\delta^{-m/2} + \sigma^{-1}a_\sigma^\delta e^{c\sigma^{-1}a_\sigma^\delta\delta^{(1-m)/2}\delta + c\sigma^{-2}(a_\sigma^\delta)^2\delta}\right)^{-2}\right].$$

Substituting $\delta = \sigma$, we conclude that there exist positive constants c and σ_0 s.t. for every $\sigma \le \sigma_0$,

$$\gamma_\sigma(q, q') \le -e^{-c/\sigma}.$$

□

5 The countable state space case

In [1] bounds are derived for γ in the finite state space case both in discrete and in continuous time. The technique there is based on the connection between γ and the gap between the two top Lyapunov exponents of the Zakai equation, and Oseledec's multiplicative ergodic theorem is intensively used. Theorem 1 above may be directly applied to provide an extension of Theorem 2 in [1] to the case of countable state space discrete time. However, as we show below, in the low observation noise limit, a sharper result is possible by a direct computation that relies neither on the estimates of sections 2 and 3, nor on Oseledec's theorem (and provides an extension of Theorem 3 of [1]).

We let $S = \mathbb{Z}$ and $T = \mathbb{Z}^+$, and denote the transition matrix by $G = (G(i,j))$. Information on its state is given through the observation process

$$y_n = \sigma^{-1}\sum_1^n g(x_i) + \nu_i, \quad n \ge 1. \tag{47}$$

The unnormalized conditional law is then given (as in (21), (22)) by the vectors

$$\rho_n^q = D_n G^* \rho_{n-1}^q, \quad \rho_0^q = q \tag{48}$$

where G^* is the transpose of G and D_n is a diagonal matrix with $\Delta_n(i) \equiv D_n(i,i) = \exp(\sigma^{-1}g(i)\Delta y_n - \sigma^{-2}g^2(i)/2)$. The following result shows that in the limit $\sigma \to 0$, the behaviour of γ_σ is at least in $\Omega(\sigma^{-2})$.

Theorem 4 *Assume that the state process $\{x_n\}_{n\in\mathbb{Z}_+}$ satisfies the following conditions. (a) $\{x_n\}_{n\in\mathbb{Z}_+}$ is ergodic. (b) $Eg^2(x_n) < \infty$. (c) $E_s \log G(x_0, x_1) > -\infty$. (d) $\|G\|_1 = \sup_j \sum_i G(i,j) < \infty$. Let q and q' be any two probability distributions on S. Then P-a.s.,*

$$\limsup_{\sigma\to 0} \sigma^2\gamma_\sigma(q, q') \le -\frac{1}{2}\, E_s \inf_{j\ne x_0} (g(x_0) - g(j))^2 \tag{49}$$

In particular, if there exists a state i_0 such that $P_s(x_0 = i_0) > 0$ and $\inf_{j \neq i_0} |g(i_0) - g(j)| \neq 0$ then the limsup in (49) is a.s. negative.

In order to prove the theorem, we introduce some notations. For $a, b \in \ell_1$, let the exterior product be defined by $a \wedge b = ab^* - ba^*$ (compare with the usual definition [5, p. 61]). Note that $W = a \wedge b$ is an antisymmetric matrix. We define for it the pth norm by

$$\|\!|W|\!\|_p = \left(\sum_{i,j} |W(i,j)|^p \right)^{1/p}, \quad p \in [1, \infty), \qquad \|\!|W|\!\|_\infty = \sup_{i,j} |W(i,j)|. \tag{50}$$

Let $\wedge^2 \ell_p$ denote the space $\{a \wedge b | a, b \in \ell_p\}$ and note that $\|\!|a \wedge b|\!\|_p < \infty$ if $a \wedge b \in \wedge^2 \ell_p$. For $p \in [1, \infty]$ and a linear operator A on $\wedge^2 \ell_p$ we denote by $\|\!|A|\!\|_p$ its operator norm w.r.t. the pth norm defined by (50). We do not distinguish between the measure $q \in \mathcal{P}$ and the vector $q \in \ell_1$ for which $\|q\|_1 = 1$. Moreover, we define the set of positive probability measures on $(S, \mathcal{S})$ by $\mathcal{P}^o = \{q \in \mathcal{P} | q(i) > 0, i \in \mathbb{Z}\}$.

