Lyapunov exponents for finite state 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

July 19, 1994. Revised June 17, 1995.

Abstract Consider the Wonham optimal filtering problem for a finite state ergodic Markov process in both discrete and continuous time, and let σ be the noise intensity for the observation. We examine the sensitivity of the solution with respect to the filter's initial conditions in terms of the gap between the first two Lyapunov exponents of the Zakai equation for the unnormalized conditional probability. This gap is studied in the limit as $\sigma \to 0$ by techniques involving considerations of nonlinear filtering and the stochastic Feynman-Kac formula. Conditions are given for the limit to be either negative or $-\infty$. Asymptotic bounds are derived in the latter case.

1 Introduction and statement of results

Let $\{X_n\}_{n=0}^{\infty}$ denote a finite state space, discrete time homogeneous Markov chain, with transition matrix G and initial distribution p_0. Without loss of generality, we take the state space of the Markov chain to consist of the set $\{1,\ldots,d\}$. Denote the law of the chain X_n on $\{1,\ldots,d\}^{\mathbb{Z}}$ by P. Throughout this paper, we assume that G leads to an ergodic non cyclic chain. That is, we assume

(A1) there exists a $k \geq 1$ such that $G^k(i,j) > 0$ for all $i,j \in \{1,\ldots,d\}$.

We denote by E_s expectations under the unique stationary measure of $\{X_n\}$.

We assume that the Markov chain X_n is observed through the sequence $\{Y_n\}_{n=1}^{\infty}$, where

$$Y_n = \delta h_{X_n} + \sqrt{\delta}\sigma\nu_n \,.$$

Here, $h : \{1,\ldots,d\} \to \mathbb{R}$ is the observation function, $\delta > 0$ is a parameter (which, for as long as one deals only with discrete time, may be taken as $\delta = 1$), σ is an observation noise parameter related to the Signal to Noise Ratio (SNR), and $\{\nu_n\}_{n=1}^{\infty}$ is a sequence of i.i.d., standard Gaussian random variables.

*The work of this author was partially supported by a US–ISRAEL BSF grant and by the fund of promotion of research at the Technion.

Let $\mathcal{Y}_n$ denote the σ–algebra generated by the observations $Y_1, \ldots, Y_n$. The nonlinear filtering problem consists of computing the conditional law $p_j(n) = P(X_n = j|\mathcal{Y}_n)$. Let D_n denote the diagonal matrix with $D_n(i,i) = \exp[\sigma^{-2}(h_i Y_n - h_i^2 \delta/2)]$, and define

$$\rho(n) = D_n G^* \rho(n-1)\,, \tag{1}$$

where G^* denotes the transpose of G, and $\rho(0) = p_0$. It is a straight forward consequence from Bayes' rule (see e.g [1, page 460] and also the continuous time case in [9]) that the vector $p(n) = (p_1(n), \ldots, p_d(n))^*$ satisfies $p(n) = \rho(n)/\langle \rho(n), \mathbf{1}\rangle$ where $\rho(n) = (\rho_1(n), \ldots, \rho_d(n))^*$, $\mathbf{1} = (1, \ldots, 1)^*$ and $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^d$.

Often, one has no access to the initial distribution p_0. A common procedure is then to initialize (1) with some initial condition $q_0 \in S^{d-1}$, where S^{d-1} denotes the $(d-1)$ – dimensional simplex. Denote by $\rho^{q_0}(n)$ the solution to (1) initialized this way, and denote by $p^{q_0}(n)$ the corresponding normalized (random) probability vector. Natural questions are then how far is $p^{q_0}(n)$ from $p^{p_0}(n)$; what are the conditions for stability in the sense that $\|p^{q_0}(n) - p^{p_0}(n)\| \to_{n\to\infty} 0$, and under these conditions, what is the rate of convergence. We emphasize that we deal here with the dependence of the optimal filter on *its* initial conditions, and not with its dependence on perturbations of the state process, $\{X_n\}$. The latter is a different problem which we do not deal with here.

Motivated by the approach taken in [4], (see [6] for a related computation in the continuous time, linear case), we couch the question in terms of Lyapunov exponents. That is, for any two $q_0 \neq q_0' \in S^{d-1}$, define

$$\gamma_\sigma^\delta(q_0, q_0', \omega) = \limsup_{n\to\infty} \frac{1}{n\delta} \log ||p^{q_0}(n) - p^{q_0'}(n)||\,.$$

Although here and in the sequel we take $||\cdot||$ to denote the Euclidean norm, note that the definition does not depend on the precise norm used, and in particular one could use here the variation (ℓ^1) norm.

We will see that, under mild conditions, $\gamma_\sigma^\delta(q_0, q_0', \omega)$ is almost surely deterministic, $\gamma_\sigma^\delta = \gamma_\sigma^\delta(q_0, q_0', \omega)$ is independent of q_0, q_0' for a.e. q_0, q_0' (when q_0, q_0' are distributed uniformly over the simplex), and is related to the gap between the top two Lyapunov exponents associated with the Zakai equation for the unnormalized conditional probability. The deterministic quantity $-1/\gamma_\sigma^\delta$ can then be interpreted as the "memory length" of the filter. Obviously, this approach is meaningful only if $\gamma_\sigma^\delta < 0$. We will identify below sufficient conditions for this to happen. An analogous continuous time question is examined as well.

We remark that in order to deal with the filter's memory length, we introduce and use tools borrowed from the theory of products of random matrices. Especially, we formulate the (qualitative) question of stability, and the (quantitative) question of memory length in terms of Lyapunov exponents of the solution of Zakai's equation. While the question of computing Lyapunov exponents is, in general, difficult, we study the above mentioned gap in the limiting cases, i.e. the regimes $\sigma \to \infty$ and $\sigma \to 0$. Under appropriate conditions, we obtain the exact order of the memory length as a function of σ in the latter case.

A natural guess is that γ_σ^δ becomes more negative as the SNR increases (i.e., as $\sigma \to 0$). As pointed out in [4] for the continuous time setup, this is not always the case, and one may even have situations where $\lim_{\sigma\to 0}\gamma_\sigma^\delta = 0$ though $\gamma_\sigma^\delta < 0$ for all positive σ. We identify below conditions for the memory length $-1/\gamma_\sigma^\delta$ to remain bounded as a function of σ, and conditions for it to decay to zero as $\sigma \to 0$.

The structure of the paper is as follows. In the rest of this section, we describe the results for the memory length in both discrete and continuous time. In particular, in both cases, we provide the uniform bounds on γ_σ^δ alluded to above, and determine under appropriate conditions the limits of γ_σ^δ under both high and low SNR. Sections 2 and 3, respectively, are devoted to proofs of the discrete and continuous time results.

We begin with the following rather straight forward consequence of Oseledec's theorem (see [2] p. 181 and [3]):

Theorem 1 *Assume (A1). Then there exists a deterministic function of σ and δ, γ_σ^δ, which admits the following:*

1. *Let q_0, q_0' be random, uniformly distributed (U) on the simplex S^{d-1}, independent of each other and of the chain $X_0, \{X_n, Y_n\}_{n=1}^\infty$. Then*

$$\gamma_\sigma^\delta(q_0, q_0', \omega) = \gamma_\sigma^\delta\,, \quad U \times U \times P - a.s.$$

2. *For any deterministic $q_0 \neq q_0'$ with all entries strictly positive, one has*

$$\gamma_\sigma^\delta(q_0, q_0', \omega) \leq \gamma_\sigma^\delta\,, \quad P - a.s.$$

As is seen in section 2, γ_σ^δ is just δ^{-1} times the difference between the two top Lyapunov exponents of solutions of (1).

We turn to study γ_σ^δ quantitatively. First is a bound which is uniform with respect to σ.

Theorem 2 *Assume that all entries of G are strictly positive. Then*

$$\gamma_\sigma^\delta \leq \frac{c}{\delta} < 0$$

for some constant c independent of h, σ, δ.

Remark: Actually, one may somewhat relax the condition that all entries of G are positive and still have the conclusion of the theorem. See Theorem 8 in Section 2 for such a statement, and its proof there (which also serves as a proof of Theorem 2) for the explicit dependence of c on the matrix G.

While the above bound relies on the nature of the law of $\{X_n\}$, and *its* mixing properties, the next bound relies on the good quality of the observation. In fact, it is shown that under a condition on $h_\cdot$, the decay rate tends to infinity as the noise parameter tends to zero. The condition required on h is that it possesses one coordinate which differs from the rest (h one to one suffices). For each $i \in \{1, \ldots, d\}$, define the set

$$\mathrm{nbr}(i) = \{j \neq i : |h_i - h_j| = \min_{k \neq i} |h_i - h_k|\} ,$$

and define $h_{\mathrm{nbr}(i)} = h_j$ where j is one of the members in the set $\mathrm{nbr}(i)$.

Theorem 3 *Assume (A1). Then*

$$\limsup_{\sigma \to 0} \sigma^2 \gamma_\sigma^\delta \leq -\frac{1}{2} E_s [h_{X_1} - h_{\mathrm{nbr}(X_1)}]^2 . \tag{2}$$

If, in addition, $\det(G) \neq 0$ *then*

$$\liminf_{\sigma \to 0} \sigma^2 \gamma_\sigma^\delta \geq -\frac{1}{2} E_s \sum_{i=1}^{d} [h_{X_1} - h_i]^2 . \tag{3}$$

Note that while the gap between the upper and lower bounds increases with the dimension d (and is nonzero as soon as $d > 2$), one may conclude from Theorem 3 that $\gamma_\sigma^\delta = \Omega(\sigma^{-2})$ as soon as there exists an i such that the set $\{j : h_j = h_i\}$ consists of a single point. The memory length is thus of the order of σ^2.

