Exponential decay rate of the filter's dependence on the initial distribution

Rami Atar*

January 15, 2009 (Revised May 7, 2009)

Abstract: We review how tools from multiplicative ergodic theory and the theory of positive operators are used in the analysis of exponential stability of the optimal nonlinear filter. Particularly, in the case of finite state, we relate the filter sensitivity to perturbations in its initial data to the Lyapunov spectral gap associated with the filtering equation, and, in a general setting, use Hilbert's metric and Birkhoff's contraction coefficient to estimate the decay rate of the error.

1. Introduction

The problem of stability of the nonlinear filter arises in the following practical context. If the transition law of a given Markov process is known, but its initial law is not available, under what conditions does one not lose optimality of the filter when initializing it with an arbitrary (thus wrong) initial data, in the limit when time tends to infinity? This question, first posed by Ocone and Pardoux [35] and Delyon and Zeitouni [20], has attracted much attention. This paper focuses on the *exponential* rate of decay of the error made by wrong initialization, and reviews results that relate this quantity to multiplicative ergodic theory (MET) on one hand, and to Hilbert's metric and Birkhoff's contraction coefficient on the other hand. MET is instrumental in establishing that, in a finite state setting, the decay rate is deterministic (roughly speaking). In fact, it identifies the rate with the Lyapunov spectral gap associated with the filtering equation. The set of tools borrowed from the theory of positive operators, are more useful in providing estimates on the decay rate, which in turn enable to establish sufficient conditions for nonvanishing thereof.

Our main goal in this exposition is to present the methods in an elementary and reasonably self-contained way, starting from a simple case and then extending the ideas to a general setting. We make no attempt to present the strongest results to date, or to survey various other techniques to tackle the problem (such as [1, 4–6, 11, 14, 16, 19, 29, 33, 40]).

The paper is organized as follows. In Section 2 we begin by describing the finite state, discrete time setting and define the decay rate; we then review relevant results from MET and make the link to the Lyapunov spectral gap. We also describe analogous results in continuous time. The short Section 3 reviews definitions regarding positive operators and explains their relation to Lyapunov exponents, based on a lemma of Peres, which leads to first estimates on the decay rate. Section 4 presents bounds on the decay rate in a general state space.

*Research supported in part by the Israel Science Foundation (Grant 1349/08), and the Technion fund for promotion of research

Notation. Given a positive integer d we write $M(d)$ for the set of $d \times d$ matrices with real entries, and $M_+(d)$ for the set of elements of $M(d)$ with nonnegative entries. The set of elements of $\mathbb{R}^d$ with nonnegative entries is denoted by $\mathbb{R}^d_+$. Vectors in $\mathbb{R}^d$ are understood to be column vectors unless otherwise specified. The ith entry of a vector x is denoted by x^i, and the (i,j)th entry of a matrix M is $M^{i,j}$. $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are the usual scalar product and norm on $\mathbb{R}^d$. For $M \in M(d)$, $\|M\| = \sup\{\|Mx\| : x \in \mathbb{R}^d, \|x\| = 1\}$ is the operator norm corresponding to $\| \cdot \|$. The transpose of a matrix M is $M^\top$. Expectation with respect to a probability measure denoted by $\mathbb{P}^A_B$ is written as $\mathbb{E}^A_B$, for any set of symbols A and B to be used (particularly $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$).

2. Finite state filtering and Lyapunov exponents

Let d be a positive integer and set $\bar{d} := \{1, 2, \ldots, d\}$. Denoting $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^d$, and

$$\mathcal{P} = \{x \in \mathbb{R}^d : x_i \geq 0, \langle x, \mathbf{1} \rangle = 1\}, \tag{1}$$

we will identify members of $\mathcal{P}$ with probability distributions over $\bar{d}$ via $(p^i)_{i \in \bar{d}} = (p(\{i\}))_{i \in \bar{d}}$. Consider a homogeneous Markov process $X = \{X_n, n \geq 0\}$ taking values in $\bar{d}$, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by G the transition matrix

$$G^{i,j} = P(X_{n+1} = j | X_n = i), \qquad i, j \in \bar{d},\ n \geq 0,$$

and by $\pi_0 \in \mathcal{P}$ the initial distribution, regarded as a column vector. Next, fix $\ell \in \mathbb{N}$ and denote $\mathcal{R} = \mathcal{B}(\mathbb{R}^\ell)$. Let a family $\widetilde{G}(i, \mathrm{d}y)$, $i \in \bar{d}$ of probability measures on $(\mathbb{R}^\ell, \mathcal{R})$ be given, and assume that for some measurable function $g : \bar{d} \times \mathbb{R}^\ell \to \mathbb{R}$ and a probability measure $\widetilde{G}_0 \in \mathcal{M}(\mathbb{R}^\ell, \mathcal{R})$,

$$\widetilde{G}(i, \mathrm{d}y) = g(i, y) \widetilde{G}_0(\mathrm{d}y), \qquad i \in \bar{d}.$$

We are given a process $\{Y_n, n \geq 1\}$ of noisy observations of X_n, satisfying

$$\mathbb{P}(Y_i \in E_i, i \in \{1, 2, \ldots, n\} | X_i = x_i, i \in \{0, 1, \ldots, n\}) = \prod_{i=1}^n \widetilde{G}(x_i, E_i),$$

for all $n \in \mathbb{N}$, $E_1, E_2, \ldots, E_n \in \mathcal{R}$ and $x_1, x_2, \ldots, x_n \in \bar{d}$. For simplicity we assume in this section that g takes positive values. X_n and Y_n are called the *state* and *observation* processes, respectively.

Example 2.1. For some $m \in \mathbb{N}$, $\widetilde{G}_0$ could be the uniform probability measure on the finite set $\{1, 2, \ldots, m\}$, and then $\widetilde{G}(i, \{j\}) = g^{i,j}$, where $\{g^{i,j}\}_{i,j}$ is a positive $d \times m$ matrix. Thus provided $X_n = i$, the conditional probability of the event $\{Y_n = j\}$ is given by $g^{i,j}$. ◇

Example 2.2. Let $\ell = 1$. Let $k : \bar{d} \to \mathbb{R}$ be a given function. We model Y_n as observations of X_n via the 'sensor' k, perturbed by Gaussian noise, by defining

$$Y_n = k(X_n) + \sigma W_n, \qquad n \geq 1. \tag{2}$$

Here $\{W_n, n \geq 1\}$ is an i.i.d. sequence of standard normals, independent of $\{X_n\}$, and the parameter $\sigma > 0$ is the level of noise. This model is seen to fit the above setup if we set $\widetilde{G}_0$ to be the $(0, \sigma^2)$-Gaussian measure on $\mathbb{R}$,

$$\widetilde{G}_0(\mathrm{d}y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \mathrm{d}y,$$

and let

$$\widetilde{G}(i,\mathrm{d}y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-k(i))^2}{2\sigma^2}} \mathrm{d}y.$$

This example will be referred to as the *one-dimensional additive Gaussian noise.* $\diamond$

We denote by $\mathcal{Y}_n$ the σ-algebra generated by the observations $\{Y_1, Y_2, \ldots, Y_n\}$, $n \geq 1$, and set $\mathcal{Y}_0$ to be the trivial σ-field. Let

$$\pi_n = (P(X_n = 1|\mathcal{Y}_n), \ldots, P(X_n = d|\mathcal{Y}_n))^\top, \qquad n \geq 0. \tag{3}$$