We first prove the following

Lemma 2 *Let $\rho, \eta \in \ell_1$ be non-negative and non-zero, and denote $p = \rho / \|\rho\|_1, q = \eta / \|\eta\|_1$. Then*

$$\frac{\|\!|\rho \wedge \eta|\!\|_\infty}{\|\rho\|_1 \|\eta\|_1} \leq \|p - q\|_1 \leq \frac{\|\!|\rho \wedge \eta|\!\|_1}{\|\rho\|_1 \|\eta\|_1}. \tag{51}$$

Proof: On one hand, it is easy to see that for any $\alpha \geq 1$ and $a, b \in \ell_2$, if $\|a\|_2 = \|b\|_2 = 1$ then $\|a - b\|_2 \leq \|a - \alpha b\|_2$. Thus, assuming w.l.o.g. that $\|p\|_2 \leq \|q\|_2$, and denoting the angle between p and q by $\widehat{p\,q}$ we get

$$\begin{aligned}
\|p - q\|_1 \quad &\geq \quad \|p - q\|_2 \; = \; \|p\|_2 \left\| \frac{p}{\|p\|_2} - \frac{\|q\|_2}{\|p\|_2} \frac{q}{\|q\|_2} \right\|_2 \\
&\geq \quad \|p\|_2 \left\| \frac{p}{\|p\|_2} - \frac{q}{\|q\|_2} \right\|_2 \; = \; \|p\|_2 \; 2 |\sin(\frac{1}{2} \widehat{p\,q})| \\
&\geq \quad \|p\|_2 \; |\sin \widehat{p\,q}| \; = \; \|p\|_2 \; \frac{\|\!|\rho \wedge \eta|\!\|_2}{\|\rho\|_2 \|\eta\|_2} \\
&\geq \quad \frac{\|\!|\rho \wedge \eta|\!\|_\infty}{\|\rho\|_1 \|\eta\|_1}.
\end{aligned}$$

The last equality is well known (see e.g., [5, p. 61]). On the other hand,

$$\begin{aligned}
\|p - q\|_1 \quad &= \quad \frac{\sum_i \left| \sum_j \rho(i)\eta(j) - \rho(j)\eta(i) \right|}{\|\rho\|_1 \|\eta\|_1} \\
&\leq \quad \frac{\sum_i \sum_j |\rho(i)\eta(j) - \rho(j)\eta(i)|}{\|\rho\|_1 \|\eta\|_1} \\
&= \quad \frac{\|\!|\rho \wedge \eta|\!\|_1}{\|\rho\|_1 \|\eta\|_1}.
\end{aligned}$$

$\square$

Proof of Theorem 4 : Given q and q' we denote $\rho_n = \rho_n^q$ and $\eta_n = \rho_n^{q'}$. In order to use the lemma, we first provide lower bounds on $\|\rho_n\|_1$ and $\|\eta_n\|_1$, and an upper bound on $\|\rho_n \wedge \eta_n\|_1$. First,

$$\rho_n(i) = \sum_{i_0,\dots,i_{n-1}} \rho_0(i_0)G(i_0,i_1)\cdots G(i_{n-1},i)\Delta_1(x_1)\cdots\Delta_n(x_n),$$

$$\|\rho_n\|_1 \geq \rho_n(x_n) \geq \rho_0(x_0)G(x_0,x_1)\cdots G(x_{n-1},x_n)\Delta_1(x_1)\cdots\Delta_n(x_n).$$

Using the ergodic theorem, we get that P-a.s.

$$\liminf_{n\to\infty} \frac{1}{n}\log\|\rho_n\|_1 \geq E_s \log G(x_0,x_1) + E_s g^2(x_0)/2\sigma^2. \tag{52}$$

Secondly, using (48), we obtain the following recursion

$$\rho_n \wedge \eta_n = \rho_n\eta_n^* - \eta_n\rho_n^* = D_n G^*(\rho_{n-1}\wedge\eta_{n-1})GD_n.$$

Write the above elementwise as

$$(\rho_n \wedge \eta_n)_{ij} = \Delta_n(i)\left(\sum_{k,l} G(k,i)(\rho_{n-1}\wedge\eta_{n-1})_{k,l}G(l,j)\right)\Delta_n(j), \tag{53}$$

and get the relation

$$\|\rho_n \wedge \eta_n\|_1 \leq \sup_{i\neq j}\Delta_n(i)\Delta_n(j)\ \|\tilde{G}\|_1\|\rho_{n-1}\wedge\eta_{n-1}\|_1$$

where $\tilde{G}$ is the operator on $\wedge^2\ell_1$ defined by $u\wedge v \mapsto G^*(u\wedge v)G$. By (53),

$$\|\tilde{G}\|_1 \leq \sup_{i,j}\sum_{k,l} G(k,i)G(l,j) = \|G\|_1^2.$$

We thus obtain, using again the ergodic theorem, that P-a.s.