In continuous time, the behavior at low SNR ($\sigma \to \infty$) is completely determined by the top, non–zero eigenvalue of G (see [4]). An analogous result is shown here to hold for the discrete time case.

Let τ be the Birkhoff contraction coefficient (see section 2, equation (11) for definition).

Theorem 4 *Assume (A1). Then*

$$\limsup_{\sigma \to \infty} \gamma_\sigma^\delta \leq \inf_{m \geq 1} \frac{1}{m\delta} \log \tau(P^m) < 0.$$

In continuous time we prove results analogous to theorems 1, 2 and 3. Though the statements are similar, the proofs are harder and involve different techniques; in particular, a naive discretization approach fails. Let $\{x_t\}$ denote a Markov chain, with state space $\{1, \cdots, d\}$, and transition matrix $\hat{G}$. We assume that $\hat{G}$ leads to an ergodic chain, that is,

$$\text{(A2)} \qquad \text{for every } \delta > 0,\ (\exp(\hat{G}\delta))(i,j) > 0 \text{ for all } i, j \in \{1, \ldots, d\}.$$

The above holds iff all states are communicating. Next, assume that $\{x_t\}$ is observed via

$$dy_t = h_{x_t} dt + \sigma d\nu_t,$$

where ν_t is a standard Wiener process independent of $\{x_t\}$, and h is as in the discrete time case. Let H denote the diagonal matrix with elements $H(i,i) = h_i$, then the Zakai equation for the problem is

$$d\rho_t = \hat{G}^* \rho_t dt + \sigma^{-2} H \rho_t dy_t \tag{4}$$

with $p_t = \rho_t / \langle \rho_t, \mathbf{1} \rangle$. Now define for every $q_0 \neq q_0' \in S^{d-1}$

$$\gamma_\sigma(q_0, q_0', \omega) = \limsup_{t\to\infty} \frac{1}{t} \log \|p_t^{q_0} - p_t^{q_0'}\| \,,$$

then a result similar to Theorem 1 holds:

Theorem 5 *Assume (A2). Then there exists a deterministic function of σ, γ_σ, which admits the following:*

1. *Let q_0, q_0' be random, uniformly distributed (U) on the simplex S^{d-1}, independent of each other and of the chain $\{x_t, y_t\}_{t=0}^{\infty}$. Then*

$$\gamma_\sigma(q_0, q_0', \omega) = \gamma_\sigma \,, \quad U \times U \times P - a.s.$$

2. *For any deterministic $q_0 \neq q_0'$, one has*

$$\gamma_\sigma(q_0, q_0', \omega) \leq \gamma_\sigma \,, \quad P - a.s.$$

A result analogous to Theorem 2 holds also:

Theorem 6 *Assume (A2), then*

$$\gamma_\sigma \leq -2 \min_{1\leq i,j\leq d \;:\; i\neq j} (g_{ij} g_{ji})^{1/2}$$

where $g_{ij} = \hat{G}(i,j)$.

Remark: In [4] it is already proved that $\gamma_\sigma < 0$ under certain conditions, though not uniformly in σ.

Finally, a result analogous to Theorem 3 holds.

Theorem 7 *Assume (A2). Then*

$$\limsup_{\sigma\to 0} \sigma^2 \gamma_\sigma \leq -\frac{1}{2} E_s [h_{x_0} - h_{\mathrm{nbr}(x_0)}]^2 \,. \tag{5}$$

Moreover,

$$\liminf_{\sigma\to 0} \sigma^2 \gamma_\sigma \geq -\frac{1}{2} E_s \sum_{i=1}^{d} [h_{x_0} - h_i]^2 \,. \tag{6}$$

2 Proofs - discrete time

Throughout, we let $T_n = D_n G^*$ and $M_n^\sigma = T_n \cdots T_1$. We denote by $a \wedge b$ the exterior product of two vectors in $\mathbb{R}^d$, and by $A \wedge B$ the exterior product of two subspaces of $\mathbb{R}^d$ (see [2] for definitions of exterior products). For a $d \times d$ matrix A, $||A||$ denotes the operator norm (with respect to the Euclidean norm on $\mathbb{R}^d$). Finally, we use throughout c to denote a constant, whose value may change from line to line, which is independent of n, σ, δ.

Proof of Theorem 1: Note first that it is enough to prove the theorem in the case that X_0 is distributed according to the stationary distribution of $\{X_n\}$. Indeed, due to (A1), the stationary distribution has all entries strictly positive, and thus all almost sure statements, once proved for X_0 distributed according to the stationary law, must translate to the case where $X_0 = j$ for any $j = 1, \ldots, d$. The case of general initial distributions follows immediately.

We may thus assume that X_0 is distributed according to its stationary law. In that case, the sequence of matrices $\{D_n G^*\}_{n=1}^\infty$ possesses a stationary law, which is also ergodic by (A1). Moreover,

$$E \log^+ ||D_n G^*|| \leq cE \max_{i=1}^{d} \sigma^{-2}(Y(n)h_i - \frac{1}{2}h_i^2\delta)^+ < \infty\,.$$

Hence, we may apply Oseledec's theorem (see, e.g., [2]) to conclude that there exists a (random) strict subspace $S_\omega^1 \subset \mathbb{R}^d$ such that if $q_0 \notin S_\omega^1$ then

$$\frac{1}{n} \log ||\rho^{q_0}(n)|| \to \lambda_1^\sigma\,, \quad P-\text{a.s.} \tag{7}$$

Here and in the sequel, λ_i^σ denotes the i-th (non-random) Lyapunov exponent associated with the product of matrices M_n^σ. As is well known (see [3]), the matrix series $((M_n^\sigma)^* M_n^\sigma)^{1/2n}$ has a (random) limit a.s., the eigenvalues of which are $e^{\lambda_i^\sigma}$. Note that $(M_n^\sigma)^* M_n^\sigma$ is a non–negative matrix, thus by the Perron-Frobenius theorem the eigenvector associated with the highest eigenvalue of $(M_n^\sigma)^* M_n^\sigma$ has all coordinates real and non-negative. The last property thus holds for $((M_n^\sigma)^* M_n^\sigma)^{1/2n}$, too, and for $\lim_{n\to\infty}(M_n^* M_n)^{1/2n}$. Since S_ω^1 must be orthogonal to the eigenvector associated with the highest eigenvalue of $\lim_{n\to\infty}(M_n^* M_n)^{1/2n}$, it follows that S_ω^1 can not include any probability vector with all entries strictly positive. As for the case where q_0 does not have all its entries strictly positive, notice that $p^{q_0}(k)$ does (where k is such that $G^k(i,j) > 0$ for all $i,j \in \{1,\ldots,d\}$). Thus (7) really holds for any $q_0 \in S^{d-1}$.

Using again Oseledec's theorem, this time for the $\mathbb{R}^d \wedge \mathbb{R}^d$ –valued process $\rho^{q_0}(n) \wedge \rho^{q_0'}(n)$, there exists a (random) strict subspace $S_\omega^2 \subset \mathbb{R}^d \wedge \mathbb{R}^d$ such that if $q_0 \wedge q_0' \notin S_\omega^2$ then

$$\frac{1}{n} \log ||\rho^{q_0}(n) \wedge \rho^{q_0'}(n)|| \to_{n\to\infty} \lambda_1^\sigma + \lambda_2^\sigma\,, \quad P-\text{a.s.} \tag{8}$$

Furthermore, for $q_0 \wedge q_0' \in S_\omega^2$, Oseledec's theorem implies

$$\limsup_{n\to\infty} \frac{1}{n} \log ||\rho^{q_0}(n) \wedge \rho^{q_0'}(n)|| \leq \lambda_1^\sigma + \lambda_2^\sigma\,, \quad P-\text{a.s.} \tag{9}$$

Next, note that there exists a dimensional constant c_d such that if a, b are two probability vectors in S^{d-1} then

$$\frac{1}{c_d}|\sin(a,b)| \leq ||a-b|| \leq c_d|\sin(a,b)|\,,$$

where (a,b) denotes the angle between the vectors a, b. Since for any two non zero vectors c, d (not necessarily normalized) one has that $|\sin(c,d)| = ||c \wedge d||/(||c|| \cdot ||d||)$, one may conclude that,

$$\limsup_{n\to\infty}\frac{1}{n}\log||p^{q_0}(n)-p^{q_0'}(n)|| = \limsup_{n\to\infty}\frac{1}{n}[\log||\rho^{q_0}(n)\wedge\rho^{q_0'}(n)|| - \log||\rho^{q_0}(n)|| - \log||\rho^{q_0'}(n)||]\,.$$

Combining this and the fact that (7) holds for any probability vectors q_0, q_0' with either (8) or (9) yields both parts of the theorem, with $\gamma_\sigma^\delta = \delta^{-1}(\lambda_2^\sigma - \lambda_1^\sigma)$. □

It is useful to state the last sentence of the proof of Theorem 1 as the following

Corollary 1

$$\gamma_\sigma^\delta = \delta^{-1}(\lambda_2^\sigma - \lambda_1^\sigma)\,. \tag{10}$$

As is clear from [4] (and is evident also in the course of the proof of Theorem 1), the gap between the first and second Lyapunov exponents will play a crucial role in our study of the stability of the nonlinear filter. Before providing the proof of Theorem 2, it is useful to recall some definitions and a result of Peres concerning this gap. We follow the notations of [7], [8].

