



# Stability of the nonlinear filter for slowly switching Markov chains

Pavel Chigansky\*

*Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel*

Received 2 February 2005; received in revised form 21 February 2006; accepted 21 February 2006

Available online 20 March 2006

Dedicated to Robert Liptser on the occasion of his 70th birthday

---

## Abstract

Exponential stability of the nonlinear filtering equation is revisited, when the signal is a finite state Markov chain. An asymptotic upper bound for the filtering error due to an incorrect initial condition is derived in the case of a slowly switching signal.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Hidden Markov models; Nonlinear filtering; Lyapunov exponents; Stability; Kullback–Leibler relative entropy

---

## 1. Introduction and the main result

Consider a discrete time Markov chain  $X = (X_n)_{n \in \mathbb{Z}_+}$  with values in a finite real alphabet  $\mathbb{S} = \{a_1, \dots, a_d\}$ , initial distribution  $v_i = P(X_0 = a_i)$  and transition probabilities  $\lambda_{ij} = P(X_n = a_j | X_{n-1} = a_i)$ . Suppose that the chain is partially observed via the noisy sequence of random variables  $Y = (Y_n)_{n \in \mathbb{Z}_+}$ , generated by

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

---

\* Tel.: +972 89344214; fax: +972 89344308.

E-mail address: [pavel.chigansky@weizmann.ac.il](mailto:pavel.chigansky@weizmann.ac.il).

where  $\xi = (\xi_n)_{n \geq 1}$  is a sequence of i.i.d. random vectors with independent entries  $\xi_n(i)$ ,  $i = 1, \dots, d$ , such that

$$P(\xi_1(i) \in B) = \int_B g_i(u) \varphi(du), \quad B \in \mathcal{B}(\mathbb{R})$$

with densities  $g_i(u)$  and a  $\sigma$ -finite reference measure  $\varphi(du)$ .

Let  $\mathcal{F}_n^Y = \sigma\{Y_1, \dots, Y_n\}$  and  $\pi_n(i) = P(X_n = a_i | \mathcal{F}_n^Y)$ . The vector  $\pi_n$  of the conditional probabilities satisfies the recursive *filtering* equation

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

where  $G(y)$ ,  $y \in \mathbb{R}$  is a diagonal matrix with entries  $g_i(y)$ ,  $\Lambda^*$  is the transposed matrix of transition probabilities and  $|x| = \sum_{i=1}^d |x_i|$  for  $x \in \mathbb{R}^d$ .

Suppose that (2) can be solved subject to a probability distribution  $\bar{\nu} \neq \nu$  and denote the corresponding solution by  $\bar{\pi}_n$ . Under certain mild conditions (to be specified later) the limit

$$\gamma := \lim_{n \rightarrow \infty} \frac{1}{n} \log |\pi_n - \bar{\pi}_n|, \quad \text{P-a.s.}$$

exists and if it is negative the filter is said to be (exponentially) stable. The *stability index*  $\gamma$  is elusive for explicit calculation and much research has focused recently on estimating  $\gamma$  in various filtering settings (see [1–3,5–7,9] and others). In particular, a Gaussian additive white noise model was considered in [1] (cf. (1))

$$Y_n = h(X_n) + \sigma \eta_n, \quad n \geq 1, \eta_1 \sim \mathcal{N}(0, 1)$$

and the following asymptotic upper bound was derived:

$$\overline{\lim}_{\sigma \rightarrow 0} \sigma^2 \gamma(\sigma) \leq -\frac{1}{2} \sum_{i=1}^d \mu_i \min_{j \neq i} (h(a_i) - h(a_j))^2, \tag{3}$$

where  $\mu$  is the stationary distribution of the chain  $X$ , assumed to be ergodic. Recall that  $X$  is ergodic if  $\mu_i = \lim_{n \rightarrow \infty} P(X_n = a_i)$ ,  $i = 1, \dots, d$ , exist, are unique and positive, which holds iff  $\Lambda^q$  has positive entries for some integer  $q \geq 1$  (see e.g. [11]).

In this note a different scaling of the problem is chosen, namely the *slow chain* limit of  $\gamma$  is considered. Let  $X_n^\varepsilon$  be a Markov chain on  $\mathbb{S}$  with transition probabilities

$$\lambda_{ij}^\varepsilon = P(X_n^\varepsilon = a_j | X_{n-1}^\varepsilon = a_i) = \begin{cases} \varepsilon \lambda_{ij}, & i \neq j \\ 1 - \varepsilon \sum_{\ell \neq i} \lambda_{i\ell}, & i = j \end{cases}$$

for an  $\varepsilon \in (0, 1)$ . Notice that  $X^\varepsilon$  is an ergodic chain with the same invariant distribution  $\mu$  as  $X$ . Denote by  $Y^\varepsilon$  the corresponding observation sequence generated by (1), with  $X$  replaced by  $X^\varepsilon$ , and let  $\pi^\varepsilon, \bar{\pi}^\varepsilon$  be the solutions of (2) subject to  $\nu, \bar{\nu}$ , with  $Y$  and  $\Lambda$  replaced by  $Y^\varepsilon$  and  $\Lambda^\varepsilon$ .

