Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Stability of the nonlinear filter for slowly switching HMM

Pavel Chigansky

***Abstract*— Stability of the nonlinear filtering equation is revisited in the case, when the signal is a finite state space Markov chain with slow transitions rate. The derived formulae reveal surprising properties of the stability exponent.**

I. INTRODUCTION

This paper deals with stability of the nonlinear filtering equation with respect to initial condition. Suppose that a discrete time Markov chain $X = (X_n)_{n\geq 0}$ with values in a finite real alphabet $\mathbb{S} = \{a_1, ..., a_d\}$, transition probabilities $\lambda_{ij} = \mathrm{P}(X_n = a_j | X_{n-1} = a_i)$ and initial distribution $\nu_i = \mathrm{P}(X_0 = a_i)$ is observed via white noise observations, generated by

$$Y_n = \sum_{i=1}^{d} \xi_n(i) \mathbf{1}_{\{X_n = a_i\}}, \quad n \geq 1, \tag{1}$$

where $\xi = (\xi_n)_{n\geq 1}$ is a sequence of i.i.d. random vectors in $\mathbb{R}^d$ with independent entries, sampled according to the probability laws

$$\mathrm{P}\big(\xi_n(i) \leq x\big) = \int_{-\infty}^{x} g_i(u)\varphi(du), \quad n \geq 1, \quad i = 1, ..., d,$$

where φ is a σ-finite measure on $\mathbb{R}$ and $g_i(u)$ are probability densities with respect to φ. The sequences X and ξ are defined on a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and are assumed to be independent. This statistical setup is usually referred as Hidden Markov Model (HMM) and is used in a variety of applications (see the recent survey [10]).

One of the basic problems related to HMM is filtering, i.e. estimation of current state of the chain on the basis of past observations. For a fixed function $f : \mathbb{S} \mapsto \mathbb{R}$, the optimal in the mean square sense filtering estimate of $f(X_n)$ given $\mathcal{F}_n^Y = \sigma\{Y_1, ..., Y_n\}$ is the conditional expectation

$$\mathrm{E}\big(f(X_n)|\mathcal{F}_n^Y\big) = \sum_{i=1}^{d} f(a_i)\mathrm{P}(X_n = a_i|\mathcal{F}_n^Y).$$

The (column) vector π_n of conditional probabilities $\pi_n(i) = \mathrm{P}(X_n = a_i|\mathcal{F}_n^Y)$ satisfies the *filtering* recursion ($n \geq 1$)

$$\pi_n = \frac{G(Y_n)\Lambda^*\pi_{n-1}}{\big|G(Y_n)\Lambda^*\pi_{n-1}\big|}, \quad \pi_0 = \nu, \tag{2}$$

where Λ^* is the transposed matrix of transition probabilities, $G(y)$, $y \in \mathbb{R}$ is a scalar matrix with the entries $g_i(u)$, $i = 1, ..., d$ on the diagonal and $|x|$ denotes the ℓ_1-norm, i.e. for $x \in \mathbb{R}^d$, $|x| = \sum_{i=1}^d |x(i)|$. As usual the probability distributions on $\mathbb{S}$ are identified with the column vectors in the simplex $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : x_i \geq 0, \sum_{i=1}^d x_i = 1\}$.

Suppose that the recursion (2) is started from a probability distribution $\bar{\nu}$ different from ν. This of course makes sense only if the filtering equation remains well defined P-a.s., in which case the pair of distributions $(\nu, \bar{\nu})$ is said to be *admissible*. A simple admissibility condition is $\nu \ll \bar{\nu}$ (holds e.g. if all the entries of $\bar{\nu}$ are positive), which is assumed hereafter. Denote by $\bar{\pi}_n$ the solution of (2) subject to $\bar{\pi}_0 = \bar{\nu}$. The filter is *asymptotically stable* with respect to an admissible pair $(\nu, \bar{\nu})$ if

$$\big|\pi_n - \bar{\pi}_n\big| \xrightarrow{n\to\infty} 0, \quad \mathrm{P} - a.s. \tag{3}$$

and is exponentially stable if the stronger convergence holds

$$\gamma := \lim_{n\to\infty} \frac{1}{n} \log \big|\pi_n - \bar{\pi}_n\big| < 0, \quad \mathrm{P} - a.s. \tag{4}$$

Existence of the limit in (4) follows from the Oseledec multiplicative ergodic theorem (MET), if the noise densities are sufficiently regular (see Theorem 2.1 below). In fact the *stability exponent* γ can only take a finite number of values, depending on $(\nu, \bar{\nu})$ (see Section IV). It is easy to see that the convergence in (3) can be superexponential in general, though no evidence is available at this point for a subexponential convergence.