We say that a matrix A possessing non-negative entries is *allowable* if it contains no columns or rows whose entries are all zero. Let S_+^{d-1} denote those elements of S^{d-1} whose entries are all strictly positive. *Hilbert's projective metric* is the metric $\bar{h}(\cdot,\cdot)$ defined on $S_+^{d-1}\times S_+^{d-1}$ defined by

$$\bar{h}(x,y) = \log\max_{1\leq i,j\leq d}\frac{x_i y_j}{x_j y_i}\,.$$

Every allowable matrix A can be seen, by normalization of the linear action of A, as an operator $A : S_+^{d-1} \to S_+^{d-1}$. We denote by $A.x$ its action on $x \in S_+^{d-1}$. Define now the *Birkhoff contraction coefficient* of an allowable matrix A by

$$\tau(A) = \sup\left\{\frac{\bar{h}(A.x, A.y)}{\bar{h}(x,y)}\Big| x, y \in S_+^{d-1}, x \neq y\right\}\,. \tag{11}$$

Lemma 1 (Peres [7]) *Let $\{T_n\}_{n\geq 1}$ be an ergodic stationary sequence of non-negative, allowable matrices, such that $E\log^+||T_1|| < \infty$. Let λ_1, λ_2 denote the top two Lyapunov exponents for the random product of the T_i. Then,*

$$\lambda_1 - \lambda_2 \geq -E\log\tau(T_1)\,,$$

where $\lambda_2 = -\infty$ if the right hand side is infinite.

Proof: See [7, Proposition 5]. □

We recall from [7] and [8] the following useful properties of the contraction coefficient $\tau(\cdot)$:

1. $\tau(AD) = \tau(DA) = \tau(A)$ for any diagonal matrix D with strictly positive diagonal terms.

2. For any matrix A with strictly positive entries, $\tau(A) < 1$.

3. Let A be allowable, and define

$$\psi(A) = \min_{i,j,k,l} \frac{a_{ik}a_{jl}}{a_{il}a_{jk}}. \tag{12}$$

Then

$$\tau(A) = \frac{1-\sqrt{\psi(A)}}{1+\sqrt{\psi(A)}}. \tag{13}$$

We are now in a position to state the extension of Theorem 2 alluded to in the introduction.

Theorem 8 *Assume that $\tau(G) < 1$. Then*

$$\gamma_\sigma^\delta \le \frac{\log \tau(G)}{\delta} < 0\,.$$

Note that Theorem 2 follows at once from Theorem 8 by using property 2 for $\tau(G)$. Moreover, it follows that c may be taken as $c = \log(1-\Psi)/(1+\Psi)$ with $\Psi = \min_{i,j} G_{ij}/\max_{i,j} G_{ij}$.

Proof of Theorem 8: Applying Theorem 1 and Corollary 1 in combination with Lemma 1 to the recursion (1), one sees that

$$\gamma_\sigma^\delta \le \delta^{-1} E \log \tau(D_1 G^*) = \delta^{-1}\log\tau(G^*) = \delta^{-1}\log\tau(G) < 0\,,$$

where the first equality follows from property 1 for $\tau(\cdot)$, the second from property 3, and the last inequality from the assumption. □

Proof of Theorem 3: Suppose equation (1) is given two initial conditions q_0, q_0' and denote

$$q_n = p_n^{q_0}, \quad q_n' = p_n^{q_0'}, \quad r_n = q_n - q_n'.$$

Now, $q_n = T_n q_{n-1}/\langle T_n q_{n-1}, \mathbf{1}\rangle$, and subtracting $\langle T_n q_{n-1}', \mathbf{1}\rangle q_n' = T_n q_{n-1}'$ from $\langle T_n q_{n-1}, \mathbf{1}\rangle q_n = T_n q_{n-1}$ one gets

$$\langle T_n q_{n-1}, \mathbf{1}\rangle r_n + \langle T_n r_{n-1}, \mathbf{1}\rangle q_n' = T_n r_{n-1}.$$

Denoting $a_n = \langle T_n q_{n-1}, \mathbf{1}\rangle$ and noticing $a_n > 0$ one then has that

$$\begin{aligned} r_n &= a_n^{-1} T_n r_{n-1} - a_n^{-1} q_n' \langle T_n r_{n-1}, \mathbf{1}\rangle = a_n^{-1}(I - q_n'\mathbf{1}^*) T_n r_{n-1} = \\ &= a_n^{-1}(I - q_n'\mathbf{1}^*) D_n G^* r_{n-1}. \end{aligned}$$

The following recursion for r_n then holds :

$$\begin{aligned} r_0 &= q_0 - q_0' \\ r_n &= a_n^{-1} T_n' G^* r_{n-1} \end{aligned} \tag{14}$$

where we denote

$$T'_n = (I - q'_n \mathbf{1}^*) D_n.$$

Now, in order to estimate the evolution of r_n one may write

$$\frac{1}{n}\log\|r_n\| \le \frac{1}{n}\sum_{i=1}^{n}\log a_i^{-1} + \frac{1}{n}\sum_{i=1}^{n}\log\|T'_i\| + \frac{1}{n}\sum_{i=1}^{n}\log\|G^*\| + \frac{1}{n}\log\|r_0\|. \tag{15}$$

Noticing that the third term is bounded by zero and the fourth tends to zero, we turn to bound the two first terms. The first term tends for any q_0, a.s., to $-\lambda_1^\sigma$, since

$$\frac{1}{n}\sum_{i=1}^{n}\log a_i^{-1} = -\frac{1}{n}\log\langle T_n T_{n-1}\cdots T_1 q_0, \mathbf{1}\rangle$$

(c.f. the discussion following (7) above). Hence, it is enough to compute the limit of the last quantity for $q_0 = p_0$. Denoting the density of $(Y_1,\ldots,Y_n)$ by $f_{Y_1^n}(\beta_1^n)$ and the distribution of $(X_1,\ldots,X_n)$ by $P\left((X_1,\ldots,X_n) = (\alpha_1,\ldots,\alpha_n)\right) = p_{X_1^n}(\alpha_1^n)$ it follows from Bayes' rule that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\log a_i^{-1} = -\frac{1}{n}\log[f_{Y_1^n}(Y_1^n)(2\pi\sigma^2\delta)^{n/2}\exp\frac{1}{2\sigma^2\delta}\sum_i Y_i^2] \\
&= -\frac{1}{n}\log[\sum_{\alpha_1^n} p_{X_1^n}(\alpha_1^n)(2\pi\sigma^2\delta)^{-n/2}\exp -\frac{1}{2\sigma^2\delta}\sum_i (Y_i - h_{\alpha_i}\delta)^2] \\
&-\frac{1}{2}\,\log 2\pi\sigma^2\delta - \frac{1}{2n\sigma^2\delta}\sum_i Y_i^2 \\
&\le -\frac{1}{n}\log[p_{X_1^n}(X_1^n)\exp -\sum_i \frac{1}{2\sigma^2}(\sigma\nu_i)^2] - \frac{1}{2\sigma^2\delta}\frac{1}{n}\sum_i Y_i^2 \\
&= -\frac{1}{n}\log p_{X_1^n}(X_1^n) - \frac{1}{2\sigma^2}\frac{1}{n}\sum_i [h_{X_i}^2\delta + 2\sqrt{\delta}\sigma\nu_i h_{X_i}].
\end{aligned} \tag{16}$$

Now we turn to the second term of inequality (15). Writing the diagonal terms of D_n as $\Delta_n^i = D_n(i,i)$ we have the following expression for T'_n:

$$T'_n = \begin{pmatrix}
\Delta_n^1(1-q_n'^1) & \Delta_n^2(-q_n'^1) & \cdots & \Delta_n^d(-q_n'^1) \\
\Delta_n^1(-q_n'^2) & \Delta_n^2(1-q_n'^2) & \cdots & \Delta_n^d(-q_n'^2) \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_n^1(-q_n'^d) & \Delta_n^2(-q_n'^d) & \cdots & \Delta_n^d(1-q_n'^d)
\end{pmatrix}.$$

It is useful to consider here the operator norm of $T'_n : \ell^1 \to \ell^1$, namely, $\|T'_n\|_1 = \max_k \sum_i |(T'_n)_{ik}|$. Fix n and suppose $X_n = j$, then $Y_n = h_j\delta + \sigma\nu_n\sqrt{\delta}$, and

$$\|T'_n\|_1 = \max_i\{\Delta_n^i[1 - q_n'^i + \sum_{l\ne i} q_n'^l]\} = 2\max_i \Delta_n^i(1 - q_n'^i).$$

Denoting the vector $b_n = (b_n^1, \ldots, b_n^d)^* := G^* q'_{n-1}$ one has that

$$q_n'^j = \frac{b_n^j \Delta_n^j}{\sum_{l=1}^d b_n^l \Delta_n^l},$$

and thus

$$1 - q_n'^j = \frac{\sum_{k \neq j} b_n^k \Delta_n^k}{\sum_{l=1}^d b_n^l \Delta_n^l} \leq \min(1, \frac{\max_{k \neq j} \Delta_n^k}{b_n^j \Delta_n^j}) \leq 1_{\{b_n^j < \alpha\}} + 1_{\{b_n^j \geq \alpha\}} \frac{\max_{k \neq j} \Delta_n^k}{\alpha \Delta_n^j}, \tag{17}$$

for every fixed $0 < \alpha < 1$. Therefore,

$$2\Delta_n^j (1 - q_n'^j) \leq 2\Delta_n^j 1_{\{b_n^j < \alpha\}} + \frac{2}{\alpha} 1_{\{b_n^j \geq \alpha\}} \max_{k \neq j} \Delta_n^k,$$

and, clearly,

$$\forall i, i \neq j \quad 2\Delta_n^i (1 - q_n'^i) \leq \frac{2}{\alpha} \max_{k \neq j} \Delta_n^k.$$