**Theorem 1.** Assume that  $X$  is ergodic and the noise densities  $g_i(u)$

- (a1) are bounded,
- (a2) have the same support and
- (a3)  $\int_{\mathbb{R}} g_i(u) \log g_j(u) \varphi(du) > -\infty$ , for all  $i, j$ .

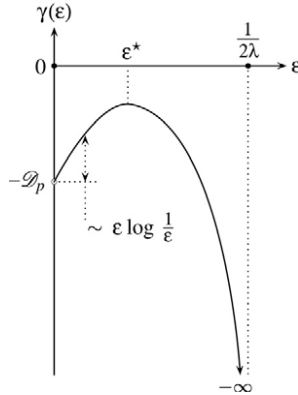


Fig. 1.  $\gamma(\varepsilon)$  for a BSC example.

Then for any pair  $(\nu, \bar{\nu})$  of probability distributions on  $\mathbb{S}$

$$\gamma(\varepsilon) \leq -\sum_{i=1}^d \mu_i \min_{j \neq i} \mathcal{D}(g_i \parallel g_j) + o(1), \quad \varepsilon \rightarrow 0, \tag{4}$$

where  $\mathcal{D}(g_i \parallel g_j) = \int_{\mathbb{R}} g_i(u) \log \frac{g_i}{g_j}(u) \varphi(du)$  are the Kullback–Leibler relative entropies. For  $d = 2$  the asymptotic (4) is precise, i.e.

$$\gamma(\varepsilon) = -\mu_1 \mathcal{D}(g_1 \parallel g_2) - \mu_2 \mathcal{D}(g_2 \parallel g_1) + o(1), \quad \varepsilon \rightarrow 0. \tag{5}$$

This theorem reveals the following interesting properties of  $\gamma(\varepsilon)$  (see Fig. 1).

1.  $\gamma(\varepsilon)$  may be discontinuous at  $\varepsilon = 0$

$$\gamma(0+) = \overline{\lim}_{\varepsilon \rightarrow 0} \gamma(\varepsilon) < \gamma(0) = 0,$$

if at least one of the entropies  $\mathcal{D}(g_i \parallel g_j)$  is strictly positive. This means that for small  $\varepsilon > 0$  the filter remains stable with virtually the same stability index as long as the chain is not “frozen” completely, while the filter, corresponding to the limit chain  $X_n^0 \equiv X_0, n \geq 1$ , may be unstable (e.g. when some but not all  $g_i(u)$ ’s coincide  $\varphi$ -a.s.). Such a behavior is not observed in the analogous “slowly varying” setting for the Kalman–Bucy filter, where the state space of the signal is continuous.

Surprising as it may seem at first glance, this phenomenon is quite natural for signals with discrete state space and can be explained as follows. The distance  $|\pi_n^\varepsilon - \bar{\pi}_n^\varepsilon|$  never increases and tends to decrease exponentially quickly whenever  $X_n^\varepsilon$  resides in a state with a distinct noise probability distribution. Since the average occupation time of this “synchronizing” state does not depend on  $\varepsilon$ , the decay remains exponential with nonzero average rate. The “dual” manifestation of this phenomenon is that the filter stability improves when the signal-to-noise ratio is increased in the setting of (3) (see [6,1]).

2. As demonstrated in the following example,  $\gamma(\varepsilon)$  may have a maximum at some  $\varepsilon^* > 0$  or, in other words, stability may improve when the chain is slowed down! This provides yet more evidence against the false intuition, directly relating stability of the filter to ergodic properties of the signal (see an extended discussion of this issue in [6,3]). The reason for such behavior stems from the delicate interplay between two stabilizing mechanisms: ergodicity of the signal and the

synchronizing effect of the observations. The first dominates the second for the faster chain, and vice versa when the chain is slow.

**Example 2.** Consider the so-called Binary Symmetric Channel (BSC) model, for which  $X_n \in \{0, 1\}$  is a symmetric chain with the jump probability  $\lambda$ , and  $Y_n = (X_n - \xi_n)^2$ , where  $\xi$  is an i.i.d.  $\{0, 1\}$  binary sequence with  $P(\xi_1 = 1) = p \in (0, 1/2)$ . Let  $X^\varepsilon$  and  $Y^\varepsilon$  denote the “slow” instances as defined above. In this case more can be said about the convergence in (5) (see the proof in Section 3 below), namely

$$\gamma(\varepsilon) \geq -\mathcal{D}_p + \frac{4\lambda(\log(2) - h(p))}{\mathcal{D}_p} \varepsilon \log \varepsilon^{-1}(1 + o(1)), \quad \varepsilon \rightarrow 0 \tag{6}$$

where  $\mathcal{D}_p := p \log \frac{p}{1-p} + (1-p) \log \frac{1-p}{p}$  and  $h(p) = -p \log p - (1-p) \log(1-p)$ . On the other hand,  $\gamma(\varepsilon) \leq \log(1 - 2\varepsilon\lambda) \rightarrow -\infty$  as  $\varepsilon \rightarrow 1/(2\lambda)$  (see e.g. Theorem 2.3 in [1]). Since the second term in the expansion of  $\gamma(\varepsilon)$  in (6) is positive and by (5)  $\gamma(\varepsilon) \rightarrow -\mathcal{D}_p$  as  $\varepsilon \rightarrow 0$ , one gets the qualitative behavior depicted in Fig. 1.  $\square$