$$\limsup_{n\to\infty}\frac{1}{n}\log\|\rho_n\wedge\eta_n\|_1 \leq E_s\log\sup_{i\neq j}\Delta_1(i)\Delta_1(j) + 2\log\|G\|_1. \tag{54}$$

Next, since $E_s(\Delta y_1)^2 = (1+E_s g^2(x_0)/\sigma^2)/2$,

$$\begin{aligned}
&E_s\log\sup_{i\neq j}\Delta_1(i)\Delta_1(j)\\
&= -\frac{1}{2}\,E_s\inf_{i\neq j}\Big[(\Delta y_1 - g(i)/\sigma)^2 + (\Delta y_1 - g(j)/\sigma)^2\Big] + (1+E_s g^2(x_0)/\sigma^2)\\
&\leq -\frac{1}{2}\,E_s\left\{\inf_{i\neq j}\Big[(\Delta y_1 - g(i)/\sigma)^2 + (\Delta y_1 - g(j)/\sigma)^2\Big]\,\Big|\,|\sigma\nu_1|<\epsilon\right\}P(|\sigma\nu_1|<\epsilon) + (1+E_s g^2(x_0)/\sigma^2)\\
&\leq -\frac{1}{2\sigma^2}E_s\left\{\inf_{j\neq x_1}(g(x_1)+\sigma\nu_1 - g(j))^2\,\Big|\,|\sigma\nu_1|<\epsilon\right\}P(|\sigma\nu_1|<\epsilon) + (1+E_s g^2(x_0)/\sigma^2)\\
&\leq -\frac{1}{2\sigma^2}E_s\left([\inf_{j\neq x_1}|g(x_1)-g(j)|-\epsilon]\vee 0\right)^2(1-e^{-\epsilon^2/2\sigma^2}) + (1+E_s g^2(x_0)/\sigma^2),
\end{aligned} \tag{55}$$

for any constant $\epsilon > 0$ and sufficiently small σ. Now the right hand side of (51) together with (52), (54) and the above imply the theorem. $\square$

We next show that in a specific class of problems, (49) is also a lower bound in a sense to be explained below. We concentrate on the class $\mathcal{U}$ of processes for which the transition matrix G has all its 2×2 minors non-negative. An important subclass of $\mathcal{U}$ is that of the nearest neighbour processes i.e., those that possess a transition matrix $G = \exp(sA)$ where $s > 0$ and A is an intensity matrix for which $|i-j| > 1$ implies $A(i,j) = 0$. Indeed, the latter condition on G implies that it is totally positive as an operator on ℓ_1, which by definition implies that the process it generates belongs to $\mathcal{U}$. For the above results and further discussion see [11, pp. 38–45]. Let us denote for $i, j, k, l \in \mathbb{Z}$

$$G_{i,j}^{k,l} = G(k,i)G(l,j) - G(l,i)G(k,j).$$

For $i < j, k < l$ the latter are the 2×2 minors of G. We further assume that the infimum in (49) is always achieved namely, there exists a function $\pi : \mathbb{Z} \to \mathbb{Z}$ such that

$$\inf_{j \neq i} |g(i) - g(j)| = |g(i) - g(\pi(i))| \text{ and } \pi(i) \neq i, \quad i \in \mathbb{Z}. \tag{56}$$

For a probability measure m on $\mathcal{P}$, we shall require the following.

Assumption A3: For every proper linear subspace Q of ℓ_1, $m(Q \cap \mathcal{P}) = 0$.