Using the equivalence of the norms $||\cdot||$ and $||\cdot||_1$, it follows that there exists a constant c, independent of n such that

$$\|T_n'\| \leq \frac{2c}{\alpha} [1_{\{b_n^j < \alpha\}} \max_k \Delta_n^k + 1_{\{b_n^j \geq \alpha\}} \max_{k \neq j} \Delta_n^k]$$

and thus, defining $h_{\max} = \max_i \{|h_i|\}$,

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \log \|T_i'\| \leq \log \frac{2c}{\alpha} + \frac{1}{n} \sum_{i=1}^n 1_{\{b_i^{X_i} < \alpha\}} \sigma^{-2} \max_k (h_k h_{X_i} \delta + h_k \sigma \sqrt{\delta} \nu_i - \frac{1}{2} h_k^2 \delta) \\
&+ \frac{1}{n} \sum_{i=1}^n 1_{\{b_i^{X_i} \geq \alpha\}} \sigma^{-2} \max_{k \neq X_i} (h_k h_{X_i} \delta + h_k \sigma \sqrt{\delta} \nu_i - \frac{1}{2} h_k^2 \delta) \\
&\leq \log \frac{2c}{\alpha} + \frac{1}{n} \sum_{i=1}^n 1_{\{b_i^{X_i} < \alpha\}} (\sigma^{-2} \delta {h_{\max}}^2 + \sigma^{-1} \sqrt{\delta} h_{\max} |\nu_i|) \\
&+ \frac{1}{n} \sum_{i=1}^n \sigma^{-2} \delta (h_{\mathrm{nbr}(X_i)} h_{X_i} - \frac{1}{2} h_{\mathrm{nbr}(X_i)}^2) + \sigma^{-1} \sqrt{\delta} h_{\max} |\nu_i|.
\end{aligned}$$

Now, $b_i^{X_i} \geq (G)_{X_{i-1} X_i} q_{i-1}'^{X_{i-1}}$ so choosing $\alpha = \frac{1}{2} \min_{u,v:(G)_{uv}>0} (G)_{uv}$ one has that $1_{\{b_i^{X_i} < \alpha\}} \leq 1_{\{q_{i-1}^{X_{i-1}} < \frac{1}{2}\}}$. Combining this with inequalities (15) and (16) one has, after taking expectation and limit, that

$$\lim_{n \to \infty} \frac{1}{n} E \log \|r_n\| \leq c_1 + c_2 \sqrt{\delta}/\sigma - \frac{\delta}{2\sigma^2} E_s (h_{X_i} - h_{\mathrm{nbr}(X_i)})^2 + c_3 \sigma^{-2} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n P(q_{i-1}^{X_{i-1}} < \frac{1}{2}).$$

Since $P(q_n'^{X_n} < \frac{1}{2}) \longrightarrow_{\sigma \to 0} 0$ uniformly in n and since again by Oseledec's theorem (see e.g. [2] p. 181),

$$\lim \frac{1}{n} \log \| \wedge^r T_n \cdots T_1 \| = \lim \frac{1}{n} E \log \| \wedge^r T_n \cdots T_1 \| \quad a.s.,$$

the first part of the theorem is proved.

The second part easily follows from the following facts. First, the spectrum of the matrix process certainly satisfies

$$\lambda_2^\sigma - \lambda_1^\sigma \geq \lambda_2^\sigma - \lambda_1^\sigma + 2\lambda_1^\sigma + \lambda_3^\sigma + \cdots + \lambda_d^\sigma - d\lambda_1^\sigma = \sum_{i=1}^d \lambda_i^\sigma - d\lambda_1^\sigma .$$

Second, since $\det T_n \cdots T_1 = \det T_n \cdots \det T_1$, the sum of the exponents can be explicitly expressed as

$$\sum_{i=1}^d \lambda_i^\sigma = \lim_{n\to\infty} \frac{1}{n} \log|\det T_n \cdots T_1| = E_s \frac{1}{\sigma^2}(Y_1 \sum_{i=1}^d h_i - \frac{1}{2}\sum_{i=1}^d h_i^2 \delta) + \log|\det G|$$
$$= \frac{\delta}{\sigma^2} E_s\,(h_{X_1} \sum_{i=1}^d h_i - \frac{1}{2}\sum_{i=1}^d h_i^2) + \log|\det G|\,,$$

while

$$\lambda_1^\sigma \leq E\,\log\|\mathrm{diag}(\Delta_1^i)_{i=1}^d\|_1 + \log\|G\|_1 \leq E\,\max_i[-\frac{\delta}{2\sigma^2}(h_{X_1}-h_i)^2 + \frac{\delta h_{X_1}^2}{2\sigma^2} + h_i\frac{\sqrt{\delta}\nu_1}{\sigma}] \leq \frac{\delta}{2\sigma^2} E_s\, h_{X_1}^2 + \frac{c\sqrt{\delta}}{\sigma}.$$

Thus we conclude that

$$\liminf_{\sigma\to 0} \sigma^2\gamma_\sigma^\delta \geq E_s\,[h_{X_1}\sum_{i=1}^d h_i - \frac{1}{2}\sum_{i=1}^d h_i^2 - \frac{d}{2}h_{X_1}^2] = E_s\,[-\frac{1}{2}\sum_i (h_{X_1} - h_i)^2].$$

$\square$

Proof of Theorem 4: The last inequality holds since by the assumption, $\exists m_0\ s.t.\ \forall m \geq m_0\ G^{m_0} > 0$. As for the first inequality, like in the proof of Theorem 1, it suffices to work under the assumption that X_0 is distributed according to the stationary distribution. One may apply Lemma 1 for the process of matrices that are derived from $\{T_n\}$ by taking products of blocks at length m, where $m \geq m_0$:

$$T_m T_{m-1} \cdots T_1,\quad T_{2m}T_{2m-1}\cdots T_{m+1},\quad \ldots .$$

Ergodicity, stationarity and integrability follow from those of $\{T_n\}$. Since $(G^*)^m$ is positive, that is,

$$\sum_{i_2,\ldots,i_{m-1}} (G^*)_{i_1 i_2} \cdots (G^*)_{i_{m-1} i_m} > 0$$

it follows that

$$(T_m \cdots T_1)_{i_1 i_m} = \sum_{i_2,\ldots,i_{m-1}} \Delta_i^{i_1}(G^*)_{i_1 i_2} \cdots \Delta_m^{i_m}(G^*)_{i_{m-1} i_m} > 0$$

and allowability follows. The Lyapunov spectrum for this series is $\{m\lambda_i^\sigma\}_{i=1}^d$, thus

$$\gamma_\sigma^\delta \leq \frac{1}{m\delta} E\,\log\tau(T_m T_{m-1}\cdots T_1). \tag{18}$$

The diagonal terms Δ_j^i for which $T_j = \mathrm{diag}(\Delta_j^i)_{i=1}^d G^*$, may be expressed as

$$\Delta_j^i = \exp \sigma^{-2}(h_i Y_j - \frac{1}{2}\, h_i^2 \delta) = \exp[\delta\sigma^{-2}(h_i h_{X_j} - \frac{1}{2}\, h_i^2) + \sqrt{\delta}\sigma^{-1} h_i \nu_j] = 1 + \alpha_j^i.$$

Thus

$$T_m T_{m-1} \cdots T_1 = (I + \mathrm{diag}(\alpha_m^i)_{i=1}^d) G^* \cdots (I + \mathrm{diag}(\alpha_1^i)_{i=1}^d) G^* = (G^*)^m + M,$$

where M is a matrix satisfying

$$\|M\| \leq \|G\|^m [(1 + \|\mathrm{diag}(\alpha_m^i)_{i=1}^d\|) \cdots (1 + \|\mathrm{diag}(\alpha_1^i)_{i=1}^d\|) \ - 1].$$

Now,

$$\|\mathrm{diag}(\alpha_j^i)_{i=1}^d\| = \max_i | \exp[\delta\sigma^{-2}(h_i h_{X_j} - \frac{1}{2}\, h_i^2) + \sqrt{\delta}\sigma^{-1} h_i \nu_j] \ - 1| \longrightarrow_{\sigma\to\infty} 0 \quad a.s.,$$

therefore

$$T_m T_{m-1} \cdots T_1 \longrightarrow_{\sigma\to\infty} (G^*)^m \quad a.s..$$

As $T_m T_{m-1} \cdots T_1$ is positive, ψ is continuous, and so is τ, yielding

$$\tau(T_m T_{m-1} \cdots T_1) \longrightarrow_{\sigma\to\infty} \tau((G^*)^m) = \tau(G^m).$$

Since $\log(\tau(\cdot)) \leq 0$, Fatou's Lemma may be applied to get

$$\limsup_{\sigma\to\infty} E \ \log \tau(T_m T_{m-1} \cdots T_1) \leq E \ \limsup_{\sigma\to\infty} \log \tau(T_m T_{m-1} \cdots T_1) = E \ \log \tau(G^m) = \log \tau(G^m)$$

and the result follows from inequality (18). □

3 Proofs - continuous time

Throughout this section, c denotes a t-independent deterministic constant (whose value may change from line to line). p_{stat}^x denotes the (unique, by (A2)) stationary law corresponding to $\hat{G}$. We use the notations x_0^t and y_0^t to denote the sub σ-fields generated, respectively, by $\{x_s, 0 \leq s \leq t\}$ and $\{y_s, 0 \leq s \leq t\}$. E_0 denotes expectations under the product measure $\mathcal{P}_x \times \mathcal{P}_y$, where $\mathcal{P}_x$ denotes the law of the Markov chain x (under the stationary measure) and $\mathcal{P}_y$ denotes the law of the observation process $\{y_t, 0 \leq t < \infty\}$.