The problem of stability is to specify the conditions in terms of Λ, g_i's and $(\nu, \bar{\nu})$ so that (3) or (4) holds. Having inspired much research in the last decade (see [3] for an up to date reference list), this problem still remains unresolved in many aspects. Being the simplest genuine nonlinear filter, the equation (2) yet exhibits much of the problem complexity.

Probably the most counterintuitive issue is that (3) is not implied in general by ergodicity of X. Recall that X is ergodic if the limits $\mu_i := \lim_{n\to\infty} \mathrm{P}(X_n = a_i)$ exist, are positive and do not depend on ν. The latter holds if and only if Λ is primitive, i.e. all the entries of Λ^q are positive for some integer $q \geq 1$. If ergodicity is viewed as a form of signal "stability", it is tempting to think that the filter cannot be "less stable" than the signal itself.

Another closely related question is whether the filtering π_n, considered as a measure valued Markov process, has a unique invariant measure. This was conjectured to hold by D.Blackwell in [4], provided X is ergodic and $Y_n = h(X_n)$ with $h : \mathbb{S} \mapsto \mathbb{R}$ and was found false by T.Kaijser in [11]. A simple counterexample in fact appeared already in [4], but was used originally to demonstrate other intriguing features of filtered Markov chains. Later it was revisited/reinvented in [11], [9], [3] and [7] to expose different instability properties of the filter (2).

Research supported by a grant of the Israeli Science Foundation
P. Chigansky is with Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel, pavel.chigansky@weizmann.ac.il

Independently H.Kunita [12] considered a more general filtering setting, with a Feller-Markov signal process, evolving on a general state space, and additive Gaussian white noise observations. The main result of [12] claimed that the filtering process has a unique invariant measure, if the signal X is ergodic and its tail σ-algebra $\cap_{t\le 0}\mathcal{F}^X_t$ is P-a.s. empty. A serious gap was revealed recently in the proof of this statement (see [3]) and its validity remains a challenging open question.

It turns out that stability of the filter requires either stronger ergodic property of X (as e.g. strong mixing) or sufficient regularity of the observation noise densities g_i, $i = 1, ..., d$. For example, if all the transition probabilities are positive $\lambda_{ij} > 0$, then (see [2], [13], [8])

$$\big|\pi_n - \bar{\pi}_n\big| \le C\left(1-\frac{\lambda_{\min}}{\lambda_{\max}}\right)^n, \quad n \ge 1,$$

with $\lambda_{\min} = \min_{ij}\lambda_{ij}$ and $\lambda_{\max} = \max_{ij}\lambda_{ij}$ and a constant C, depending on ν and $\bar{\nu}$. If only $\min_j \lambda_{ij} > 0$ for some row i and the chain is ergodic, then ([7])

$$\lim_{n\to\infty}\frac{1}{n}\log\big|\pi_n - \bar{\pi}_n\big| \le -\frac{\lambda_*}{\lambda_{\max}},$$

with $\lambda_* := \sum_{i=1}^d \mu_i \min_j \lambda_{ij}$. The latter is the weakest known assumption on X for (4) to hold *regardless* of g_i's.

On the other hand, ergodicity of X guarantees exponential stability, if the noise densities are bounded and have the same support. In particular, the authors of [2] considered the additive Gaussian observation model

$$Y_n = h(X_n) + \sigma\xi_n,$$

where h is a $\mathbb{S}\mapsto\mathbb{R}$ function, $\sigma > 0$ is a constant and ξ_n is a standard sequence of i.i.d Gaussian random variables. Among other interesting results, the following upper bound was derived

$$\overline{\lim_{\sigma\to 0}}\,\sigma^2\gamma_\sigma \le -\frac{1}{2}\sum_{i=1}^d \mu_i \min_{j\ne i}\big(h(a_i)-h(a_j)\big)^2. \qquad (5)$$

Roughly speaking it suggests that the stability exponent γ_σ, viewed as a function of σ, decreases as $\sigma \to 0$, improving stability of the filter, if the image of $\mathbb{S}$ under h has at least one unique point.

In this paper we address asymptotic behavior of γ as a function of a parameter, controlling the transitions rate of the chain. This is precisely formulated in Section II, along with the continuous time analogous setting. In Section III several surprising (at the first glance) consequences of the main results are discussed and Section IV outlines the main idea of the proof.

