STABILITY OF NONLINEAR FILTERS: A SURVEY

PAVEL CHIGANSKY

ABSTRACT. Filtering deals with the optimal estimation of signals from their noisy observations. The standard setting consists of a pair of random processes $(X, Y) = (X_t, Y_t)_{t\geq 0}$, where the *signal* component X is to be estimated at a current time $t > 0$ on the basis of the trajectory of Y, observed up to t. Under the minimal mean square error criterion, the optimal estimate of X_t is the conditional expectation $E(X_t|Y_{[0,t]})$. If both X and (X, Y) are Markov processes, then the conditional distribution $\pi_t(A) = P(X_t \in A|Y_{[0,t]})$, $A \subseteq R$ satisfies a recursive equation, called filter, which realizes the optimal fusion of the a priori statistical knowledge about the signal and the a posteriori information borne by the observation path.

The filtering equation is to be initialized by the probability distribution ν of the signal at time $t = 0$. Suppose ν is unknown and another reasonable probability distribution $\bar{\nu}$ is used to start the filter. As the corresponding solution $\bar{\pi}_t(\cdot)$ differs from the optimal $\pi_t(\cdot)$, the natural question of stability arises: what are the conditions in terms of the signal/observation parameters to guarantee $\lim_{t\to\infty} \|\pi_t - \bar{\pi}_t\| = 0$ in an appropriate sense ? The article discusses the recent progress in solving this stability problem, which turns to be quite interesting and, sometimes, counterintuitive.

CONTENTS

Date: August, 2006.

Key words and phrases. nonlinear filtering, hidden Markov models, stability, Lyapunov exponents, Hilbert projective metric, Birkhoff contraction coefficient, ergodicity, time reversal.

These lecture notes have been prepared for the series of talks, delivered by the author at Colóquio Interinstitucional (CBPF - IMPA - UFF - UFRJ) Modelos Estocásticos e Aplicações and Laboratório Nacional de Computacão Cientifica, Petropolis, Brazil in August, 2006 upon the invitation of Prof. Jack Baczynski.

1. A fast-forward introduction

1.1. **Hidden Markov Models.** Consider a Markov chain $X = (X_n)_{n\ge 0}$ with values in a finite alphabet $\mathbb{S} = \{a_1, ..., a_d\}$, the transition probabilities $\lambda_{ij} := \mathsf{P}\big(X_n = a_j | X_{n-1} = a_i\big)$ and initial distribution $\nu_i := \mathsf{P}(X_0 = a_i)$, $i, j = 1, ..., d$. Let the observation sequence $Y = (Y_n)_{n\ge 1}$ be generated by

$$Y_n = \sum_{i=1}^d \mathbf{1}_{\{X_n = a_i\}} \xi_n(i), \quad n \ge 1, \tag{1.1}$$

where $\xi = (\xi_n)_{n\ge 1}$ is a sequence of i.i.d. random vectors, independent of X. This setting is usually viewed as a model of a noisy channel, which emits a realization of the random variable $\xi_n(i)$, each time the symbol a_i is transmitted. The entries of the vector ξ_1 are assumed to be independent and to have known probability densities $g_i(y)$, $i = 1, ..., d$ with respect to some reference σ-finite measure $\psi(dy)$ (e.g. the Lebesgue measure on $\mathbb{R}$ or purely atomic measure).

Having observed the trajectory of Y up to time $n \ge 0$, it is required to estimate the state of the signal X_n on the basis of the observations in an optimal way. The main building block in the solution of this estimation problem under various optimization criteria are conditional probabilities $\pi_n(i) = \mathsf{P}(X_n = a_i | \mathscr{F}^Y_n)$, $i = 1, ..., d$, where $\mathscr{F}^Y_n = \sigma\{Y_1, ..., Y_n\}$. For example, the *maximum a posteriori probability* (MAP) estimate of X_n given $\mathscr{F}^Y_n$ is

$$\widehat{X}^{\mathrm{map}}_n := \mathrm{argmax}_{a_i \in \mathbb{S}} \pi_n(i)$$

and it is optimal in the sense of minimizing the error probability of guessing the state of X_n given the realization of the trajectory $\{Y_1, ..., Y_n\}$:

$$\inf_{\zeta_n \in \mathbb{L}_\infty(\Omega, \mathscr{F}^Y_n, \mathsf{P})} \mathsf{P}\big(X_n \ne \zeta_n\big) = \mathsf{P}\big(X_n \ne \widehat{X}^{\mathrm{map}}_n\big) = 1 - \mathsf{E} \max_{a_i \in \mathbb{S}} \pi_n(i) \tag{1.2}$$

Another criterion is minimizing *the mean square error* (MSE), under which the optimal estimate is the conditional expectation $\widehat{X}^{\mathrm{mse}}_n = \mathsf{E}\big(X_n | \mathscr{F}^Y_n\big) = \sum_{i=1}^d a_i \pi_n(i)$:

$$\inf_{\zeta_n \in \mathbb{L}_2(\Omega, \mathscr{F}^Y_n, \mathsf{P})} \mathsf{E}\big(X_n - \zeta_n\big)^2 = \mathsf{E}\big(X_n - \widehat{X}^{\mathrm{mse}}_n\big)^2 = \mathsf{E}\bigg(\sum_{i=1}^d a_i^2 \pi_n(i) - \Big(\sum_{i=1}^d a_i \pi_n(i)\Big)^2\bigg) \tag{1.3}$$

The vector of conditional probabilities π_n satisfies the following *filtering* equation (essentially the recursive Bayes formula):

$$\pi_n = \frac{G(Y_n)\Lambda^* \pi_{n-1}}{\big|G(Y_n)\Lambda^* \pi_{n-1}\big|}, \quad \pi_0 = \nu, \tag{1.4}$$

where $G(y)$ is a diagonal matrix with entries $g_i(y)$, $i = 1, ..., d$, Λ^* is the transposed matrix of the transition probabilities and $|\cdot|$ denotes the ℓ_1-norm, i.e. $|x| = \sum_i |x_i|$, $x \in \mathbb{R}^d$.

The model described above is a particular instance of the so called Hidden Markov Models. The finite state space is special, because many related statistical problems have efficient closed form solutions. For example, the aforementioned state estimation problem is completely solved in an efficient way by iterating the equation (1.4). Another familiar special case is the linear Gaussian systems, for which the conditional distribution of X_n

given $\mathscr{F}^Y_n$ is Gaussian, with the mean and covariance satisfing the celebrated Kalman filtering equations. In the general case, the *filtering problem*, i.e. calculating the conditional distribution of X_n given $\mathscr{F}^Y_n$, is more complicated and less efficient. The solution is given in the form of an infinite dimensional recursive equation for conditional distribution or its density. Usually it is used as the basis for efficiently realizable approximate algorithms (particle filters, etc.) The first part of this minicourse is intended as a self contained brief presentation of the filtering theory both in discrete and continuous time settings.

The second part will focus on a more recent research in filtering, namely stability of the nonlinear filtering equation with respect to its initial conditions, ergodicity of the filtering process, robustness with respect to the model parameters, etc. To make things more concrete and transparent, we will use the equation (1.4) as a toy model. In spite of its seemingly simple structure, this equation nevertheless features much of the essential complexity of the problem. On the other hand, it is one of the few genuine nonlinear filtering equations of significant practical importance.

1.2. **Ergodicity of the filtering process.** Does the estimation error converge to a steady state ? Do the limits as $n \to \infty$ of the performance indices in (1.2) and (1.3) exist and if yes, are they independent of ν ?

Clearly the answers to both questions would be affirmative, if the distribution of π_n converges to a unique distribution over $\mathcal{M}$. Using the properties of conditional expectations, one can verify that the random sequence $\pi = (\pi_n)_{n\ge 0}$ is a Markov process with values in the simplex $\mathcal{S}^{d-1}$. Then the question reduces to whether $\pi = (\pi_n)_{n\ge 0}$ is an ergodic process, i.e. it has an invariant measure $\mathcal{M}$ and this measure is unique. While the existence of such measures even in more general situations can be often established using the Markov property of the filter, the uniqueness issue turns to be quite nontrivial and in fact still lacks a complete answer!

The common intuition suggests that the filtering process π inherits ergodicity from the signal X itself. Recall that, by definition, a finite state Markov chain X is ergodic if the limits $\mu_i := \lim_{n\to\infty} \mathsf{P}(X_n = a_i)$ exist, are positive and do not depend on ν. A chain is ergodic if and only if its transition matrix is q-primitive, i.e. there is an integer $q \ge 1$, such that the entries of Λ^q are positive.

Ergodicity of π for ergodic chains X was conjectured by D.Blackwell in [11] (1957), who studied these models in a particularly simple (or as we now realize quite nontrivial!) case, when the observation sequence is formed by a deterministic function $h : \mathbb{S} \mapsto \mathbb{R}$ of the signal, i.e. $Y_n = h(X_n)$. The original motivation of D.Blackwell was the search for a simple formula for the entropy rate of Y. He did find a formula, but it turned to be far from being simple, as it involved averaging with respect to $\mathcal{M}$ (the invariant measure of π) and this in turn had a remarkably complicated structure (e.g. it may be singular with respect to the Lebesgue measure on $\mathcal{S}^{d-1}$ yet having no atoms). D.Blackwell was not concerned primarily with the uniqueness of this measure, as he dealt with the stationary (X, Y). However to find $\mathcal{M}$ one had to solve the corresponding integral equation and this is where the uniqueness matter showed up.

This conjecture was proven to be false by T.Kaijser in [37] (1975), who pointed out that an appropriate counterexample was already there in [11], Blackwell's own paper! This counterexample turns to be quite illuminating as it demonstrates several relevant surprising features - we will use its particular version, independently rediscovered in [31] (see also [8]).

Example 1.1. Consider a chain with four states $\mathbb{S} = \{1,2,3,4\}$, the following transition matrix

$$\Lambda = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

and initial distribution ν. The entries of Λ^3 are positive and hence the chain is ergodic. Assume that the observation sequence is defined by $Y_n = \mathbf{1}_{\{X_n \in \{1,3\}\}}$. Notice that, having observed the trajectory of Y till time n, one can recover exactly the transitions of X between the groups of states $\{1,3\} \leftrightarrow \{2,4\}$. However it is impossible to tell which one of the states within $\{1,3\}$ (or $\{2,4\}$) the chain actually resides. The filtering recursion (1.4) in this case reads[1]

$$\begin{aligned} \pi_n(1) &= \big(\pi_{n-1}(4) + \pi_{n-1}(1)\big) Y_n \\ \pi_n(2) &= \big(\pi_{n-1}(1) + \pi_{n-1}(2)\big)(1 - Y_n) \\ \pi_n(3) &= \big(\pi_{n-1}(2) + \pi_{n-1}(3)\big) Y_n \\ \pi_n(4) &= \big(\pi_{n-1}(3) + \pi_{n-1}(4)\big)(1 - Y_n) \end{aligned} \tag{1.5}$$

subject to $\pi_0 = \nu$. It is not hard to see that π_n may take values among the following eight vectors

$$\phi_1 = \begin{pmatrix} \nu_1+\nu_4 \\ 0 \\ \nu_2+\nu_3 \\ 0 \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} 0 \\ \nu_1+\nu_4 \\ 0 \\ \nu_2+\nu_3 \end{pmatrix}, \quad \phi_3 = \begin{pmatrix} \nu_2+\nu_3 \\ 0 \\ \nu_1+\nu_4 \\ 0 \end{pmatrix}, \quad \phi_4 = \begin{pmatrix} 0 \\ \nu_2+\nu_3 \\ 0 \\ \nu_1+\nu_4 \end{pmatrix},$$

$$\phi_5 = \begin{pmatrix} 0 \\ \nu_1+\nu_2 \\ 0 \\ \nu_3+\nu_4 \end{pmatrix}, \quad \phi_6 = \begin{pmatrix} \nu_3+\nu_4 \\ 0 \\ \nu_1+\nu_2 \\ 0 \end{pmatrix}, \quad \phi_7 = \begin{pmatrix} 0 \\ \nu_3+\nu_4 \\ 0 \\ \nu_1+\nu_2 \end{pmatrix}, \quad \phi_8 = \begin{pmatrix} \nu_1+\nu_2 \\ 0 \\ \nu_3+\nu_4 \\ 0 \end{pmatrix}$$

and that Y_n form an i.i.d. symmetric binary sequence. Hence the invariant measure of π is uniform over these eight points of $\mathcal{S}^{d-1}$:

$$\mathcal{M}(du) = \frac{1}{8}\delta_{\{\phi_1\}}(du) + \dots + \frac{1}{8}\delta_{\{\phi_8\}}(du),$$

and clearly depends on ν. Sufficient conditions for ergodicity of π in Blackwell's setting were derived by T.Kaijser [37] and recently significantly improved by F.Kochman and J.Reeds, [44]. It is still unclear whether the conditions of [44] are also necessary.

Virtually the same kind of question was independently addressed by H.Kunita in [45] (1971) in the continuous time setting. The signal was assumed to be an ergodic Markov

[1]formally the reference measure $\psi(dy) = \delta_{\{0\}}(dy) + \delta_{\{1\}}(dy)$ and the densities are $g_1(y) = g_3(y) = y$ and $g_2(y) = g_4(y) = 1 - y$

process with values in a compact real subset $\mathbb{S} \subseteq \mathbb{R}$, the transition semigroup $(P_t)_{t\in\mathbb{R}_+}$ and the initial distribution ν. The observation process was assumed to satisfy

$$Y_t = \int_0^t h(X_s)ds + B_t,$$

with the Brownian motion (Wiener process) B, independent of X and bounded function h. The conditional measure process $\pi_t(A) := \mathsf{P}(X_t \in A|\mathscr{F}_t^Y)$ satisfies a stochastic partial integro-differential equation (see (2.11) below), which is initialized by the distribution ν. H.Kunita posed the question ($\pi_t(f) := \int_{\mathbb{S}} f(x)\pi_t(dx)$)

$$\text{Does the limit } \lim_{t\to\infty} \mathsf{E}\big(f(X_t) - \pi_t(f)\big)^2 \text{ exist and is it unique ?} \tag{1.6}$$

for any continuous and bounded function f. The main result of [45] is that this limit exists and is unique, if the signal is a Feller-Markov process, whose tail σ-algebra $\mathscr{F}_\infty^X = \bigcap_{t\ge 0} \mathscr{F}_{[t,\infty)}^X$ is P-a.s. empty. Recently a serious gap in the proof of this claim has been discovered in [8] (2004) (see Section 3.3 below) and currently its validity remains a challenging open problem.