**2. The proof of Theorem 1**

Hereafter the assumptions of Theorem 1 are in force and the following notations are used: probability measures on  $\mathbb{S}$  are identified with (column) vectors in  $\mathcal{S}^{d-1} = \{x \in \mathbb{R} : x_i \geq 0, \sum_{i=1}^d x_i = 1\}$ ,  $\mu(f) := \sum_{i=1}^d f(a_i)\mu_i$  for  $f : \mathbb{S} \mapsto \mathbb{R}$  and  $\mu \in \mathcal{S}^{d-1}$ ,  $\mu(A) := \mu(\mathbf{1}_{\{A\}})$  for  $A \subseteq \mathbb{S}$ . For a random sequence  $Z = (Z_n)_{n \in \mathbb{Z}}$  and  $m \geq k$  the notation  $\mathcal{F}_{[k,m]}^Z = \sigma\{Z_k, \dots, Z_m\}$  is used and  $\mathcal{F}_n^Z := \mathcal{F}_{[1,n]}^Z$  for brevity. Convergence of random sequences is understood in the P-a.s. sense unless stated otherwise.

The proof relies on the following idea from [1]. Recall that  $\pi_n = \rho_n/|\rho_n|$ ,  $n \geq 0$ , where  $\rho_n$  is the solution of the Zakai linear equation ( $\bar{\pi}_n$  is obtained similarly)

$$\rho_n = G(Y_n)A^* \rho_{n-1}, \quad \rho_0 = \nu. \tag{7}$$

Let  $\rho_n \wedge \bar{\rho}_n := \frac{1}{2}(\rho_n \bar{\rho}_n^* - \bar{\rho}_n \rho_n^*)$  denote the exterior product of  $\rho_n$  and  $\bar{\rho}_n$ . The elementary inequality

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

implies

$$\gamma := \lim_{n \rightarrow \infty} \frac{1}{n} \log |\pi_n - \bar{\pi}_n| = \lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n \wedge \bar{\rho}_n| - \lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n| - \lim_{n \rightarrow \infty} \frac{1}{n} \log |\bar{\rho}_n|. \tag{8}$$

Since  $g_i(u)$ 's are bounded, the limits in the right hand side exist by virtue of the Oseledec Multiplicative Ergodic Theorem (MET). Moreover, since  $(G(Y_n)A^*)_{n \geq 1}$  are matrices with nonnegative entries, the Perron–Frobenius theorem implies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n| = \lim_{n \rightarrow \infty} \frac{1}{n} \log |\bar{\rho}_n| := \lambda_1, \quad \forall \nu, \bar{\nu} \in \mathcal{S}^{d-1},$$

where  $\lambda_1$  is the top Lyapunov exponent corresponding to (7). Similarly MET implies  $\lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n \wedge \bar{\rho}_n| \leq \lambda_1 + \lambda_2$  and thus one concludes that  $\gamma \leq \lambda_2 - \lambda_1 \leq 0$ , i.e. the

filter stability index is controlled by the Lyapunov spectral gap of (7). The reader is referred to [1] for further details.

The statement of Theorem 1 follows from (8) and asymptotic expressions derived in Lemmas 3 and 4 below.

2.1. Asymptotic expression for  $\lambda_1(\varepsilon)$

**Lemma 3.** For any  $\varepsilon > 0$  the Markov process  $(X^\varepsilon, \pi^\varepsilon)$  has a unique stationary invariant measure  $\mathcal{M}^\varepsilon$ . The top Lyapunov exponent is given by

$$\lambda_1(\varepsilon) = \int_{\mathcal{S}^{d-1}} \sum_{i=1}^d (A^{\varepsilon*} u)_i \int_{\mathbb{R}} g_i(y) \log |G(y)A^{\varepsilon*}u| \varphi(dy) \mathcal{M}_\pi^\varepsilon(du), \tag{9}$$

where  $\mathcal{M}_\pi^\varepsilon$  is the  $\pi$ -marginal of  $\mathcal{M}^\varepsilon$ . For each  $\mathcal{J}_j = \{a_\ell : \mathcal{D}(g_j \parallel g_\ell) = 0\}$

$$\lim_{\varepsilon \rightarrow 0} \int \left( \mathbf{1}_{\{x \in \mathcal{J}_j\}} - \sum_{\ell: a_\ell \in \mathcal{J}_j} u_\ell \right)^2 \mathcal{M}^\varepsilon(dx, du) = 0 \tag{10}$$

and in particular

$$\lim_{\varepsilon \rightarrow 0} \lambda_1(\varepsilon) = \sum_{i=1}^d \mu_i \int_{\mathbb{R}} g_i(y) \log g_i(y) \varphi(dy). \tag{11}$$