Example 1 *Let $a_i, i \in \mathbb{Z}$ be i.i.d. real-valued non-negative bounded random variables, with a common law that is absolutely continuous w.r.t. the Lebesgue measure, and let $c_i, i \in \mathbb{Z}$ be positive constants s.t. $\sum_i c_i < \infty$. Let $z_i = c_i a_i / \sum_j c_j a_j, i \in \mathbb{Z}$ and define m to be the measure on $\mathcal{P}$ induced by the random vector $\{z_i\}_{i \in \mathbb{Z}}$. Then m satisfies Assumption A3.*

Proof: Define μ to be the measure on ℓ_∞ induced by $\{a_i\}_{i \in \mathbb{Z}}$. Let Q be a proper subspace of ℓ_1. Then there exists a non-zero $w \in \ell_\infty$ s.t. if $q \in Q$ then $\sum w_i q_i = 0$. Let k be such that $w_k \neq 0$. Then $m(\{z \in \mathcal{P} | \sum_i w_i z_i = 0\}) = \mu(\{a \in \ell_\infty | w_k a_k = -\sum_{i \neq k} w_i c_i a_i\})$ which vanishes, since a_k is μ-independent of $a_i, i \neq k$, and its law is absolutely continuous w.r.t. the Lebesgue measure. $\square$

We have the following result.

Theorem 5 *Let $\{x_n\}_{n \in \mathbb{Z}_+}$ belong to $\mathcal{U}$ and let it satisfy (a)–(d) of theorem 4. Assume moreover that there exists a $\pi : \mathbb{Z} \to \mathbb{Z}$ satisfying (56), such that $E_s \log |G_{x_1, \pi(x_1)}^{x_0, \pi(x_0)}| > -\infty$. Let q and q' be random, distributed according to two given probability measures m and m', respectively, on $\mathcal{P}$ satisfying Assumption A3. Then $m \times m' \times P$-a.s.,*

$$\liminf_{\sigma \to 0} \sigma^2 \liminf_{n \to \infty} \frac{1}{n} \log d_{TV}(p_n^q, p_n^{q'}) \geq -\frac{1}{2} E_s (g(x_0) - g(\pi(x_0)))^2. \tag{57}$$

In particular, $m \times m' \times P$-a.s., (49) holds with equality.

Proof: We shall use the left hand side of (51) and develop bounds analogous to (52) and (54) but in opposite direction. The first bound follows easily from (48) namely,

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log \|\rho_n\|_1 &\le \log \|G\|_1 + E_s \log \sup_i \Delta_1(i) \\ &\le \log \|G\|_1 + \frac{1}{2}\,(1 + E_s g^2(x_0)/\sigma^2). \end{aligned} \tag{58}$$

The lower bound on $\|\rho_n \wedge \eta_n\|_\infty$ is more complicated. Assume first that q and q' are such that all the entries above the diagonal of the matrix $q \wedge q'$ are positive. By (53), positivity of the 2×2 minors and the positivity assumption on the initial condition, it follows that for $i < j$ and $k < l$

$$(\rho_n \wedge \eta_n)_{i,j} \ge \Delta_n(i)\Delta_n(j) G_{i,j}^{k,l} (\rho_{n-1} \wedge \eta_{n-1})_{k,l}. \tag{59}$$

Note however, that for $i > j$ we have $(\rho_n \wedge \eta_n)_{i,j} = -(\rho_n \wedge \eta_n)_{j,i}$, and moreover, $G_{i,j}^{k,l} = -G_{i,j}^{l,k} = -G_{j,i}^{k,l}$. It thus follows from (59) that for any i, j, k, l we have

$$|(\rho_n \wedge \eta_n)_{i,j}| \ge \Delta_n(i)\Delta_n(j)|G_{i,j}^{k,l}|\,|(\rho_{n-1} \wedge \eta_{n-1})_{k,l}|. \tag{60}$$

Substituting $i = x_n, j = \pi(x_n), k = x_{n-1}$ and $l = \pi(x_{n-1})$, passing to the limit, we get that P-a.s.,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \|\rho_n \wedge \eta_n\|_\infty &\ge E_s \log \Delta_1(x_1)\Delta_1(\pi(x_1)) + E_s \log |G_{x_1,\pi(x_1)}^{x_0,\pi(x_0)}| \qquad (61) \\ &= \frac{1}{\sigma^2} E_s[\frac{1}{2}\, g^2(x_0) + g(x_0) g(\pi(x_0)) - \frac{1}{2}\, g^2(\pi(x_0))] + E_s \log |G_{x_1,\pi(x_1)}^{x_0,\pi(x_0)}|. \end{aligned}$$

