APRIL 1988 LIDS-P-1764

LYAPUNOV EXPONENTS FOR FILTERING PROBLEMS

by

Bernard Delyon
Ofer Zeitouni

Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, MA 02139

ABSTRACT

The dependence of the optimal nonlinear filter on it's initial conditions is considered for continuous time linear filtering and for finite state space nonlinear filtering. Partial results are obtained in the high signal to noise ration case, together with a characterization of the Lyapunov exponent in the (easier) low SNR case.

This work was supported in part by the Air Force Office of Scientific Research under grant AFOSR-85-0227B and by the Weizmann Poctdoctoral Fellowship.

1. INTRODUCTION

In this note, we consider the dependence of the conditional density in the nonlinear filtering problem on the initial a-priori distribution of the state. From a practical point of view, one is often interested in knowing how long does he have to wait to reach near optimality when initiating the optimal filter with the wrong initial conditions.

Our purpose in this note is mainly to expose the problem. Despite our efforts, we have at the moment only very partial results, mainly of an asymptotic nature, which will be explained below. Even those results, however, exhibit interesting features: it turns out that, at the limit of high signal to noise ratios, structural properties of the process and mainly of its observations dominate the memory length.

The models we consider are of one of the following types:

(a) **Finite state space in white noise.** The state process x is a continuous time, ergodic, stationary Markov chains with values $\{1,2,...,k\}$, and with an infinitesimal generator $G = \{g_{ij}\}_{1\leq i,j\leq k}$, where if $P_{ij}(\varepsilon) = P(x(t+\varepsilon) = j \mid (x(t) = i)$, then

$$P_{ii}(\varepsilon) = 1 + g_{ii}\varepsilon + o(\varepsilon) \tag{1.1}$$

$$P_{ij}(\varepsilon) = g_{ij}\varepsilon + o(\varepsilon) \quad j \neq i$$

The observation process $\{y_t, t\geq 0\}$ satisfies

$$dy_t = h(x_t)dt + N_0^{1/2}db_t \tag{1.2}$$

where b_t is a one dimensional Brownian motion independent of x_t and $h = \{h(i)\}_{1\leq i\leq k}$ is a given vector.

(b) **Rational Gaussian process in white noise.** The state process $x_t \in \mathbf{R}^n$ satisfies the linear, stochastic differential equation

$$dx_t = Ax_tdt + Bdw_t, \;\; p(x_0) = N(m_0, P_0) \tag{1.3}$$

with w_t being an n-dimensional Brownian motion, A,B given constant matrices of appropriate dimensions, and $p(x_0)$, the initial a-priori density of x_0, is normal with mean m_0 and covariance matrix P_0. The observation process $y_t \varepsilon \mathbf{R}^m$ satisfies

$$dy_t = Cx_t dt + N_0^{1/2} db_t \tag{1.4}$$

where here b_t is an m-dimensional Brownian motion independent of x_t and C is again a constant given matrix.

(We remark that a natural candidate for model instead of (1.3), (1.4) is the general diffusion nonlinear filtering problem; a remark concerning it can be found in the end of section 3).

We define now the notion of "memory length". In both cases, let

$$p_t^{p_0}(x) \triangleq P_{p_0}(x_t = x | y_s,\ 0 \le s \le t) \tag{1.5}$$

denote the conditional distribution (density in case b) of x_t, given an initial a-priori distribution (density in case b) p_0.

The "memory length" γ is defined as follows:

$$\text{(case a)} \quad \gamma \triangleq \sup_{p_0, p_0'} \limsup_{t \to \infty} \frac{1}{t} \log \| P_t^{p_0} - P_t^{p_0'} \| \tag{1.6a}$$

where $\| \ \|$ denotes the Euclidean norm and

$$\text{(case b)} \quad \gamma \triangleq \sup_{m_0, m_0' \varepsilon K} \limsup_{t \to \infty} \frac{1}{t} \log \| \hat{x}_t(m_0', p_0) - \hat{x}_t(m_0, p_0) \| \tag{1.6b}$$

where $x_t(m_0, p_0)$ is the conditional mean starting from $N(m_0, p_0)$ and K is some compact set. In both cases, γ is a reminiscent of the usual definition of Lyapunov exponents. In (1.6b), we consider x_t because, the conditional density being Gaussian, it best characterizes the distribution. We could also allow P_0 to change, for brevity we do not do that here, even though the analysis is exactly the same.