1.3. **Stability of the filtering equation.** A different but of course related question of "steady state" behavior was posed by B.Delyon and O.Zeitouni in [31] (1989). Suppose that the actual distribution of X_0, needed to initialize the recursion (1.4), is unknown. It is then reasonable to start the filter from some other probability distribution (e.g. uniform on $\mathbb{S}$) and ask whether the obtained solution, denoted hereafter by $\bar{\pi}_n$, will be close to the optimal one π_n for large enough n. It is not immediately clear that an arbitrary probability distribution can be used to start (1.4), without causing an ambiguity and in fact some care should be taken to avoid this kind of pathology. As we will see later, the condition $\nu \ll \bar{\nu}$ (i.e. $\bar{\nu}_i = 0 \implies \nu_i = 0$) is sufficient for $\bar{\nu}$ to be a valid initialization for (1.4). The question is what are the conditions in terms of the model parameters, i.e. Λ, $g_i(u)$'s and $(\nu, \bar{\nu})$, for the filter to be stable in the sense

$$\big\|\pi_n - \bar{\pi}_n\big\| := \sum_{i=1}^d |\pi_n(i) - \bar{\pi}_n(i)| \xrightarrow[n\to\infty]{\mathsf{P}-a.s.} 0, \tag{1.7}$$

where $\|\cdot\|$ denotes the total variation norm, i.e. $\|x\| = \sum_{i=1}^d |x_i|$.

The aforementioned counterexample shows that the filtering equation may not be stable, even if the signal is ergodic, namely for (1.5)

$$\big\|\pi_n - \bar{\pi}_n\big\| \ge C > 0, \quad \forall n,$$

where C is a constant depending on $(\nu, \bar{\nu})$.

The relation between the stability of (1.4) and ergodicity of the process π has been established by D.Ocone and E.Pardoux in [59]: in a quite general setting, the affirmative answer to (1.6) implies the stability of the filter in the sense (cf. (1.7))

$$\lim_{t\to\infty} \mathsf{E}\big(\pi_t(f) - \bar{\pi}_t(f)\big)^2 = 0, \quad \text{for continuous and bounded functions } f, \tag{1.8}$$

if $\nu \ll \bar{\nu}$ and the signal X is ergodic. Unfortunately in view of the gap in the proof of (1.6) in [45] this does not provide any useful information about (1.8) in terms of the model. Clearly such type of convergence is weaker than (1.7), however currently (1.8) is only known as an implication of this stronger convergence (with the only exceptions for some special cases as in [27], [25]).

The most significant progress in establishing (1.7) has been accomplished during the last decade by addressing the problem in even stronger form, namely studying the the limit[2]:

$$\gamma := \lim_{n\to\infty} \frac{1}{t} \log \big\|\pi_n - \bar{\pi}_n\big\|. \tag{1.9}$$

Negativity of this limit, if exists, implies (1.7). Moreover the value of γ quantifies the rate of convergence. The problem was first addressed in this form yet in [31], but the real progress has been made by R.Atar and O.Zeitouni in [5], [6] (1997). These papers introduced two different approaches to calculation of γ: the Hilbert projective distance and Lyapunov exponents techniques. Deferring the detailed discussion of the method until later, let us briefly review the consequences as applied to the finite state filtering problem under consideration.

The limit γ in (1.9) exits under mild conditions (essentially ergodicity of X and certain integrability of the noise densities). Moreover it is a random variable, which takes values in a finite set of real numbers, including $\{-\infty\}$, depending on[3] the initial conditions $(\nu, \bar{\nu})$. Though exact calculation of γ seems to be impossible[4], certain information about it can be gained in the form of upper and lower bounds.

Without any restrictions on the noise densities $g_i(u)$, the following upper bound[5] holds

$$\gamma \le -\frac{\lambda_*}{\lambda^*}, \tag{1.10}$$

where $\lambda_* := \min_{i,j} \lambda_{ij}$ and $\lambda^* := \max_{i,j} \lambda_{ij}$. The latter means that the filter is stable, if all the transition probabilities are strictly positive, i.e. $\lambda_* > 0$. The latter property, sometimes referred in the literature as the *mixing* property[6], is clearly much more stronger than just ergodicity, and thus the filter does inherit stability from the signal regardless of the observations structure, but of a rather strong type. In fact (1.10) holds even non-asymptotically:

$$\|\pi_n - \bar{\pi}_n\| \le C \exp\Big(-\frac{\lambda_*}{\lambda^*} n\Big), \quad n \ge 1,$$

with $C > 0$ depending on $(\nu, \bar{\nu})$ (more bounds of the same spirit were reported in [51, 50], [24]). On the other hand, Example 1.1 shows that just ergodicity of X is not enough. Then how "much" ergodicity is really required to guarantee filter stability? The exact answer is not known yet. It turns out that if one of the rows of Λ has all positive entries and the chain is ergodic, then

$$\gamma \le -\frac{\lambda_\diamond}{\lambda^*}, \tag{1.11}$$

[2]all the statements involving random variables are understood to hold P-a.s. as usually

[3]in the "telegraphic signal" case $d = 2$, γ is independent of $(\nu, \bar{\nu})$. In the general case $d > 2$, the actual dependence on the initial condition remains unclear

[4]except for the case $d = 2$ in continuous time - see [22]

[5]in fact a slightly more tight bound holds, but we prefer to give its simple version at this point to emphasize the pros and cons

[6]$\lambda_* > 0$ does indeed imply that X is a mixing in the usual sense, but this condition is not necessary for the chain to be mixing, even when the state space is continuum

with $\lambda_\diamond := \sum_{i=1}^d \mu_i \min_{j\ne i}\lambda_{ij}$ (recall that μ is the stationary distribution of X_n). The proof of (1.11) requires a completely different argument (certain conditional time reversal in P.Baxendale et al [8], P.Ch. and R.Liptser [24]).

Another appealing fact is that the filter is stabilized by noise. Namely, let Y be generated by

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

where σ is a constant, $h : \mathbb{S} \mapsto \mathbb{R}$ is a deterministic function and $\xi = (\xi_n)_{n\ge 1}$ is a Gaussian i.i.d. sequence. Then the results of [6] imply that for any $\sigma \ne 0$, $\gamma < 0$ and thus the filter is stable. Moreover the following asymptotic bounds as $\sigma \to 0$ hold (hereafter we write $\gamma(\cdot)$ to emphasize its dependence on the relevant parameter)

$$\varlimsup_{\sigma\to 0} \sigma^2\gamma(\sigma) \le -\frac{1}{2}\sum_{i=1}^d \mu_i \min_{j\ne i}\big(h(a_i)-h(a_j)\big)^2 \tag{1.12}$$

$$\varliminf_{\sigma\to 0} \sigma^2\gamma(\sigma) \ge -\frac{1}{2}\sum_{i=1}^d \mu_i \sum_{j=1}^d \big(h(a_i)-h(a_j)\big)^2. \tag{1.13}$$

The upper bound (1.12) suggests that the filtering stability is improved as the noise intensity decreases, if there is at least one unique point in the image of $\mathbb{S}$ under h. Indeed, Blackwell's counterexample hints that $\gamma(\sigma)$ may converge to zero as $\sigma \to 0$, which was numerically tested yet in [31]. In the strong noise regime the filter turns to be as stable as the signal itself:

$$\lim_{\sigma\to\infty}\gamma(\sigma) \le \inf_{m\ge 1}\frac{1}{m}\log\tau(\Lambda^m) < 0 \tag{1.14}$$

where $\tau(\cdot)$ is the *Birkhoff contraction coefficient* (see Section 3.1 below), which is strictly less than 1 for matrices with positive entries (recall that if X ergodic Λ^m has positive entries for some integer $m \ge 1$).

Another interesting feature of γ is revealed in the slow switching regime. Let X^ε denote the Markov chain whose transition probabilities are defined via the following scaling (with $\varepsilon \in (0,1)$)

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

Clearly smaller values of ε correspond to the chain with less frequent transitions. This setting is in a sense more flexible than the noise scaling, since it allows the greater generality[7] of the observations model (1.1). A slight adjustment of the arguments from [6] shows that $\gamma(\varepsilon)$ remains negative for any $\varepsilon > 0$ under the assumption that $g_i(u)$ are bounded and has the same support. More effort is required (essentially the Furstenberg-Khasminskii formulae, see [23]) to show that

$$\gamma(\varepsilon) \le -\sum_{i=1}^d \mu_i \min_{j\ne i}\mathscr{D}(g_i \parallel g_j) + o(1), \quad \varepsilon \to 0,$$

[7] scaling noise by a multiplicative constant σ, does not always makes sense: e.g. when ξ_n is purely atomic

where $\mathscr{D}(g_i \parallel g_j) = \int_{\mathbb{R}} g_i(u) \log \dfrac{g_i}{g_j}(u)\varphi(du)$ are the Kullback-Leibler relative entropies. This suggests that for small ε, the filter remains stable, if at least one entropy is positive. This effect seems to be an attribution of the finite state space, since it is absent in the Kalman filtering setting.

For $d = 2$ this asymptotic is precise, i.e.

$$\gamma(\varepsilon) = -\mu_1 \mathscr{D}(g_1 \parallel g_2) - \mu_2 \mathscr{D}(g_2 \parallel g_1) + o(1), \quad \varepsilon \to 0.$$

and $\gamma(\varepsilon)$ turns to be not necessarily monotonic in ε. Namely for a symmetric binary chain with transition probability λ and $Y_n = (X_n - \xi_n)^2$ with an i.i.d. binary noise sequence ξ, $\mathsf{P}(\xi_1 = 1) = p$,

$$\gamma(\varepsilon) \geq -\mathscr{D}_p + \frac{4\lambda\big(\log(2) - h(p)\big)}{\mathscr{D}_p} \varepsilon \log \varepsilon^{-1}\big(1 + o(1)\big), \quad \varepsilon \to 0.$$

where $\mathscr{D}_p := p \log \dfrac{p}{1-p} + (1-p) \log \dfrac{1-p}{p}$ and $h(p) = -p \log p - (1-p) \log(1-p)$. As the second order term is positive, the formula (3.22) suggests that the limit $-\mathscr{D}_p$ is approached from above. On the other hand, it can be easily seen that $\gamma(\varepsilon') = -\infty$ for $\varepsilon' = 1/(2\lambda)$. Hence the function $\gamma(\varepsilon)$ has a global maximum at some positive $\varepsilon^\star$ (see Figure 1 on page 27), which means that the filtering stability may improve as the signal is slowed down beyond certain value of ε!

1.4. **What does this survey leave out ?** Limited by the course time scale, this survey does not elaborate some of the results available in the literature (though the author does try his best to provide a complete bibliography). Here is a brief account of things, which have been omitted.

The results and methods mentioned in the Introduction translate without much effort to the settings with Markov ergodic signals on compact (not necessarily finite) domains (see e.g. [29], [30], [28]). It is possible to push some of the methods to noncompact/nonergodic settings: some clever arguments appeared in [3], [16], [15], [52], [53] (and more recent variations on this theme in [32], [74], [60], [61], [43], [42]). However none of the results is even close to the powerful controllability/observability stability criteria, available in the Kalman-Bucy case. Thus the final word in this story is still missing and a completely fresh idea may be required to fill this gap.

On the other hand some results, which directly rely or repeat the arguments from [45], are to be revised: [59], [14], [13], [12], [9], [72], [73], [46], [49].

There are some "out of mainstream" interesting results, indicating that (1.8) (or even weaker stability) may hold for certain function, even when stability in the total variation norm as (1.7) fails or unknown (see [27], [58], [25]). Sometimes stronger results are possible in specific situations as e.g. for Beneŝ filters in [57], the noise free signal dynamics [21], etc. (see also [4], [7]). A variational approach of a functional analysis flavor was recently suggested by W.Stannat in a series of papers [69, 71, 70].

The stability with respect to initial conditions is naturally related to the robustness of the filtering equation with respect to the model parameters *over the infinite horizon.* The related results appeared in [53, 52], [60], [19, 18, 17]. Recently the continuous time case has been addressed in [26].

The rest of this article is organized as follows. Section 2 gives a sketchy overview of the nonlinear filtering theory, which is intended to give a self contained background for the reader, unfamiliar with it. Section 3 gives a detailed presentation of three approaches to filtering stability and ergodicity, mentioned in the Introduction.

2. Nonlinear filtering: a brief overview

2.1. **Filtering in discrete time.** The more general filtering problem is formulated as follows. Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be a Markov sequence with values in $\mathbb{R}^d$, transition probability density $\lambda(x,u)$:

$$\mathsf{P}\big(X_n \in \varGamma | \mathscr{F}^X_{n-1}\big) = \int_\varGamma \lambda(X_{n-1},u)\psi(du), \quad \forall \varGamma \in \mathscr{B}(\mathbb{R}^d), \quad \mathsf{P}-a.s.$$

where $\psi(du)$ is another σ-finite measure on $\mathbb{R}^d$ and initial probability density ν, i.e.

$$\mathsf{P}(X_0 \in \varGamma) = \int_\varGamma \nu(x)\psi(dx), \quad \forall \varGamma \in \mathscr{B}(\mathbb{R}^d).$$

Sometimes, when no confusion is caused, we will write $\lambda(x,du)$ for $\lambda(x,u)\psi(du)$ and $\lambda(x,\varGamma)$ for $\int_\varGamma \lambda(x,u)\psi(du)$ for brevity and similarly, denote by ν the measure $\nu(x)\psi(dx)$ rather than the density itself.

The observation process $Y = (Y_n)_{n\in\mathbb{Z}_+}$ is assumed to form an i.i.d. random sequence[8], conditioned on X, i.e. for $n \ge 1$

$$\mathsf{P}\big(Y_n \in \varGamma | \mathscr{F}^X_n \vee \mathscr{F}^Y_{n-1}\big) = \int_\varGamma g(X_n,y)\varphi(dy), \quad \forall \varGamma \in \mathscr{B}(\mathbb{R}^\ell), \quad \mathsf{P}-a.s.,$$

where $g(x,y)$ is the observation probability density with respect to a σ-finite reference measure $\varphi(dy)$ on $\mathbb{R}^\ell$. $g(x,y)$ is sometimes referred as *the likelihood* function.