The stochastic process $\{\pi_n, n \geq 0\}$ takes values in $\mathcal{P}$. It represents the conditional law of X_n given $\mathcal{Y}_n$, and is often referred to as the *nonlinear filter* associated with the processes X_n and Y_n. A use of Bayes' rule shows that π_n satisfies the recursion

$$\pi_n = \frac{D_n G^\top \pi_{n-1}}{\langle D_n G^\top \pi_{n-1}, \mathbf{1}\rangle}, \qquad n \geq 1,$$

where D_n is the diagonal matrix

$$D_n^{i,i} = g(i, Y_n), \qquad i \in \bar{d}.$$

An equivalent way of writing this recursion is via

$$\rho_n = D_n G^\top \rho_{n-1}, \qquad n \geq 1, \qquad \rho_0 = \pi_0, \tag{4}$$

in which case π_n is given by $\rho_n / \langle \rho_n, \mathbf{1}\rangle$. Thus the filter can be expressed as a normalized version of the solution to a simple linear recursion. The process ρ_n is often referred to as the *unnormalized conditional measure* of X_n given $\mathcal{Y}_n$. Often in practice, the initial law π_0 is not available, and one uses the recursion (3) with wrong initial data. Given $p \in \mathcal{P}$, let π_n^p denote the solution to the recursion with initial data p. We will use the term *exact filter* for $\pi_n^{\pi_0}$, and refer to π_n^p as the *filter initialized at* p. As posed by Ocone and Pardoux [35], it is interesting to ask whether the filter "overcomes" the error made by choosing wrong initial data, as $n \to \infty$ (a question much related to earlier work by Kunita [24], [25] and Stettner [38], [39] on convergence in law of the exact filter to a unique measure, under suitable ergodic assumptions about the state process). Delyon and Zeitouni [20] suggested that, because of the multiplicative form of (4), it is natural to ask when the error resulting from wrong initialization decays exponentially, and to study the rate of decay via multiplicative ergodic theory (MET). Denote by d_{TV} the total variation distance between measures, and note that one has $d_{TV}(p,q) = \|p-q\|_1$ where $\|\cdot\|_1$ denotes ℓ_1 norm and one uses the identification alluded to above. Let

$$\gamma(p,q) = \limsup_{n\to\infty} \frac{1}{n} \log \|\pi_n^p - \pi_n^q\|_1. \tag{5}$$

The main point of this section is to recall results based on MET, stating that the quantity γ is, loosely speaking, deterministic and independent of p and q. If $\gamma < 0$ then the filter can be said to be *exponentially stable with respect to perturbations in the initial condition.* Also, it is natural to interpret $-1/\gamma$ as the memory length of the filter, and thus it is useful to quantify γ.

We will need some basic results from MET (the reader is referred to [15, 17, 28] for further reading on the subject). Let $\{T_n, n \geq 1\}$ be a stationary ergodic sequence of $d \times d$ matrices, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying $\mathbb{E}[\log^+ \|T_1\|] < \infty$. Denote $T_{(n)} = T_n T_{n-1} \cdots T_1$. Oseledec's theorem states that there exist deterministic constants

$$-\infty \leq \lambda_d \leq \lambda_{d-1} \leq \cdots \leq \lambda_1 < \infty,$$

and a full $\mathbb{P}$-measure event, Ω_1, on which the following holds.

(i) The sets

$$V(i) := \{x \in \mathbb{R}^d : \lim \frac{1}{n} \log \|T_{(n)} x\| \leq \lambda_i\}$$

are subspaces, and $\dim V(i) = \#\{j : \lambda_j \leq \lambda_i\}$ *(in particular,* $V(1) = \mathbb{R}^d$*).*

(ii) With $V(d+1) = \{0\}$*, one has for every* $x \in \mathbb{R}^d \setminus \{0\}$

$$\lim_{n\to\infty} \log \|T_{(n)} x\| = \lambda_i,$$

where i is the unique j for which $x \in V(j) \setminus V(j+1)$.

(iii) The sequence of matrices $(T_{(n)}^\top T_{(n)})^{1/(2n)}$ *converges to a matrix T whose eigenvalues are* $e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_d}$*. For $i \in \bar{d}$ such that $V(i) \neq V(i+1)$, the orthogonal complement of $V(i+1)$ in $V(i)$ is the eigenspace of T corresponding to* e^{λ_i}*.*

The constants λ_i are called the *Lyapunov exponents* associated with $\{T_n\}$, under $\mathbb{P}$.

One defines the ith exterior power $\wedge^i M$ of a matrix M as the linear operator on the ith exterior power of $\mathbb{R}^d$, for which

$$\wedge^i M(e_{j_1} \wedge e_{j_2} \wedge \cdots \wedge e_{j_i}) = (Me_{j_1}) \wedge (Me_{j_2}) \wedge \cdots \wedge (Me_{j_i})$$

(see e.g. [10]). The only two facts that we need about exterior products here are, first, that given i vectors $x_1, x_2, \ldots, x_i \in \mathbb{R}^d$, the quantity $\bar{v}(x_1, x_2, \ldots, x_i) := \|x_1 \wedge x_2 \wedge \cdots \wedge x_i\| = (\det[\{\langle x_j, x_k\rangle\}_{j,k}])^{1/2}$ equals the i-dimensional volume of the parallelogram generated by these vectors; particularly, $\bar{v}(x_1, x_2)^2 = \|x_1 \wedge x_2\|^2 = \|x_1\|^2 \|x_2\|^2 - \langle x_1, x_2\rangle^2$. And second, that (as in fact a corollary of Oseledec's theorem) the following holds $\mathbb{P}$-a.s.

$$\lim_{n\to\infty} \frac{1}{n} \log \| \wedge^i T_{(n)}\| = \sum_{j=1}^{i} \lambda_j, \qquad i \in \bar{d}, \tag{6}$$

where $\|\cdot\|$ is used here to denote the corresponding operator norm.

For a nonzero vector x in $\mathbb{R}^d_+$, write $\bar{x}$ for $x/\langle x, \mathbf{1}\rangle$. For x and y such vectors,

$$\bar{x} - \bar{y} = \frac{x}{\langle x, \mathbf{1}\rangle} - \frac{y}{\langle y, \mathbf{1}\rangle} = \frac{\|y\|_1 x - \|x\|_1 y}{\|x\|_1 \|y\|_1},$$

hence

$$\|\bar{x} - \bar{y}\|_1 \leq \frac{\sum_{i,j=1}^d |x^i y^j - y^i x^j|}{\|x\|_1 \|y\|_1}.$$

Also,

$$\bar{v}(x, y)^2 = \|x \wedge y\|^2 = \|x\|^2 \|y\|^2 - \langle x, y\rangle^2 = \frac{1}{2} \sum_{i,j} (x^i y^j - x^j y^i)^2,$$

which gives

$$\|\bar{x} - \bar{y}\|_1 \le c_d \frac{\bar{v}(x,y)}{\|x\|_1 \|y\|_1}. \tag{7}$$

This inequality will help us establish a bound on $\gamma(p,q)$ in term of exponential growth rates of the three objects $\bar{v}(\rho_n^p, \rho_n^q)$, $\|\rho_n^p\|_1$ and $\|\rho_n^q\|_1$. To see that the latter are quite simple to quantify by Lyapunov exponents, consider first an arbitrary sequence $\{T_n\}$ of matrices from $M_+(d)$, satisfying the assumptions of Oseledec's theorem. Let $x, y \in \mathcal{P}$, and denote $x_n = T_{(n)}x$ and $y_n = T_{(n)}y$. By the Perron-Frobenius theorem, the matrix $T_{(n)}^\top T_{(n)} \in M_+(d)$ has an eigenvector $u_n \in \mathbb{R}_+^d$ corresponding to the largest eigenvalue, say μ_n. Consequently $\mu_n^{1/(2n)}$ and u_n are eigenvalue and eigenvector of $(T_{(n)}^\top T_{(n)})^{1/(2n)}$, and thus by item (iii) of Oseledec's theorem, the limit matrix T has an eigenvector $u_0 \in \mathbb{R}_+^d$ corresponding to its largest eigenvalue, e^{λ_1}. If i is such that $V(i) \neq V(1) = \mathbb{R}^d$ then it follows from item (iii) that $V(i)$ is orthogonal to u_0, hence, provided that x is strictly positive, $x \notin V(i)$. Thus by item (ii), for all such x, one has $\mathbb{P}$-a.s.,

$$\lim \frac{1}{n} \log \|x_n\| = \lambda_1.$$

Combining this with (6) and (7), we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \|\bar{x}_n - \bar{y}_n\|_1 \le \lambda_1 + \lambda_2 - 2\lambda_1 = \lambda_2 - \lambda_1, \tag{8}$$

provided $x, y \in \mathcal{P}_P := \{z \in \mathcal{P} : z^i > 0, i \in \bar{d}\}$. $\lambda_1 - \lambda_2$ is often referred to as the *spectral gap.*