II. MAIN RESULTS

A. Discrete time setup

For a fixed $\varepsilon \in (0,1)$, let $X^\varepsilon = (X^\varepsilon_n)_{n\ge 1}$ be the *slow* Markov chain on $\mathbb{S}$ with transition probabilities

$$\lambda^\varepsilon_{ij} := \mathrm{P}\big(X^\varepsilon_n = a_j | X^\varepsilon_{n-1} = a_i\big) = \begin{cases} 1-\varepsilon\sum_{\ell\ne i}\lambda_{i\ell}, & i = j \\ \varepsilon\lambda_{ij}, & i \ne j \end{cases}$$

and initial distribution ν. Clearly ε controls the transitions rate of the chain, preserving however its invariant distribution μ. The corresponding observation process Y^ε is generated by (1) with X replaced by X^ε and the filtering processes π^ε and $\bar{\pi}^\varepsilon$ solve (2), driven by Y^ε instead of Y and with Λ replaced with the matrix Λ^ε of the "slow" transition probabilities λ^ε_{ij}.

Theorem 2.1: Assume

$\mathbf{a_1}$) X is ergodic

$\mathbf{a_2}$) the densities $g_i(u)$ are bounded and all the corresponding measures are equivalent. Moreover

$$\int_{\mathbb{R}} g_i(u)\log g_j(u)\varphi(du) > -\infty, \quad \forall i,j$$

Then for an admissible $(\nu,\bar{\nu})$,

$$\lim_{n\to\infty}\frac{1}{n}\log|\pi^\varepsilon_n - \bar{\pi}^\varepsilon_n| \le -\sum_{i=1}^d \mu_i \min_{j\ne i}\mathcal{D}(g_i \parallel g_j) + o(1), \quad \varepsilon\to 0, \quad (6)$$

where[1]

$$\mathcal{D}(g_i\parallel g_j) = \int_{\mathbb{R}} g_i(u)\log\frac{g_i}{g_j}(u)\varphi(du)$$

are the Kullback-Leibler divergences. In the two dimensional case $d = 2$ the asymptotic of (6) is precise

$$\lim_{n\to\infty}\frac{1}{n}\log|\pi^\varepsilon_n - \bar{\pi}^\varepsilon_n| = -\mu_1\mathcal{D}(g_1\parallel g_2) - \mu_2\mathcal{D}(g_2\parallel g_1) + o(1), \quad \varepsilon\to 0. \quad (7)$$

□

B. Continuous time setup

Let $X = (X_t)_{t\ge 0}$ be a continuous time Markov chain with values in $\mathbb{S} = \{a_1, ..., a_d\}$, transition intensities matrix Λ and initial distribution ν. Note that the same notation is used for transition intensities in continuous time and transition probabilities in discrete time. Recall that the chain is ergodic if and only if $\exp(\Lambda)$ has positive entries, in which case the invariant distribution μ is the solution of $\Lambda^*\mu = 0$, $\mu\in\mathcal{S}^{d-1}$.

The observation process is generated by

$$Y_t = \int_0^t h(X_s)ds + \sigma W_t, \quad t\ge 0, \qquad (8)$$

where $h : \mathbb{S}\mapsto\mathbb{R}$ is a fixed function, $\sigma > 0$ is a constant and $W = (W_t)_{t\ge 0}$ is a Wiener process, independent of X.

[1]as usual $0\log 0 = 0$ is understood and all random (in)equalities are claimed to hold P-a.s.

The vector of the conditional probabilities $\pi_t(i) = \mathrm{P}(X_t = a_i | \mathcal{F}_t^Y)$, where $\mathcal{F}_t^Y = \sigma\{Y_s, s \le t\}$, satisfies the Shiryaev-Wonham filtering Itô equation ([15], [16])

$$d\pi_t = \Lambda^* \pi_t dt + \big(\mathrm{diag}(\pi_t) - \pi_t \pi_t^*\big) h \big(dY_t - \pi_t^* h dt\big), \quad (9)$$

subject to $\pi_0 = \nu$, where $\mathrm{diag}(x)$ denotes a scalar matrix with $x \in \mathbb{R}^d$ on the diagonal and h is the column vector with entries $h(a_1)$,...,$h(a_d)$. It is not difficult to see that (9) has a unique strong solution if started from $\pi_0 = \bar{\nu}$, which is denoted by $\bar{\pi}_t$.

In the continuous time case the slow chain is obtained by the time scaling $X_t^\varepsilon := X_{t\varepsilon}$, which is again a Markov chain with generator $\varepsilon\Lambda$. Let Y^ε satisfy (8) with X replaced by X^ε and π_t^ε and $\bar{\pi}_t^\varepsilon$ be the solutions of (9), driven by Y^ε and with Λ replaced by $\varepsilon\Lambda$, subject to ν and $\bar{\nu}$ respectively.