Proof of Theorem 5: Aside from the conditions needed for Oseledec's Theorem, that are proved below, the proof is identical to that of Theorem 1. Notice that equation (4) is bilinear, thus there exists a multiplicative process denoted $U = \{U_t\}_{t \in \mathbb{R}_+}$ such that $\rho_t = U_t \rho_0$. Assuming x_0 is distributed according to its stationary law, the shift transformation θ_t is measure preserving w.r.t. $\{x, \nu\}$ and thus w.r.t. U. Ergodicity of U follows from that of $\{x, \nu\}$, and separability follows from continuity. It follows from Theorem 2.1 in [5] that $U_t : \mathbb{R}^d \to \mathbb{R}^d$ is a homeomorphism, and thus invertible. For Oseledec's Theorem to hold, one needs to show also integrability of (see [2] p. 181)

$$u_1 = \sup_{0 \leq t \leq 1} \log^+ \|U_t\|, \quad u_2 = \sup_{0 \leq t \leq 1} \log^+ \|U_t^{-1}\|.$$

To show u_1 is integrable, it suffices to show $\sup_{0\le t\le 1}\|U_t\|$ is. Note that by the Kallianpur-Striebel formula, U_t is a non-negative matrix, and hence the unit vector w maximizing $\|U_t w\|$ has non-negative entries. Thus, it suffices, by considering the projection on w, to show integrability of $\sup_{0\le t\le 1}\|U_t v\|$ for some v all of whose entries are positive. Under (A2) all entries of p^x_{stat} are positive, so v may be chosen to be p^x_{stat}. Since this vector is also the initial distribution of x, it follows that $\|U_t p^x_{\mathrm{stat}}\|_1 = \langle \rho_t, \mathbf{1}\rangle$. By the Kallianpur-Striebel formula,

$$\langle \rho_t, \mathbf{1}\rangle = E_0[\exp\int_0^t (h_{x_s} dy_s - \frac{1}{2}\, h^2_{x_s} ds)|y_0^t] \le E_0[\exp h_{\max}\sum_i |\Delta y_i|],$$

where $\Delta y_i = y_{\tau_{i+1}} - y_{\tau_i}, \tau_0 = 0, \tau_i = \min\{t > \tau_{i-1} : x_t \ne x_{\tau_{i-1}}\} \wedge 1$, and integrability follows from the existence of exponential moments of the normal distribution and the exponential law of $\tau_i - \tau_{i-1}$. As for u_2, denote, for a symmetric matrix A, by $\lambda_i(A)$ the ith biggest eigenvalue of A, then

$$\begin{aligned}\log^+\|U_t^{-1}\| \le \|U_t^{-1}\| &= [\lambda_1(U_t^{-1*}U_t^{-1})]^{1/2} = [\lambda_d(U_tU_t^*)]^{-1/2}\\ &\le \frac{[\lambda_1(U_tU_t^*)]^{(d-1)/2}}{(|\det U_tU_t^*|)^{1/2}} \le \frac{\|(U_tU_t^*)\|^{(d-1)/2}}{|\det U_t|} \le \frac{\|U_t\|^{d-1}}{|\det U_t|}.\end{aligned} \tag{19}$$

Now, U_t solves the following Stratonovich equation

$$dU_t = (\hat{G}^* - \frac{1}{2}\,\sigma^{-2}H^2)U_t dt + \sigma^{-2}HU_t \circ dy_t\,,$$

thus

$$|\det U_t| = \exp\int_0^t \mathrm{trace}(\hat{G}^* - \frac{1}{2}\,\sigma^{-2}H^2)ds + \mathrm{trace}(\sigma^{-2}H)\circ dy_s$$

and combining this with inequality (19) and the Cauchy-Schwartz inequality, the integrability of u_2 follows. □

A corollary analogous to Corollary 1 follows:

Corollary 2 *Let λ_i^σ denote the Lyapunov exponents associated with the multiplicative process U_t. Then,*

$$\gamma_\sigma = \lambda_2^\sigma - \lambda_1^\sigma\ . \tag{20}$$

Proof of Theorem 6: By Oseledec's theorem,

$$\lim_t \frac{1}{t}\log\|\wedge^r\, U_t\| = \lambda_1^\sigma + \cdots + \lambda_r^\sigma \quad a.s.$$

This limit equals the limit on the discrete time series $\{n\delta\}$ for some $\delta > 0$, so if one looks at the series of linear operators $\{\hat{A}^\delta_n\}$ that satisfy

$$\rho^{p_0}(n\delta) = \hat{A}^\delta_n \rho^{p_0}((n-1)\delta)\,, \quad \rho^{p_0}(0) = p_0\,, \tag{21}$$

then

$$\lim_n \frac{1}{n\delta}\log\|\wedge^r\, \hat{A}^\delta_n\cdots\hat{A}^\delta_1\| = \lambda_1^\sigma + \cdots + \lambda_r^\sigma \quad a.s.$$

Stationarity, ergodicity and integrability of $\{\hat{A}_n^\delta\}$ follow from those of the continuous time process, so the assumptions of Oseledec's theorem hold, and there exists a Lyapunov spectrum for (21) denoted $\{\lambda_i^{\sigma,\delta}\}_{i=1}^d$. The relation between the spectra is $\frac{1}{\delta}\lambda_i^{\sigma,\delta} = \lambda_i^\sigma$ and thus by (20),

$$\gamma_\sigma = \frac{1}{\delta}(\lambda_2^{\sigma,\delta} - \lambda_1^{\sigma,\delta}).$$

It is useful to consider here the well–known representation of the solution to the Zakai equation as

$$\rho^{p_0}(t) = L_t f_t$$

where

$$L_t = \mathrm{diag}\{\exp \sigma^{-2}(h_i y_t - \frac{1}{2}\, h_i^2 t)\}_{i=1}^d$$

and $f : \mathbb{R}^+ \mapsto \mathbb{R}^d$ is a $C^1[0,\infty)$ function satisfying

$$\begin{cases} \dot{f}_t = L_t^{-1}\hat{G}^* L_t f_t \\ f_0 = p_0. \end{cases}$$

Denote by $\{A_i^\delta\}$ the matrices for which $f_{n\delta} = A_n^\delta f_{(n-1)\delta}$. Now, by property 1 , $\tau(\hat{A}_1^\delta) = \tau(L_\delta A_1^\delta) = \tau(A_1^\delta)$. As U_t is a homeomorphism, $\hat{A}_1^\delta$ is invertible and so is A_1^δ, which is thus also allowable. One therefore has by Lemma 1

$$\gamma_\sigma \leq \frac{1}{\delta} E\, \log \tau(A_1^\delta),$$

and by Fatou's lemma,

$$\gamma_\sigma \leq \limsup_{\delta \to 0} \frac{1}{\delta} E\, \log \tau(A_1^\delta) \leq E\, \limsup_{\delta \to 0} \frac{1}{\delta} \log \tau(A_1^\delta). \tag{22}$$

Since f_t belongs to $C^1[0,\infty)$, it follows that

$$f_\delta = (I + L_0^{-1}\hat{G}^* L_0 \delta + M^\delta) f_0,$$

where M^δ is a $d \times d$ matrix with $\|M^\delta\| = o(\delta)$. It suffices to prove the theorem for $\hat{G}$ for which $\forall i,j,\ i \neq j,\ g_{ij} > 0$. Under this condition, ψ may be written as

$$\psi(A_\sigma^\delta) = \psi(I + \hat{G}^*\delta + M^\delta) = \min_{1 \leq i,j,k,l \leq d} \frac{(1_{\{i=j\}} + g_{ji}\delta + m_{ij}^\delta)(1_{\{l=k\}} + g_{kl}\delta + m_{lk}^\delta)}{(1_{\{i=k\}} + g_{ki}\delta + m_{ik}^\delta)(1_{\{l=j\}} + g_{jl}\delta + m_{lj}^\delta)}$$

where $m_{ij}^\delta = (M^\delta)_{ij}$. There exists a δ_0 such that for every $0 < \delta < \delta_0$, the minimum is achieved on $i = k \neq l = j$, thus

$$\psi(A_\sigma^\delta) = \min_{i,j: i \neq j} g_{ij} g_{ji} \delta^2 + o(\delta^2)$$

and

$$\psi^{1/2}(A_\sigma^\delta) = \min_{i,j: i \neq j} (g_{ij} g_{ji})^{1/2} \delta + o(\delta),$$

and thus

$$\frac{1}{\delta}\log\tau(A_\sigma^\delta)=\frac{1}{\delta}\log\frac{1-\psi^{1/2}(A_\sigma^\delta)}{1+\psi^{1/2}(A_\sigma^\delta)}=\frac{1}{\delta}[-2\min_{i,j:i\neq j}(g_{ij}g_{ji})^{1/2}\delta+o(\delta)]\longrightarrow_{\delta\to 0}-2\min_{i,j:i\neq j}(g_{ij}g_{ji})^{1/2},$$

and the result follows from inequality (22). □

Proof of Theorem 7: It seems natural to approach the continuous time case as a limit of the discrete time problem. Note however that a change in the order of limits is needed to carry out this approach, and justifying this change of order seems challenging. We thus take below a different route. Although the general idea is similar to the discrete time case, extra care is needed due to the fact that trajectories of the x process do not possess positive probability, and an appropriate version of the Feynman-Kac formula is needed.

The first part of the theorem is a direct consequence of the following three lemmas, whose proof is deferred:

Lemma 2 *Assume (A2) holds. Then* $\limsup_{\sigma\to 0}\sigma^2\lambda_1^\sigma\leq\frac{1}{2}\,Eh_{x_0}^2$.