This analysis is not directly applicable to the filtering situation described above because the matrix process is not necessarily stationary. Of course if $\{X_n\}$ is a stationary ergodic process then so is the matrix process $T_n := D_n G^\top$. We will assume that $\{X_n\}$ is an irreducible aperiodic chain. Thus there exists a unique invariant measure for the chain, π_S, and we denote by $\mathbb{P}_S$ the corresponding measure on $(\Omega, \mathcal{F})$. In the special case where π_0 equals π_S, the sequence $\{T_n\}$ is stationary and ergodic. For general π_0 we let the term *the Lyapunov exponents associated* $\{T_n\}$ mean the Lyapunov exponents associated with $\{T_n\}$ under $\mathbb{P}_S$. The result (8) developed above can in fact be improved in the following way [2]:

Theorem 2.1. *Assume that the chain $\{X_n\}$ is irreducible and aperiodic. Assume $\mathbb{E}_S[\log^+ \|D_1 G^\top\|] < \infty$. Let U denote the uniform measure on $\mathcal{P}$. Then $\mathbb{P}$-a.s., for $U \times U$-a.e. (p,q),*

$$\gamma(p,q) = \lambda_2 - \lambda_1,$$

where $\{\lambda_i\}$ are the Lyapunov exponents associated with $\{D_n G^\top\}$. Moreover, $\mathbb{P}$-a.s., we have for every $(p,q) \in \mathcal{P} \times \mathcal{P}$

$$\gamma(p,q) \le \lambda_2 - \lambda_1.$$

Remark 2.1. This result is proved in [2, Section 2] for the case of additive Gaussian noise, but the proof holds in the generality presented here. Also, the second assertion of Theorem 2.1 above is written in [2] in a slightly weaker way, namely that for every p and q, we have, $\mathbb{P}$-a.s., $\gamma(p,q) \le \lambda_2 - \lambda_1$, but the form presented here is valid according to the same proof (indeed, a review of that proof shows that, in the claim made in (8) and (9) of [2], the full $\mathbb{P}$-measure event on which the inequality holds is what we have denoted by Ω_1 in the above statement of Oseledec's Theorem, which in particular does not depend on the initial conditions p and q).

For a continuous time analog of this result, consider a Markov process X taking values in $\bar{d}$ with intensity matrix $\hat{G}$, and initial distribution π_0. Let the observation process be given by

$$Y_t = \int_0^t k(X_s)\mathrm{d}s + \sigma W_t,$$

where W is a standard Brownian motion independent of X and denote by π_t the conditional law of X_t given $\mathcal{Y}_t := \sigma\{Y_s : s \in [0,t]\}$. Denoting by K the diagonal $d \times d$ matrix with $K^{i,i} = k(i)$, and letting $\{\rho_t\}$ be the unique solution to the stochastic differential equation

$$\mathrm{d}\rho_t = \hat{G}^\top \rho_t \mathrm{d}t + \sigma^{-2} K \rho_t \mathrm{d}Y_t, \qquad t \geq 0, \qquad \rho_0 = \pi_0,$$

one has $\pi_t = \rho_t / \langle \rho_t, \mathbf{1}\rangle$ (see e.g. [41, equation (5)] and use Ito's lemma). Note that this can be written in terms of the $M_+(d)$-valued process $\{T_t\}$ solving

$$\mathrm{d}T_t = \hat{G}^\top T_t \mathrm{d}t + \sigma^{-2} K T_t \mathrm{d}Y_t, \qquad t \geq 0, \qquad T_0 = I, \tag{9}$$

where $I \in M(d)$ is the identity matrix. Namely, one has $\rho_t = T_t \pi_0$. For general initial data, $p \in \mathcal{P}$, set $\rho_t^p = T_t p$ and $\pi_t^p = \rho_t^p / \langle \rho_t^p, \mathbf{1}\rangle$.

Oseledec theorem has an analogue in continuous time. We present here the version [15, Section IV.2]. Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be endowed with a semigroup $\{\theta_t, t \geq 0\}$ of measure preserving transformations. Let $\{T_t, t \geq 0\}$ be a process on this space, taking values in $GL(d, \mathbb{R})$ (the group of linear automorphisms of $\mathbb{R}^d$). Then $\{T_t\}$ is said to be *multiplicative* if $T_0 = I$, and $T_{t+s} = (T_s \circ \theta_t) T_t$, for all $s, t \geq 0$. Assume that $\{\theta_t\}$ is ergodic, $\{T_t\}$ is a separable multiplicative process, and that $\mathbb{E}[\sup_{t \in [0,1]} \log^+ \|T_t^k\|] < \infty$ for both $k = 1$ and $k = -1$. Then items (i) and (ii) of the discrete version of the theorem, that appears above, hold upon replacing $T_{(n)}$ by T_t, $\frac{1}{n}$ by $\frac{1}{t}$, and $\lim_n$ by $\lim_t$. The way this result is used in the present context is by considering the standard shift transformation. As in the discrete time case, where the process may not be stationary, the shift transformation may not be measure preserving; however, under an irreducibility assumption, it is so in the case when π_0 equals the invariant measure π_S. For general π_0 we use the same convention and define Lyapunov exponents for the non-stationary process to be the ones for the stationary counterpart.

Analogously to (5) let

$$\gamma(p,q) = \limsup_{t \to \infty} \frac{1}{t} \log \|\pi_t^p - \pi_t^q\|_1. \tag{10}$$

The assumptions of the above version of the MET can be verified, and one has the following.

Theorem 2.2. *[2, 20] Assume that the process X is an irreducible continuous time Markov chain on $\bar{d}$. Then the conclusions of Theorem 2.1 hold for γ of (10) and the Lyapunov exponents $\{\lambda_i\}$ associated with the process (9).*

The above results assert that the decay rate γ is, roughly stated, deterministic and independent of the initial condition, and identify it with the (negative) Lyapunov spectral gap. The question of exponential stability can thus be posed as that of determining whether the gap is positive. Unfortunately, the Lyapunov spectrum is in general hard to calculate, if not impossible. However, one can sometimes obtain bounds on the gap from which such information can be extracted. This will be the point of the next section.

3. Hilbert's projective metric in finite state space

We present here Hilbert's projective metric, and its contraction properties under the action of positive matrices, used to obtain a bound on the Lyapunov spectral gap. This will enable us to attain conditions under which the gap, and in view of last section's results, the decay rate, is nonzero. It will also give rise to quantitative information on the gap.