**Proof.** The process  $(X^\varepsilon, \pi^\varepsilon)$  is Markov and by (a1) it is also Feller, and thus at least one invariant measure  $\mathcal{M}^\varepsilon$  exists. Its uniqueness can be deduced (as in Theorem 7.1 in [4]) from the stability property  $\lim_{n \rightarrow \infty} |\pi_n^\varepsilon - \bar{\pi}_n^\varepsilon| = 0, \forall v, \bar{v} \in \mathcal{S}^{d-1}$ , which in turn holds under the assumption (a2) by the arguments used in the proof of Theorem 2.3 in [1] (see also Theorem 4.1 in [3]). Concentration properties of  $\mathcal{M}_\pi^\varepsilon$  have been studied in [8], when all the noises are distinct, i.e.  $\mathcal{D}(g_i \parallel g_j) > 0$  for all  $i \neq j$ , which is not necessarily the case here.

Let  $\tilde{X}^\varepsilon$  be the stationary chain (i.e.  $\tilde{X}_0 \sim \mu$ ) and  $\tilde{\pi}^\varepsilon$  the corresponding optimal filtering process, generated by (2) subject to  $\tilde{\pi}_0^\varepsilon = \mu$ . For an  $f : \mathbb{S} \rightarrow \mathbb{R}$  and  $n, m \geq 0$  ( $\tilde{Y}^\varepsilon$  denotes the observations corresponding to  $\tilde{X}^\varepsilon$ ),

$$\begin{aligned} \mathbb{E} (f(\tilde{X}_{n+m}^\varepsilon) - \tilde{\pi}_{n+m}^\varepsilon(f))^2 &= \mathbb{E} \left( f(\tilde{X}_{n+m}^\varepsilon) - \mathbb{E} \left( f(\tilde{X}_{n+m}^\varepsilon) \mid \mathcal{F}_{n+m}^{\tilde{Y}^\varepsilon} \right) \right)^2 \\ &\leq \mathbb{E} \left( f(\tilde{X}_{n+m}^\varepsilon) - \mathbb{E} \left( f(\tilde{X}_{n+m}^\varepsilon) \mid \mathcal{F}_{[m+1, n+m]}^{\tilde{Y}^\varepsilon} \right) \right)^2 \\ &\stackrel{\dagger}{=} \mathbb{E} \left( f(\tilde{X}_n^\varepsilon) - \mathbb{E} \left( f(\tilde{X}_n^\varepsilon) \mid \mathcal{F}_n^{\tilde{Y}^\varepsilon} \right) \right)^2 \\ &= \mathbb{E} (f(\tilde{X}_n^\varepsilon) - \tilde{\pi}_n^\varepsilon(f))^2, \end{aligned}$$

where stationarity of  $(\tilde{X}^\varepsilon, \tilde{Y}^\varepsilon)$  has been used in  $\dagger$ . This means that the filtering error for the stationary signal does not increase with time. Then by the uniqueness of  $\mathcal{M}^\varepsilon$  for any fixed  $m \geq 0$ ,

$$\int (f(x) - u(f))^2 \mathcal{M}^\varepsilon(dx, du) = \lim_{n \rightarrow \infty} \mathbb{E} (f(\tilde{X}_n^\varepsilon) - \tilde{\pi}_n^\varepsilon(f))^2 \leq \mathbb{E} (f(\tilde{X}_m^\varepsilon) - \tilde{\pi}_m^\varepsilon(f))^2. \tag{12}$$

Define

$$\widehat{\pi}_n^\varepsilon(i) = \frac{\mu_i \prod_{k=1}^n g_i(\widetilde{Y}_k^\varepsilon)}{\sum_{j=1}^d \mu_j \prod_{k=1}^n g_j(\widetilde{Y}_k^\varepsilon)}, \quad i = 1, \dots, d$$

and let  $A_m^\varepsilon = \{\widetilde{X}_k^\varepsilon = \widetilde{X}_0, \forall k \leq m\}$ , the event that  $\widetilde{X}_k^\varepsilon$  does not jump on  $[0, m]$ . Notice that on the set  $A_m^\varepsilon$ , the observation process is independent of  $\varepsilon$ , namely

$$\widetilde{Y}_k^\varepsilon \equiv \widetilde{Y}_k^0 = \sum_{i=1}^d \mathbf{1}_{\{\widetilde{X}_0=a_i\}} \xi_k(i), \quad k = 1, \dots, m.$$

Then by optimality of  $\widetilde{\pi}^\varepsilon$

$$\begin{aligned} \mathbb{E} (f(\widetilde{X}_m^\varepsilon) - \widetilde{\pi}_m^\varepsilon(f))^2 &\leq \mathbb{E} (f(\widetilde{X}_m^\varepsilon) - \widehat{\pi}_m^\varepsilon(f))^2 \\ &= \mathbb{E} \mathbf{1}_{\{A_m^\varepsilon\}} (f(\widetilde{X}_0) - \widehat{\pi}_m^0(f))^2 + \mathbb{E} \mathbf{1}_{\{\Omega \setminus A_m^\varepsilon\}} (f(\widetilde{X}_m^\varepsilon) - \widehat{\pi}_m^\varepsilon(f))^2 \\ &\leq \mathbb{E} (f(\widetilde{X}_0) - \widehat{\pi}_m^0(f))^2 + 4d^2 \max_{a_i \in \mathbb{S}} |f(a_i)|^2 (1 - \mathbb{P}(A_m^\varepsilon)) \\ &\xrightarrow{\varepsilon \rightarrow 0} \mathbb{E} (f(\widetilde{X}_0) - \widehat{\pi}_m^0(f))^2. \end{aligned}$$