In both cases (a), (b), as $N_0 \to \infty$, γ will tend to the closest to zero (in real part) negative eigenvalue of G in case a and pole in case b. As $N_0 \to 0$, ("high signal to noise"), surprisingly, $\gamma \to -\infty$ necessarily (this is however the case when

$n=m=1$, $C\neq 0$ in case (b) and $k=2$, $h_1\neq h_2$ is case (a)). For case a, we provide an example where $\gamma \to 0$ as $N_0 \to 0$, whereas in case (b), the complete analysis of section 3 shows that if the "transfer function" of (1.3), (1.4) possesses zeroes on the imaginary axis, $\gamma \to 0$ as $N_0 \to 0$, and otherwise $\gamma \to -\infty$ as $N_0 \to 0$. Thus, structural properties of the system involved determine it's "forgiveness" of initial mistakes, even in high signal to noise ratios!

The remain of the paper in organized as follows: in section 2, case a is presented; the analysis as $N_0 \to \infty$, an example with $\gamma \to 0$ as $N_0 \to 0$ and the k=2 cases are considered. The difficulties in the general case are also pointed out. In Section 3, the Gaussian case b is presented, with the full asymptotic analysis of $N_0 \to 0$.

Acknowledgements. We wish to thank Prof. A. Willsky for many fruitful discussions.

2. FINITE STATE SPACE CASE

In this section, we consider case a. We start by considering the 'high-noise" behavior:

Theorem 1: If the process x_t is ergodic, there exist two positive constants ε and K_0 such that , for $N_0 > K_0$,

$$\lim_{t \to \infty} \frac{1}{t} \log \| p_t^{p_0} - p_t^{p_0'} \| \leq - \varepsilon \qquad \text{a.e.} \tag{2.1}$$

Moreover, as $N_0 \to \infty$, $\varepsilon \to \lambda_{max}(G)$, the largest non zero real part of the eigenvalues of G.

Proof. Let H denote the diagonal matrix with $H_{ii} = h(i)$. Then, from Wonham [1],

$$dp_t = p_t G dt - N_0^{-1} \langle p_t, h \rangle p_t (H - \langle p_t, h \rangle I) dt$$
$$+ N_0^{-1} p_t (H - \langle p_t, h \rangle I) dy_t \tag{2.2}$$

Note that for proving (2.1), it is enough to consider the equation satisfied by the derivative q_t of p_t with respect to the initial conditions p_0 in any direction $d = (d_1 ... d_k)$. From (2), one has:

$$\begin{aligned} dq_t &= q_t G\, dt - N_0^{-1} \langle q_t, h \rangle p_t (H - \langle p_t, h \rangle I) dt \\ &\quad - N_0^{-1} \langle p_t, h \rangle q_t (H - \langle p_t, h \rangle I) dt + N_0^{-1} \langle p_t, h \rangle p_t \langle q_t, h \rangle dt \\ &\quad + N_0^{-1} q_t (H - \langle p_t, h \rangle I) dy_t - N_0^{-1} p_t \langle q_t, h \rangle dy_t = \\ &= q_t G\, dt - N_0^{-1} \langle q_t, h \rangle p_t (H - \bar{h} I) dt + N_0^{-1/2} q_t (H - \bar{h}\, I) d\nu \\ &\quad - N_0^{-1/2} \langle q_t, h \rangle p_t d\nu_t \end{aligned}$$

where $d\nu_t = \dfrac{dy_t - \langle p_t, h \rangle dt}{N_0}$ is the innovation white noise and $\bar{h} = \langle h, p_t \rangle$.

$$dq_t = q_t(G + N_0^{-1} A_t)dt + N_0^{-1/2} q_t B_t dv_t \qquad (2.3)$$
$$q_0 = d$$

where A_t and B_t are two matrix valued measurable processes, and there exists a constant α depending only on h such that:

$$\|A_t\| < \alpha$$
$$\|B_t\| < \alpha \qquad \text{all } t, N_0$$

almost surely ($\| \ \|$ denotes the usual norm of matrices).