We need some notation regarding positive matrices (we follow Seneta [37]). A matrix that is an element of $M_+(d)$ is said to be *allowable* if it contains no columns or rows whose entries are all zero. *Hilbert's projective metric* is the mapping $h : \mathcal{P}_P^2 \to \mathbb{R}_+$ defined by

$$h(x, y) = \log \max_{1 \le i,j \le d} \frac{x_i y_j}{x_j y_i}. \tag{11}$$

An allowable $d \times d$ matrix M can be seen, by normalization of the action of M, as an operator $M : \mathcal{P}_P \to \mathcal{P}_P$. We denote by $M.x$ its action on $x \in \mathcal{P}_P$. This definition turns out to be very useful mainly due to the fact that h makes any allowable matrix a contraction. Namely, $\tau(M) \le 1$ where τ is the *Birkhoff contraction coefficient* of an allowable matrix M, defined by

$$\tau(M) = \sup \Big\{ \frac{h(M.x, M.y)}{h(x, y)} : x, y \in \mathcal{P}_P, x \ne y \Big\}. \tag{12}$$

An explicit formula for τ in terms of M is available [37], namely

$$\tau(M) = \frac{1 - \sqrt{\psi(M)}}{1 + \sqrt{\psi(M)}}, \qquad \text{where} \qquad \psi(M) = \min_{i,j,k,l} \Big\{ \frac{M^{i,k} M^{j,l}}{M^{i,l} M^{j,k}} : M^{i,l} M^{j,k} \ne 0 \Big\}. \tag{13}$$

Here are two additional elementary properties. As follows directly from (11) and (12), if D is a diagonal matrix with $D^{i,i} > 0$ for $i \in \bar{d}$ then $\tau(MD) = \tau(DM) = \tau(M)$. If $M \in M_+(d)$ is a matrix whose entries are all positive, then, by (13), $\tau(M) < 1$.

We borrow the following from [36].

Lemma 3.1. *Let $\{T_n\}$ be a stationary ergodic sequence of nonnegative, allowable matrices, and assume $\mathbb{E}[\log^+ \|T_1\|] < \infty$. Let λ_1 and λ_2 denote the top two Lyapunov exponents associated with the sequence. Then*

$$\lambda_1 - \lambda_2 \ge -\mathbb{E}[\log \tau(T_1)],$$

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

Combined with Theorems 2.1 and 2.2, this gives a direct relation between the decay rate and the contraction coefficient, not involving the Lyapunov spectrum.

Corollary 3.1. *Under the assumptions of Theorem 2.1, $\mathbb{P}$-a.s., for every $p, q \in \mathcal{P}$, $\gamma(p, q) \le \mathbb{E}_S[\log \tau(T_1)]$, where $T_1 = D_1 G^\top$. Under the assumptions of Theorem 2.2, $\gamma(p, q) \le \mathbb{E}_S[\log \tau(T_1)]$, where T_t is given in* (9).

By the foregoing discussion on properties of τ, we have in the discrete setting $\tau(T_1) = \tau(D_1 G^\top) = \tau(G^\top) = \tau(G)$. If $G^{i,j} > 0$ for all i, j then $\tau(G) < 1$, and as a consequence we have the following.

Theorem 3.3. *[2] In the discrete time setting, assume that $G^{i,j} > 0$ for all $i, j \in \bar{d}$. Then $\mathbb{P}$-a.s., for all $p, q \in \mathcal{P}$, $\gamma(p, q) \le \log \tau(G) < 0$.*

This provides an estimate that depends only on the transition law of $\{X_n\}$. We refer to [16], this volume, for a bound that holds in greater generality, under which G may be a more general element of $M_+(d)$.

The approach is useful in obtaining estimates on the decay rate in the small noise (large signal-to-noise ratio) asymptotics. For the following result recall the setting of one-dimensional additive Gaussian noise (Example 2.2). For $i \in \bar{d}$ denote

$$\delta(i) = \min_{j \neq i} |k(i) - k(j)|.$$

Denote by γ_σ the negative spectral gap, emphasizing the dependence on σ.

Theorem 3.4. *[2] Consider the setting of Example 2.2. Then under the assumptions of Theorem 2.1,*

$$\limsup_{\sigma \to 0} \sigma^2 \gamma_\sigma \leq -\frac{1}{2} \mathbb{E}_S[\delta(X_1)^2]. \tag{14}$$

If, in addition, $\det G \neq 0$, *we have*

$$\liminf_{\sigma \to 0} \sigma^2 \gamma_\sigma \geq -\frac{1}{2} \mathbb{E}_S \Big[\sum_{i=1}^{d} (k(X_1) - k(i))^2 \Big]. \tag{15}$$

Note that δ is not identically zero if and only if there exists at least one i for which $k(i)$ is distinct from $k(j)$, all $j \neq i$. Thus the upper bound presented above is meaningful only under this condition; and when the condition holds, the combination of both bounds establish that the order of magnitude of γ_σ is σ^{-2}.

Although in general it is an open question whether either the upper or lower bounds can be improved, let us mention that (14) holds with equality when one assumes that X_n is a 'nearest neighbor' process, in the following sense: the transition matrix is given by $G = \exp(sA)$ for some $s > 0$ and A is an intensity matrix for which $|i - j| > 1$ implies $A(i, j) = 0$ (some additional technical conditions are required; see [3, Theorem 5]).

Analogous bounds hold in continuous time.

Theorem 3.5. *[2] Consider the filtering problem in continuous time, and let the assumptions of Theorem 2.2 hold. Then* $\mathbb{P}$*-a.s., for all* $p, q \in \mathcal{P}$,

$$\gamma(p, q) \leq -2 \min_{i,j \in \bar{d}: i \neq j} \sqrt{\hat{G}^{i,j} \hat{G}^{j,i}}.$$

Moreover, the following bounds hold:

$$\limsup_{\sigma \to 0} \sigma^2 \gamma_\sigma \leq -\frac{1}{2} \mathbb{E}_S[\delta(X_0)^2], \qquad \liminf_{\sigma \to 0} \sigma^2 \gamma_s \geq -\frac{1}{2} \mathbb{E}_S \Big[\sum_{i=1}^{d} (k(X_0) - k(i))^2 \Big].$$

4. Hilbert's projective metric in general state space

The fact established in Corollary 3.1 deserves a deeper look. We will see that this result has an easy proof not involving Lyapunov exponents or MET, valid in fact in far greater generality.

Thus in what follows we shall study the filtering problem in a general setting, and present an extended definition of h, and present an extended version of Corollary 3.1.

Let $\mathbb{S}$ be a Polish space, and let $\mathcal{S}$ denote the corresponding Borel σ-field. Fix a positive integer ℓ. We will introduce a Markov process X_n taking values in $\mathbb{S}$ and an observation process Y_n taking values in $\mathbb{R}^\ell$. We will, in fact, present a Markovian family. To this end, assume we are given a probability kernel $G : \mathbb{S}^2 \to \mathbb{R}$ (that is, for every $\alpha \in \mathbb{S}$, $G(\alpha, \cdot)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$ and for every $E \in \mathcal{S}$, $G(\cdot, E)$ is a measurable map). Also, we are given a probability kernel $\widetilde{G} : \mathbb{S} \times \mathbb{R}^\ell \to \mathbb{R}$. We define for every $\alpha \in \mathbb{S}$ a probability measure $\mathbb{P}^{(\alpha)}$ on $(\Omega, \mathcal{F})$ via

$$\begin{aligned}&\mathbb{P}^{(\alpha)}(X_1 \in E_1, X_2 \in E_2, \dots, X_n \in E_n, \quad Y_1 \in F_1, Y_2 \in F_2, \dots, Y_n \in F_n)\\ &= \int_{E_1\times\cdots\times E_n\times F_1\times\cdots\times F_n} G(\alpha, \mathrm{d}x_1) \prod_{i=2}^{n} G(x_{i-1}, \mathrm{d}x_i) \prod_{i=1}^{n} \widetilde{G}(x_i, \mathrm{d}y_i),\end{aligned}$$

for $n \in \mathbb{N}$, $E_i \in \mathcal{S}$, $F_i \in \mathcal{R}$, $i \le n$, where, throughout, we denote $\mathcal{R} = \mathcal{B}(\mathbb{R}^\ell)$. For $p \in \mathcal{M}(\mathbb{S}, \mathcal{S})$, let

$$\mathbb{P}^p = \int_{\mathbb{S}} \mathbb{P}^{(\alpha)} p(\mathrm{d}\alpha).$$

Fix a probability measure $\pi_0 \in \mathcal{M}(\mathbb{S}, \mathcal{S})$, and let $\mathbb{P} := \mathbb{P}^{\pi_0}$. Then under $\mathbb{P}$, X_n is a Markov process starting from initial measure π_0, and Y_n is an observation process. As before, we write $\mathcal{Y}_n$ for the σ-field generated by $(Y_1, Y_2, \dots, Y_n)$, $n \in \mathbb{N}$. The exact filter is thus

$$\pi_n(\varphi) = \mathbb{E}[\varphi(X_n)|\mathcal{Y}_n], \qquad n \ge 0.$$

We introduce a reference measure on $(\Omega, \mathcal{F})$. To this end we will need the assumption: *There exists a probability measure $\widetilde{G}_0$ on $(\mathbb{R}^\ell, \mathcal{R})$ and a measurable mapping $g : \mathbb{S} \times \mathbb{R}^\ell \to \mathbb{R}$ with respect to $\mathcal{S} \otimes \mathcal{R}$, such that, for every $\alpha \in \mathbb{S}$,*

$$\widetilde{G}(\alpha, F) = \int_F g(\alpha, y) \widetilde{G}_0(\mathrm{d}y), \qquad F \in \mathcal{R}.$$