For  $f(x) := \mathbf{1}_{\{x \in \mathcal{J}_j\}}$  the latter and (12) imply

$$\overline{\lim}_{\varepsilon \rightarrow 0} \int \left( \mathbf{1}_{\{x \in \mathcal{J}_j\}} - \sum_{\ell: a_\ell \in \mathcal{J}_j} u_\ell \right)^2 \mathcal{M}^\varepsilon(dx, du) \leq \mathbb{E} (f(\widetilde{X}_0) - \widehat{\pi}_m(f))^2 \xrightarrow{m \rightarrow \infty} 0,$$

where the convergence holds since  $\{\widetilde{X}_0 \in \mathcal{J}_j\} \in \mathcal{F}_\infty^{\widetilde{Y}^0} = \bigvee_{n \geq 1} \mathcal{F}_n^{\widetilde{Y}^0}$  by definition of  $\mathcal{J}_j$  and since  $\widehat{\pi}_m^0(i), i = 1, \dots, d$ , are the optimal estimates of  $\mathbf{1}_{\{\widetilde{X}_0=a_i\}}$  given  $\mathcal{F}_m^{\widetilde{Y}^0}$ .

Once the existence of the ergodic stationary pair  $(X^\varepsilon, \pi^\varepsilon)$  is established<sup>1</sup> one may use it to realize the limit  $\lambda_1$  by means of the approach due to H. Furstenberg and R. Khasminskii (see e.g. [10]). The idea is to study the growth rate of  $\rho_n^\varepsilon$  by projecting it on the unit sphere  $(\mathcal{S}^{d-1}$  in this case):

$$|\rho_n^\varepsilon| = |G(Y_n^\varepsilon)A^{\varepsilon*} \rho_{n-1}^\varepsilon| = |\rho_{n-1}^\varepsilon| \left| G(Y_n^\varepsilon)A^{\varepsilon*} \frac{\rho_{n-1}^\varepsilon}{|\rho_{n-1}^\varepsilon|} \right| = |\rho_{n-1}^\varepsilon| |G(Y_n^\varepsilon)A^{\varepsilon*} \pi_{n-1}^\varepsilon|.$$

<sup>1</sup> Such a pair can be generated by taking both  $X_0$  and  $\pi_0$  randomly distributed according to  $\mathcal{M}^\varepsilon$  and its definition can be extended to the negative times by the usual arguments. Note that this is different from  $(\widetilde{X}^\varepsilon, \widetilde{\pi}^\varepsilon)$  used in the proof of  $\mathcal{M}^\varepsilon$  concentration.

Then by the law of large numbers (LLN) for ergodic processes (the required integrability conditions are provided by (a1) and (a3))

$$\begin{aligned}
 \lambda_1(\varepsilon) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n^\varepsilon| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n \log |G(Y_n^\varepsilon) A^{\varepsilon*} \pi_{n-1}^\varepsilon| = \mathbb{E} \log |G(Y_1^\varepsilon) A^{\varepsilon*} \pi_0^\varepsilon| \\
 &= \mathbb{E} \sum_{i=1}^d \mathbf{1}_{\{X_1^\varepsilon = a_i\}} \log |G(\xi_1(i)) A^{\varepsilon*} \pi_0^\varepsilon| \\
 &= \mathbb{E} \sum_{i=1}^d \mathbb{P} \left( X_1^\varepsilon = a_i \mid \mathcal{F}_{(-\infty, 0]}^{Y^\varepsilon} \right) \log |G(\xi_1(i)) A^{\varepsilon*} \pi_0^\varepsilon| \\
 &= \mathbb{E} \sum_{i=1}^d (A^{\varepsilon*} \pi_0^\varepsilon)_i \log |G(\xi_1(i)) A^{\varepsilon*} \pi_0^\varepsilon|. \tag{13}
 \end{aligned}$$

The latter expression is nothing but (9). The asymptotic (11) follows from  $\Lambda^\varepsilon = I + O(\varepsilon)$  and the concentration (10) of  $\mathcal{M}^\varepsilon$  as  $\varepsilon \rightarrow 0$ , since  $g_i(u)$ 's coincide  $\varphi$ -almost surely for all  $a_i \in \mathcal{J}_j$  for any  $j$  and the  $X$ -marginal of  $\mathcal{M}^\varepsilon$  is given by  $\mathcal{M}_X^\varepsilon(dx) = \sum_{i=1}^d \mu_i \delta_{a_i}(dx)$ .  $\square$