As a special case, this formulation includes the recursion

$$\begin{aligned} X_n &= a(X_{n-1}) + b(X_{n-1})\eta_n \\ Y_n &= c(X_n) + d(X_n)\xi_n, \end{aligned}$$

where η and ξ are independent i.i.d. sequences and $a(\cdot)$, $b(\cdot)$,$c(\cdot)$ and $d(\cdot)$ are functions of appropriate dimensions. For example, for the scalar linear Gaussian problem (i.e. when $a(x) := ax$, $c(x) := cx$, $b(x) := b$ and $d(x) := d$ and when the noises ξ and η are Gaussian)

$$\lambda(x,u) = \frac{1}{\sqrt{2\pi}b}\exp\left\{\frac{(u-ax)^2}{2b^2}\right\}$$

and

$$g(x,y) = \frac{1}{\sqrt{2\pi}d}\exp\left\{\frac{(y-cx)^2}{2d^2}\right\}$$

with ψ and φ being the Lebesgue measures on $\mathbb{R}$.

The filtering equation for the general problem propagates the conditional density (with respect to ψ) of X_n given $\mathscr{F}^Y_n$, $n \ge 1$

$$\pi_n(x) = \frac{g(x,Y_n)\int_{\mathbb{R}^d}\lambda(u,x)\pi_{n-1}(u)\psi(du)}{\int_{\mathbb{R}^d} g(x,Y_n)\int_{\mathbb{R}^d}\lambda(u,x)\pi_{n-1}(u)\psi(du)\psi(dx)}, \quad \pi_0(x) = \nu(x), \tag{2.1}$$

[8]as before $Y_0 \equiv 0$ is assumed, or in other words $\mathscr{F}^Y_0 = \{\emptyset, \Omega\}$

where

$$\pi_n(\Gamma) := \int_\Gamma \pi_n(x)\psi(dx) = \mathsf{P}\big(X_n \in \Gamma | \mathscr{F}^Y_n\big), \quad \Gamma \in \mathscr{B}(\mathbb{R}).$$

2.1.1. *Finite dimensional filters.* The equation (2.1) is infinite dimensional in general, meaning that its solution cannot be parameterized by a finite set of *sufficient statistics*. For this reason, the general filtering equation is of limited practical value and is usually used as the basis for various approximations.

However there are two important classes of systems for which the filter turns to be finite dimensional: the aforementioned finite state Markov chains and the familiar Kalman's linear Gaussian setting. In the former case, the filtering distribution is just the vector of conditional probabilities satisfying the $d-1$ dimensional recursion (1.4). In the latter case, i.e. when the signal/observation pair is generated by

$$\begin{aligned} X_n &= AX_{n-1} + B\eta_n \\ Y_n &= CX_n + D\xi_n \end{aligned} \tag{2.2}$$

with independent i.i.d. Gaussian noises $\eta = (\eta_n)_{n\ge 1}$ and $\xi = (\xi_n)_{n\ge 1}$, deterministic matrices A,B,C and D of appropriate dimensions and Gaussian initial condition X_0, independent of η and ξ, the conditional density is Gaussian:

$$\pi_n(x) = \frac{1}{(2\pi)^{n/2}\det(P_n)} \exp\left\{-\frac{1}{2}(x-\widehat{X}_n)P_n^{-1}(x-\widehat{X}_n)^*)\right\}$$

with the conditional mean $\mathsf{E}(X_n|\mathscr{F}^Y_n) = \widehat{X}_n$ and covariance $\mathsf{E}(X_n-\widehat{X}_n)(X_n-\widehat{X}_n)^* = P_n$ satisfying the Kalman recursions:

$$\begin{aligned} \widehat{X}_n &= A\widehat{X}_{n-1} + P_{n|n-1}C^*\big(CP_{n|n-1}C^* + DD^*\big)^{-1}\Big(Y_n - CA\widehat{X}_{n-1}\Big) \\ P_{n|n-1} &= AP_{n-1}A^* + BB^* \\ P_n &= P_{n|n-1} - P_{n|n-1}C^*\big(CP_{n|n-1}C^* + DD^*\big)^{-1}CP_{n|n-1}. \end{aligned}$$

These two settings are virtually the only practically important instances of (2.1) (in fact, some other estimation problems for these models turn to be finite dimensional as well - see e.g. [33]).

2.1.2. *The reference measure point of view - the Zakai equation.* The equation (2.1) is nonlinear, however its solution is obtained by solving the linear Zakai type equation:

$$\rho_n(x) = g(x, Y_n)\int_{\mathbb{R}^d} \lambda(u,x)\rho_{n-1}(u)\psi(du), \quad n \ge 0, \tag{2.3}$$

subject to $\rho_0(x) = \nu(x)$, via normalization:

$$\pi_n(x) = \frac{\rho_n(x)}{\int_{\mathbb{R}^d}\rho_n(u)\psi(du)}. \tag{2.4}$$

This can be readily verified by induction, however the following "representation" formulae turn to be useful on their own, in particular in the stability problems under consideration. Let $g(y)$ be a probability density with respect to φ, such that

$$\text{both} \quad \frac{g(y)}{g(x,y)} \quad \text{and} \quad \frac{g(x,y)}{g(y)}$$

are well defined for each $(x,y)\in\mathbb{R}^d\times\mathbb{R}^\ell$ (possibly with the convention $0/0\equiv 0$). We assume that such a density exists, (which will usually be the case - e.g. any non-vanishing density would do for the Kalman model), though its specific choice is of no importance. For a fixed $n\ge 1$, introduce the random variable

$$Z_n=\prod_{m=1}^n\frac{g(Y_m)}{g(X_m,Y_m)}$$

and define the measure $\tilde{\mathsf{P}}$ by means of the Radon-Nikodym derivative $\dfrac{d\tilde{\mathsf{P}}}{d\mathsf{P}}=:Z_n(\omega)$. Since $Z_n>0$ P-a.s. and

$$\begin{aligned}&\tilde{\mathsf{P}}(\Omega)=\mathsf{E}Z_n=\mathsf{E}\mathsf{E}\left(\prod_{m=1}^n\frac{g(Y_m)}{g(X_m,Y_m)}\Big|\mathscr{F}^X_n\right)=\\&\mathsf{E}\int_{\mathbb{R}^\ell\times\ldots\times\mathbb{R}^\ell}\prod_{m=1}^n\frac{g(y_m)}{g(X_m,y_m)}\prod_{m=1}^n g(X_m,y_m)\varphi(dy_1)\ldots\varphi(dy_m)=\\&\mathsf{E}\int_{\mathbb{R}^\ell\times\ldots\times\mathbb{R}^\ell}\prod_{m=1}^n g(y_m)\varphi(dy_1)\ldots\varphi(dy_m)=1,\end{aligned}$$

$\tilde{\mathsf{P}}$ is a probability measure. Moreover under $\tilde{\mathsf{P}}$, X and Y are independent, X is distributed as under P and Y is an i.i.d. sequence with Y_1 having distribution $g(y)\psi(dy)$. Indeed, for any bounded functionals $\alpha_n:(\mathbb{R}^d)^n\mapsto\mathbb{R}$ and $\beta_n:(\mathbb{R}^\ell)^n\mapsto\mathbb{R}$

$$\begin{aligned}&\tilde{\mathsf{E}}\alpha_n(X)\beta_n(Y)=\mathsf{E}\alpha_n(X)\beta_n(Y)Z_n=\mathsf{E}\alpha_n(X)\mathsf{E}\Big(\beta_n(Y)\prod_{m=1}^n\frac{g(Y_m)}{g(X_m,Y_m)}\Big|\mathscr{F}^X_n\Big)=\\&\mathsf{E}\alpha_n(X)\mathsf{E}\Big(\int_{\mathbb{R}^\ell}\ldots\int_{\mathbb{R}^\ell}\beta_n(y)\prod_{m=1}^n\frac{g(y_m)}{g(X_m,y_m)}\prod_{m=1}^n g(X_m,y_m)\varphi(dy_1)\ldots\varphi(dy_n)\Big|\mathscr{F}^X_n\Big)=\\&\mathsf{E}\alpha_n(X)\int_{\mathbb{R}^\ell}\ldots\int_{\mathbb{R}^\ell}\beta_n(y)\prod_{m=1}^n g(y_m)\varphi(dy_1)\ldots\varphi(dy_n)=\tilde{\mathsf{E}}\alpha_n(X)\tilde{\mathsf{E}}\beta_n(Y).\end{aligned}$$

The following lemma derives the transformation of the conditional expectations under equivalent change of measure.

Lemma 2.1. *Let* P *and* $\tilde{\mathsf{P}}$ *be a pair of equivalent*[9] *measures on* $(\Omega,\mathscr{F})$ *and* $\mathscr{G}$ *be a sub-*σ*-algebra of* $\mathscr{F}$*, then for any random variable* α *with* $\tilde{\mathsf{E}}|\alpha|<\infty$

$$\tilde{\mathsf{E}}\big(\alpha|\mathscr{G}\big)=\frac{\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}{\mathsf{E}\big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}.$$

Proof. One has to check that for any $\mathscr{G}$-measurable bounded random variable θ

$$\tilde{\mathsf{E}}\left(\alpha-\frac{\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}{\mathsf{E}\big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}\right)\theta=0.$$

[9] i.e. $\mathsf{P}\ll\tilde{\mathsf{P}}$ and $\tilde{\mathsf{P}}\ll\mathsf{P}$

The latter indeed holds:

$$\tilde{\mathsf{E}}\frac{\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}{\mathsf{E}\big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}\theta = \mathsf{E}\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)\frac{\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}{\mathsf{E}\big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}\theta = \mathsf{E}\mathsf{E}\Big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)\big|\mathscr{G}\Big)\frac{\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}{\mathsf{E}\big(\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)}\theta =$$

$$\mathsf{E}\mathsf{E}\big(\alpha\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega)|\mathscr{G}\big)\theta = \mathsf{E}\alpha\theta\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}}(\omega) = \mathsf{E}\alpha\theta.$$

□

Since $\dfrac{d\mathsf{P}}{d\tilde{\mathsf{P}}} = Z_n^{-1}$, by this lemma for any $\Gamma \in \mathscr{B}(\mathbb{R}^d)$

$$\pi_n(\Gamma) = \mathsf{P}\big(X_n \in \Gamma|\mathscr{F}_n^Y\big) = \frac{\tilde{\mathsf{E}}\big(\mathbf{1}_{\{X_n\in\Gamma\}}Z_n^{-1}|\mathscr{F}_n^Y\big)}{\tilde{\mathsf{E}}\big(Z_n^{-1}|\mathscr{F}_n^Y\big)} =$$

$$\frac{\tilde{\mathsf{E}}\Big(\mathbf{1}_{\{X_n\in\Gamma\}}\prod_{m=1}^n \dfrac{g(X_m,Y_m)}{g(Y_m)}\Big|\mathscr{F}_n^Y\Big)}{\tilde{\mathsf{E}}\Big(\prod_{m=1}^n \dfrac{g(X_m,Y_m)}{g(Y_m)}\Big|\mathscr{F}_n^Y\Big)} = \frac{\tilde{\mathsf{E}}\big(\mathbf{1}_{\{X_n\in\Gamma\}}\prod_{m=1}^n g(X_m,Y_m)|\mathscr{F}_n^Y\big)}{\tilde{\mathsf{E}}\big(\prod_{m=1}^n g(X_m,Y_m)|\mathscr{F}_n^Y\big)} \overset{\dagger}{=}$$

$$\frac{\int \mathbf{1}_{\{x_n\in\Gamma\}}\prod_{m=1}^n g(x_m,Y_m)\mu^X(dx)}{\int \prod_{m=1}^n g(x_m,Y_m)\mu^X(dx)},$$

where the independence of X and Y under $\tilde{\mathsf{P}}$ was used and $\mu^X(dx)$ denotes the probability distribution[10] of $X = (X_n)_{n\in\mathbb{Z}_+}$. Using the Markov property of X, the numerator of the latter expression is found to satisfy the equation (2.3) (recall that under $\tilde{\mathsf{P}}$, X and Y are independent). Namely, let

$$\int \mathbf{1}_{\{x_n\in\Gamma\}}\prod_{m=1}^n g(x_m,Y_m)\mu^X(dx) =: \rho_n(\Gamma), \quad n \ge 1,$$

then

$$\rho_n(\Gamma) = \int \mathbf{1}_{\{x_n\in\Gamma\}}\prod_{m=1}^n g(x_m,Y_m)\mu^X(dx) = \tilde{\mathsf{E}}\Big(\mathbf{1}_{\{X_n\in\Gamma\}}\prod_{m=1}^n g(X_m,Y_m)\Big|\mathscr{F}_n^Y\Big) =$$

$$\tilde{\mathsf{E}}\Big[\prod_{m=1}^{n-1} g(X_m,Y_m)\tilde{\mathsf{E}}\Big(\mathbf{1}_{\{X_n\in\Gamma\}}g(X_n,Y_n)\Big|\mathscr{F}_n^Y\vee\mathscr{F}_{n-1}^X\Big)\Big|\mathscr{F}_n^Y\Big] =$$

$$\tilde{\mathsf{E}}\Big[\prod_{m=1}^{n-1} g(X_m,Y_m)\int_\Gamma g(u,Y_n)\lambda(X_{n-1},u)\psi(du)\Big|\mathscr{F}_n^Y\Big] =$$

$$\int_\Gamma g(u,Y_n)\Big(\int_{\mathbb{R}}\lambda(x,u)\rho_{n-1}(dx)\Big)\psi(du).$$

ρ_n is a measure valued random sequence, called *unnormalized* conditional distribution.

---

[10] more formally the induced measure on $(\mathbb{R}^d)^\infty$

2.2. **Filtering in continuous time.** Both the problem formulation and its solution is much more delicate in the continuous time, due to the lack of the "white noise" analogue for the discrete time i.i.d. sequence. This goal of this section is to give a very superficial guide to the subject. For the complete and consistent presentation the reader is referred to [55, 54] (other texts are [38], [33])

Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a Markov process with trajectories in the space of functions $[0,T] \mapsto \mathbb{R}$ continuous from the right and having limits from the left (abbreviated in French as cadlag functions). This space is denoted by $\mathbb{D}_{[0,T]}$ and is to a complete separable metric (i.e. Polish) space, when endowed with the Skorokhod metric. The transition semigroup and the initial distribution of the process are assumed to be known. The observation process $Y = (Y_t)_{t\in\mathbb{R}_+}$ is given by

$$Y_t = \int_0^t h(X_s)ds + B_t, \tag{2.5}$$

where h is a continuous $\mathbb{R} \mapsto \mathbb{R}$ function and $B = (B_t)_{t\in\mathbb{R}_+}$ is a Brownian motion, independent of X.