Define for $\alpha \in \mathbb{S}$ the reference probability measure $\mathbb{P}_0^{(\alpha)}$

$$\begin{aligned}&\mathbb{P}_0^{(\alpha)}(X_1 \in E_1, X_2 \in E_2, \dots, X_n \in E_n, \quad Y_1 \in F_1, Y_2 \in F_2, \dots, Y_n \in F_n)\\ &= \int_{E_1\times\cdots\times E_n\times F_1\times\cdots\times F_n} G(\alpha, \mathrm{d}x_1) \prod_{i=2}^{n} G(x_{i-1}, \mathrm{d}x_i) \prod_{i=1}^{n} \widetilde{G}_0(\mathrm{d}y_i),\end{aligned}$$

for $n \in \mathbb{N}$, $E_i \in \mathcal{S}$, $F_i \in \mathcal{R}$, $i \le n$. Denote $\mathbb{P}_0^p = \int \mathbb{P}^{(\alpha)} p(\mathrm{d}\alpha)$ and $\mathbb{P}_0 = \mathbb{P}_0^{\pi_0}$. Consider the stochastic process $\{\Lambda_n\}$, $n \ge 0$, where $\Lambda_0 = 1$ and

$$\Lambda_n = \prod_{i=1}^{n} g(X_i, Y_i), \qquad n \ge 1.$$

Then clearly $\mathbb{P}^p(B) = \int_B \Lambda_n \mathrm{d}\mathbb{P}_0^p$, for $B \in \sigma\{X_i, Y_i, i \le n\}$. A use of Bayes rule shows

$$\pi_n(\varphi) = \frac{\mathbb{E}_0[\varphi(X_n)\Lambda_n|\mathcal{Y}_n]}{\mathbb{E}_0[\Lambda_n|\mathcal{Y}_n]}.$$

We thus let

$$\rho_n(\varphi) = \mathbb{E}_0[\varphi(X_n)\Lambda_n|\mathcal{Y}_n], \tag{16}$$

and note that $\pi_n = \rho_n/\rho_n(1)$. Toward writing a recursion for ρ_n, we select a regular conditional probability distribution

$$\mathbb{P}_0^{(\alpha)}[\,\cdot\,|\mathcal{Y}_n, X_n = \beta],$$

satisfying, for $A \in \mathcal{S}$, $B \in \sigma\{X_i, Y_i, i \le n\}$,

$$\mathbb{P}_0^{(\alpha)}[B \cap \{X_n \in A\}|\mathcal{Y}_n] = \int_A \mathbb{P}_0^{(\alpha)}(B|\mathcal{Y}_n, X_n = \beta)\mathbb{P}_0^{(\alpha)}(X_n \in \mathrm{d}\beta).$$

Letting

$$I_n(\alpha, \beta) = \mathbb{E}_0^{(\alpha)}[\Lambda_n|\mathcal{Y}_n, X_n = \beta],$$

we can write ρ_n (16) as $\rho_n^{\pi_0}$, where for $p \in \mathcal{M}(\mathbb{S}, \mathcal{S})$,

$$\rho_n^p(\varphi) = \iint \varphi(\beta)G_n(\alpha, \mathrm{d}\beta)I_n(\alpha, \beta)p(\mathrm{d}\alpha), \tag{17}$$

and we denote $G_n(\alpha, B) = \mathbb{P}_0^{(\alpha)}(X_n \in B)$. Let $\mathcal{V}$ denote the vector space of finite signed measures on $(\mathbb{S}, \mathcal{S})$. Denote by $J_{0,n}$ the (random) mapping from $\mathcal{V}$ to itself, mapping $p \in \mathcal{V}$ to ρ_n^p according to (17). Denoting by θ_n the shift transformation, let also

$$J_{m,n} = J_{0,n-m} \circ \theta_m, \qquad 0 \le m \le n.$$

By conditioning, it is clear that for $p \in \mathcal{P}$, $0 \le m \le n$,

$$\rho_n^p = J_{m,n}J_{0,m}\,p, \tag{18}$$

and consequently, $J_{0,n} = J_{m,n}J_{0,m}$. This gives rise to the recursion

$$\rho_n^p = J_{n-1,n}\rho_{n-1}^p, \qquad n \ge 1, \qquad \rho_0^p = p. \tag{19}$$

Set $\pi_n^p = \rho_n^p/\rho_n^p(1)$, $p \in \mathcal{P}$ and

$$\gamma(p, q) = \limsup_{n\to\infty} \frac{1}{n} \log d_{TV}(\pi_n^p, \pi_n^q), \qquad p, q \in \mathcal{P}. \tag{20}$$

A closer look at the recursion (19) shows that it is possible that the measure ρ_n becomes zero for some n. Here is an example. Consider the degenerate chain $X_n = X_0$, $n \ge 0$, where X_0 is a random variable on $\{1, 2\}$. Assume the observation process is given by $Y_n = X_n$ for $n \ge 1$. With the notation of the previous section, $G = I$ and $D_n^{i,i} = 1_{\{Y_n=i\}}$. If π_0 consists of an atom at 1 then $\mathbb{P}$-a.s., for all n,

$$T_n = D_nG^\top = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus if p is an atom at 2 then $\rho_1^p = T_1p = 0$.

In view of the foregoing discussion, we must define π_n^p and γ more carefully. Thus if for some n_0 $\rho_{n_0}^p = 0$ we let $\pi_n = 0$ for all $n \ge n_0$. We extend d_{TV} to $\mathcal{P} \cup \{0\}$ by defining

$d_{TV}(0,\lambda) = d_{TV}(\lambda,0) = d_{TV}(0,0) = 1$ for $\lambda \in \mathcal{P}$. Thus $\gamma(p,q) = 0$ on the event that ρ_n^p takes the value zero for some n.