2.2. Asymptotic bound for  $\lambda_1(\varepsilon) + \lambda_2(\varepsilon)$

**Lemma 4.** For any  $v, \bar{v} \in \mathcal{S}^{d-1}$ ,

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n^\varepsilon \wedge \bar{\rho}_n^\varepsilon| \\
 &\leq \sum_{i=1}^d \mu_i \max_{k \neq m} \int_{\mathbb{R}} g_i(u) \log(g_m(u) g_k(u)) \varphi(du) + o(1), \quad \varepsilon \rightarrow 0. \tag{14}
 \end{aligned}$$

In the case  $d = 2$ ,

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \frac{1}{n} \log |\rho_n^\varepsilon \wedge \bar{\rho}_n^\varepsilon| &= \log(1 - \varepsilon \lambda_{12} - \varepsilon \lambda_{21}) + \mu_1 \int_{\mathbb{R}} g_1(u) \log(g_1(u) g_2(u)) \varphi(du) \\
 &\quad + \mu_2 \int_{\mathbb{R}} g_2(u) \log(g_1(u) g_2(u)) \varphi(du). \tag{15}
 \end{aligned}$$

**Proof.** The process  $Q_n^\varepsilon := \rho_n^\varepsilon \wedge \bar{\rho}_n^\varepsilon$  evolves in the space of antisymmetric matrices (with zero diagonal) and satisfies the linear equation

$$Q_n^\varepsilon = G(Y_n^\varepsilon) A^{\varepsilon*} Q_{n-1}^\varepsilon A^\varepsilon G(Y_n^\varepsilon), \quad Q_0^\varepsilon = v \wedge \bar{v},$$

or in the componentwise notation

$$Q_n^\varepsilon(i, j) = \sum_{1 \leq k \neq \ell \leq d} g_k(Y_n^\varepsilon) \lambda_{ki}^\varepsilon Q_{n-1}^\varepsilon(k, \ell) \lambda_{\ell j}^\varepsilon g_\ell(Y_n^\varepsilon), \quad i \neq j.$$

Unlike in the case of (7), it is not clear whether the limit  $\lim_{n \rightarrow \infty} \frac{1}{n} \log |Q_n^\varepsilon|$  depends on  $v, \bar{v}$  or whether  $\Pi_n^\varepsilon = Q_n^\varepsilon / |Q_n^\varepsilon|$  has any useful concentration properties as  $\varepsilon \rightarrow 0$ . However the

technique used in the previous section still gives the upper bound. With a fixed integer  $r \geq 1$ ,

$$\begin{aligned} |Q_n^\varepsilon| &= |Q_{n-r}^\varepsilon| \left| \{G(Y_n^\varepsilon) \Lambda^{\varepsilon*} \dots \{G(Y_{n-r+1}^\varepsilon) \Lambda^{\varepsilon*} \Pi_{n-r}^\varepsilon \Lambda^\varepsilon G(Y_{n-r+1}^\varepsilon)\} \dots \Lambda^\varepsilon G(Y_n^\varepsilon)\} \right| \\ &\leq |Q_{n-r}^\varepsilon| \left( \sum_{i \neq j} |\Pi_{n-r}^\varepsilon(i, j)| \prod_{m=n-r+1}^n g_i(Y_m^\varepsilon) g_j(Y_m^\varepsilon) + c_1(r) \varepsilon \right) \\ &\leq |Q_{n-r}^\varepsilon| \left( \max_{i \neq j} \prod_{m=n-r+1}^n g_i(Y_m^\varepsilon) g_j(Y_m^\varepsilon) + c_1(r) \varepsilon \right), \quad n \geq r \end{aligned}$$

with a constant  $c_1(r) > 0$ , depending only on  $r$  (due to assumption (a1)). By the MET the limit  $\lim_{n \rightarrow \infty} \frac{1}{n} \log |Q_n^\varepsilon|$  exists P-a.s. and hence (recall the definitions of  $Y^\varepsilon$  and  $A_r^\varepsilon$  in the proof of Lemma 3)

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log |Q_n^\varepsilon| &= \lim_{\ell \rightarrow \infty} \frac{1}{\ell r} \log |Q_{\ell r}^\varepsilon| \\ &\leq \lim_{\ell \rightarrow \infty} \frac{1}{\ell} \sum_{k=1}^{\ell} \frac{1}{r} \log \left( \max_{i \neq j} \prod_{m=kr-r+1}^{kr} g_i(Y_m^\varepsilon) g_j(Y_m^\varepsilon) + c_1(r) \varepsilon \right) \\ &\stackrel{\dagger}{=} \frac{1}{r} \mathbf{E} \log \left( \max_{i \neq j} \prod_{m=1}^r g_i(\tilde{Y}_m^\varepsilon) g_j(\tilde{Y}_m^\varepsilon) + c_1(r) \varepsilon \right) \\ &\leq \frac{1}{r} \mathbf{E} \mathbf{1}_{\{A_r^\varepsilon\}} \log \left( \max_{i \neq j} \prod_{m=1}^r g_i(\tilde{Y}_m^\varepsilon) g_j(\tilde{Y}_m^\varepsilon) + c_1(r) \varepsilon \right) + c_2(r) (1 - \mathbf{P}_\mu(A_r^\varepsilon)) \\ &\leq \frac{1}{r} \sum_{\ell=1}^d \mu_\ell \mathbf{E} \log \left( \max_{i \neq j} \prod_{m=1}^r g_i(\xi_m(\ell)) g_j(\xi_m(\ell)) + c_1(r) \varepsilon \right) + c_3(r) (1 - \mathbf{P}_\mu(A_r^\varepsilon)) \\ &\xrightarrow{\varepsilon \rightarrow 0} \sum_{\ell=1}^d \mu_\ell \mathbf{E} \max_{i \neq j} \frac{1}{r} \log \prod_{m=1}^r g_i(\xi_m(\ell)) g_j(\xi_m(\ell)), \end{aligned}$$