The general filtering equation for the conditional distribution of X_t, given $\mathscr{F}_t^Y = \{Y_s, s \le t\}$ can be derived either via the martingale representation theorem or the reference measure approach. We omit the discussion of the former approach (see e.g. Chapter 8, [55]) and outline the main idea of the later. The key tool is the Girsanov change of measure.

Theorem 2.2 (I.Girsanov)**.** *Consider a Brownian motion B, defined on a filtered probability space*[11] $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathsf{P})$ *and define*

$$Z_t := \exp\left(\int_0^t \alpha_s dB_s - \frac{1}{2}\int_0^t \alpha_s^2 ds\right),$$

where α_s is a random process, such that[12] *the Itô integral with respect to B is well defined. Assume*[13] *that $\mathsf{E}Z_T = 1$ and define a probability measure $\tilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ by means of the Radon-Nikodym derivative $\dfrac{d\tilde{\mathsf{P}}}{d\mathsf{P}} = Z_T(\omega)$. Then the process $V_t = B_t - \int_0^t \alpha_s ds$ is a Brownian motion on $(\Omega, \mathscr{F}, \mathscr{F}_t, \tilde{\mathsf{P}})$.*

Proof. (sketch) Since the trajectories of V are continuous functions, the claim holds by the Levý theorem if for any $\lambda \in \mathbb{R}$

$$\tilde{\mathsf{E}}\Big(e^{i\lambda(V_t - V_s)}\big|\mathscr{F}_s\Big) = e^{-\frac{1}{2}\lambda^2(t-s)^2}.$$

[11] all the processes are adapted to $\mathscr{F}_t$: in particular one may take $\mathscr{F}_t := \mathscr{F}_t^B \vee \mathscr{F}_t^\alpha$

[12] essentially α should be adapted to $\mathscr{F}_t^B$ and satisfy $\mathsf{P}\Big(\int_0^T \alpha_s^2 ds < \infty\Big) = 1$

[13] by the Itô formula Z_t satisfies the SDE $Z_t = 1 + \int_0^t Z_s\alpha_s dB_s$. This however does not guarantee that $\mathsf{E}Z_T = 1$, i.e. the expectation of the stochastic integral may be nonzero. If the Novikov condition is satisfied, i.e. $\mathsf{E}\exp\left(\frac{1}{2}\int_0^T \alpha_s^2 ds\right) < \infty$, then semimartingale Z_t is also a martingale, i.e. $\mathsf{E}Z_T = 1$ and thus it can be used to define a change of probability measure

For simplicity let's verify the latter for $s=0$ (the proof for $s>0$ is essentially the same). By the Itô formula $U_t := e^{i\lambda V_t}$ satisfies

$$dU_t = i\lambda U_t dV_t - \frac{\lambda^2}{2}U_t dt.$$

Since $\mathsf{E}Z_T = 1$,

$$\tilde{E}e^{i\lambda V_t} = \mathsf{E}Z_T e^{i\lambda V_t} = \mathsf{E}e^{i\lambda V_t}\mathsf{E}(Z_T|\mathscr{F}_t) = \mathsf{E}e^{i\lambda V_t}Z_t = \mathsf{E}U_t Z_t$$

Once again using the Itô formula we get:

$$\begin{aligned} d\big(U_t Z_t\big) =& U_t dZ_t + Z_t dU_t + i\lambda U_t Z_t \alpha_t dt = \\ & U_t Z_t \alpha_t dB_t + i\lambda U_t Z_t dV_t - \frac{\lambda^2}{2}U_t Z_t dt + i\lambda U_t Z_t \alpha_t dt = \\ & U_t Z_t \alpha_t dB_t + i\lambda U_t Z_t dB_t - i\lambda U_t Z_t \alpha_t dt - \frac{\lambda^2}{2}U_t Z_t dt + i\lambda U_t Z_t \alpha_t dt, \end{aligned} \tag{2.6}$$

and taking the expectation from both sides one gets the ODE for $\Psi(t) := \tilde{\mathsf{E}}e^{i\lambda V_t} = \mathsf{E}U_t Z_t$

$$\frac{d}{dt}\Psi(t) = -\frac{1}{2}\lambda^2\Psi(t), \quad \Psi(0) = 1,$$

whose solution gives the required answer. $\square$

Back to the filtering problem at hand, introduce

$$Z_t = \exp\Big(-\int_0^t h(X_s)dB_s - \frac{1}{2}\int_0^t h^2(X_s)ds\Big)$$

and define the *reference measure* by

$$\frac{d\tilde{\mathsf{P}}}{d\mathsf{P}} := Z_T(\omega).$$

If e.g. $h(X_s)$ is bounded, then $\tilde{\mathsf{P}}$ is a probability measure and by the Girsanov theorem with $\alpha_t := -h(X_t)$, the process Y_t is a Brownian motion under the probability $\tilde{\mathsf{P}}$. Moreover independence of X and B under P translates to the independence of X and Y under $\tilde{\mathsf{P}}$. In fact, a stronger version of the Girsanov's theorem holds in this case, namely the process Y_t is a Brownian motion under the conditional measure given $\mathscr{F}^X_T$. By Lemma 2.1

$$\tilde{\mathsf{E}}\Big(e^{i\lambda V_t}|\mathscr{F}^X_T\Big) = \frac{\mathsf{E}\big(U_t Z_T|\mathscr{F}^X_T\big)}{\mathsf{E}(Z_T|\mathscr{F}^X_T)}.$$

Recall that if B is independent of a σ-algebra $\mathscr{G} \subseteq \mathscr{F}$, then[14] $\mathsf{E}\Big(\int_0^t \alpha_s dB_s\big|\mathscr{G}\Big) = 0$. The martingale Z_t satisfies

$$Z_t = 1 - \int_0^t Z_s h(X_s)dB_s \tag{2.7}$$

[14] this is easily verified for simple functions α and extended to the general case by a limiting procedure

and since B is independent of $\mathscr{F}_T^X$, we have $\mathsf{E}\big(Z_T|\mathscr{F}_T^X\big)=1$. Similarly

$$\mathsf{E}\Big(U_tZ_T|\mathscr{F}_T^X\Big)=\mathsf{E}\big(U_t\mathsf{E}(Z_T|\mathscr{F}_T^X\vee\mathscr{F}_t^B)\big|\mathscr{F}_T^X\Big)=$$
$$\mathsf{E}\Big(U_t\mathsf{E}\big(Z_t-\int_t^T h(X_s)Z_sdB_s|\mathscr{F}_T^X\vee\mathscr{F}_t^B\big)\big|\mathscr{F}_T^X\Big)=\mathsf{E}\big(U_tZ_t|\mathscr{F}_T^X\big)$$

and in turn by (2.6)

$$\tilde{\mathsf{E}}\Big(e^{i\lambda V_t}|\mathscr{F}_T^X\Big)=\mathsf{E}\Big(U_tZ_t|\mathscr{F}_T^X\Big)=e^{-\frac{1}{2}\lambda^2t^2}.$$

Hence for any measurable bounded functionals $F:\mathbb{D}_{[0,T]}\mapsto\mathbb{R}$ and $G:\mathbb{C}_{[0,T]}\mapsto\mathbb{R}$

$$\tilde{\mathsf{E}}F(X)G(Y)=\tilde{\mathsf{E}}F(X)\tilde{\mathsf{E}}\big(G(Y)|\mathscr{F}_T^X\big)=\tilde{\mathsf{E}}F(X)\mathsf{E}G(B),$$

i.e. X and Y are independent under $\tilde{\mathsf{P}}$. Finally

$$\tilde{\mathsf{E}}F(X)=\mathsf{E}Z_TF(X)=\mathsf{E}F(X)\mathsf{E}\big(Z_T|\mathscr{F}_T^X\big)=\mathsf{E}F(X)$$

i.e. the distribution of X remains unaltered under the change of measure.

To summarize, under the reference measure $\tilde{\mathsf{P}}$, X and Y are independent, Y is a Brownian motion and X has the same distribution as under P. Note that since $\mathsf{P}\sim\tilde{\mathsf{P}}$,

$$\frac{d\mathsf{P}}{d\tilde{\mathsf{P}}}(\omega)=Z_T^{-1}=\exp\Big(\int_0^t h(X_s)dB_s+\frac{1}{2}\int_0^t h^2(X_s)ds\Big)=$$
$$\exp\Big(\int_0^t h(X_s)dY_s-\frac{1}{2}\int_0^t h^2(X_s)ds\Big)$$

Then by Lemma 2.1,

$$\mathsf{P}\big(X_t\in\Gamma|\mathscr{F}_t^Y\big)=\frac{\tilde{\mathsf{E}}\Big(\mathbf{1}_{\{X_t\in\Gamma\}}\exp\Big(\int_0^t h(X_s)dY_s-\frac{1}{2}\int_0^t h^2(X_s)ds\Big)\Big|\mathscr{F}_t^Y\Big)}{\tilde{\mathsf{E}}\Big(\exp\Big(\int_0^t h(X_s)dY_s-\frac{1}{2}\int_0^t h^2(X_s)ds\Big)\Big|\mathscr{F}_t^Y\Big).}$$

Since under $\tilde{\mathsf{P}}$, X and Y are independent this can be rewritten in a more compact form, known as the Kallianpur-Striebel representation formula

$$\mathsf{P}\big(X_t\in\Gamma|\mathscr{F}_t^Y\big)=\frac{\int_{\mathbb{D}_{[0,T]}}\mathbf{1}_{\{x_t\in\Gamma\}}\Phi_t\big(x,Y(\omega)\big)\mu^X(dx)}{\int_{\mathbb{D}_{[0,T]}}\Phi_t\big(x,Y(\omega)\big)\mu^X(dx)}\tag{2.8}$$

where $\mu^X(dx)$ is the probability measure induced by X on $\mathbb{D}_{[0,T]}$ and [15]

$$\Phi_t\big(x,Y(\omega)\big):=\exp\Big(\int_0^t h(x_s)dY_s-\frac{1}{2}\int_0^t h^2(x_s)ds\Big).$$

Now the Markov property of X can be used to deduce a recursive equation for the *unnormalized* conditional distribution:

$$\rho_t(\Gamma)=\tilde{\mathsf{E}}\big(\mathbf{1}_{\{X_t\in\Gamma\}}\Phi_t(X,Y)|\mathscr{F}_t^Y\big).$$

[15] i.e. $\Phi_t(X,Y)=Z_t^{-1}$

For definiteness, consider the simplest case, when X_t is a Markov chain with values in a finite alphabet $\mathbb{S} = \{a_1, ..., a_d\}$, with transition intensities λ_{ij}:

$$\mathsf{P}\big(X_{t+\delta} = a_j | X_t = a_i\big) = \begin{cases} \lambda_{ij}\delta + o(\delta), & i \ne j \\ 1 - \sum_{\ell \ne i} \lambda_{i\ell}\delta + o(\delta), & i = j \end{cases}$$

and initial distribution ν: $\nu_i = \mathsf{P}(X_0 = a_i)$. Introduce *indicators* vector process J_t with the entries $J_t(i) = \mathbf{1}_{\{X_t = a_i\}}$ and let ρ_t be the vector of unnormalized conditional probabilities:

$$\rho_t := \tilde{\mathsf{E}}\big(J_t Z_t^{-1} | \mathscr{F}_t^Y\big).$$

Clearly there is a one-to-one correspondence between X and J (assuming that all a_i's are different). The process J_t is a semimartingale with the decomposition

$$J_t = J_0 + \int_0^t \Lambda^* J_s ds + M_t,$$

where Λ^* is the transposed matrix of transition intensities and M is a (purely discontinuous) martingale (with respect to $\mathscr{F}_t^X$). Recall that Z_t^{-1} satisfies the equation

$$dZ_t^{-1} = Z_t^{-1} h(X_t) dY_t$$

and apply the Itô formula to the product $J_t Z_t^{-1}$:

$$\begin{aligned} d(J_t Z_t^{-1}) = J_t dZ_t^{-1} + Z_t^{-1} dJ_t = J_t h(X_t) Z_t^{-1} dY_t + \Lambda^* J_t Z_t^{-1} dt + Z_t^{-1} dM_t = \\ H(J_t Z_t^{-1}) dY_t + \Lambda^*(J_t Z_t^{-1}) dt + Z_t^{-1} dM_t \quad (2.9) \end{aligned}$$

where the equality $J_t h(X_t) = H J_t$ with[16] $H = \mathrm{diag}(h)$ was used. Since M is independent of $\mathscr{F}^Y$ under $\tilde{\mathsf{P}}$,

$$\tilde{\mathsf{E}}\Big(\int_0^t Z_s^{-1} dM_s \big| \mathscr{F}_T^Y\Big) = 0.$$

Conditioning both sides of (2.9) on $\mathscr{F}_t^Y$, one gets the *Zakai* equation for ρ_t:

$$d\rho_t = \Lambda^* \rho_t dt + H \rho_t dY_t, \quad \rho_0 = \nu.$$

This is a linear (more exactly bilinear) SDE with respect to ρ_t, driven by the observation process Y_t. It is the continuous time analogue of the recursion (2.3). The conditional probabilities $\pi_t := \mathsf{E}(J_t | \mathscr{F}_t^Y)$ are recovered from ρ_t by normalization $\pi_t = \rho_t / \|\rho_t\|$, with $\|\rho_t\| = \sum_{i=1}^d \rho_t(i)$.

The vector π_t satisfies the nonlinear (Wonham) SDE, obtained by applying the Itô formula to $\rho_t / \|\rho_t\|$:

$$d\pi_t = \Lambda^* \pi_t dt + \big(\mathrm{diag}(\pi_t) - \pi_t \pi_t^*\big) h \big(dY_t - \pi_t(h) dt\big), \quad \pi_0 = \nu, \tag{2.10}$$

where h is a vector of the values of h on $\mathbb{S}$, $\pi_t(h) = \langle \pi_t, h \rangle = \sum_{i=1}^d h(a_i) \pi_t(i)$ and $\mathrm{diag}(\pi_t)$ is a diagonal matrix with entries $\pi_t(i)$, $i = 1, ..., d$. This is the analogue of the HMM filter (1.4) in continuous time setting.