Note that the hyperspace $E = \{q, \langle q,1\rangle = 0\}$ (where $1 = (1,....,1)^T$) is stable under G, A_t and B_t because

$$G1 = 0$$

$$A_t 1 = (hp_t)(H - \langle p_t,h\rangle I)1 = (hp_t)(h - \langle p_t, h\rangle 1) = 1$$

and if $q \in E$:

$$qB_t 1 = q(H - \langle p_t,h\rangle I)1 - \langle q,h\rangle\langle p_t,1\rangle$$
$$= \langle q,h\rangle - \langle q,h\rangle = 0$$

But the constraint on p_0:

$$\sum p_0(i) = 1$$

implies that q_0 has to be choosen in E, and q_t remains in E. Choosing any fixed orthogonal base in E, (2.3) can be rewritten as:

$$dq'_t = q'_t(G' + N_0^{-1} A'_t)dt + N_0^{-1/2} q'_t B'_t dv_t \qquad (2.4)$$

where $q'_t = (q'_t(1),...,q'_t(k-1))$ is a representation of q_t in this base; G', A'_t and B'_t are the matrices associated to the restriction to E of the applications represented by G, A_t and B_t in the whole space. Moreover, the spectrum of G' is the spectrum of G without the eigenvalue zero, and then, by the assumption of ergodicity, all the eigenvalues of G' have negative real part.

Let S be a symetric positive matrix and denote:

$$r_t = q'_t S q'^T_t$$
$$u_t = r_t^{-1/2} q'_t$$

then:

$$\begin{aligned} dr_t &= q'_t(G'S + SG'^T + N_0^{-1}(A'_tS + SA'^T_t))q'^T_t\, dt \\ &+ N_0^{-1/2} q'_t (B'_tS + SB'^T_t) q'^T_t dv_t \\ &+ N_0^{-1} q'_t B'_t SB'^T_t q'^T_t dt \\ &= r_t u_t (G'S + S^T G'^T + N_0^{-1}(2A'_tS + B'_tSB'^T_t))u^T_t\, dt \\ &+ 2N_0^{-1/2} r_t u_t B'_t S u^T_t dv_t \end{aligned}$$

using Ito's formula, we get:

$$\begin{aligned} d \log r_t = u_t(G'S &+ S^T G'^T + N_0^{-1}(2A'_tS + B'_tSB'^T_t))\, u^T_t\, dt \\ &- 2N_0^{-1}(u_t B'_t S u_t)^2 dt + 2N_0^{-1/2} u_t B'_t S u^T_t dv_t \end{aligned}$$

and then:

$$\begin{aligned} \frac{1}{t}\log r_t \le \frac{1}{t} \log r_0 &+ \frac{1}{t}\int_0^t u_s(G'S + SG'^T)u^T_s\, ds \\ &+ \frac{1}{t} N_0^{-1} \int_0^t u_s(2A'_s S + B'_sSB'^T_s)u^T_s\, ds \\ &+ \frac{1}{t} 2N_0^{-1/2} \int_{-0}^t u_s B'_s S u^T_s\, dv_s \end{aligned} \tag{2.5}$$

Let λ be the real part of the largest eigenvalue of G' and choose $\mu > \lambda$, then the matrix $H = G' - \mu I$ is still a stability matrix and there exist symetric S such that:

$$HS + SH^T = -I$$

for this choice of S, the second term of (2.5) is bounded by 2μ and the third term is smaller than $N^{-2}_0 c_1(S)$ (where $c_1(S)$ is a constant depending only on S). For the last term, consider the time change:

$$\tau_t = \int_0^t (u_s B'_0 S u_s^T)^2 ds$$

then , we know that there exists a brownian motion B_t such that:

$$\int_0^t u_s B'_s S u_s^T dv_s = B_{\tau_t}$$

and then :

$$\frac{1}{t} \left| \int_0^t u_s\ B'_s\ Su_s^T\ dv_s \right| = \frac{1}{t} \left| B_{\tau_t} \right| \leq c_2(s) \left| \frac{B_{\tau_t}}{\tau_t} \right|$$

where $c_2(S)$ is constant depending only on S. This proves that the last term tends to zero as t tends to infinity (if τ_t is bounded, use the last equality, if τ_t tends to infinity, use the last inequality). Finally, we get

$$\limsup_{t \to \infty} \frac{1}{t} \log r_t \leq 2\mu + N_0^{-1} c_1(S)$$

and then, by the equivalence of the norms of $\mathbf{R}^k$:

$$\limsup_{t \to \infty} \frac{1}{t} \log \| q_t \| \leq \mu + N_0^{-1} c_1(S)$$

which is arbitrarily close to λ if N_0 is large enough.