Observe that in the derivation of the bounds (54), (55) and (61), only bounds on the operators D_n and G are involved, not on q or q'. Hence the null sets of P on which these bounds fail to hold are independent of q and q'. We thus summarize (54), (55) and (61) as follows. There exists a set $\Omega_0 \in \mathcal{B}$ of full P-measure, such that for all q and q' with $q \wedge q'$ positive above the diagonal, and $\omega \in \Omega_0$,

$$\begin{aligned} \liminf_{\sigma\to 0} \sigma^2 \liminf_{n\to\infty} \frac{1}{n} \log \|\rho_n \wedge \eta_n\|_\infty &= \limsup_{\sigma\to 0} \sigma^2 \limsup_{n\to\infty} \frac{1}{n} \log \|\rho_n \wedge \eta_n\|_\infty \\ &= E_s[g^2(x_0) + 2g(x_0)g(\pi(x_0)) - g^2(\pi(x_0))]/2 \\ &= C. \end{aligned} \tag{62}$$

Hence, for q, q' as assumed, (57) holds by the l.h.s. of (51), (58) and (62).

We extend the above result to general q and q' in a few steps.

Lemma 3 *For any $q \in \mathcal{P}^o$ there exists a $q_0 \in \mathcal{P}^o$ s.t. all entries above the diagonal of $q \wedge q_0$ are positive.*

Lemma 4 *There exists a set $\Omega_1 \in \mathcal{B}$ of full P-measure s.t. the following holds. If $q \in \mathcal{P}^o$ and $Q'(\omega, q)$ is the subset of $\mathcal{P}$ of all q' for which (62) does not hold, then for every $\omega \in \Omega_1$, $Q'(\omega, q)$ is the intersection of a proper linear subspace of ℓ_1 and $\mathcal{P}$.*

In view of Lemma 4 and Assumption A3, for every $\omega \in \Omega_1$ and $q \in \mathcal{P}^o$, (57) fails to hold only when q' belongs to a null set of m' (possibly depending on q and ω). Since from Assumption A3, $m(\mathcal{P} \setminus \mathcal{P}^o) = m(\bigcup_n \{q \in \mathcal{P} : q(n) = 0\}) \leq \sum_n m(\{q \in \mathcal{P} : q(n) = 0\}) = 0$, the proof is complete. □

Proof of Lemma 3: Define the sequence $a(n), n \in \mathbb{Z}$ as follows. Let $a(0) = 1$; for $n \geq 1$ and $k \leq -1$ let

$$a(n) = e^{1/n^2} a(n-1) q(n)/q(n-1), \quad a(k) = e^{-1/k^2} a(k+1) q(k)/q(k+1). \tag{63}$$

Then for any $n \in \mathbb{Z}$, $a(n) \leq c q(n)/q(0)$ so that $a(n)$ is summable. Let now $q_0(n) = a(n)/\sum_i a(i)$. Then for $j = i+1$ we have from (63) that $q(i) q_0(j) > q(j) q_0(i)$. By induction, the latter holds for any $i < j$, and the Lemma follows. □

Proof of Lemma 4: Denote the matrix $\rho_n \wedge \eta_n$ for which $\rho_0 = q$ and $\eta_0 = q'$ by $W_n(q, q')$. Since (54) and (55) hold, there exists a full P-measure set $\Omega'_0 \in \mathcal{B}$, such that for all $\omega \in \Omega'_0$ and $q, q' \in \mathcal{P}$,

$$\limsup_{\sigma \to 0} \sigma^2 \limsup_{n \to \infty} \frac{1}{n} \log \|W_n(q, q')\|_\infty \leq C,$$

where C is as in (62). Fix $q \in \mathcal{P}^o$. If $Q'(\omega, q)$ is empty, the lemma holds trivially. Otherwise, let $p_1, p_2 \in Q'(\omega, q)$. Then (62) does not hold for $W_n(q, p_i)$, $i = 1, 2$ (i.e., when $\rho_n = \rho_n^q$ and $\eta_n = \rho_n^{p_i}$). Hence, for all $\omega \in \Omega'_0$,

$$\limsup_{\sigma \to 0} \sigma^2 \limsup_{n \to \infty} \frac{1}{n} \log \|W_n(q, p_i)\|_\infty < C, \quad i = 1, 2.$$

Let $\lambda \in \mathbb{R}$ be such that $q' = q'(\lambda) = \lambda p_1 + (1-\lambda) p_2 \in \mathcal{P}$. Then $W_n(q, q') = \lambda W_n(q, p_1) + (1 - \lambda) W_n(q, p_2)$, and it follows that for $\omega \in \Omega'_0$,

$$\limsup_{\sigma \to 0} \sigma^2 \limsup_{n \to \infty} \frac{1}{n} \log \|W_n(q, q')\|_\infty < C.$$