Theorem 2.2: If X is ergodic, then

$$\lim_{t\to\infty} \frac{1}{t} \log |\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon| \le \\ -\frac{1}{2\sigma^2} \sum_{i=1}^d \mu_i \min_{j\ne i} \big(h(a_i) - h(a_j)\big)^2 + o(\varepsilon), \ \varepsilon \to 0. \quad (10)$$

For telegraphic signal $d = 2$

$$\lim_{t\to\infty} \frac{1}{t} \log |\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon| = -(\lambda_{12} + \lambda_{21})\varepsilon - \\ \frac{1}{2}(\Delta h)^2 + (\Delta h)^2 \frac{\int_0^1 q(x) x(1-x) dx}{\int_0^1 q(x) dx} \quad (11)$$

where $\Delta h = \big(h(a_1) - h(a_2)\big)/\sigma \ne 0$ and

$$q(x) = \frac{\exp\Big(-\frac{2\lambda_{21}\varepsilon}{(\Delta h)^2 x(1-x)} + \frac{2(\lambda_{12}-\lambda_{21})\varepsilon}{(\Delta h)^2}\Big(\log\frac{x}{1-x} + \frac{1}{1-x}\Big)\Big)}{x^2(1-x)^2}.$$

In particular,

$$\lim_{t\to\infty} \frac{1}{t} \log |\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon| = -\frac{1}{2}(\Delta h)^2 + \\ \big(1 + o(1)\big)\varepsilon \log \varepsilon^{-1} \frac{4\lambda_{12}\lambda_{21}}{\lambda_{12} + \lambda_{21}}, \quad \varepsilon \to 0. \quad (12)$$

□

Remark 2.3: Notice the similarity between (10) and (5). In fact in continuous time both asymptotics $\sigma \to 0$ and $\varepsilon \to 0$ are related by an appropriate time scaling. The quadratic rate in (5) stems from the Gaussian distribution of the observation noise.

III. SURPRISING BEHAVIOR

The results of Theorems 2.1 and 2.2 reveal some interesting and surprising details, which are best demonstrated if compared to the linear Kalman-Bucy setting. Consider the linear filtering problem for the signal X_t solving

$$X_t = X_0 - \int_0^t X_s ds + W_t,$$

subject to a standard Gaussian r.v. X_0. The slow signal $X_t^\varepsilon := X_{t\varepsilon}$ satisfies

$$X_t^\varepsilon = X_0 - \varepsilon \int_0^t X_t^\varepsilon ds + \sqrt{\varepsilon}\widetilde{W}_t,$$

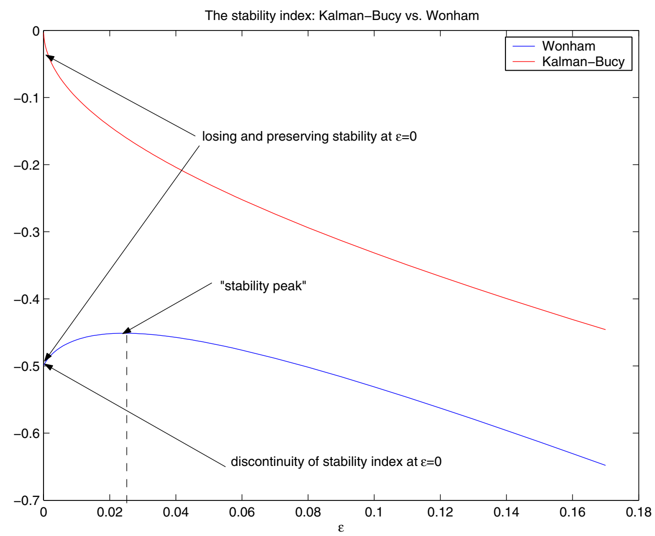

Fig. 1. Different behavior of the stability exponent for linear and nonlinear filter

where $\widetilde{W}_t := \varepsilon^{-1/2} W_{t\varepsilon}$ is a Wiener process. Note that X and X^ε have the same stationary invariant measure. Now define

$$Y_t^\varepsilon = \int_0^t X_s^\varepsilon ds + V_t$$

with V being a Wiener process, independent of W and X_0. For a Gaussian X_0, the conditional distribution $\pi_t^\varepsilon(dx)$ of X_t^ε given $\mathcal{F}_t^{Y^\varepsilon}$, is Gaussian with the mean M_t^ε and variance P_t^ε, satisfying the Kalman-Bucy equations (e.g. [14])

$$\begin{aligned} dM_t^\varepsilon &= -\varepsilon M_t^\varepsilon dt + P_t^\varepsilon \big(dY_t^\varepsilon - M_t^\varepsilon dt\big) \\ \dot{P}_t^\varepsilon &= -2\varepsilon P_t^\varepsilon dt + \varepsilon - \big(P_t^\varepsilon\big)^2, \end{aligned} \quad (13)$$