Lemma 3 *Assume (A2) holds. Then* $\limsup_{\sigma\to 0}\sigma^2(\lambda_1^\sigma+\lambda_2^\sigma)\leq\frac{1}{2}Eh_{x_0}^2+Eh_{x_0}h_{\mathrm{nbr}(x_0)}-\frac{1}{2}Eh_{\mathrm{nbr}(x_0)}^2$.

Lemma 4 *Assume (A2) holds. Then* $\liminf_{\sigma\to 0}\sigma^2\lambda_1^\sigma\geq\frac{1}{2}\,Eh_{x_0}^2$.

Given Lemma 2 above, the proof of (6) is similar to the proof of (3) in the discrete time setup, with trace $\hat{G}$ playing the role of $\log|\det G|$ there. □

Proof of Lemma 2: Using (A2), and denoting by e_i the unit vectors in $\mathbb{R}^d$, it holds that $\cos(e_i,p_{\mathrm{stat}}^x)\geq c>0$ for some c independent of t. Therefore, since U_t is nonnegative, and using c_1 to denote another positive deterministic constant independent of t, $\|U_tp_{\mathrm{stat}}^x\|\geq\min_i\cos(e_i,p_{\mathrm{stat}}^x)\max_i\|U_te_i\|\geq c_1\|U_t\|$ and

$$\lambda_1^\sigma=\lim_{t\to\infty}\frac{1}{t}E\,\log\|U_t\|\leq\lim_{t\to\infty}\frac{1}{t}E\log\|U_tp_{\mathrm{stat}}^x\|\ .$$

Let $\tilde{x}$ denote the realization of the x process under E_0, initialized at the stationary measure. Then, by the Kallianpur-Striebel formula and Oseledec's theorem,

$$\lambda_1^\sigma\leq\lim_{t\to\infty}\frac{1}{t}E\log E_0\left[\exp\left(\sigma^{-2}(\int_0^t h(\tilde{x}_s)dy_s-\frac{1}{2}\,h^2(\tilde{x}_s)ds)\right)|y_0^t\right]\ . \tag{23}$$

Fix $\delta>0$ and define $\Delta_iy=y_{(i+1)\delta}-y_{i\delta}$, $|\Delta y|_{\max}^{i,\delta}=\max\{|y_t-y_{t'}|\,:\,t,t'\in[i\delta,(i+1)\delta)\}$; let $\{\tau_i\}$ be the jumping times of $\{\tilde{x}_t\}$, $|\Delta y|_{\max}^{\tau_i,\delta}=\max\{|y_t-y_{t'}|\,:\,t,t'\in[\tau_i-\delta,\tau_i+\delta)\}$. Define similarly $\Delta_i\nu,|\Delta\nu|_{\max}^{i,\delta}$, $|\Delta\nu|_{\max}^{\tau_i,\delta}$, and let $h_{\max}=\max_i|h_i|$. Let $i_t=[t/\delta]$ and $N_t=\max\{i:\tau_i\leq t\}=\#\{\tau_i\leq$

$t\}$. We control the integral in (23) by its discrete time skeleton, with errors occuring only around jump times. That is,

$$\begin{aligned}
&\int_0^t \left(h(\tilde{x}_s)dy_s - \frac{1}{2}\, h^2(\tilde{x}_s)ds\right) && (24)\\
&\leq \sum_{i=0}^{i_t} \left(h(\tilde{x}_{i\delta})\Delta_i y - \frac{1}{2}\, h^2(\tilde{x}_{i\delta})\delta\right) + \sum_{i=1}^{N_t} (2h_{\max}|\Delta y|_{\max}^{\tau_i,\delta} + {h_{\max}}^2\delta)\\
&\leq \frac{1}{2} \sum_{i=0}^{i_t} \frac{(\Delta_i y)^2}{\delta} + \sum_{i=1}^{N_t} \left[2h_{\max}\left(2h_{\max}\delta + \sigma|\Delta\nu|_{\max}^{\tau_i,\delta}\right) + h_m^2\delta\right] .
\end{aligned}$$

Thus, by Jensen's inequality,

$$\begin{aligned}
&\frac{1}{t} E \log E_0 \left[\exp\left(\sigma^{-2}\int_0^t h(\tilde{x}_s)dy_s - \frac{1}{2}\, h^2(\tilde{x}_s)ds\right) |y_0^t\right]\\
&\leq \frac{1}{t}\frac{1}{2\sigma^2\delta} E\sum_{i=0}^{i_t}(\Delta_i y)^2 + \frac{1}{t}\log E_0\left[\exp\left(\sigma^{-2}\sum_{i=1}^{N_t}(5{h_{\max}}^2\delta + 2h_{\max}\sigma|\Delta\nu|_{\max}^{\tau_i,\delta})\right)\right] . && (25)
\end{aligned}$$

On the other hand, using stationarity,

$$\begin{aligned}
&E(\Delta_i y)^2 = E(\int_0^\delta h(x_s)ds + \sigma\nu_\delta)^2 = E(\int_0^\delta h(x_s)ds)^2 + \sigma^2\delta\\
&\leq E_s h^2(x_0)\delta^2 + \sigma^2\delta + E\left[1_{\{x_t \text{ jumps in } [0,\delta)\}}(2h_{\max}\delta)^2\right]\\
&= E_s h^2(x_0)\delta^2 + \sigma^2\delta + \delta^2 C'_\delta . && (26)
\end{aligned}$$

with $C'_\delta \underset{\delta \to 0}{\longrightarrow} 0$.

Conditioning on N_t, one has

$$\begin{aligned}
\frac{1}{t}\log E_0 \exp\left(\sigma^{-2}\sum_{i=1}^{N_t}\left(5{h_{\max}}^2\delta + 2h_{\max}\sigma|\Delta\nu|_{\max}^{\tau_i,\delta}\right)\right) &\leq \frac{1}{t}\log E_0\exp(N_t c\sigma^{-2}\delta)\\
&\leq \frac{1}{t}\log\sum_{n=0}^{\infty}\exp(nc\sigma^{-2}\delta)c\frac{(\mu t)^n}{n!}e^{-\mu t}\\
&= \frac{1}{t}\log c + \mu(e^{c\delta/\sigma^2} - 1)\,, && (27)
\end{aligned}$$

where $\mu = \max_i \sum_{j\neq i}\hat{G}_{ij}$. Combining (23), (25), (26) and (27),

$$\sigma^2\lambda_1^\sigma \leq \frac{1}{2}\, E_s h^2(x_0) + \frac{\sigma^2}{2\delta} + \frac{C'_\delta}{2} + \sigma^2\mu\left(e^{c\delta/\sigma^2} - 1\right) . \qquad (28)$$

Now take $\sigma^2/\delta = \epsilon$ and $\delta, \sigma \to 0$, then take infimum over $\{\epsilon > 0\}$ to get

$$\limsup_{\sigma\to 0}\sigma^2\lambda_1^\sigma \leq \frac{1}{2}\, E_s h^2(x_0)\,. \qquad \square$$

Proof of Lemma 3: We use the same notations as in Lemma 2. Let $d\rho_t = \hat{G}^* \rho_t dt + \sigma^{-2} H \rho_t dy_t$, $d\eta_t = \hat{G}^* \eta_t dt + \sigma^{-2} H \eta_t dy_t$ (the difference between ρ_t and η_t lies in possibly different initial conditions). In the sequel, we suppress the index t. Write $\rho \wedge \eta = \frac{1}{2}(\rho\eta^* - \eta\rho^*)$, then

$$
\begin{aligned}
d\rho\eta^* &= \hat{G}^* \rho\eta^* dt + \sigma^{-2} H\rho\eta^* dy_t + \rho\eta^* \hat{G} dt + \sigma^{-2} \rho\eta^* H\, dy_t + \sigma^{-2} H\rho\eta^* H\, dt\,, \qquad (29)\\
d(\rho\wedge\eta) &= \left[\hat{G}^*(\rho\wedge\eta) - \left(\hat{G}^*(\rho\wedge\eta)\right)^*\right] dt + \sigma^{-2}\left[H(\rho\wedge\eta) - (H(\rho\wedge\eta))^*\right] dy_t + \sigma^{-2} H(\rho\wedge\eta) H\, dt\,.
\end{aligned}
$$

Let the $(d-1)d$ - dimensional vector

$$
\underline{\alpha} = \begin{pmatrix} \alpha_{12} \\ \alpha_{13} \\ \vdots \\ \alpha_{1d} \\ \alpha_{21} \\ \alpha_{23} \\ \vdots \\ \alpha_{d(d-1)} \end{pmatrix}
$$

be defined by

$$
\rho \wedge \eta = \begin{pmatrix} 0 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1d} \\ \alpha_{21} & 0 & \alpha_{23} & \cdots & \alpha_{2d} \\ \vdots & & \ddots & & \\ \alpha_{d1} & & & & 0 \end{pmatrix},
$$

then (29) can be written as

$$
d\underline{\alpha} = \bar{G}^* \,\underline{\alpha}\, dt + \sigma^{-2} \bar{H}_1 \,\underline{\alpha}\, dt + \sigma^{-2} \bar{H}_2 \,\underline{\alpha}\, dy_t \qquad (30)
$$

with $\bar{G}(i,j) \geq 0$ for all $i \neq j$,

$$
\bar{H}_1 = \begin{pmatrix} h_1 h_2 & & & \\ & h_1 h_3 & & \\ & & \ddots & \\ & & & h_d h_{d-1} \end{pmatrix}, \quad \bar{H}_2 = \begin{pmatrix} h_1 + h_2 & & & \\ & h_1 + h_3 & & \\ & & \ddots & \\ & & & h_d + h_{d-1} \end{pmatrix}.
$$