The process $\bar{B}_t := Y_t - \int_0^t \pi_t(h) ds$ is the *innovation* Brownian motion with respect to $\mathscr{F}_t^Y$ under the original measure P. This fact can be established in a much more greater generality and is the key to the martingale derivation of the equation for π_t (which we

[16] the function h on $\mathbb{S}$ is naturally identified with the vector of $h(a_1)$,...,$h(a_d)$

omit here). This approach allows to derive a general filtering *Fujisaki-Kallianpur-Kunita* equation for the conditional distribution of X_t with respect to $\mathscr{F}_t^Y$, which is the general continuous time analogue of the filtering recursion (2.1). The FKK equation is the fundamental result in filtering, defining the evolution of the conditional distribution π_t. Being a measure valued functional stochastic equation, it has a rather complicated form, which we do not describe here. The finite dimensional Wonham filter (2.10) is the particularly simple instance of this equation, which is of practical importance.

In the case when X is a diffusion process

$$X_t = a(X_t)dt + b(X_t)dW_t,$$

where W is a Brownian motion, independent of B and a and b are functions with appropriate properties, the FKK equation reduces to the functional Kushner-Stratonovich SPDE (cf. (2.10)) for the conditional density $\pi_t(x)$ with respect to the Lebesgue measure (if it exists!):

$$d_t\pi_t(x) = \mathcal{L}^*\pi_t dt + \big(h(x) - \pi_t(h)\big)\pi_t(x)\big(dY_t - \pi_t(h)dt\big), \tag{2.11}$$

where $\mathcal{L}^*$ is the infinitesimal operator corresponding to X and $\pi_t(h) := \int_{\mathbb{R}} h(x)\pi_t(x)dx$. The corresponding unnormalized conditional distribution in this case has the density[17] $\rho_t(x)$, which satisfies the linear Zakai SPDE

$$d_t\rho_t(x) = \mathcal{L}^*\rho_t(x)dt + h(x)\rho_t(x)dY_t, \quad \rho_0(x) = \nu(x). \tag{2.12}$$

Conditions for existence and uniqueness of the solutions of the filtering equations as well as their properties are far from being obvious and is a subject of extensive research both in the past and now. In the particular case of linear system (A, B, C and D are constant matrices; W and B are independent vector Brownian motions)

$$\begin{aligned} dX_t &= AX_tdt + BdW_t \\ dY_t &= CX_tdt + DdB_t \end{aligned}$$

subject to a Gaussian X_0, the conditional density $\pi_t(x)$ is Gaussian with the mean $\widehat{X}$ and covariance P_t, satisfying the familiar Kalman-Bucy equations:

$$\begin{aligned} d\widehat{X}_t &= A\widehat{X}_tdt + P_tC^*(DD)^{-1}(dY_t - h\widehat{X}_tdt) \\ \dot{P}_t &= AP_t + P_tA^* + BB^* - P_tC^*(DD)^{-1}CP_t, \end{aligned}$$

subject to $\widehat{X}_0 = \mathsf{E}X_0$ and $P_0 = \mathsf{E}(X_0 - \widehat{X}_0)(X_0 - \widehat{X}_0)^*$. Finite dimensional realizations are available for various functionals of (X, Y) in the LQG and finite state settings (see the book [33]). Other finite dimensional cases of the filtering equation are known (most notably the Beneŝ diffusion case), but their practical value is limited.

[17] here a measure and its density are denoted by the same letter, as e.g. here $\rho_t(dx) = \rho_t(x)dx$ or $\nu(dx) = \nu(x)dx$.

3. Filtering stability and ergodicity

This section deals with the stability problem of the nonlinear filtering equation and the ergodic properties of its solution. We will restrict the following discussion to the simple (and yet nontrivial!) case of finite state Markov chains in discrete time and will show how most of the results, presented in the introduction are derived. Further extensions and other results, reported in the literature, are gathered in the bibliography for the readers reference. The main reason for the choice of such simplified treatment is its transparency and concreteness. Since the "curse of dimensionality" is resolved in this case due to the finite dimensionality of the signal (and not due to the very special properties of the Gaussian processes as in the Klamaa-Bucy setting), many of the results extend to more general settings without major difficulty - essentially to the models with ergodic signals on a compact state space. The nonergodic/noncompact case is more difficult and virtually all currently available results combine one of the methods with a "compactification" trick. Regretfully none of them recaptures the well known powerful controllability/observability conditions of the Kalman-Bucy linear Gaussian case.

3.1. **The Hilbert projective metric approach.**

3.1.1. *Nonnegative operators acting on nonnegative measures.* This section presents the ideas from the theory of nonnegative operators, introduced into the filtering stability problem by R.Atar and O.Zeitouni in [6], [5].

Let $\mathbb{S} \subseteq \mathbb{R}^d$ be a measurable set[18] and $\mathcal{M}_+$ be the space of nonnegative measures on $\big(\mathbb{S}, \mathscr{B}(\mathbb{S})\big)$ with the partial order relation $p \preceq q$ if $p(A) \le q(A)$ for any measurable $A \subseteq \mathbb{S}$. The measures p and q are *comparable* if $c_1 p \preceq q \preceq c_2 p$ for some constants $c_1, c_2 > 0$. The *Hilbert projective distance* is defined

$$h(p,q) = \log \frac{\sup_{A, q(A)>0} p(A)/q(A)}{\inf_{A, q(A)>0} p(A)/q(A)}, \quad p, q \in \mathcal{M}_+ \text{ are comparable}$$

and $h(p,q) = \infty$ otherwise. Clearly two comparable measures p and q are equivalent and vise versa and hence ($\|\cdot\|_\infty$ is the supremum norm)

$$h(p,q) = \log\Big(\Big\|\frac{dp}{dq}\Big\|_\infty \Big\|\frac{dq}{dp}\Big\|_\infty\Big).$$

It is easy to see that $h(p,q)$ is a nonnegative symmetric function, satisfying the triangle inequality

$$h(p,q) \le h(p,r) + h(r,q), \quad p,q,r \in \mathcal{M}_+.$$

Also $h(p,q) = 0$ iff $p = cq$ for some $c > 0$. The latter property turns $(\mathcal{M}_+, h)$ into a pseudo-metric space. Notice also that on the space of probability (i.e. normalized to 1) measures, h defines a metric on the part of the domain where it takes finite values. This is not as innocent as it may seem - e.g. this metric is infinite for Gaussian measures with different means!

The following two properties are important for our purposes:

[18] the host space is taken to be $\mathbb{R}^d$ here for definiteness - more general spaces are possible. Also $\mathbb{S}$ can be $\mathbb{R}^d$ itself. The distinction is made to emphasize in the sequel to distinguish the compact and noncompact state spaces

(p-1) $h(p,q) = h(c_1p, c_2q)$ for any $p, q \in \mathcal{M}_+$ and any scalars $c_1, c_2 > 0$

(p-2) $\|p - q\| \le \frac{2}{\log(3)} h(p,q)$ for any $p, q \in \mathcal{S}^{d-1}$

The first property is obvious from the definition. The proof of the second one is given e.g. in Lemma 1 in [5].

Let K be a linear positive operator, mapping $\mathcal{M}_+$ to itself. The Birkhoff contraction coefficient is defined as

$$\tau(K) := \sup_{0<h(p,q)<\infty} \frac{h(Kp, Kq)}{h(p,q)}. \tag{3.1}$$

$\tau(K)$ has the following expression in terms of h-diameter $H(K) := \sup_{p,q\in\mathcal{M}_+} h\big(Kp, Kq\big)$ of K:

$$\tau(K) = \tanh\Big(\frac{H(K)}{4}\Big). \tag{3.2}$$

Moreover, if the operator K is of the integral form

$$(Kp)(du) := \int_{\mathbb{S}} \kappa(x,u)p(dx)\psi(du),$$

where $\kappa(x,u)$ is a nonnegative function (kernel) and ψ is a σ-finite measure, then

$$H(K) = \log \operatorname*{ess\,sup}_{x,u,x',u'} \frac{\kappa(x,u)\kappa(x',u')}{\kappa(x,u')\kappa(x',u)}, \tag{3.3}$$

where $0/0 = 1$ is assumed and the sup is strict over x and y and ψ-essential over u and u'. Proofs of these facts can be found in Theorem 3 of Chapter XVI in [10] or Theorem 1 in [35].

Remark 3.1. In the filtering context, we will be particularly interested in the operators with even more specific structure, namely

$$K_g p := g(u) \int_{\mathbb{S}} \kappa(x,u)p(dx)\psi(du),$$

where $g(u)$ is a nonnegative function (the likelihood). Note that if $g(u)$ is strictly positive, then $H(K_g) = H(K)$. However, K_g is still a strict contraction, regardless of the properties of g, if the kernel $\kappa(x,u)$ satisfies the "mixing" condition, i.e. if some constants κ_* and κ^*

$$0 < \kappa_* \le \kappa(x,u) \le \kappa^* < \infty.$$

In this case, (3.3) implies

$$H(K) \le \log\Big(\frac{\kappa^*}{\kappa_*}\Big)^2 \tag{3.4}$$

and consequently

$$\tau(K) \le \frac{\kappa^* - \kappa_*}{\kappa^* + \kappa_*} =: \hat{\tau}(K) < 1. \tag{3.5}$$

As observed in [15] and [52], these inequalities remain valid for K_g as well, since it remains to be mixing with respect to a different reference measure, namely $\psi_g(du) := g(u)\psi(du)$:

$$\tau(K_g) \le \hat{\tau}(K) < 1. \tag{3.6}$$

Example 3.2. For the finite state space $\mathbb{S} = \{a_1, ..., a_d\}$, the above notions read as follows. The space $\mathcal{M}_+$ is just the nonnegative cone of $\mathbb{R}^d$ or in other words $\mathcal{M}_+$ can be identified with the vectors from $\mathbb{R}^d$ with nonnegative entries. The Hilbert distance between $p, q \in \mathcal{M}_+$ is defined as[19]

$$h(p,q) := \begin{cases} \log \dfrac{\max_i p_i/q_i}{\min_j p_j/q_j}, & \text{if } p \sim q \\ \infty, & \text{otherwise} \end{cases} \tag{3.7}$$

with the convention $0/0 = 1$.

Any nonnegative operator on $\mathcal{M}_+$ can be represented by a $d \times d$ matrix $A = (a_{ij})$ with nonnegative entries. If A is an allowable matrix, i.e. none of its columns or rows contains only zeros, then

$$H(A) = \log \max_{i,j,k,\ell} \frac{a_{ik}a_{j\ell}}{a_{i\ell}a_{jk}}.$$

Notice that if a contains at least one zero entry, then $H(K) = \infty$ and $\tau(A) = 1$. Otherwise

$$\tau(A) = \tanh\Big(\frac{H(A)}{4}\Big) = \frac{1 - \sqrt{\psi(A)}}{1 + \sqrt{\psi(A)}}, \quad \psi(A) := \min_{i,j,k,\ell} \left\{ \frac{a_{ik}a_{j\ell}}{a_{i\ell}a_{jk}} \Big| a_{il}a_{jk} \neq 0 \right\}. \tag{3.8}$$

Hence $\tau(A)$ is strictly less than unity if and only if all the entries of an allowable matrix are nonzero. In particular, with $a_* := \min_{ij} a_{ij}$ and $a^* := \max_{ij} a_{ij}$,

$$\log \tau(A) \le -\frac{a_*}{a^*}$$

and, as the formula (3.8) suggests, for any diagonal matrix D with strictly positive diagonal entries

$$\tau(A) = \tau(AD) = \tau(DA). \tag{3.9}$$

If D has some zero diagonal entries, but all the elements of A are positive, then

$$\tau(AD) \le \frac{a^* - a_*}{a^* + a_*}.$$

3.1.2. *Application to the filtering problem.* The following construction will be convenient to use in the sequel. Without loss of generality, we assume that (X, Y) are coordinate processes on the canonical probability space $(\Omega, \mathscr{F}) = \big(\mathbb{R}^\infty \times \mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty) \times \mathscr{B}(\mathbb{R}^\infty)\big)$. We denote by P and $\bar{\mathsf{P}}$ the probability measures, under which (X, Y) is a Markov process with the given transition semigroup (i.e. X is a finite state Markov chain with transitions probabilities matrix Λ and Y is an i.i.d. sequence conditioned on $\mathscr{F}^X$) and X_0 is has distribution ν and $\bar{\nu}$ respectively. Let P^Y and $\bar{\mathsf{P}}^Y$ be the distributions of $Y = (Y_n)_{n \ge 1}$ under P and $\bar{\mathsf{P}}$ and P^Y_n and $\bar{\mathsf{P}}^Y_n$ be their restrictions on $\mathscr{F}^Y_n = \sigma\{Y_1, ..., Y_n\}$ (i.e. these are just the probability distributions of the vector $\{Y_1, ..., Y_n\}$ under P and $\bar{\mathsf{P}}$). Finally let P^s the probability measure under which (X, Y) is a stationary process, i.e. when $X_0 \sim \mu$.

The n first iterations of (1.4) define a functional $\Psi_n(y) : \mathbb{R}^n \mapsto \mathcal{S}^{d-1}$, which is well defined on a set of full P-probability. By the Markov property of (X, Y), the assumption $\nu \ll \bar{\nu}$ implies $\mathsf{P} \ll \bar{\mathsf{P}}$ and hence $\Psi_n(y)$ is a well defined $\bar{\mathsf{P}}$-a.s. as well. Note also that for

[19] $p \sim q$ stands for equivalence (in the sense of mutual absolute continuity) relation between the measures p and q. In the finite case, this means that p and q should not vanish at the same indices

ergodic X, the invariant measure μ is positive and hence $\nu \ll \mu$ for any $\nu \in \mathcal{S}^{d-1}$. This e.g. implies $\mathsf{P} \ll \mathsf{P}^s$.

Consider the filtering equation (1.4) and let $\pi = (\pi_n)_{n\ge 0}$ and $\bar\pi = (\bar\pi_n)_{n\ge 0}$ be the exact and "wrong" filtering processes. The following result is essentially Theorem 2.3 from [6] (see also [51, 50], [65])

Theorem 3.3. *Assume $\nu \sim \bar\nu$ and*

(**a-1**) *X is an ergodic chain*

(**a-2**) *there exists an integer $r \ge 1$ so that all the entries of the product matrix*

$$\Lambda^* G(Y_{r-1})...G(Y_1)\Lambda^*$$

are strictly positive with nonzero P^s-probability

(**a-3**) *the noise densities $g_i(u)$ are bounded*

Then $(X,\pi) = (X_n, \pi_n)_{n\ge 0}$ is an ergodic Markov process and

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \|\pi_n - \bar\pi_n\| < 0, \quad \mathsf{P} - a.s. \tag{3.10}$$

*In particular, if all the entries of Λ are positive (i.e. (**a-2**) holds with $r = 1$)*

$$\|\pi_n - \bar\pi_n\| \le \frac{2}{\log(3)} h(\nu, \bar\nu) \Big(\frac{\lambda^* - \lambda_*}{\lambda^* + \lambda_*}\Big)^n, \tag{3.11}$$

with $\lambda_* = \min_{ij} \lambda_{ij}$ and $\lambda^* = \max_{ij} \lambda_{ij}$.