subject to $M_0^\varepsilon = 0$ and $P_0^\varepsilon = 1$. Suppose that the "incorrect" initial density is also Gaussian with mean 1 and unit variance. Then the incorrect density $\bar{\pi}_t^\varepsilon(dx)$ remains Gaussian as well with the mean $\bar{M}_t^\varepsilon$, satisfying (13) subject to $\bar{M}_0^\varepsilon = 1$ and the same variance P_t^ε. The total variation distance

$$|\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon| = \int_{\mathbb{R}} |\pi_t^\varepsilon(x) - \bar{\pi}_t^\varepsilon(x)| dx$$

is governed by $\Delta_t^\varepsilon := |M_t^\varepsilon - \bar{M}_t^\varepsilon|$ in this case, given by

$$\Delta_t^\varepsilon = \exp\Big(-\int_0^t (\varepsilon + P_s^\varepsilon) ds\Big)$$

and hence

$$\gamma_\varepsilon := \lim_{t\to\infty} t^{-1} \log |\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon| \propto -\varepsilon - P^\varepsilon \xrightarrow{\varepsilon\to 0} 0$$

where $P^\varepsilon = \lim_{t\to\infty} P_t^\varepsilon = -\varepsilon + \sqrt{\varepsilon^2 + \varepsilon}$.

Note the following differences between the stability exponent behavior as a function of ε (plotted at Figure 1) in the linear and nonlinear cases

1) Slowing down the diffusion signal in the Kalman-Bucy case leads to loss of exponential stability ($0 = \gamma_0 = \gamma_{0+}$), while the nonlinear filter remains stable $\gamma_{0+} < 0$ for arbitrary small $\varepsilon > 0$ if at least one of

the summands in (10) is positive. This also means that γ_ε is discontinuous at $\varepsilon = 0$ for the nonlinear filter. Indeed, in the limit model (X^0, Y^0), corresponding to $\varepsilon = 0$, the signal X^0 is just a random variable with values in $\mathbb{S}$ and distribution ν, which can be formally seen as a non-ergodic Markov chain. The filter is the equation with $\Lambda \equiv 0$ and is easily seen to be unstable: suppose $h(a_1) = ... = h(a_{d-1}) \neq h(a_d)$ and that both ν and $\bar{\nu}$ have positive entries. The states a_1,...,a_{d-1} cannot be distinguished from the observations on the set $\{X_0 \neq a_d\}$:

$$\lim_{t\to\infty} \frac{\bar{\pi}_t(i)}{\sum_{j=1}^{d-1} \bar{\pi}_t(j)} = \frac{\bar{\nu}_i}{\sum_{j=1}^{d-1} \bar{\nu}_j}, \quad i \neq d$$

which implies never vanishing dependence on $\bar{\nu}$. So $\gamma_0 = 0$ with positive probability, while

$$\gamma_{0+} \leq -\frac{1}{2\sigma^2}\mu_d(h(a_d) - h(a_1))^2 < 0.$$

2) While γ_ε is monotonous for the linear filter, it exhibits a maximum for the Shiryaev-Wonham filter. In other words, slowing down the signal may improve stability!

Strict positivity of γ_{0+} can be intuitively explained as follows: when the chain occupies a state a_i, having a unique image under h, the vectors π_t^ε and $\bar{\pi}_t^\varepsilon$ tend to concentrate the probability mass at the i-th entry, thus "synchronizing" between π_t^ε and $\bar{\pi}_t^\varepsilon$. One can verify that the distance $|\pi_t^\varepsilon - \bar{\pi}_t^\varepsilon|$ does not increase as a function of time and so it decays at least when the "synchronizing" states are revisited. The occupation time for these states increases as $\varepsilon \to 0$, but the time between revisits decreases, so that on average the filter spends the same proportion of time in the "synchronizing" states and so stability is preserved. The stabilization for diffusions is essentially different: note, for example, that the filtering error corresponding to estimation of a constant Gaussian signal decreases linearly, while for a constant random signal taking finite number of values it decreases exponentially. This provides at least partial intuition for (1).

The property (2) seems to be the manifestation of the interplay between two different stabilizing mechanisms: the ergodicity of the signal vanishes as $\varepsilon \to 0$, stripping the aforementioned "synchronizing" effect of the observations.

IV. SKETCH OF THE PROOF

The proof implements the Lyapunov exponents approach suggested by R.Atar and O.Zeitouni in [2]. The main novel technical element in this paper is application of the Furstenberg-Khasminskii stability approach, which simplifies the derivations and gives an additional insight into the problem. We give a sketch of the proof in discrete time, leaving out the technical details, fully developed in [6] and [5].