We consider next the case $N_0 \to 0$. For $k = 2$, equation (2.4) is one dimensional, $A'_t < 0$ and therefore,

$$\gamma \overset{N_0 \to 0}{\longrightarrow} -\infty \tag{2.6}$$

One is led to think that this situation is generic; our conjecture is that indeed it is, under suitable "structural conditions", which we don't know at this point to specify. The problem is that eq. (2.4) is a Bilinear equation with non-constant coefficients, and the known methods of computing the Lyapunov exponent fail. The following counter example demonstrates clearly that (2.4) does not hold in general without restrictions; actually, in this example, $\gamma \to 0$!

Consider the four state process with transition matrix:

$$G = \begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \\ 1 & & & -1 \end{pmatrix} \tag{2.7}$$

c.f. fig. 1. The observation vector is

$$h(x_1) = h(x_3) = h, \quad h(x_2) = h(x_4) = 0 \tag{2.8}$$

Note that the fact that $h(x_2) = 0$ is not significant, as the addition of a d.c. term to h(x) does not change the filter's structure. The fact that the observations are the same for different states is of extreme significant, as the following intuitive argument demonstrates: indeed, note that as $N_0 \to \infty$, by theorem 1, the Lyapunov exponent of the optimal filter converges to 1, and the conditional distribution converges to the stationary distribution regardless of the initial conditions. However, as $N_0 \to 0$, consider the two initial conditions:

(1,0,0,0) and (0,0,1,0)

The fact that $N_0 \to 0$ allows us to track accurately the transitions in the system, but reveals nothing as to the initial conditions and thus the state estimate will highly depend on p(0). Thus, we expect that the Lyapunov exponent will go to zero as $N_0 \to 0$; the following analysis will show that this is indeed the case. For simplicity, we make the change of variables:

$$\Delta(t) = \begin{pmatrix} \dfrac{p_1(t)}{p_1(t)+p_3(t)} - \dfrac{1}{2} \\ \dfrac{p_4(t)}{p_2(t)+p_4(t)} - \dfrac{1}{2} \end{pmatrix}, \text{ where } \left.\frac{p_1}{p_1+p_3}\right|_{p_1=p_3=0} = \frac{1}{2} \tag{2.9}$$

$\Delta(t)$ reflects the "mismatch" with the stationary solution $p_1 = p_2 = p_3 = p_4 = 1/4$, where $\Delta=0$; we analyze the Lyapunov exponent of $\Delta(t)$, which is easily seen to have the same behavior as that of $x(t)$. It is easily seen that v_t is the same under initial conditions d_1 or d_3 (which correspond to $\Delta_1(0) = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$, $\Delta_3(0) = \begin{pmatrix} -1/2 \\ 0 \end{pmatrix}$). Note however that, by an easy computation based on (2.2), $\Delta(t)$ satisfies:

$$\dot{\Delta}_1(t) = \begin{bmatrix} -\gamma_t & \gamma_t \\ -1/\gamma_1 & -1/\gamma_t \end{bmatrix} \Delta(t)$$

where

$$d\gamma_t = [1 + \left(\frac{h_1^2/N_0}{\gamma_t+1} - 1 \right) \gamma_t^2] dt - \frac{h_1}{N_0^{1/2}} \gamma_t dv_t$$

and γ_t is highly oscillatory. Analysing directly eq. (3.4) is rather difficult; fortunately enough, (3.5) is easy to simulate; for

$$\lambda \triangleq \frac{1}{t} \lim \log \|\Delta(t)\|,$$

the results are summarized in table 1, which agrees with the heuristic analysis above.

N_0	10^4	1	0.2	0.1	0.05	0.02
λ	−1	−0.89	−0.53	−0.21	−0.085	−0.021

Table 1. Lyapunov exponent as function of $N_0^{1/2}$

3. THE GAUSSIAN CASE

In this section, we analyze the "memory length" of the optimal filter for the Kalman filtering problem (Case b). Surprisingly enough, there are cases where the memory length does not tend to zero when the signal to noise ratio is high, even in the fully observable case, c.f. below.

We assume throughout that the pair (A,C) is observable and that (A,B) is stabilizable (c.f. [2]).