Proof. Recall that $\pi_n = \rho_n/\|\rho_n\|$ and $\bar\pi_n = \bar\rho_n/\|\bar\rho_n\|$ where ρ_n and $\bar\rho_n$ are solutions of the Zakai equation (cf. (2.4) and (2.3))

$$\rho_n = G(Y_n)\Lambda^* \rho_{n-1}, \tag{3.12}$$

subject to $\rho_0 = \nu$ and $\bar\rho_0 = \bar\nu$ respectively. Hence

$$\rho_n = G(Y_n)\Lambda^* ... G(Y_1)\Lambda^* \nu, \quad \bar\rho_n = G(Y_n)\Lambda^* ... G(Y_1)\Lambda^* \bar\nu.$$

First consider the case, when $\lambda_* > 0$. The matrix $G(Y_n)\Lambda^*$ is mixing in the sense of Remark 3.1, i.e. the measures $\sum_{j=1}^d \lambda_{ij}\delta_{a_j}(du)$ have positive density with respect to the reference measure $\sum_j g_j(Y_n)\delta_{a_j}(du)$. Hence

$$\|\pi_n - \bar\pi_n\| \le \frac{2}{\log 3} h(\pi_n, \bar\pi_n) = \frac{2}{\log 3} h\Big(\frac{\rho_n}{\|\rho_n\|}, \frac{\bar\rho_n}{\|\bar\rho_n\|}\Big) \stackrel{\dagger}{=}$$
$$\frac{2}{\log 3} h(\rho_n, \bar\rho_n) \le \frac{2}{\log 3} h(\nu, \bar\nu) \hat\tau^n(\Lambda) = \frac{2}{\log 3} h(\nu, \bar\nu) \Big(\frac{\lambda^* - \lambda_*}{\lambda^* + \lambda_*}\Big)^n \tag{3.13}$$

where the equality † holds by the scaling invariance of the Hilbert distance (p-1) and the latter inequality follows by iterations of (3.1) (recall the definition of $\hat\tau(\cdot)$ in (3.5)). This proves (3.11).

Notice that by (3.1) and (3.2),

$$h(\rho_n, \bar\rho_n) \le \tau\big(G(Y_n)\Lambda^*\big) h(\rho_{n-1}, \bar\rho_{n-1}) \le h(\rho_{n-1}, \bar\rho_{n-1}),$$

i.e. the sequence $h(\rho_n, \bar\rho_n)$ is non-increasing and its $\overline{\lim}$ in (3.10) can be realized along any subsequence. Let r be the integer defined in (**a-2**), define

$$T_\ell(Y) := \Lambda^* G(Y_{(\ell+1)r-1})...G(Y_{\ell r+1})\Lambda^*, \quad \ell = 0, 1, ...$$

Then as in (3.13)

$$\|\pi_n - \bar{\pi}_n\| \le \frac{2}{\log 3} h(\nu, \bar{\nu}) \prod_{i=0}^{[n/r]} \hat{\tau}(T_i(Y))$$

Let P^s denote the probability measure on $(\Omega, \mathscr{F})$, under which (X, Y) is stationary, i.e. $X_0 \sim \mu$ (the invariant measure of the chain). Then under P^s

$$\varlimsup_{n\to\infty} \log \frac{1}{n} \log \|\pi_n - \bar{\pi}_n\| = \varlimsup_{\ell\to\infty} \log \frac{1}{\ell r} \log \|\pi_{\ell r} - \bar{\pi}_{\ell r}\| \le \frac{1}{r} \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{i=0}^{\ell} \log \hat{\tau}(T_i(Y)) \le$$
$$\frac{1}{r} \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{i=0}^{\ell} \big(\log \hat{\tau}(T_i(Y)) \vee -1\big) = \frac{1}{r} \mathsf{E}^s \big(\log \hat{\tau}(T_0(Y)) \vee -1\big),$$

where the latter convergence holds P^s-a.s. by the Birkhoff-Khinchine LLN for the stationary process (X, Y) (the clipping by -1 is needed along with boundedness of $g_i(y)$'s for the appropriate integrability). Since $\nu \ll \mu$, $\mathsf{P} \ll \mathsf{P}^s$ and hence this bound holds under P as well. Since the entries of $T_0(Y)$ are positive with positive probability, $\mathsf{E}^s\big(\log \hat{\tau}(T_0(Y)) \vee -1\big) < 0$ and the assertion (3.10) follows.

The process (X, π) is Markov and is also Feller and thus it has at least one invariant measure (by the classic Krylov-Bogolyubov argument - see e.q. [45]). Its uniqueness is deduced in Theorem 7.1 in [18] from the stability property (3.10). □

Remark 3.4. Notice that the assumption (**a-2**) is violated for the Blackwell's chain from Example 1.1.

Corollary 3.5. *Assume that X is ergodic and $g_i(u)$ are bounded and has the same support (e.g. do not vanish on $\mathbb{R}^d$). Then* (3.10) *holds.*

Proof. In this case $\mathsf{P}^Y_n \sim \bar{\mathsf{P}}^Y_n$ and the condition $\nu \ll \bar{\nu}$ is void. Since X is ergodic, its transition matrix Λ is m-primitive, i.e. Λ^m has strictly positive entries for some integer $m \ge 1$. Since $g_i(u)$ are supported on the same set the diagonal of $G(Y_n)$ is P-a.s. positive and thus (**a-2**) is satisfied with $r := m$. □

3.1.3. *The "mixing" condition in the general setting.* The Hilbert metric approach is applicable to the general filtering equation (2.1) along the same lines.

Theorem 3.6 (typical claim in the spirit of [5])**.** *Suppose the signal evolves on a subset $\mathbb{S} \subseteq \mathbb{R}^d$, i.e. $\mathsf{P}(X_n \notin \mathbb{S}) = 0$ for all $n \ge 0$ ($\mathbb{S} = \mathbb{R}^d$ is also allowed). Assume that $h(\nu, \bar{\nu}) < \infty$ and there exists a reference measure ψ on $\mathbb{S}$, with respect to which the transition law of the signal has a uniformly positive and bounded density, i.e.*

$$0 < \lambda_* \le \lambda(x, u) \le \lambda^* < \infty. \tag{3.14}$$

Then

$$\|\pi_n - \bar{\pi}_n\| \le \frac{2}{\log 3} h(\nu, \bar{\nu}) \Big(\frac{\lambda^* - \lambda_*}{\lambda^* + \lambda_*}\Big)^n, \quad n \ge 1. \tag{3.15}$$

Proof. Let ρ_n and $\bar{\rho}_n$ be the solutions of the Zakai equation (2.3), then taking into account (2.4) and the properties of the Hilbert metric:

$$\|\pi_n - \bar{\pi}_n\| \le \frac{2}{\log 3} h(\pi_n, \bar{\pi}_n) = \frac{2}{\log 3} h(\rho_n, \bar{\rho}_n) \le \frac{2}{\log 3} h(\nu, \bar{\nu}) \prod_{k=1}^{n} \tau\big(\Lambda_n(Y)\big),$$

where the operators $\Lambda_n(Y)$ are given by

$$\Lambda_n(Y)p := g(u, Y_n)\Big(\int_{\mathbb{S}} \lambda(x,u)p(dx)\Big)\psi(du).$$

The assertion (3.15) is true, since (see (3.6)) $H(\Lambda_n) \le \big(\lambda^*/\lambda_*\big)^2$ and consequently

$$\tau(\Lambda_n(Y)) \le \frac{\lambda^* - \lambda_*}{\lambda^* + \lambda_*}.$$

□

The "mixing" condition (3.14) is quite natural for compact sets $\mathbb{S}$: in this case the Lebesgue measure can be typically chosen as ψ to provide (3.14). It is not hard to see that in this case X is also mixing in the usual sense and a fortiori ergodic. If $\mathbb{S}$ is noncompact it is usually not clear how to choose the reference measure ψ to satisfy the mixing condition. For example, consider the signal generated by

$$X_n = h(X_{n-1}) + \eta_n, \quad n \ge 1$$

where h is a bounded function, say $|h(x)| \le 1$ and $\eta = (\eta_n)_{n\ge 1}$ is an i.i.d. sequence with

$$\frac{d}{dx}\mathsf{P}(\eta_1 \le x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R}.$$

Clearly $\mathbb{S} \equiv \mathbb{R}$ in this case and if one chooses $\psi(du) = du$, then $\lambda(x,u) = \frac{1}{2}e^{-|u-h(x)|}$, so that no λ_* required by (3.14) exists. In this case $H(\Lambda_n(Y)) = \infty$ and $\tau(\Lambda_n(Y)) = 1$. However if one chooses $\psi(du) := e^{-|u|}du$, then

$$\lambda(x,u) = e^{-|u-h(x)|+|u|} \ge e^{-|h(x)|} \ge e^{-1} := \lambda_*$$

and

$$\lambda(x,u) \le e^{-\big||u|-|h(x)|\big|+|u|} \le e^{|h(x)|} \le e^{1} := \lambda^*$$

and hence (3.14) holds. This trick does not always work and most notably fails for the Gaussian case.

Thus applicability *per se* of the "mixing" condition (3.14) for noncompact state spaces is limited. Some further extensions of this technique are possible via "compactification". For example, suppose that the transition density $\lambda(x,u)$ is positive on any compact and that the likelihood $g(x, Y_n)$ is supported on a compact $C_n \subseteq \mathbb{R}$ P-a.s. for any $n \ge 1$. Then

(Example 3.10 in [52], essentially the same trick was used already in [15])

$$
\begin{aligned}
&g(x,Y_n)\int_{\mathbb{R}}\lambda(x,u)\pi_{n-1}(du) = g(x,Y_n)\int_{\mathbb{R}}\lambda(x,u)\Big(g(u,Y_{n-1})\int_{\mathbb{R}}\lambda(u,u')\pi_{n-2}(du')\psi(du)\Big)\\
&= g(x,Y_n)\int_{\mathbb{R}}\mathbf{1}_{\{C_n\}}(x)\lambda(x,u)\mathbf{1}_{\{C_{n-1}\}}(u)\Big(g(u,Y_{n-1})\int_{\mathbb{R}}\lambda(u,u')\pi_{n-2}(du')\psi(du)\Big)\\
&= g(x,Y_n)\int_{\mathbb{R}}\tilde{\lambda}(x,u)\Big(g(u,Y_{n-1})\int_{\mathbb{R}}\lambda(u,u')\pi_{n-2}(du')\psi(du)\Big)\\
&= g(x,Y_n)\int_{\mathbb{R}}\tilde{\lambda}(x,u)\pi_{n-1}(du)
\end{aligned}
$$

where

$$
\tilde{\lambda}(u,x) = \begin{cases}\lambda(u,x) & (u,x)\in C_{n-1}\times C_n\\ 1 & \text{otherwise.}\end{cases}
$$

Since $\lambda(u,x)$ is positive on any compact in $\mathbb{R}\times\mathbb{R}$, $\tilde{\lambda}(u,x)$ is mixing (with respect to the same reference measure and with the mixing constants depending on C_n and C_{n-1}) and consequently the filter is stable. For further developments of this idea see [52, 53], [60], [74], [32], [43, 42].

Remark 3.7. Essentially the same results can be obtained using other characterization of contraction, notably Dobrushin's ergodic coefficient as in [29].

Remark 3.8. Though technically more involved, the same treatment can be done in continuous time case - see [5], [30].

3.2. **The Lyapunov exponents approach.** The key idea of this approach, pioneered in [6], is the following simple inequality (the scalar case is treated for clarity)

$$
\begin{aligned}
\|\pi_n-\bar{\pi}_n\| = \int_{\mathbb{R}}|\pi_n(x)-\bar{\pi}_n(x)|\psi(dx) = \int_{\mathbb{R}}\Big|\frac{\rho_n(x)}{\|\rho_n\|}-\frac{\bar{\rho}_n(x)}{\|\bar{\rho}_n\|}\Big|\psi(dx) = \\
\int_{\mathbb{R}}\Big|\frac{\rho_n(x)}{\int_{\mathbb{R}}\rho_n(y)\psi(dy)}-\frac{\bar{\rho}_n(x)}{\int_{\mathbb{R}}\bar{\rho}_n(z)\psi(dz)}\Big|\psi(dx) = \\
\frac{\int_{\mathbb{R}}\big|\rho(x)\int_{\mathbb{R}}\bar{\rho}_n(z)\psi(dz)-\bar{\rho}_n(x)\int_{\mathbb{R}}\rho_n(y)\psi(dy)\big|\psi(dx)}{\|\rho_n\|\|\bar{\rho}_n\|} \le \\
\frac{\int_{\mathbb{R}\times\mathbb{R}}\big|\rho(x)\rho_n(y)-\bar{\rho}_n(x)\rho_n(y)\big|\psi(dx)\psi(dy)}{\|\rho_n\|\|\bar{\rho}_n\|} := \frac{\|\rho_n\wedge\bar{\rho}_n\|}{\|\rho_n\|\|\bar{\rho}_n\|}.
\end{aligned}
\tag{3.16}
$$

Hence

$$
\overline{\lim_{n\to\infty}}\frac{1}{n}\log\|\pi_n-\bar{\pi}_n\| \le \overline{\lim_{n\to\infty}}\frac{1}{n}\log\|\rho_n\wedge\bar{\rho}_n\| - \underline{\lim}_{n\to\infty}\frac{1}{n}\log\|\rho_n\| - \underline{\lim}_{n\to\infty}\frac{1}{n}\log\|\bar{\rho}_n\| \tag{3.17}
$$

This suggests that stability of (2.1) is controlled by the growth rates of the solutions of corresponding Zakai equation (2.3). The limits in the right hand side are not trivial for calculations and usually only some qualitative information can be extracted: e.g. the asymptotic behavior as functions of various system parameters, etc. The treatment of the finite state space is in fact closely related to the theory of Lyapunov exponents for linear random dynamical systems (see e.g. monograph [2]).