where the LLN was used in  $\dagger$  and the  $c_i(r)$  stand for  $r$ -dependent constants. Applying the LLN once again one gets for each  $\ell$

$$\begin{aligned} \frac{1}{r} \log \prod_{m=1}^r g_i(\xi_m(\ell)) g_j(\xi_m(\ell)) &= \frac{1}{r} \sum_{m=1}^r \log g_i(\xi_m(\ell)) g_j(\xi_m(\ell)) \\ &\xrightarrow{r \rightarrow \infty} \int_{\mathbb{R}} g_\ell(u) \log(g_i(u) g_j(u)) \varphi(du), \quad \text{P-a.s.} \end{aligned}$$

Since “max” is a continuous function

$$\max_{i \neq j} \frac{1}{r} \log \prod_{m=1}^r g_i(\xi_m(\ell)) g_j(\xi_m(\ell)) \xrightarrow{r \rightarrow \infty} \max_{i \neq j} \int_{\mathbb{R}} g_\ell(u) \log(g_i(u) g_j(u)) \varphi(du)$$

and by the uniform integrability, provided by assumption (a3),

$$\mathbf{E} \max_{i \neq j} \frac{1}{r} \log \prod_{m=1}^r g_i(\xi_m(\ell)) g_j(\xi_m(\ell)) \xrightarrow{r \rightarrow \infty} \max_{i \neq j} \int_{\mathbb{R}} g_\ell(u) \log(g_i(u) g_j(u)) \varphi(du).$$



Putting all parts together one gets the bound (14). In the case  $d = 2$ , the process  $Q_n^\varepsilon$  is one dimensional and all the calculations can be carried out exactly, leading to the expression (15).  $\square$

### 3. Proof of (6)

When the observation process  $Y_n^\varepsilon$  takes values in a discrete alphabet  $\mathbb{S}' = \{b_1, \dots, b_{d'}\}$ , the conditional densities (with respect to the point measure  $\varphi(dy) = \sum_{i=1}^{d'} \delta_{b_i}(dy)$ ) are of the form

$$g_i(y) = \sum_{j=1}^{d'} p_{ij} \mathbf{1}_{\{y=b_j\}}, \quad \sum_{j=1}^{d'} p_{ij} = 1, \quad p_{ij} \geq 0,$$

and hence by (13) ( $\pi_{1|0}^\varepsilon := A^{\varepsilon*} \pi_0^\varepsilon$  for brevity)

$$\begin{aligned} \lambda_1(\varepsilon) &= \mathbb{E} \log |G(Y_1^\varepsilon) A^{\varepsilon*} \pi_0^\varepsilon| = \mathbb{E} \sum_{j=1}^{d'} \mathbf{1}_{\{Y_1^\varepsilon=b_j\}} \log \left( \sum_{i=1}^{d'} p_{ij} \pi_{1|0}^\varepsilon(i) \right) \\ &= \mathbb{E} \sum_{j=1}^{d'} \mathbb{P} \left( Y_1^\varepsilon = b_j | \mathcal{F}_{(-\infty,0]}^{Y^\varepsilon} \right) \log \mathbb{P} \left( Y_1^\varepsilon = b_j | \mathcal{F}_{(-\infty,0]}^{Y^\varepsilon} \right) =: -\mathcal{H}(Y^\varepsilon), \end{aligned} \tag{16}$$

where  $\mathcal{H}(Y^\varepsilon)$  is known as the entropy rate of the stationary process  $Y^\varepsilon = (Y_n^\varepsilon)_{n \in \mathbb{Z}}$ .

Consider now the special case when  $X^\varepsilon$  and  $Y^\varepsilon$  take values in  $\mathbb{S} = \{0, 1\}$  and  $p = \mathbb{P}(Y_n^\varepsilon = i | X_n^\varepsilon = j)$  for  $i \neq j$ . The vector  $\pi_n^\varepsilon$  is one dimensional and hence  $\mathbb{P}(Y_1^\varepsilon = 1 | \mathcal{F}_{(-\infty,0]}^{Y^\varepsilon}) = (1 - p)\pi_{1|0}^\varepsilon + p(1 - \pi_{1|0}^\varepsilon)$ , where

$$\pi_{1|0}^\varepsilon := \mathbb{P} \left( X_1^\varepsilon = 1 | \mathcal{F}_{(-\infty,0]}^{Y^\varepsilon} \right) = (1 - \varepsilon \lambda_{10}) \pi_0^\varepsilon + \varepsilon \lambda_{01} (1 - \pi_0^\varepsilon) \tag{17}$$

and  $\pi_0^\varepsilon := \mathbb{P}(X_0^\varepsilon = 1 | \mathcal{F}_{(-\infty,0]}^{Y^\varepsilon})$  are redefined for brevity.