The goal is now to show that γ can be bounded in terms of contraction coefficient for linear operators, in a fashion similar to Section 3. To this end we need an extended definition of Hilbert's metric. Define on $\mathcal{V}$ the partial order $\preccurlyeq$ by $\lambda \preccurlyeq \mu$ for $\lambda, \mu \in \mathcal{V}$ if $\lambda(A) \le \mu(A)$ for every $A \in \mathcal{S}$. Denote by $\mathcal{C}_0 \subset \mathcal{V}$ the cone of members λ of $\mathcal{V}$ for which $0 \preccurlyeq \lambda$, where 0 is the zero measure, and by $\mathcal{C}$ the collection of members of $\mathcal{C}$ excluding the zero measure. Two elements $\lambda, \mu \in \mathcal{C}$ are said to be *comparable* if there exist $0 < \alpha, \beta < \infty$ for which $\alpha\lambda \preccurlyeq \mu \preccurlyeq \beta\lambda$. Define $h : \mathcal{C}^2 \to [0,\infty]$ by

$$h(\lambda,\mu) = \log \frac{\sup_{A\in\mathcal{S}_\mu} \lambda(A)/\mu(A)}{\inf_{A\in\mathcal{S}_\mu} \lambda(A)/\mu(A)} \qquad \text{if } \lambda,\mu \in \mathcal{C} \text{ are comparable,} \tag{21}$$

and $h(\lambda,\mu) = \infty$ otherwise, where we denoted $\mathcal{S}_\mu = \{A \in \mathcal{S} : \mu(A) > 0\}$. The function h, called Hilbert's metric, is a pseudo-metric on $\mathcal{C}$, and a metric on the space of members $\lambda \in \mathcal{P}$ that are comparable to a given $\lambda_0 \in \mathcal{P}$ [8, Ch. 16]. A linear operator mapping $\mathcal{V}$ into itself is *positive* if it maps $\mathcal{C}$ into itself. As shown by Birkhoff [7] and Hopf [22], any positive linear operator L on $\mathcal{V}$ is a contraction with respect to Hilbert's metric, and

$$\tau(L) := \sup_{0<h(\lambda,\mu)<\infty} \frac{h(L\lambda, L\mu)}{h(\lambda,\mu)} = \tanh \frac{H(L)}{4}, \tag{22}$$

where

$$H(L) = \sup_{\lambda,\mu\in\mathcal{C}} h(L\lambda, L\mu), \tag{23}$$

and $\tau = 1$ in case when $H = \infty$ (see also [8] and [30] for these and various additional useful facts on the Hilbert metric). The formula above for τ is an extension of the formula (13) of Section 3.

We extend h to $\mathcal{C}_0^2$ by letting $h(0,\lambda) = h(\lambda,0) = h(0,0) = \infty$ for $\lambda \in \mathcal{C}$. Further, we let $\tau(L) = 1$ for any linear operator L mapping $\mathcal{C}_0$ into itself that is not positive. Such an operator will be called *weakly positive*.

To apply this to the filtering equations, note that $J_{m,n}$ are weakly positive, and by (18) that, for $p,q \in \mathcal{P}$, one has for any $n,m \in \mathbb{N}$,

$$h(\rho_{nm}^p, \rho_{nm}^q) \le h(p,q) \prod_{i=1}^n \tau(J_{im-m,im}).$$

Now, by definition of h, $h(c_1\lambda, c_2\mu) = h(\lambda,\mu)$ for any $c_1, c_2 \in (0,\infty)$, and thus $h(\pi_n^p, \pi_n^q) = h(\rho_n^p, \rho_n^q)$. Note also that, since $\tau \le 1$, $h(\rho_n^p, \rho_n^q)$ is monotone in n. By these considerations, as soon as $h(p,q) < \infty$, a bound on a quantity similar to γ (20) follows, namely

$$\gamma^h(p,q) := \limsup_n \frac{1}{n} \log h(\pi_n^p, \pi_n^q) \le \limsup_n \frac{1}{mn} \sum_{i=1}^n \log \tau(J_{im-m,im}).$$

In fact, it is easy to prove [3, Lemma 1]

$$d_{TV}(\lambda,\mu) \le \frac{2}{\log 3}\, h(\lambda,\mu), \qquad \lambda,\mu \in \mathcal{P},$$

whence $\gamma \le \gamma^h$. We summarize this in the following

Lemma 4.2. *If $p, q \in \mathcal{P}$ are comparable, or more generally, if $\inf_n h(\rho_n^p, \rho_n^q) < \infty$ $\mathbb{P}$-a.s., then for any positive integer m, $\mathbb{P}$-a.s.,*

$$\gamma(p,q) \le \limsup_n \frac{1}{mn} \sum_{i=1}^n \log \tau(J_{im-m,im}) =: \mathit{\Gamma}_m. \tag{24}$$

It is natural to apply this lemma to cases where J has some ergodic properties, so that $\mathit{\Gamma}_m$ can be expressed as expectation of a single term. For example, consider the case where for some $\pi_S \in \mathcal{P}$, the process $\{X_n\}$ is stationary ergodic under $\mathbb{P}_S := \mathbb{P}^{\pi_S}$. Then under the same law, so is the sequence $\{(X_n, Y_n)\}$, and in turn also the process $\{J_n\}$. In this case, $\mathit{\Gamma}_m$ is $\mathbb{P}_S$-a.s. equal to $\bar{\mathit{\Gamma}}_m := m^{-1}\mathbb{E}_S[\log \tau(J_{0,m})]$. Next, if say $\pi_0 \ll \pi_S$ then also $\mathbb{P} \ll \mathbb{P}_S$ and thus it is true also $\mathbb{P}$-a.s. that $\mathit{\Gamma}_m = \bar{\mathit{\Gamma}}_m$. More generally, the same conclusion will be valid, provided that $\mathbb{P}$ and $\mathbb{P}_S$ agree on the tail σ-field, because $\mathit{\Gamma}_m$ is measurable on this σ-field. This is recorded in the following.

Theorem 4.6. *[3] Assume there exists $\pi_S \in \mathcal{P}$ for which the corresponding law $\mathbb{P}_S$ makes $\{X_n\}$ stationary and ergodic. Assume also that the restrictions of both $\mathbb{P}$ and $\mathbb{P}_S$, to the tail σ-field, agree. Then $\mathbb{P}$-a.s.,*

$$\gamma(p,q) \le \frac{1}{m}\mathbb{E}_S[\log \tau(J_{0,m})], \qquad p, q \in \mathcal{P},\ m \in \mathbb{N}.$$

Let us exhibit a situation where the above gives rise to an exponential stability result. Consider the case where the state process satisfies the strong mixing condition. Namely, for some $\lambda \in \mathcal{P}$ and constants $0 < c_1, c_2 < \infty$,

$$c_1 \lambda \preccurlyeq G(x, \cdot) \preccurlyeq c_2 \lambda, \qquad x \in \mathbb{S}. \tag{25}$$

Note that G is a positive operator, and by (22), (23) that $\tau(G) < 1$. We claim that without any assumptions on the observation process (i.e., on $\widetilde{G}_0$ and g), one has $\tau(J_{0,1}) \le c < 1$, for some constant c.

Theorem 4.7. *[3] Assume (25) holds for some constants $0 < c_1, c_2 < \infty$ and $\lambda \in \mathcal{P}$. Then $\tau(J_{0,1}) \le c_3 := (c_2 - c_1)/(c_2 + c_1)$, $\mathbb{P}^{(x)}$-a.s., for any $x \in \mathbb{S}$. Consequently, under the hypotheses of Theorem 4.6, $\gamma(p,q) \le \log c_3 < 0$, for all $p, q \in \mathcal{P}$, $\mathbb{P}$-a.s.*

Proof. Let $\bar{a}(\gamma) = \int_{\mathbb{S}} g(\beta, \gamma)\lambda(\mathrm{d}\beta)$. We first show that for every α, $\mathbb{P}^{(\alpha)}$-a.s. one has $\bar{a}(Y_1) > 0$. To see this, let B denote the set $\{\gamma \in \mathbb{R}^\ell : \bar{a}(\gamma) = 0\}$. Then

$$\begin{aligned}
\mathbb{P}^{(\alpha)}(Y_1 \in B) &= \iint \mathbf{1}_B(\gamma) G(\alpha, \mathrm{d}\beta) \widetilde{G}(\beta, \mathrm{d}\gamma) \\
&\le c_2 \iint \mathbf{1}_B(\gamma) \lambda(\mathrm{d}\beta) g(\beta, \gamma) \widetilde{G}_0(\mathrm{d}\gamma) \\
&= c_2 \int \mathbf{1}_B(\gamma) \bar{a}(\gamma) \widetilde{G}_0(\mathrm{d}\gamma) = 0.
\end{aligned}$$

Next, given $A \in \mathcal{S}$ let $a_A := \int_A g(\beta, Y_1)\lambda(\mathrm{d}\beta)$ and note by (17) that $c_1 a_A \le \rho_1^p(A) \le c_2 a_A$ for every $p \in \mathcal{P}$. Hence $c_1/c_2 \le \rho_1^p(A)/\rho_1^q(A) \le c_2/c_1$, provided $a_A > 0$. Thus, provided there exists A for which $a_A > 0$, by (21), $h(\rho_1^p, \rho_1^q) \le 2\log(c_2/c_1)$, and by (22), (23), $\tau(J_{0,1}) \le$

$\tanh[\frac{1}{4}\log(c_2/c_1)] = (c_2 - c_1)/(c_2 + c_1) = c_3$. In view of the first paragraph, one has, in fact, $a_{\mathbb{S}} = \bar{a}(Y_1) > 0$ $\mathbb{P}^{(\alpha)}$-a.s. for arbitrary α, and we conclude that $\tau(J_{0,1}) \le c_3$, $\mathbb{P}^{(\alpha)}$-a.s.