Let's start with the classical Oseldec's Multiplicative Ergodic Theorem (MET) (cited here from [1])

Theorem 3.9. *Let $A_1, A_2, ...$ be a stationary ergodic sequence of $d \times d$ matrices such that ($\log^+(x) = \max(0, \log(x))$)*

$$\mathsf{E}\log^+ \|A_1\| < \infty. \tag{3.18}$$

Then there exist constants, the Lyapunov exponents, $-\infty \le \lambda_d \le \lambda_{d-1} \le ... \le \lambda_1 < \infty$ *with the following properties*

(a) *With probability one the random sets*

$$V(q,\omega) := \{v \in \mathbb{R}^d : \lim_{n\to\infty} \frac{1}{n} \log \|A_n ... A_1 v\| \le \lambda_q\}$$

are subspaces. The map $\omega \mapsto V(q,\omega)$ is measurable from the probability space into the Grassmann manifold, and if θ is the shift on the probability space form which $A_i(\theta\omega) = A_{i+1}(\omega)$, then

$$V(q,\theta\omega) = A_1(\omega)V(q,\omega).$$

(b) $\dim V(q) = \#\{i : \lambda_i \le \lambda_q\}$

(c) *Set $V(d+1) = \{0\}$ and let $i_1 = 1 < i_2 < ... < i_{p+1} = d+1$ be the unique indices at which λ_1 jumps, i.e. $\lambda_1 = \lambda_2 = ... = \lambda_{i_2-1} > \lambda_{i_2} = \lambda_{i_2+1} = ... = \lambda_{i_3-1} > \lambda_{i_3}...$ Then for $v \in V(i_{s-1}) \setminus V_{i_s}$ one has*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A_n ... A_1 v\| = \lambda_{i_{s-1}}, \quad 2 \le s \le p+1.$$

(d) *The sequence of matrices*

$$\big(A_1^* ... A_n^* A_n ... A_1\big)^{1/(2n)}$$

converges almost surely to a limit matrix B with eigenvalues $\mu_1 = e^{\lambda_1}, ..., \mu_d = e^{\lambda_d}$. The orthogonal complement of $V(i_s)$ in $V(i_{s-1})$ is the eigenspace of B corresponding to $\mu_{i_{s-1}}$

(e) *If $\overline{\lim}_{n\to\infty} \frac{1}{n} \log \|A_n ... A_1\| > 0$ and* $\det(A_1) = 1$ *with probability one, then $\lambda_d < 0 < \lambda_1$ so that $V(d)$, the subspace corresponding to λ_d, is a proper non-empty subspace of $\mathbb{R}^d$.*

One of the main messages of this theorem is that the solutions of random linear recursions grow with one of d possible exponential rates, which are deterministic (!). For any fixed deterministic initial condition, the Lyapunov exponent is determined at random (since $V(q,\omega)$ are random).

As was mentioned before, one may study the stability problem under the stationary probability P^s, in the sense that any P^s-a.s. statement will automatically hold P-a.s. as $\mathsf{P} \ll \mathsf{P}^s$ (since $\nu \ll \mu$). Under P^s the solution of the Zakai equation (3.12) is exactly in the scope of the MET, namely we deal with the stationary sequence of random matrices $A_n := G(Y_n)\Lambda^*$. The condition (3.18) is satisfied if all the densities $g_i(y)$ are e.g. bounded or sufficiently integrable. Hence the limits

$$\lim_{n\to\infty} \frac{1}{n} \log |\rho_n|, \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \log |\bar{\rho}_n|$$

exist. In our case, the matrix $A_n^*A_n$ has nonnegative entries and hence, by Perron-Frobenius theorem, its largest eigenvalue is real and the corresponding eigenvector has nonnegative entries. The same holds for $(A_n^*A_n)^{1/(2n)}$ as well as for $\lim_{n\to\infty}(A_n^*A_n)^{1/(2n)}$ (which exists by (d)). Hence by (d) the orthogonal compliment of $V(i_2)$ in $V(1)$ contains a vector with nonnegative entries, which means that it contains all the vectors with strictly positive entries (since they have nonzero projection on a vector with nonzero vector with nonnegative entries). If we also assume e.g. (**a-2**) from Theorem 3.3, then eventually both $G(Y_n)\Lambda^*...G(Y_1)\Lambda^*$ will enter the interior of $\mathcal{S}^{d-1}$ and thus in fact for any $p\in\mathcal{S}^{d-1}$ (including the boundary) $\lim_{n\to\infty}\frac{1}{n}\log|\prod_{k=1}^n G(Y_k)\Lambda^* p| = \lambda_1$. In particular, the two latter limits in the right hand side of (3.17) coincide and equal the top Lyapunov exponent corresponding to (3.12):

$$\lambda_1 = \frac{1}{n}\log|\rho_n| = \frac{1}{n}\log|\bar\rho_n|.$$

The *exterior* product $\rho_n\wedge\bar\rho_n$ (i.e. the matrix with entries $\rho_n(i)\bar\rho_n(j)-\rho_n(j)\bar\rho_n(i)$) can be associated in this case with the antisymmetric matrix $\rho_n\bar\rho_n^*-\bar\rho_n\rho_n^*$, which satisfies the equation

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

This is a *linear* random recursion and hence is also in the scope of MET. In fact, in this case one has [20]

$$\lim_{n\to\infty}\frac{1}{n}\log\|\rho_n\wedge\bar\rho_n\| \le \lambda_1+\lambda_2, \tag{3.19}$$

where λ_2 is the second Lyapunov exponent of (3.12). In this case, the generated random flow is not positive anymore and hence only inequality can be claimed. Then (3.17) suggests that the stability of the filter is controlled by the Lyapunov spectral gap of (3.12):

$$\gamma := \lim_{n\to\infty}\frac{1}{n}\log\|\pi_n-\bar\pi_n\| \le \lambda_1+\lambda_2-\lambda_1-\lambda_1 = \lambda_2-\lambda_1 \le 0.$$

The main difficulty is now to calculate (usually impossible) or estimate this gap. Theorem 3.10 below was inspired by Theorem 1.7 in [6], which proves the asymptotics (1.12) and (1.13), in the case when $g_i(u)$ are Gaussian with means $h(a_i)$ and variance σ^2. The authors used Feynman-Kac type formulae to derive these bounds. The presentation here follows [23], which takes a more classic route due to H.Furstenberg and R.Khasminskii.

With $\varepsilon\in(0,1)$, define the *slow* Markov chain $X^\varepsilon=(X_n^\varepsilon)_{n\ge0}$ with the transition probabilities

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

and initial distribution ν. The observation sequence Y^ε and the filtering processes π^ε and $\bar\pi^\varepsilon$ are defined by (1.1) and (1.4) with X replaced by X^ε and Λ by Λ^ε. The chain X^ε is ergodic if X is and its invariant measure equals μ, independently of ε. Clearly ε controls the transitions rate of X^ε - the smaller ε the less frequent are its transitions. To emphasize the dependence on ε, $\gamma(\varepsilon)$ is written for γ, etc.

[20]the exterior product $\|a\wedge b\|$ is twice the area formed by the vectors $a,b\in\mathbb{R}^d$. The area between A^na and A^nb, where A is a fixed deterministic matrix is known to grow not faster than the sum of the largest absolute values of its eigenvalues. The formula (3.19) can be seen as its (very nontrivial!) analog. In fact, similar formulae are available for the k-th exterior products.

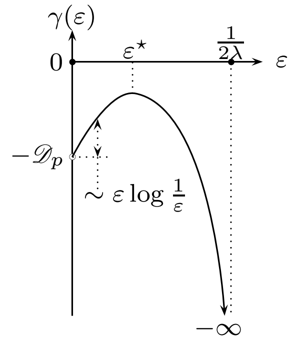

FIGURE 1. $\gamma(\varepsilon)$ for the BSC example

Theorem 3.10. *Assume that X is ergodic and the noise densities $g_i(u)$*

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

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

$$\gamma(\varepsilon) \le -\sum_{i=1}^{d} \mu_i \min_{j\neq i} \mathscr{D}(g_i \parallel g_j) + o(1), \quad \varepsilon \to 0, \tag{3.20}$$

where $\mathscr{D}(g_i \parallel g_j) = \int_{\mathbb{R}} g_i(u) \log \dfrac{g_i}{g_j}(u)\varphi(du)$ are the Kullback-Leibler relative entropies. For $d = 2$ the asymptotic (3.20) *is precise, i.e.*

$$\gamma(\varepsilon) = -\mu_1 \mathscr{D}(g_1 \parallel g_2) - \mu_2 \mathscr{D}(g_2 \parallel g_1) + o(1), \quad \varepsilon \to 0. \tag{3.21}$$

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

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

$$\gamma(0+) = \overline{\lim_{\varepsilon\to 0}}\, \gamma(\varepsilon) < \gamma(0) = 0,$$

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

Surprising as it may seem at first glance, this phenomenon is quite natural for signals with discrete state space and can be explained as follows. The distance $\|\pi_n^\varepsilon - \bar{\pi}_n^\varepsilon\|$ never increases and tends to decrease exponentially fast whenever X_n^ε resides in a state with distinct noise probability distribution. Since the average occupation time of this "synchronizing" state does not depend on ε, the decay remains exponential with nonzero average rate. The "dual" manifestation of this phenomenon is that the filter stability improves, when the signal-to-noise ratio is increased in the setting of (1.12) (see [31, 6]).

2. As demonstrated in the following example, $\gamma(\varepsilon)$ may have a maximum at some $\varepsilon^\star > 0$ or, in other words, stability may improve when the chain is slowed down! This provides yet another evidence against the false intuition, directly relating stability of the filter to ergodic properties of the signal (as was explained in the introduction). The reason for such behavior stems from the delicate interplay between two stabilizing mechanisms: ergodicity of the signal and synchronizing effect of the observations. The first dominates the second for the faster chain, and vise versa when the chain is slow.

Example 3.11. Consider the so called Binary Symmetric Channel (BSC) model, for which $X_n \in \{0,1\}$ is a symmetric chain with the jump probability λ and $Y_n = (X_n - \xi_n)^2$, where ξ is an i.i.d. $\{0,1\}$ binary sequence with $\mathsf{P}(\xi_1 = 1) = p \in (0, 1/2)$. Let X^ε and Y^ε denote the "slow" instances as defined above. In this case more can be said about the convergence in (3.21) (see the proof below), namely

$$\gamma(\varepsilon) \ge -\mathscr{D}_p + \frac{4\lambda\big(\log(2) - h(p)\big)}{\mathscr{D}_p}\varepsilon \log \varepsilon^{-1}\big(1 + o(1)\big), \quad \varepsilon \to 0. \tag{3.22}$$

where $\mathscr{D}_p := p \log \dfrac{p}{1-p} + (1-p)\log \dfrac{1-p}{p}$ and $h(p) = -p\log p - (1-p)\log(1-p)$. On the other hand, $\gamma(\varepsilon) \le \log(1 - 2\varepsilon\lambda) \to -\infty$ as $\varepsilon \to 1/(2\lambda)$ (at $\varepsilon = 1/(2\lambda)$ the chain is just an i.i.d. sequence). Since the second term in the expansion of $\gamma(\varepsilon)$ in (3.22) is positive and by (3.21) $\gamma(\varepsilon) \to -\mathscr{D}_p$ as $\varepsilon \to 0$, one gets the qualitative behavior depicted in Figure 1. □

The statement of Theorem 3.10 follows from (3.17) and asymptotic expressions derived in Lemmas 3.12 and 3.13 below.

3.2.1. *Asymptotic expression for* $\lambda_1(\varepsilon)$.

Lemma 3.12. *For any* $\varepsilon > 0$ *the Markov process* $(X^\varepsilon, \pi^\varepsilon)$ *has a unique stationary invariant measure* $\mathcal{M}^\varepsilon$. *The top Lyapunov exponent is given by*

$$\lambda_1(\varepsilon) = \int_{\mathcal{S}^{d-1}} \sum_{i=1}^d \big(\Lambda^{\varepsilon *} u\big)_i \int_{\mathbb{R}} g_i(y) \log \big|G(y)\Lambda^{\varepsilon *} u\big| \varphi(dy) \mathcal{M}^\varepsilon_\pi(du), \tag{3.23}$$

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

$$\lim_{\varepsilon \to 0} \int \big(\mathbf{1}_{\{x \in \mathcal{J}_j\}} - \sum_{\ell : a_\ell \in \mathcal{J}_j} u_\ell\big)^2 \mathcal{M}^\varepsilon(dx, du) = 0 \tag{3.24}$$

and in particular

$$\lim_{\varepsilon \to 0} \lambda_1(\varepsilon) = \sum_{i=1}^d \mu_i \int_{\mathbb{R}} g_i(y) \log g_i(y) \varphi(dy). \tag{3.25}$$

Proof. Ergodicity of $(X^\varepsilon, \pi^\varepsilon)$ essentially follows from the stability (3.10) and was already mentioned in Theorem 3.3 above (Corollary 3.5, see also [22]). Concentration properties of $\mathcal{M}^\varepsilon_\pi$ have been studied in [41], when all the noises are distinct, i.e. $\mathscr{D}(g_i \parallel g_j) > 0$ for all $i \neq j$, which is not necessarily the case here.