We may regard now $\underline{\alpha}$ as a $d(d-1)$-dimensional vector indexed by ij with $i \neq j$. Viewed this way, the matrix $\bar{G}$ has off-diagonal entries

$$
\bar{G}_{ij,\ell m} = \begin{cases} \hat{G}_{\ell i} & j = m \neq \ell \\ \hat{G}_{mj} & i = \ell \neq m \\ 0 & j \neq m \,\text{and}\, i \neq \ell \end{cases}
$$

The matrix $\bar{G}$ is not necessarily a transition-rate matrix. However, there exists a transition-rate matrix $\tilde{G}$ which is equal to $\bar{G}$ off the diagonal. Thus (30) may be written:

$$
d\underline{\alpha} = \tilde{G}^* \,\underline{\alpha}\, dt + \sigma^{-2} \tilde{H}_1 \,\underline{\alpha}\, dt + \sigma^{-2} \tilde{H}_2 \,\underline{\alpha}\, dy_t\,, \qquad (31)
$$

with $\tilde{G}^* + \sigma^{-2}\tilde{H}_1 = \bar{G}^* + \sigma^{-2}\bar{H}_1$, $\tilde{H}_2 = \bar{H}_2$. It follows that

$$\tilde{H}_1 = \begin{pmatrix} h_1h_2 + \sigma^2\Delta g_{12} & & & \\ & h_1h_3 + \sigma^2\Delta g_{13} & & \\ & & \ddots & \\ & & & h_dh_{d-1} + \sigma^2\Delta g_{d(d-1)} \end{pmatrix}.$$

Note that while we are primarily interested in solutions to (31) which are in the anti-symmetric subspace $\underline{\alpha}_{ij} = -\underline{\alpha}_{ji}$, (31) makes perfect sense for arbitrary vectors in $\mathbb{R}^{d(d-1)}$. This point of view is particularly useful when computing upper bounds on Lyapunov exponents.

We now use $\tilde{h}_i(j\,k)$ ($\bar{h}_i(j\,k)$) to denote the $j\,k$-th element on the diagonal of $\tilde{H}_i$ (respectively, $\bar{H}_i$), $i = 1, 2$. Let $S = \{jk : j, k \in \{1, \dots, d\}, j \neq k\}$. Associate to the Markovian generator $\tilde{G}$ the S-valued process $\{\tilde{x}_t\}$, independent of $\{x_0^t\}$ and of $\{y_0^t\}$. We now introduce an auxiliary assumption on $\tilde{G}$, which will be relaxed later on.

(A3) Let (A2) hold. In addition, for every i, j, i', j', $i \neq j, i' \neq j'$, there exist sequences $\{\alpha_k\}_{k=1}^{\ell}$, $\{\beta_k\}_{k=1}^{\ell}$, taking values in $\{1, \dots, d\}$, such that: $(\alpha_1, \beta_1) = (i, j), (\alpha_\ell, \beta_\ell) = (i', j'), \alpha_k \neq \beta_k, \hat{G}_{\alpha_k, \alpha_{k+1}} \neq 0, \hat{G}_{\beta_k, \beta_{k+1}} \neq 0$ and $\{\alpha_{k+1} = \alpha_k \text{ or } \beta_{k+1} = \beta_k\}$. (We refer to the existence of such sequences by saying that the path $ij \to i'j'$ exists).

Note that (A3) implies that all states in S are communicating under $\tilde{G}$. (A3) trivially holds (for $d \geq 3$) if $\hat{G}_{ij} \neq 0$ for all i, j, but (A3) is not implied by (A2), as the example $d = 3, \hat{G}_{13} = \hat{G}_{31} = 0$ shows.

By the stochastic Feynman-Kac formula of nonlinear filtering, (using, e.g., an argument similar to Lemma 2.1 of [10]), if $\underline{\alpha}_{ij}(t = 0) = P(\tilde{x}_0 = (ij))$ then

$$\langle \underline{\alpha}, \mathbf{1} \rangle = \sum_{i,j} \underline{\alpha}_{ij} = E_0 \left[\exp\left(\sigma^{-2}(\int_0^t \tilde{h}_2(\tilde{x}_s) dy_s - \frac{1}{2}\, \tilde{h}_2^2(\tilde{x}_s) ds + \tilde{h}_1(\tilde{x}_s) ds) \right) | y_0^t \right].$$

Let A_t denote the linear map $\underline{\alpha}_0 \to \underline{\alpha}_t$. Let $p_{\text{stat}}^{\tilde{x}}$ denote the stationary distribution of $\tilde{x}$, which by (A3) has all entries strictly positive. Mimicking the argument used in the proof of Lemma 2, one has by positivity and Oseledec's theorem that

$$\begin{aligned} \lambda_1^\sigma + \lambda_2^\sigma &= \lim_{t\to\infty} \frac{1}{t} \log \sup_{\{\rho_0, \eta_0 : ||\rho_0 \wedge \eta_0|| = 1\}} \| \rho_t \wedge \eta_t \| \leq \lim_{t\to\infty} \frac{1}{t} E \log ||A_t|| \qquad (32) \\ &\leq \lim_{t\to\infty} \frac{1}{t} E \log \| A_t p_{\text{stat}}^{\tilde{x}} \| \\ &= \lim_{t\to\infty} \frac{1}{t} E \log E_0 \left[\exp\left(\sigma^{-2}(\int_0^t \tilde{h}_2(\tilde{x}_s) dy_s - \frac{1}{2}\, \tilde{h}^2(\tilde{x}_s) ds + \tilde{h}_1(x_s) ds) \right) | y_0^t \right]. \end{aligned}$$

Define $\hat{n} : \mathbb{R} \to \{1, \dots, d\}$ by $\hat{n}(a) = \text{argmin}_i |h_i - a|$. Then there exists a constant $r_0 > 0$ such that if $|h_i - a| \leq r_0$ then $\text{argmin}_{i \neq \hat{n}(a)} |h_i - a| = \text{nbr}(\hat{n}(a))$. Now, denoting $g_m = \max |\Delta g(\cdot)|$,

$$J_t := \int_0^t \tilde{h}_2(\tilde{x}_s) dy_s - \frac{1}{2} \int_0^t \tilde{h}_2^2(\tilde{x}_s) ds + \int_0^t \tilde{h}_1(\tilde{x}_s) ds \qquad (33)$$

$$
\begin{aligned}
&= \int_0^t \bar h_2(\tilde x_s)dy_s + \int_0^t \sigma^2\Delta g(\tilde x_s)dy_s - \int_0^t \frac{1}{2}\,\bar h_2^2(\tilde x_s)ds + \int_0^t \bar h_2(\tilde x_s)\sigma^2\Delta g(\tilde x_s)ds \\
&\quad - \int_0^t \frac{1}{2}\,\sigma^4\Delta g^2(\tilde x_s)ds + \int_0^t \bar h_1(\tilde x_s)ds \\
&\le \sum_{i=0}^{i_t}\left\{\bar h_2(\tilde x_{i\delta})\Delta_i y - \frac{1}{2}\,\bar h_2^2(\tilde x_{i\delta})\delta + \bar h_1(\tilde x_{i\delta}) + \sigma^2 g_m|\Delta y|_{\max}^{i,\delta} + 2h_{\max}\sigma^2 g_m\delta\right\} \\
&\quad + \sum_{i=1}^{N_t}(2h_{\max}|\Delta y|_{\max}^{\tau_i,\delta} + 2{h_{\max}}^2\delta) \le J_t^1 + J_t^2
\end{aligned}
$$

where

$$
\begin{aligned}
J_t^1 &= \sum_{i=0}^{i_t}\delta\left\{\left[h\left(\hat n\left(\frac{\Delta_i y}{\delta}\right)\right) + h\left(\mathrm{nbr}\left(\hat n\left(\frac{\Delta_i y}{\delta}\right)\right)\right)\right]\frac{\Delta_i y}{\delta}\right. \\
&\quad \left. -\frac{1}{2}\,h^2\left(\hat n\left(\frac{\Delta_i y}{\delta}\right)\right) - \frac{1}{2}\,h^2\left(\mathrm{nbr}\left(\hat n\left(\frac{\Delta_i y}{\delta}\right)\right)\right)\right\} + \\
&\quad + \sum_{\{i\le i_t:\,|\frac{\Delta_i y}{\delta} - h_j| > r_0\ \forall j\}}\left\{4h_{\max}|\Delta_i y|_{\max}^{i,\delta} + h_m^2\delta\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
J_t^2 &= \sum_{i=1}^{i_t}\left\{\sigma^2 g_m(2h_{\max}\delta + \sigma|\Delta\nu|_{\max}^{i,\delta}) + 2h_{\max}g_m\sigma^2\delta\right\} \\
&\quad + \sum_{i=1}^{N_t}\left\{2h_{\max}\left(2h_{\max}\delta + \sigma|\Delta\nu|_{\max}^{\tau_i,\delta}\right) + 2{h_{\max}}^2\delta\right\}.
\end{aligned}
$$

Having J_t^1 measurable with respect to y_0^t, it follows using Jensen's inequality that

$$
\frac{1}{t}E\log E_0\left[\exp\left(\sigma^{-2}J_t\right)|y_0^t\right] \le \frac{1}{t}\frac{1}{\sigma^2}EJ_t^1 + \frac{1}{t}\log E_0\exp\left(\sigma^{-2}J_t^2\right). \tag{34}
$$