The second assertion of the theorem is immediate from the first one. □

It is interesting that Lemma 4.2 may also be useful in non-ergodic situations. We refer the reader to [13] for a result based on a similar argument in a setup where the state process is transient (this result was improved in a subsequent paper [14] by other techniques; see also [31] and [34] for additional treatments of transient cases).

We now consider a continuous time Markov process on a Polish space, observed in white noise. The precise setting is as follows.

Equip the space $\Omega^1 = D(\mathbb{R}_+, \mathbb{S})$, of càdlàg mappings from $\mathbb{R}_+$ to $\mathbb{S}$, with the Skorohod J_1 topology, and let $\mathcal{B}^1$ denote the corresponding Borel σ-field. Let $\Omega^2 = C(\mathbb{R}_+, \mathbb{R}^\ell)$ be equipped with the uniform-on-compacts topology, and denote by $\mathcal{B}^2$ the corresponding Borel σ-field. Let $\Omega = \Omega^1 \times \Omega^2$ and $\mathcal{B} = \mathcal{B}^1 \otimes \mathcal{B}^2$. For $\omega = (\omega_1, \omega_2) \in \Omega$, let the processes X and W be defined via $X_t(\omega) = \omega_1(t)$ and $W_t(\omega) = \omega_2(t)$. For $\alpha \in \mathbb{S}$ let $\mathbb{P}^{(\alpha)}$ denote a probability measure on $(\Omega, \mathcal{B})$ under which W and X are independent, W is a standard ℓ-dimentional Brownian motion, and

$$\mathbb{P}^{(\alpha)}(X_{t_1} \in E_1, X_2 \in E_2, \dots, X_{t_n} \in E_n) = \int_{E_1 \times \cdots \times E_n} G_{t_1}(\alpha, \mathrm{d}x_1) \cdots G_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n),$$

for $n \in \mathbb{N}$, $0 < t_1 < t_2 < \cdots < t_n$ and $E_i \in \mathcal{S}$, $i \le n$. Here, G_t is a given Feller-Markov semigroup. As before, with $p \in \mathcal{P}$, associate $\mathbb{P}^p = \int \mathbb{P}^{(\alpha)} p(\mathrm{d}\alpha)$ and set $\mathbb{P} = \mathbb{P}^{\pi_0}$ for some fixed π_0.

To describe the observation process let a measurable function $k : \mathbb{S} \to \mathbb{R}^\ell$ be given. We shall assume $\mathbb{E}^{(\alpha)}[\int_0^t \|k(X_s)\|^2 \mathrm{d}s] < \infty$, $\alpha \in \mathbb{S}$, $t \ge 0$. The process Y_t is defined via

$$Y_t = \int_0^t k(X_s)\mathrm{d}s + W_t, \qquad t \ge 0. \tag{26}$$

Let $\mathcal{Y}_t = \sigma\{Y_s : s \in [0, t]\}$, and set

$$\pi_t^p(\varphi) = \mathbb{E}^p[\varphi(X_t)|\mathcal{Y}_t], \qquad t \ge 0.$$

We note on passing that, under various smoothness assumptions on the coefficients [26, Theorem 6.3.3], π_t^p solves the *Kushner-Stratonovich equation*

$$\pi_t(\varphi) = p(\varphi) + \int_0^t \pi_s(\mathcal{L}\varphi)\mathrm{d}s + \int_0^t \langle \pi_s(\varphi k) - \pi_s(\varphi)\pi_s(k), \mathrm{d}Y_s - \pi_s(k)\mathrm{d}s \rangle,$$

for φ in the domain of $\mathcal{L}$, where $\mathcal{L}$ is the generator of the semigroup G_t; similarly, one has $\pi_t^p = \rho_t^p / \rho_t^p(1)$ where ρ_t solves the *Zakai equation*

$$\rho_t(\varphi) = p(\varphi) + \int_0^t \rho_s(\varphi)\mathrm{d}s + \int_0^t \langle \rho_s(\varphi k), \mathrm{d}Y_s \rangle.$$

Consequently, this study can be viewed as one of sensitivity of solutions to these equations with respect to perturbations in their initial conditions.

Let us now describe the flow in a way similar to the discrete time case. To this end let

$$\Lambda_t = \exp\Big\{\int_0^t \langle k(X_s), \mathrm{d}Y_s\rangle - \frac{1}{2}\int_0^t \|k(X_s)\|^2 \mathrm{d}s\Big\}, \qquad t \geq 0.$$

Define

$$\rho_t^p(\varphi) = \mathbb{E}_0^p[\varphi(X_t)\Lambda_t|\mathcal{Y}_t],$$

where under $\mathbb{P}_0^p$, X and W are independent, but each of these processes has the same law as under $\mathbb{P}^p$. Then $\pi_t^p = \rho_t^p/\rho^p(1)$, $\mathbb{P}^p$-a.s. Let $J_{0,t}$ denote the linear, weakly positive transformation sending p to ρ_t^p. Then a result analogous to Theorem 4.6 holds, by similar considerations.

Theorem 4.8. *[3] Let assumptions analogous to those of Theorem 4.6 hold for the continuous time setting described above. Then $\mathbb{P}$-a.s.,*

$$\gamma(p, q) \leq \frac{1}{t}\mathbb{E}_S[\log \tau(J_{0,t})], \qquad p, q \in \mathcal{P}, t > 0.$$

The above result is applicable in the case of a diffusion on a compact manifold. Particularly, let X_t be a diffusion process on a compact manifold M of dimension m. To state the assumptions, we embed the manifold in $\mathbb{R}^d$, some $d \in \mathbb{N}$, and assume that the process is given as the solution to the stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \bar{\sigma}(X_t)\mathrm{d}\bar{W}_t, \qquad X_0 = x,$$

where $\bar{W}$ is independent of the observation noise W. It is assumed that the semigroup associated with X_t is strictly elliptic on M, and thus (as follows from [18, Ch. 3]), given $t > 0$ there exist constants $0 < c_1 < c_2 < \infty$ such that

$$c_1\lambda \preccurlyeq G_t(\alpha, \cdot) \preccurlyeq c_2\lambda, \qquad \alpha \in \mathbb{S},$$

where λ is the surface measure on M. Arguments similar to the ones described above in the discrete setting, now based on Theorem 4.8, lead to the following [3].

Theorem 4.9. *Under the assumptions above on the process X_t, and assuming also that k of (26) is twice continuously differentiable, one has that $\mathbb{P}$-a.s., $\gamma(p, q) \leq -c$, for all $p, q \in \mathcal{P}$, where $c > 0$ is a deterministic constant.*

We now make some further remarks on small noise asymptotics. First, let us mention that the upper bound (14) continues to hold in a setting of countable state space; moreover, under suitable assumptions, a lower bound is also valid, that is different from (15) but sufficient to deduce that the order of magnitude of γ_σ is σ^{-2}. These fact were proved by other methods in [3, Section 5].