Thus for every $q \in \mathcal{P}^o$ and $\omega \in \Omega'_0$, $Q'(\omega, q)$ is the intersection of a linear subspace of ℓ_1 and $\mathcal{P}$. Let now q_0 be as in Lemma 3. Thus by (62), whenever $\omega \in \Omega_0$, one has that $q_0 \notin Q'(\omega, q)$. This implies that whenever $\omega \in \Omega_1 = \Omega_0 \cap \Omega'_0$, the subspace must be proper. □

Example 2 *If G is the transition matrix of a finite state Markov chain such that all its entries and 2×2 minors are positive, then all assumptions of Theorems 4, 5 hold.*

Acknowledgments We thank P. Bougerol and F. Ledrappier for referring us to [9]. We are also grateful to the anonymous referee for several important comments.

References

[1] Atar, R., Zeitouni, O. *Lyapunov Exponents for Finite State Nonlinear Filtering,* to appear, Siam J. Contr. Optim., 1996.

[2] Birkhoff, G. *Extensions of Jentzsch's Theorem,* Trans. Am. Math. Soc., 1957, Vol 85, pp. 219–227.

[3] ——, *Lattice Theory,* Am. Math. Soc. Publ. 25, 3rd ed., 1967.

[4] Bougerol, P. *Theorems Limite pour les Systemes Lineaire a Coefficients Markoviens,* Prob. Theory Rel. Fields, 1988, Vol 78, pp. 192–221.

[5] Bougerol, P., Lacroix, J. *Products of Random Matrices with Applications to Schrödinger Operators,* Birkhäuser, 1985.

[6] Davies, E. B. *Heat Kernels and Spectral Theory,* Cambridge University Press, 1989.

[7] Delyon, B., Zeitouni, O. *Lyapunov exponents for filtering problems,* Applied Stochastic Analysis, edited by Davis, M. H. A. and Elliot, R. J., Gordon and Breach Science Publishers, London, pp. 511–521, 1991.

[8] Flandoli, F., Schaumlöffel, K. *Stochastic Parabolic Equations in Bounded Domains: Random Evolution Operator and Lyapunov Exponents,* Stochastics and Stochastic Reports, 1990, Vol. 29, pp. 461–485.

[9] Hopf, E. *An Inequality for Positive Linear Integral Operators,* Journal of Mathematics and Mechanics, 1963, Vol 12, No. 5, pp. 683–692.

[10] Karatzas, I., Shreve, S. E. *Brownian Motion and Stochastic Calculus,* Springer-Verlag, 1991.

[11] Karlin, S. *Total Positivity,* Stanford University Press, Stanford, California, 1968.

[12] Kunita, H. *Asymptotic Behavior of the Nonlinear Filtering Errors of Markov Processes,* J. Multivariate Anal., 1971, Vol. 1, pp. 365–393.

[13] Liptser, R. S., Shiryayev, A. N. *Statistics of Random Processes,* Nauka, Moscow, 1974, English ed., Springer-Verlag, New-york, 1977.

[14] Ocone, D., Pardoux, E. *Asymptotic Stability of the Optimal Filter with respect to its Initial Condition*, Siam J. Contr. Optim., 1996, Vol. 34, No. 1, pp. 226–243.

[15] Parthasarathy, K. R. *Probability Measures on Metric Spaces,* Academic Press, New York and London, 1967.

[16] Peres, Y. *Domains of Analytic Continuation for the Top Lyapunov Exponent,* Ann. Inst. Henri Poincaré, 1992, Vol 28, no. 1, pp. 131–148.

[17] Seneta, E. *Non-negative Matrices and Markov Chains,* Springer-Verlag, 1981.

[18] Stettner, L. *On Invariant Measures of Filtering Processes,* Stochastic Differential Systems, Proc. 4th Bad Honnef Conf., 1988, Leture Notes in Control and Inform. Sci. 126, edited by Christopeit, N., Helmes, K. and Kohlmann, M., Springer, pp. 279–292, 1989.

[19] —— *Invariant Measures of the Pair: State, Approximate Filtering Process,* Colloq. Math., 1991, LXII, pp. 347–351.

[20] Zakai, M. *On the Optimal Filtering of Diffusion Processes,* Z. Wahr. Verw. Geb., 1969, Vol 11, pp. 230–243.