The key idea of the method in [2] is to study the two point motion stability of the equation (2) via the linear Zakai equation for unnormalized conditional distribution ρ_n:

$$\rho_n = G(Y_n)\Lambda^* \rho_{n-1}, \quad \rho_0 = \nu. \tag{14}$$

As is well known (and easily verified by induction in this case) $\pi_n = \rho_n/|\rho_n|$ and $\bar{\pi}_n = \bar{\rho}_n/|\bar{\rho}_n|$, where ρ_n and $\bar{\rho}_n$ are the solutions of (14), subject to $\rho_0 = \nu$ and $\bar{\rho}_0 = \bar{\nu}$ respectively. Moreover

$$\frac{1}{2}\frac{|\rho_n \wedge \bar{\rho}_n|}{|\rho_n||\bar{\rho}_n|} \leq |\pi_n - \bar{\pi}_n| \leq \frac{|\rho_n \wedge \bar{\rho}_n|}{|\rho_n||\bar{\rho}_n|}, \tag{15}$$

where $a \wedge b$ denotes the exterior product of $a, b \in \mathbb{R}^d$, which is identified with the $d \times d$ matrix with entries $a_i b_j - a_j b_i$. Thus

$$\lim_{n\to\infty} \frac{1}{n}\log|\pi_n - \bar{\pi}_n| = \lim_{n\to\infty} \frac{1}{n}\log|\rho_n \wedge \bar{\rho}_n| - \lim_{n\to\infty} \frac{1}{n}\log|\rho_n| - \lim_{n\to\infty} \frac{1}{n}\log|\bar{\rho}_n|,$$

provided the limits in the right hand side exist and are finite. Note that it suffices to verify existence of the limits under stationary distribution of (X, Y), i.e. when $\nu := \mu$, since by Markov property $\nu \ll \mu$ implies $\mathrm{Q}^\nu \ll \mathrm{Q}^\mu$, where Q^ν and Q^μ denote the measures induced by (X, Y), when $X_0 \sim \nu$ or $X_0 \sim \mu$ respectively. This allows to work with the stationary pair (X, Y) and thus to appeal to the Oseledec MET (see e.g. [1]), which, roughly speaking, states that the product of stationary random $d \times d$ matrices A_n with $\mathrm{E}\log^+|A_1| < \infty$ grows exponentially, so that for any $x \in \mathbb{R}^d$, the limit $\lim_{n\to\infty} n^{-1}\log|A_n...A_1 x|$ exists and may take one of d values $-\infty \leq \lambda_d \leq ... \leq \lambda_1 < \infty$, called the Lyapunov exponents corresponding to $A = (A_n)_{n\geq 1}$.

With bounded g_i's and stationary X, the assumptions of MET hold and as shown in [2], the solution of (14), corresponding to *any* initial condition $\bar{\nu} \in \mathcal{S}^{d-1}$, always "picks up" the top (largest) Lyapunov exponent:

$$\lambda_1 := \lim_{n\to\infty} \frac{1}{n}\log|\rho_n| = \lim_{n\to\infty} \frac{1}{n}\log|\bar{\rho}_n|, \quad \mathrm{P}-a.s.$$

It is easy to see that the matrix process $\rho_n \wedge \bar{\rho}_n$ also satisfies a linear equation (as e.g. (23) below) and thus is in the scope of MET: $\lim_{n\to\infty} \frac{1}{n}\log|\rho_n \wedge \bar{\rho}_n|$ exists and may take one of a finite number of values. In fact, another consequence of the Oseledec MET:

$$\lim_{n\to\infty} \frac{1}{n}\log|\rho_n \wedge \bar{\rho}_n| \leq \lambda_1 + \lambda_2,$$

where λ_2 is the second Lyapunov exponent of (14), implies that

$$\gamma = \lim_{n\to\infty} \frac{1}{n}\log|\pi_n - \bar{\pi}_n| \leq \lambda_1 + \lambda_2 - \lambda_1 - \lambda_1 = \lambda_2 - \lambda_1 \leq 0, \quad \mathrm{P}-a.s., \tag{16}$$

which means that the stability exponent of the filter (2) is controlled by the Lyapunov spectral gap of (14). The results claimed in Section II are obtained by estimating λ_1 and $\lambda_1 + \lambda_2$.