Next, in a continuous state space, the following example was studied in [3] (the proof is based on the estimate from Theorem 4.6, with $m = 2$).

Theorem 4.10. *Let $\mathbb{S} = [0, 1]$. Assume that $G(\alpha, \mathrm{d}\beta) = \bar{G}(\alpha, \beta)\ell(\mathrm{d}\beta)$, where ℓ is the Lebesgue measure and $\bar{G}$ is three times continuously differentiable on $\mathbb{S}^2$. Assume also that the observations are of the form (2) (from Example 2.2), and that k is C^4 on $\mathbb{S}$, while the derivative of k is bounded away from zero. Then $\mathbb{P}$-a.s., for every $p, q \in \mathcal{P}$,*

$$\limsup_{\sigma\to 0} \frac{\gamma(p, q)}{\log \frac{1}{\sigma}} \leq -1. \tag{27}$$

A direct computation in the Gaussian case reveals the behavior $1/\sigma$ rather than $\log(1/\sigma)$. In view of this it is plausible that the above results should be possible to improve upon. A similar situation occurs in the case of a diffusion on $\mathbb{R}$, where [1] bounds the rate by $\log(1/\sigma)$, whereas an analogous analysis of the Kalman filter on $\mathbb{R}$ in continuous time [27] shows dependence of the form $1/\sigma$. Under some restricted assumptions, the behavior $1/\sigma$ is established in [4], but the question is open in any reasonable generality.

The techniques involving Hilbert metric appear to work well in a variety of setting where the state space is compact, but other than in some trivial cases, they usually fail when the space is noncompact. An exception is the contribution [23], where such techniques are used in conjunction with very clever considerations to establish stability properties of the filter under mixing assumptions on the state process, which lies in $\mathbb{R}^d$. See also [21] for a refinement of this result.

Finally we would like to point out the usefulness of Hilbert metric techniques in treating a more general problem, namely the sensitivity of the filter to perturbations in the transition kernel as well as the initial condition. The result is borrowed from [12]. To this end, let us go back to the setting of Theorem 4.7. Namely, we assume that (25) holds for some $0 < c_1 < c_2 < \infty$ and $\lambda \in \mathcal{P}$. In addition, assume we are given a sequence G_m of probability kernels that approximate G in the following sense: G and G_m all admit transition probability densities, $g(\cdot,\cdot)$ with respect to λ, on $(\mathbb{S}, \mathcal{S})$; for every m, $g(\cdot,\cdot)$ and $g_m(\cdot,\cdot)$ are zero and positive on the same sets; and $\log g_m$ converge to $\log g$ on the set $\{(x,y) \in \mathbb{S}^2 : g(x,y) > 0\}$. We are also given $\pi_m \in \mathcal{P}$, converging in total variation to π_0. Denote by $\pi_n^{(m)}$ the filter that uses the initial data π_m and the transition kernel G_m, and as before, denote by π_n the exact filter. Furthermore, assume that the observation process is of the form $Y_n = k(X_n) + W_n$, where W_n are $\mathbb{R}^\ell$-valued, i.i.d., with a bounded density with respect to the Lebesgue measure. The proof of the following result is based solely on elementary properties of the Hilbert metric.

Theorem 4.11. *[12] Under the above assumptions one has*

$$\lim_{m\to\infty} \sup_{n\in\mathbb{N}} d_{TV}(\pi_n^{(m)}, \pi_n) = 0.$$

References

[1] R. Atar. Exponential stability for non-linear filtering of diffusion processes in a noncompact domain. *Annals of Probability*, 26 No. 4, 1552–1574 (1998)

[2] R. Atar and O. Zeitouni. Lyapunov exponents for finite state nonlinear filtering. *Siam J. Contr. Opt.* 35, No. 1, 36–55 (1997)

[3] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut H. Poincare, Probabilities et Statistiques* 33 No. 6, 697–725 (1997)

[4] R. Atar, F. Viens and O. Zeitouni. Robustness of Zakai's equation via Feynman-Kac representations. *Stochastic analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming*, edited by W. M. McEneaney, G. Yin and Q. Zhang, 339–352; Birkhauser, (1999)

[5] P. Baxendale, P. Chigansky and R. Liptser. Asymptotic stability of the Wonham filter: ergodic and nonergodic signals. *SIAM J. Control Optim.*, 43(2):643–669 (electronic), 2004.

[6] A. G. Bhatt, A. Budhiraja, and R. L. Karandikar. Markov property and ergodicity of the nonlinear filter. *SIAM J. Control Optim.*, 39(3):928–949 (electronic), 2000.

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

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

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

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

[11] A. Budhiraja. Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(6):919–941, 2003.

[12] A. Budhiraja and H. J. Kushner. Robustness of nonlinear filters over the infinite time interval. *SIAM J. Control Optim.* Volume 36, Issue 5, pp. 1618-1637 (1998)

[13] A. Budhiraja and D. Ocone. Exponential stability of discrete-time filters for bounded observation noise. *Systems Control Lett.* 30 (1997), no. 4, 185–193.

[14] A. Budhiraja and D. Ocone. Exponential stability in discrete-time filtering for non-ergodic signals. *Stochastic Process. Appl.* 82 (1999), no. 2, 245–257.

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

[16] P. Chigansky, R. Liptser and R. Van Handel. Intrinsic methods in filter stability. This volume.

[17] J. E. Cohen, H. Kesten and C. M. Newman. Oseledec's multiplicative ergodic theorem: a proof. in *Random Matrices and Their Applications*, edited by J. E. Cohen, H. Kesten, C. M. Newman, Am. Math. Soc., Providence, pp. 23–30, 1986.

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

[19] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(2):155–194, 2001.

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

[21] R. Douc, G. Fort, E. Moulines and P. Priouret. Forgetting of the initial distribution for hidden Markov models. To appear in *Stoch. Proc. Appl.*

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

[23] M. L. Kleptsyna and A. Yu. Veretennikov. On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields* 141 (2008), no. 3-4, 411–444.

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

[25] H. Kunita. Ergodic properties of nonlinear filtering processes. *Spatial Stochastic Processes*, K. S. Alexander and J. C. Watkins, eds., Birkhguser, Boston, 1991.

[26] H. Kunita. *Stochastic flows and stochastic differential equations.* Cambridge Studies in Advanced Mathematics, 24. Cambridge University Press, Cambridge, 1997

[27] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* Wiley-Interscience, 1972.

[28] F. Ledrappier. Quelques propriétés des exposants caractéristiques. *École d'été de probabilités de Saint-Flour*, XII—1982, 305–396, Lecture Notes in Math., 1097, Springer, Berlin,

1984.

[29] F. Le Gland and L. Mevel. Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems*, 13(1):63–93, 2000.

[30] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* 14 (2004), no. 1, 144-187.

[31] F. Le Gland and N. Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Process. Appl.* 106 (2003), no. 2, 279-316.

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

[33] D. Ocone. Entropy inequalities and entropy dynamics in nonlinear filtering of diffusion processes. In *Stochastic analysis, control, optimization and applications*, Systems Control Found. Appl., pages 477–496. Birkhäuser Boston, Boston, MA, 1999.

[34] N. Oudjane and S. Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.* 23 (2005), no. 3, 421-448.

[35] D. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim.* 34 (1996), no. 1, 226–243.

[36] Y. Peres. Domains of analytic continuation for the top Lyapunov exponent. *Ann. Inst. Henri Poincaré,* 28 (1992), pp. 131–148.

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

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

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

[40] R. van Handel. The stability of conditional markov processes and markov chains in random environments. Preprint, 2008.

[41] W. M. Wonham. Some applications of stochastic differential equations to optimal nonlinear filtering. *SIAM J. Control Opt.,* 2 (1965), pp. 347–368.

Department of Electrical Engineering
Technion–Israel Institute of Technology
Haifa 32000, Israel