Now,

$$
\frac{1}{t}EJ_t^1 \le \frac{1}{\delta}E_s\left\{(h(x_0) + h(\mathrm{nbr}(x_0)))(\delta h(x_0) + \sigma\Delta\nu) - \frac{\delta}{2}h^2(x_0) - \frac{\delta}{2}h^2(\mathrm{nbr}(x_0))\right\} + c\delta + ce^{-r_0^2\delta/2\sigma^2}, \tag{35}
$$

where the second term is due to the fact that the probability of having a jump in the x process on any δ-interval is of order δ, and the last term is due to the Gaussian law of ν. Next,

$$
\begin{aligned}
\frac{1}{t}\log E_0\left(\exp\sigma^{-2}J_t^2\right) &\le \frac{1}{t}\log E_0\,E_0\left[\exp(\sigma^{-2}J_t^2)|N_t\right] \\
&\le \frac{1}{t}\log E_0\,E_0\left[\exp\left\{\sum_{i=1}^{N_t}\left(c\delta + c\sigma|\Delta\nu|_{\max}^{i,\delta}\right) + \sum_{i=1}^{N_t}c\delta\sigma^{-2} + c\sigma^{-1}|\Delta\nu|_{\max}^{\tau_i\delta}\right\}|N_t\right] \\
&= \frac{1}{t}\log E_0\exp\left(\frac{t}{\delta}c\delta + \frac{t}{\delta}c^2\sigma^2\delta + N_t c\delta\sigma^{-2} + N_t c^2\sigma^{-2}\delta\right)
\end{aligned}
$$

$$
\begin{aligned}
&= c + c^2\sigma^2 + \frac{1}{t}\log E\, \exp cN_t\delta\sigma^{-2} \\
&\le c + c^2\sigma^2 + \lambda(e^{c\delta/\sigma^2} - 1)\,.
\end{aligned} \tag{36}
$$

Finally, combining (34), (35) and (36),

$$
\begin{aligned}
\sigma^2(\lambda_1^\sigma + \lambda_2^\sigma) &\le E_s\left[(h(x_0) + h(\mathrm{nbr}(x_0)))h(x_0) - \frac{1}{2}\,h^2(x_0) - \frac{1}{2}\,h^2(\mathrm{nbr}(x_0))\right] + c\delta \\
&+ ce^{-r_0^2\delta/2\sigma^2} + \sigma^2 c + \sigma^4 c^2 + \sigma^2 c\Big(e^{c\delta/\sigma^2} - 1\Big)\,.
\end{aligned}
$$

Take now $\sigma^2/\delta = \epsilon$, $\delta, \sigma \to 0$, then take $\epsilon \to 0$ to conclude the Lemma under (A3).

As mentioned above, (A3) is not implied by (A2). In order to prove the lemma under (A2) alone, note first that the above proof can be repeated assuming only that $\tilde G$ possesses a stationary distribution (not necessarily unique) with strictly positive entries, that is, $\tilde x$ has no transient states. We claim that the latter fact is actually implied by (A2). It suffices to show that for every two states $ij,\ kl \in S$, if $(\tilde G^m)_{ij,kl} > 0$ for some m, then there exists an n such that $(\tilde G^n)_{kl,ij} > 0$, or, in the terminology we use in the sequel, if there is a path $ij \to kl$ then the path $kl \to ij$ exists. Note next that it suffices to show the above for $j = l$ and for $i, k \neq j$ such that $\hat G_{ik} > 0$ (that is $i \to k$ in one step). Suppose then, that $ij \to kj$ in one step. It needs to be shown that $kj \to ij$. If there exists a path $k \to i$ that does not contain j, then the claim is proved. Otherwise, since $\hat G$ is communicating, there exists a path $k \to j \to i$ such that $k \to j$ does not contain i and $j \to i$ does not contain k. Thus the following path exists too:

$$
kj \to ki \to ji \to jk \to ik \to ij,
$$

and it follows that $\tilde x$ has no transient states. This concludes the proof of the lemma under (A2). $\square$

Proof of Lemma 4:

The top Lyapunov exponent satisfies

$$
\lambda_1^\sigma = \lim_{t\to\infty} \frac{1}{t} E\log\|U_t\| \ge \lim_{t\to\infty}\frac{1}{t}E\log\|U_t p^x_{\mathrm{stat}}\|\,.
$$

Again, we compute $\|U_t p^x_{\mathrm{stat}}\|$ using the Kallianpur-Striebel formula. Fix a $\delta > 0$ and let I_i be the interval $[\delta j, \delta(j+1))$ such that $\tau_i \in I_i$. Then, denoting, as above, by $\tilde x$ a copy of the process x which is independent of y_0^t under E_0,

$$
\begin{aligned}
&\frac{1}{t}E\log E_0\left[\exp\left(\sigma^{-2}\Big(\int_0^t h(\tilde x_s)dy_s - \frac{1}{2}\int_0^t h^2(\tilde x_s)ds\Big)\right)|y_0^t\right] \\
&\ge \frac{1}{t}E\log E_0\left[E_0\left[1_{\{\tilde x_s = x_s\,\forall s\notin\bigcup_i I_i,\, s<t\}}\exp\left(\sigma^{-2}\Big(\int_0^t h(\tilde x_s)dy_s - \frac{1}{2}\int_0^t h^2(\tilde x_s)ds\Big)\right)|x_0^t, y_0^t\right]|y_0^t\right] \\
&\ge \frac{1}{t}E\log E_0\left[E_0\left[1_{\{\tilde x_s = x_s\,\forall s\notin\bigcup_i I_i,\, s<t\}}\exp\left(\sigma^{-2}\,B_t(x,y)\right)|x_0^t, y_0^t\right]|y_0^t\right]
\end{aligned}
$$

where

$$B_t(x,y) = \int_0^t h(x_s)dy_s - \frac{1}{2}\int_0^t h^2(x_s)ds - \sum_i (c|\Delta y|_{\max}^{\tau_i,\delta} + c\delta).$$

Thus, by Jensen's inequality,

$$\frac{1}{t}E\log\|U_t p_{\text{stat}}^x\| \geq \frac{1}{t}E\log E_0\left[1_{\{\tilde{x}_s = x_s\ \forall s\notin\bigcup_i I_i,\, s<t\}}\exp\left(\sigma^{-2}B_t(x,y)\right)|x_0^t, y_0^t\right]$$
$$= \frac{1}{t}\log E\left[1_{\{\tilde{x}_s = x_s\ \forall s\notin\bigcup_i I_i,\, s<t\}}|x_0^t\right] + \frac{1}{t}E\sigma^{-2}B_t(x,y).$$

Now,

$$\begin{aligned}
&\frac{1}{t}E\log E\left[1_{\{\tilde{x}_s = x_s\ \forall s\notin\bigcup_i I_i,\, s<t\}}|x_0^t\right]\\
&= \frac{1}{t}E\log E\left[1_{\{\tilde{x}_{j\delta} = x_{j\delta}\ \forall j<t/\delta\ s.t.\ j\delta\notin\bigcup_i I_i\}}|x_0^t\right]\\
&\geq \frac{1}{t}E\log E\left[1_{\{\tilde{x}_{j\delta} = x_{j\delta}\ \forall j<t/\delta\}}|x_0^t\right].
\end{aligned}$$

The last quantity tends to $-\frac{1}{\delta}\mathcal{H}(\{x_{j\delta}\})$ where $\mathcal{H}(\{x_{j\delta}\})$ is the entropy rate for $\{x_{j\delta}\}$. Moreover,

$$\lim_{t\to\infty}\frac{1}{t}EB_t(x,y) = \frac{1}{2}\,Eh^2(x_0) - c\delta,$$

and thus we have shown

$$\sigma^2\lambda_1^\sigma \geq \frac{1}{2}\,Eh^2(x_0) - c\delta - \frac{\sigma^2}{\delta}\log d.$$

Now taking $\sigma\to 0$ and then $\delta\to 0$ yields the result. $\square$

Remark: The vector $\rho\wedge\eta$ has only $(d-1)d/2$ degrees of freedom, the same as the vector $\underline{\alpha}$. We have used a $(d-1)d$ dimensional vector $\underline{\alpha}$ in order to write (30) as a matrix with non-negative off-diagonal entries. For general (non-filtering) situations, this might not be possible.

Acknowledgement We thank P. Bougerol for bringing [7] to our attention.

References

[1] R. W. Brockett, *Nonlinear Systems and Nonlinear Estimation Theory,* in *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* pp. 441–477, edited by M. Hazwinkel and J. C. Willems, D. Reidel Publishing Company, Dordrecht, 1981.

[2] R. Carmona, J. Lacroix, *Spectral Theory of Random Schrödinger Operators,* Birkhäuser, 1990.

[3] J. E. Cohen, H. Kesten, C. M. Newman, *Oseledec's Multiplicative Ergodic Theorem: a Proof,* in *Random Matrices and Their Applications, Contemporary Math.,* Vol 50 pp. 23–30, edited by J. E. Cohen, H. Kesten, C. M. Newman, Am. Math. Soc., Providence, 1986.

[4] B. Delyon, O. Zeitouni, *Memory Length of Optimal Filters,* Applied Stochastic Analysis, edited by M. H. A. Davis and R. J. Elliot, Gordon and Breach Science Publishers, 1991.

[5] H. Kunita, *Stochastic Partial Differential Equations Connected with the Nonlinear Filtering,* Lecture Notes in Mathematics 972, Springer, Berlin, pp. 101–168.

[6] D. Ocone, E. Pardoux, *Asymptotic Stability of the Optimal Filter with respect to its Initial Conditions*, preprint.

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

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

[9] W. M. Wonham, *Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering,* SIAM J. Control, 2, 1965, pp. 347–368.

[10] O. Zeitouni, B. Z. Bobrovsky, *On the Reference Probability Approach to the Equations of Non-Linear Filtering,* Stochastics, Vol. 19, 1986, pp. 133 – 149.

AMS 1991 Subject classification: Primary 93E11. Secondary 60J57.
KEYWORDS: Lyapunov exponents. Nonlinear filtering. Wonham's equation. Feynman–Kac formula.