An additional insight on the structure of λ_1 can be gained by applying the following argument due to H.Furstenberg and R.Khasminskii (see e.g. [1])

$$|\rho_n| = \big|G(Y_n)\Lambda^*\rho_{n-1}\big| = \Big|G(Y_n)\Lambda^*\frac{\rho_{n-1}}{|\rho_{n-1}|}\Big||\rho_{n-1}| = \big|G(Y_n)\Lambda^*\pi_{n-1}\big||\rho_{n-1}|, \tag{17}$$

which implies

$$\lambda_1 = \lim_{n\to\infty} \frac{1}{n} \log|\rho_n| = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \big|G(Y_m)\Lambda^* \pi_{m-1}\big|.$$

Though by virtue of MET the limit always exists, it is not a priori clear that it is realized by averaging with respect to a unique stationary measure of (Y, π) (or equivalently of (X, ξ, π)) in the spirit of the law of large numbers. In fact (Y, π) may have several invariant measures, which average $|G(y)\Lambda^*\pi|$ to the same number. This is exactly what happens in the aforementioned Blackwell's example. Under $(\mathbf{a_1})$ and $(\mathbf{a_2})$ of Theorem 2.1 the law of large numbers does hold:

Lemma 4.1: (Lemma 4.1 in [5]) Under the assumptions of Theorem 2.1

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log|\pi_n - \bar{\pi}_n| < 0, \quad \mathrm{P}-a.s., \tag{18}$$

Moreover (X, π) is a Feller-Markov process with the unique stationary invariant measure $\Psi(dx, du)$ on $\mathbb{S}\times\mathcal{B}(\mathcal{S}^{d-1})$, such that for any continuous and bounded f

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m, \pi_m) = \lim_{n\to\infty} \mathrm{E} f(X_n, \pi_n) = \int_{\mathcal{S}^{d-1}} \sum_{i=1}^{d} u_i f(a_i, u) \Psi_\pi(du) \tag{19}$$

where $\Psi_\pi(du) = \sum_{i=1}^{d} \Psi(\{a_i\}, du)$ is the π-marginal. □

Proof: (sketch) First (18) is established by means of another approach, introduced in [2], based on the fact that the positive stochastic flow generated by (14) is a contraction with respect to the Hilbert projective metric for measures. The process (X, π) is shown then to be Markov and Feller and as such to have at least one invariant measure. Next, using (18), the uniqueness is verified and the standard arguments lead to the Birkhoff-Khintchine type of the law of large numbers, i.e. the first equality in (19). The second equality follows, since $\mathrm{E}f(X_n, \pi_n) = \mathrm{E}\sum_{i=1}^{d} \mathbf{1}_{\{X_n = a_i\}} f(a_i, \pi_n) = \mathrm{E}\sum_{i=1}^{d} \pi_n(i) f(a_i, \pi_n)$. ■

Using this lemma one obtains the Furstenberg-Khasminskii formula for λ_1:

$$\lambda_1 = \mathrm{E}_s \sum_{i=1}^{d} \pi_{1|0}(i) \int_{\mathbb{R}} g_i(u) \log\big|G(u)\pi_{1|0}\big| \varphi(du) \tag{20}$$

where E_s denotes the expectation with respect to $\Psi_\pi(du)$ and $\pi_{1|0} := \Lambda^*\pi_0$.

At this point we return to the asymptotic problem at hand: it turns out that the Lyapunov exponent λ_1^ε corresponding to ρ_n^ε, has a simple asymptotic expression as $\varepsilon \to 0$, which is the consequence of the following concentration property of the stationary conditional law of π_0:

Lemma 4.2: (Lemma 4.3 in [5]) Under the assumptions of Theorem 2.1

$$\mathrm{P_s}\text{-}\lim_{\varepsilon\to 0} \sum_{a_i \in \mathcal{J}_j} \pi_0^\varepsilon(i) = \mathbf{1}_{\{X_0 \in \mathcal{J}_j\}} \tag{21}$$

for all $j = 1, ..., d$, where $\mathcal{J}_j = \big\{a_\ell : \mathcal{D}(g_j \parallel g_\ell) = 0\big\}$. □

Proof: (The main idea) The limit (21) basically claims that those states of the chain, which are not "merged" by identical noise densities, can be determined precisely as $\varepsilon \to 0$. ■

Combination of (20) and (21) leads to

$$\lambda_1^\varepsilon = \sum_{i=1}^{d} \mu_i \int_{\mathbb{R}} g_i(u) \log g_i(u) \varphi(du) + o(1), \quad \varepsilon \to 0 \tag{22}$$