Let  $h(x) := -x \log x - (1 - x) \log(1 - x)$ ,  $x \in [0, 1]$  and  $\ell_p(q) = (1 - p)q + p(1 - q)$ , and define

$$H(p, q) := h(\ell_p(q)) \quad p, q \in [0, 1],$$

where  $0 \log 0 \equiv 0$  is understood. Since  $h(x) \leq \log(2)$  with equality at  $x = 1/2$  and  $\ell_p(1/2) = 1/2$ ,  $H(p, q) \leq \log(2)$  for all  $p, q \in [0, 1]$  with equality at  $q = 1/2$ . Since  $h(x)$  is a concave function, symmetric around  $x = 1/2$ ,

$$H(p, q) = h((1 - p)q + p(1 - q)) \geq qh(1 - p) + (1 - q)h(p) = h(p), \quad p \in [0, 1],$$

with equality at  $q = 0$  and  $q = 1$ . Finally for any fixed  $p \in [0, 1]$ ,  $q \mapsto H(p, q)$  inherits concavity and symmetry from  $h(x)$ . These properties imply the following lower bound:

$$H(p, q) \geq h(p) + \frac{\log(2) - h(p)}{1/2} \min(q, 1 - q), \quad p, q \in [0, 1]. \tag{18}$$

By Theorem 1 in [8] for the symmetric chain  $X^\varepsilon$  with jump probability  $\lambda$  and  $p \neq 1/2$ ,

$$\begin{aligned} \mathbb{E} \min(\pi_0^\varepsilon, 1 - \pi_0^\varepsilon) &= \mathbb{P}(X_0^\varepsilon \neq \operatorname{argmax}_i \pi_0^\varepsilon(i)) \\ &= \frac{\lambda}{\mathcal{D}_p} \varepsilon \log \varepsilon^{-1} (1 + o(1)), \quad \varepsilon \rightarrow 0, \end{aligned} \tag{19}$$

where  $\mathcal{D}_p := p \log \frac{p}{1-p} + (1-p) \log \frac{1-p}{p}$ . The expression for  $\mathcal{H}(Y^\varepsilon)$  in the case  $d = 2$  reads

$$\mathcal{H}(Y^\varepsilon) = EH(p, \pi_{1|0}^\varepsilon) = EH(p, \pi_0^\varepsilon) + O(\varepsilon), \quad \varepsilon \rightarrow 0$$

where the latter asymptotic form follows from (17), since  $H(p, q)$  is differentiable in  $q$ .

Now (18) and (19) imply

$$\mathcal{H}(Y^\varepsilon) \geq h(p) + 2(\log(2) - h(p)) \frac{\lambda}{\mathcal{D}_p} \varepsilon \log \varepsilon^{-1} (1 + o(1)), \quad \varepsilon \rightarrow 0,$$

and (6) follows from (8), (15) and (16).  $\square$

## Acknowledgements

The author is grateful to Rami Atar for elaborations on [1] and useful comments regarding the results of this paper. The research was supported by a grant from the Israel Science Foundation.

## References

- [1] R. Atar, O. Zeitouni, Lyapunov exponents for finite state nonlinear filtering, *SIAM J. Control Optim.* 35 (1) (1997) 36–55.
- [2] R. Atar, O. Zeitouni, Exponential stability for nonlinear filtering, *Ann. Inst. H. Poincaré Probab. Statist.* 33 (1997) 697–725.
- [3] P. Baxendale, P. Chigansky, R. Liptser, Asymptotic stability of the Wonham filter: Ergodic and nonergodic signals, *SIAM J. Control Optim.* 43 (2) (2004) 643–669.
- [4] A. Budhiraja, H.J. Kushner, Approximation and limit results for nonlinear filters over an infinite time interval, *SIAM J. Control Optim.* 37 (6) (1999) 1946–1979.
- [5] P. Chigansky, R. Liptser, Stability of nonlinear filters in nonmixing case, *Ann. Appl. Probab.* 14 (4) (2004) 2038–2056.
- [6] B. Delyon, O. Zeitouni, Lyapunov exponents for filtering problem, in: M.H.A. Davis, R.J. Elliot (Eds.), *Applied Stochastic Analysis*, Gordon & Breach, New York, 1991, pp. 511–521.
- [7] P. Del Moral, A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms, *Ann. Inst. H. Poincaré Probab. Statist.* 37 (2) (2001) 155–194.
- [8] R. Khasminskii, O. Zeitouni, Asymptotic filtering for finite state Markov chains, *Stochastic Processes. Appl.* 63 (1996) 1–10.
- [9] F. Le Gland, L. Mevel, Exponential forgetting and geometric ergodicity in hidden Markov models, *Math. Control Signals Systems* 13 (2000) 63–93.
- [10] R.Z. Khasminskii, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.
- [11] J.R. Norris, *Markov chains*, in: *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, 1998.