Our results are summarized below:

Theorem 2:

a) For $N_0 \to \infty$, $\gamma \to \lambda_{max}(A)$, where $\lambda_{max}(A)$ is defined as in Theorem 1.

b) For $N_0 \to 0$, let $\phi(s) = \det(sI-A)$,

$$H(s) = C(sI-A)^{-1}B$$

Let

$$\Theta \triangleq \{\theta | \mathrm{Re}\ \theta \le 0,\ \phi(s)\,\phi(-s)\,\det[H^T(s)H(s)] = 0\}$$

Then $\gamma \to \theta_{max}(\Theta)$, where $\theta_{max}(\Theta)$ is the element of Θ with largest real part.

Proof. The optimal filter equations are (c.f. [5]):

$$d\hat{x} = Ax_t dt + K(t)[dy_t - C\hat{x}_t dt] \tag{3.1}$$

where $K(t) = P(t)\,C^T/N_0$ and

$$\dot{P}(t) = AP(t) + P(t)A^T + BB^T - \frac{PC^TCP}{N_0} \tag{3.2}$$

Under our assumptions, $P(t) \to P_\infty$. Note that (3.1) implies that, if we denote by $\hat{x}_t(x_0)$ and $\hat{x}_t(x_0')$ the output of the filter with $\hat{x}_0(x_0) = x_0$, $\hat{x}_0(x_0') = x_0'$ and by $\Delta_t \triangleq \hat{x}_t(x_0) - \hat{x}_t(x_0')$, then

$$\dot{\Delta}_t = A\Delta_t - \frac{P(t)C^TC}{N_0}\Delta_t$$

from which (a) follows from the boundedness of P(t) and $N_0 \to \infty$.

To see (b), we consider first the case of $P(0) = P_\infty$; in that case, (b) is a rephrasing of [2, theorem 4.13]. In the general case ($P_0 = 0$), let $T = \inf\{t \mid \left\| \frac{P(t)-P_\infty}{N_0} \right\| \, \|C^TC\| < \varepsilon \text{ for all } t > T\}$. For $t \geq T$ one has

$$\dot{\Delta}_t = \left(A - \frac{P_\infty C^T C}{N_0} \right) \Delta_t - \frac{(P(t) - P_\infty)}{N_0} C^T C \Delta$$

$$= (A - \frac{P_\infty}{N_0} C^T C)\Delta_t - \tilde{K}_t \Delta_t \tag{3.6}$$

and $\|\tilde{K}_t\| < \varepsilon$.

By the argument below (2.4) (taking there $B_t = 0$, $N_0^{-1} A' = K_t$ and $G' = A - \frac{P_\infty C^T C}{N_0}$), one obtains that

$$\lambda_{max}\left(A - \frac{P_\infty C^T C}{N_0} \right) - \varepsilon C_1$$

$$< \lambda < \lambda_{max}\left(A - \frac{P_\infty C^T C}{N_0} \right) + \varepsilon C_1$$

where C_1 depends on G' and is independent of ε and where $\lambda_{max}(A - \frac{P_\infty C^T C}{N_0})$ denotes the largest real part of the eigenvalues of $A - \frac{P_\infty C^T C}{N_0}$ which is negative by the stability of the optimal filter ([2]). Taking $\varepsilon \to 0$ leads to $\lambda = \lambda_{max}(A - \frac{P_\infty C^T C}{N_0})$. Taking now $N_0 \to 0$ yields the theorem. □

We remark that the theorem implies that even for a stable, controllable and fully observable system, the limiting Lyapunov exponent can approach zero even with "good measurements": simply, take a system with a transfer matrix zero on the imaginary axis.

A remark on the general nonlinear filtering problem for diffusions seems in order here: in many cases, the optimal nonlinear filter is well approximated when

$N_0 \to 0$ by a linear system: c.f. [3], [4]. In those cases, also the "memory length" γ will exhibit the behavior as above. We omit the details here.

REFERENCES

[1] Wonham, W.H., Some Application of Stochastic Differential Equations to Optimal Nonlinear Filtering - J. SIAM Control, Ser. A, Vol. 2, No. 3, 1965.

[2] Kwakernaak, H. and Sivan, R., Linear Optimal Control Systems, Wiley, 1972.

[3] Bobrovsky, B.Z. and Zakai, M., "A lower bound on the estimation error for Certain Diffusion Processes", IEEE Trans. Inf. Theory, Vol. IT-22, 1976, pp. 45-52.

[4] Zeitouni, O. and Dembo, A., "On the Maximal Achievable Accuracy in Nonlinear Filtering Problems", to appear, IEEE Trans. Aut. Control, 1988.

[5] Gelb, A., Applied Optimal Estimation. MIT Press, Cambridge, MA, 1974.