The matrix process $Z_n^\varepsilon = \rho_n^\varepsilon \wedge \bar{\rho}_n^\varepsilon$ satisfies linear recursion

$$Z_n^\varepsilon = G(Y_n^\varepsilon)\Lambda_\varepsilon^* Z_{n-1}^\varepsilon \Lambda_\varepsilon G(Y_n^\varepsilon), \quad Z_0^\varepsilon = \nu \wedge \bar{\nu},$$

and evolves in the space of antisymmetric matrices (with zero diagonal). In particular we have for $i \neq j$

$$Z_n^\varepsilon(i,j) = \sum_{1\le k\neq \ell\le d} g_k(Y_n^\varepsilon)\lambda_{ki}^\varepsilon Z_{n-1}^\varepsilon(k,\ell)\lambda_{\ell j}^\varepsilon g_\ell(Y_n^\varepsilon). \tag{23}$$

Similarly to (17), with $\Pi_n^\varepsilon := Z_n^\varepsilon/|Z_n^\varepsilon|$ and a fixed integer $r \ge 1$

$$\begin{aligned} |Z_n^\varepsilon| = |Z_{n-r}^\varepsilon| \Big(\sum_{i\neq j} |\Pi_{n-r}^\varepsilon(i,j)| \cdot \\ \prod_{m=n-r+1}^{n} g_i(Y_m^\varepsilon) g_j(Y_m^\varepsilon) + C_{r,n}\varepsilon \Big) \le \\ |Z_{n-r}^\varepsilon| \Big(\max_{i\neq j} \prod_{m=n-r+1}^{n} g_i(Y_m^\varepsilon) g_j(Y_m^\varepsilon) + C_{r,n}\varepsilon \Big), \ n \ge r \end{aligned}$$

where $C_{r,n}$ is a random sequence, independent of $\varepsilon > 0$ and growing not faster then linearly with r. Iterating the latter recursion and taking $\varepsilon \to 0$ on gets

$$\lim_{n\to\infty} \frac{1}{n} \log|\rho_n \wedge \bar{\rho}_n| = \lim_{n\to\infty} \frac{1}{n} \log|Z_n^\varepsilon| \le \sum_{\ell=1}^{d} \mu_\ell \mathrm{E} \max_{i\neq j} \frac{1}{r} \sum_{m=1}^{r} \log g_i\big(\xi_m(\ell)\big) g_j\big(\xi_m(\ell)\big) + o(\varepsilon). \tag{24}$$

Finally by the law of large numbers and uniform integrability of the summands, one gets

$$\mathrm{E} \max_{i\neq j} \frac{1}{r} \sum_{m=1}^{r} \log g_i\big(\xi_m(\ell)\big) g_j\big(\xi_m(\ell)\big) \xrightarrow{r\to\infty} \max_{i\neq j} \int_{\mathbb{R}} g_\ell(u) \log\big(g_i(u) g_j(u)\big) \varphi(du) \tag{25}$$

Then the upper bound (6) is obtained by inserting (22) and (24), (25) into (16).

In the special case of $d = 2$, the equation (23) is one dimensional, for which precise limit is obtained, leading to (7). The continuous time case is treated in a similar way. The closed form expression in (11) is found, using the explicit formula for the stationary distribution of π, obtained by solving the appropriate Kolmogorov equation (see the details in [5]).

REFERENCES

[1] L. Arnold, *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998

[2] R. Atar, O. Zeitouni, *Lyapunov exponents for finite state nonlinear filtering*. SIAM J. Control Optim. 35 (1997), no. 1, 36–55.

[3] P. Baxendale, P. Chigansky, R. Liptser, *Asymptotic stability of the Wonham filter: ergodic and nonergodic signals*, SIAM J. Control Optim. 43 (2004), no. 2, 643–669

[4] D. Blackwell, The entropy of functions of finite-state Markov chains. 1957 Transactions of the first Prague conference on information theory, Prague, 1956 pp. 13–20 Publishing House of the Czechoslovak Academy of Sciences, Prague

[5] P. Chigansky, *Stability of the nonlinear filter for slowly switching Markov chains*, submitted (preprint `math.PR/0411596` at `http://www.arxiv.org/`)

[6] P. Chigansky, *A formula for the top Lyapunov exponent of the Zakai equation*, submitted (preprint `math.PR/0404515` at `http://www.arxiv.org/`)

[7] P. Chigansky, R. Liptser, *Stability of nonlinear filters in non-mixing case*, (2004) Ann. App. Prob. Vol. 14, No. 4

[8] 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.

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

[10] Y. Ephraim, N.Merhav, *Hidden Markov processes*. Special issue on Shannon theory: perspective, trends, and applications. IEEE Trans. Inform. Theory 48 (2002), no. 6, 1518–1569

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

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

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

[14] R.Lipster, A.Shiryaev, *Statistics of random processes: theory and applications*, I,II, Springer-Verlag, 2nd Ed., 2001

[15] A.N. Shiryaev, *Optimal methods in quickest detection problems*, Teor. Verojatnost. i Primenen. 8 1963, pp. 26–51

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