Let $\widetilde{X}^\varepsilon$ be the stationary chain (i.e. $\widetilde{X}_0 \sim \mu$) and $\widetilde{\pi}^\varepsilon$ the corresponding optimal filtering process, generated by (2.1) subject to $\widetilde{\pi}^\varepsilon_0 = \mu$. For an $f : \mathbb{S} \to \mathbb{R}$ and $n, m \geq 0$ ($\widetilde{Y}^\varepsilon$ denotes the observations corresponding to $\widetilde{X}^\varepsilon$)

$$\begin{aligned}&\mathsf{E}\big(f(\widetilde{X}^\varepsilon_{n+m}) - \widetilde{\pi}^\varepsilon_{n+m}(f)\big)^2 = \mathsf{E}\Big(f(\widetilde{X}^\varepsilon_{n+m}) - \mathsf{E}\big(f(\widetilde{X}^\varepsilon_{n+m})\big|\mathscr{F}^{\widetilde{Y}^\varepsilon}_{n+m}\big)\Big)^2 \leq \\ &\mathsf{E}\Big(f(\widetilde{X}^\varepsilon_{n+m}) - \mathsf{E}\big(f(\widetilde{X}^\varepsilon_{n+m})\big|\mathscr{F}^{\widetilde{Y}^\varepsilon}_{[m+1,n+m]}\big)\Big)^2 \stackrel{\dagger}{=} \mathsf{E}\Big(f(\widetilde{X}^\varepsilon_{n}) - \mathsf{E}\big(f(\widetilde{X}^\varepsilon_{n})\big|\mathscr{F}^{\widetilde{Y}^\varepsilon}_{n}\big)\Big)^2 = \\ &\mathsf{E}\big(f(\widetilde{X}^\varepsilon_n) - \widetilde{\pi}^\varepsilon_n(f)\big)^2,\end{aligned}$$

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

$$\begin{aligned}&\int \big(f(x) - u(f)\big)^2 \mathcal{M}^\varepsilon(dx, du) = \\ &\qquad\qquad \lim_{n\to\infty} \mathsf{E}\big(f(\widetilde{X}^\varepsilon_n) - \widetilde{\pi}^\varepsilon_n(f)\big)^2 \leq \mathsf{E}\big(f(\widetilde{X}^\varepsilon_m) - \widetilde{\pi}^\varepsilon_m(f)\big)^2. \end{aligned}\tag{3.26}$$

Define

$$\widehat{\pi}^\varepsilon_n(i) = \frac{\mu_i \prod_{k=1}^n g_i(\widetilde{Y}^\varepsilon_k)}{\sum_{j=1}^d \mu_j \prod_{k=1}^n g_j(\widetilde{Y}^\varepsilon_k)}, \quad i = 1, ..., d$$

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

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

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

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

For $f(x) := \mathbf{1}_{\{x\in\mathcal{J}_j\}}$ the latter and (3.26) implies

$$\varlimsup_{\varepsilon\to 0} \int \big(\mathbf{1}_{\{x\in\mathcal{J}_j\}} - \sum_{\ell: a_\ell\in\mathcal{J}_j} u_\ell\big)^2 \mathcal{M}^\varepsilon(dx, du) \leq \mathsf{E}\big(f(\widetilde{X}_0) - \widehat{\pi}_m(f)\big)^2 \xrightarrow[m\to\infty]{} 0,$$

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

Once the existence of ergodic stationary pair $(X^\varepsilon, \pi^\varepsilon)$ is established[21] one may use it to realize the limit λ_1 by means of the approach due to H.Furstenberg and R.Khasminskii

[21] such pair can be generated by taking both X_0 and π_0 randomly distributed according to $\mathcal{M}^\varepsilon$ and its definition can be extended to the negative times by the usual arguments. Note that this is different from $(\widetilde{X}^\varepsilon, \widetilde{\pi}^\varepsilon)$ used in the proof of $\mathcal{M}^\varepsilon$ concentration

(see e.g. [40]). The idea is to study the growth rate of ρ^ε_n by projecting it on the unit sphere ($\mathcal{S}^{d-1}$ in this case):

$$|\rho^\varepsilon_n| = \big|G(Y^\varepsilon_n)\Lambda^{\varepsilon*}\rho^\varepsilon_{n-1}\big| = |\rho^\varepsilon_{n-1}|\Big|G(Y^\varepsilon_n)\Lambda^{\varepsilon*}\frac{\rho^\varepsilon_{n-1}}{|\rho^\varepsilon_{n-1}|}\Big| = |\rho^\varepsilon_{n-1}|\big|G(Y^\varepsilon_n)\Lambda^{\varepsilon*}\pi^\varepsilon_{n-1}\big|.$$

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

$$\begin{aligned}
&\lambda_1(\varepsilon) = \lim_{n\to\infty}\frac{1}{n}\log|\rho^\varepsilon_n| = \lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^n \log\big|G(Y^\varepsilon_n)\Lambda^{\varepsilon*}\pi^\varepsilon_{n-1}\big| = \mathsf{E}\log\big|G(Y^\varepsilon_1)\Lambda^{\varepsilon*}\pi^\varepsilon_0\big| = \\
&\mathsf{E}\sum_{i=1}^d \mathbf{1}_{\{X^\varepsilon_1=a_i\}}\log\big|G\big(\xi_1(i)\big)\Lambda^{\varepsilon*}\pi^\varepsilon_0\big| = \mathsf{E}\sum_{i=1}^d \mathsf{P}\big(X^\varepsilon_1=a_i|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]}\big)\log\big|G\big(\xi_1(i)\big)\Lambda^{\varepsilon*}\pi^\varepsilon_0\big| = \\
&\mathsf{E}\sum_{i=1}^d \big(\Lambda^{\varepsilon*}\pi^\varepsilon_0\big)_i\log\big|G\big(\xi_1(i)\big)\Lambda^{\varepsilon*}\pi^\varepsilon_0\big|. \qquad (3.27)
\end{aligned}$$

The latter expression is nothing but (3.23). The asymptotic (3.25) follows from $\Lambda^\varepsilon = I + O(\varepsilon)$ and the concentration (3.24) of $\mathcal{M}^\varepsilon$ as $\varepsilon\to 0$, since $g_i(u)$'s coincide φ-almost surely for all $a_i\in\mathcal{J}_j$ for any j and the X-marginal of $\mathcal{M}^\varepsilon$ is given by $\mathcal{M}^\varepsilon_X(dx) = \sum_{i=1}^d \mu_i\delta_{a_i}(dx)$. □

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

Lemma 3.13. *For any $\nu,\bar\nu\in\mathcal{S}^{d-1}$*

$$\lim_{n\to\infty}\frac{1}{n}\log|\rho^\varepsilon_n\wedge\bar\rho^\varepsilon_n| \le \sum_{i=1}^d \mu_i\max_{k\ne m}\int_{\mathbb{R}} g_i(u)\log\big(g_m(u)g_k(u)\big)\varphi(du) + o(1), \quad \varepsilon\to 0. \qquad (3.28)$$

In the case $d=2$

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log|\rho^\varepsilon_n\wedge\bar\rho^\varepsilon_n| &= \log(1-\varepsilon\lambda_{12}-\varepsilon\lambda_{21}) + \\
&\mu_1\int_{\mathbb{R}} g_1(u)\log\big(g_1(u)g_2(u)\big)\varphi(du) + \mu_2\int_{\mathbb{R}} g_2(u)\log\big(g_1(u)g_2(u)\big)\varphi(du). \qquad (3.29)
\end{aligned}$$

Proof. The process $R^\varepsilon_n := \rho^\varepsilon_n\wedge\bar\rho^\varepsilon_n$ evolves in the space of antisymmetric matrices (with zero diagonal) and satisfies the linear equation

$$R^\varepsilon_n = G(Y^\varepsilon_n)\Lambda^{\varepsilon*}R^\varepsilon_{n-1}\Lambda^\varepsilon G(Y^\varepsilon_n), \quad R^\varepsilon_0 = \nu\wedge\bar\nu,$$

or in the componentwise notation

$$R^\varepsilon_n(i,j) = \sum_{1\le k\ne\ell\le d} g_k(Y^\varepsilon_n)\lambda^\varepsilon_{ki}R^\varepsilon_{n-1}(k,\ell)\lambda^\varepsilon_{\ell j}g_\ell(Y^\varepsilon_n), \quad i\ne j.$$

Unlike in the case of (3.12), it is not clear whether the limit $\lim_{n\to\infty}\frac{1}{n}\log|R^\varepsilon_n|$ depends on $\nu,\bar\nu$ or $\Pi^\varepsilon_n = R^\varepsilon_n/|R^\varepsilon_n|$ has any useful concentration properties as $\varepsilon\to 0$. However the technique used in the previous section still gives the upper bound. With a fixed integer $r \geq 1$

$$
\begin{aligned}
|R^\varepsilon_n| = & |R^\varepsilon_{n-r}| \Big| \Big\{ G(Y^\varepsilon_n)\Lambda^{\varepsilon *} ... \Big\{ G(Y^\varepsilon_{n-r+1})\Lambda^{\varepsilon *}\Pi^\varepsilon_{n-r}\Lambda^\varepsilon G(Y^\varepsilon_{n-r+1}) \Big\} ... \Lambda^\varepsilon G(Y^\varepsilon_n) \Big\} \Big| \leq \\
& |R^\varepsilon_{n-r}| \Big(\sum_{i\neq j} \big|\Pi^\varepsilon_{n-r}(i,j)\big| \prod_{m=n-r+1}^{n} g_i(Y^\varepsilon_m) g_j(Y^\varepsilon_m) + c_1(r)\varepsilon \Big) \leq \\
& |R^\varepsilon_{n-r}| \Big(\max_{i\neq j} \prod_{m=n-r+1}^{n} g_i(Y^\varepsilon_m) g_j(Y^\varepsilon_m) + c_1(r)\varepsilon \Big), \quad n \geq r
\end{aligned}
$$

with a constant $c_1(r) > 0$, depending only on r (due to assumption (a1)). By the MET the limit $\lim_{n\to\infty} \frac{1}{n} \log |R^\varepsilon_n|$ exists P-a.s and hence (recall the definitions of $\widetilde{Y}^\varepsilon$ and A^ε_r on page 29)

$$
\begin{aligned}
& \lim_{n\to\infty} \frac{1}{n} \log |R^\varepsilon_n| = \lim_{\ell\to\infty} \frac{1}{\ell r} \log |R^\varepsilon_{\ell r}| \leq \\
& \leq \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{k=1}^{\ell} \frac{1}{r} \log \Big(\max_{i\neq j} \prod_{m=kr-r+1}^{kr} g_i(Y^\varepsilon_m) g_j(Y^\varepsilon_m) + c_1(r)\varepsilon \Big) \stackrel{\dagger}{=} \\
& \frac{1}{r} \mathsf{E} \log \Big(\max_{i\neq j} \prod_{m=1}^{r} g_i(\widetilde{Y}^\varepsilon_m) g_j(\widetilde{Y}^\varepsilon_m) + c_1(r)\varepsilon \Big) \leq \\
& \frac{1}{r} \mathsf{E} \mathbf{1}_{\{A^\varepsilon_r\}} \log \Big(\max_{i\neq j} \prod_{m=1}^{r} g_i(\widetilde{Y}^\varepsilon_m) g_j(\widetilde{Y}^\varepsilon_m) + c_1(r)\varepsilon \Big) + c_2(r)\big(1 - \mathsf{P}_\mu(A^\varepsilon_r)\big) \leq \\
& \frac{1}{r} \sum_{\ell=1}^{d} \mu_\ell \mathsf{E} \log \Big(\max_{i\neq j} \prod_{m=1}^{r} g_i\big(\xi_m(\ell)\big) g_j\big(\xi_m(\ell)\big) + c_1(r)\varepsilon \Big) + c_3(r)\big(1 - \mathsf{P}_\mu(A^\varepsilon_r)\big) \xrightarrow{\varepsilon\to 0} \\
& \sum_{\ell=1}^{d} \mu_\ell \mathsf{E} \max_{i\neq j} \frac{1}{r} \log \prod_{m=1}^{r} g_i\big(\xi_m(\ell)\big) g_j\big(\xi_m(\ell)\big),
\end{aligned}
$$

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

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

Since "max" is a continuous function

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

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

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

Putting all parts together one gets the bound (3.28). In the case $d=2$, the process R^ε_n is one dimensional and all the calculations can be carried out exactly, leading to the expression (3.29). □

3.2.3. *Proof of* (3.22). When the observation process Y^ε_n takes values in a discrete alphabet $\mathbb{S}'=\{b_1,...,b_{d'}\}$, the conditional densities (with respect to the point measure $\varphi(dy)=\sum_{i=1}^{d'}\delta_{b_i}(dy)$) are of the form

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

and hence by (3.27) ($\pi^\varepsilon_{1|0}:=\Lambda^{\varepsilon *}\pi^\varepsilon_0$ for brevity)

$$\lambda_1(\varepsilon)=\mathsf{E}\log\big|G(Y^\varepsilon_1)\Lambda^{\varepsilon *}\pi^\varepsilon_0\big|=\mathsf{E}\sum_{j=1}^{d'}\mathbf{1}_{\{Y^\varepsilon_1=b_j\}}\log\Big(\sum_{i=1}^{d}p_{ij}\pi^\varepsilon_{1|0}(i)\Big)=$$
$$\mathsf{E}\sum_{j=1}^{d'}\mathsf{P}\big(Y^\varepsilon_1=b_j|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]}\big)\log\mathsf{P}\big(Y^\varepsilon_1=b_j|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]}\big)=:-\mathscr{H}(Y^\varepsilon),\quad (3.30)$$

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

Consider now the special case, when X^ε and Y^ε take values in $\mathbb{S}=\{0,1\}$ and $p=\mathsf{P}(Y^\varepsilon_n=i|X^\varepsilon_n=j)$ for $i\neq j$. The vector π^ε_n is one dimensional and hence $\mathsf{P}\big(Y^\varepsilon_1=1|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]}\big)=(1-p)\pi^\varepsilon_{1|0}+p(1-\pi^\varepsilon_{1|0})$, where

$$\pi^\varepsilon_{1|0}:=\mathsf{P}\big(X^\varepsilon_1=1|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]}\big)=(1-\varepsilon\lambda_{10})\pi^\varepsilon_0+\varepsilon\lambda_{01}(1-\pi^\varepsilon_0)\quad (3.31)$$

and $\pi^\varepsilon_0:=\mathsf{P}(X^\varepsilon_0=1|\mathscr{F}^{Y^\varepsilon}_{(-\infty,0]})$ are redefined for brevity.

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

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

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

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

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

$$H(p,q)\geq h(p)+\frac{\log(2)-h(p)}{1/2}\min(q,1-q),\quad p,q\in[0,1].\quad (3.32)$$

By Theorem 1 in [41] for the symmetric chain X^ε with jump probability λ and $p\neq 1/2$

$$\mathsf{E}\min(\pi_0^\varepsilon, 1-\pi_0^\varepsilon) = \mathsf{P}\big(X_0^\varepsilon \neq \mathrm{argmax}_i \pi_0^\varepsilon(i)\big) = \frac{\lambda}{\mathscr{D}_p}\varepsilon\log\varepsilon^{-1}\big(1+o(1)\big), \quad \varepsilon\to 0, \tag{3.33}$$

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

$$\mathscr{H}(Y^\varepsilon) = \mathsf{E}H(p,\pi_{1|0}^\varepsilon) = \mathsf{E}H(p,\pi_0^\varepsilon) + O(\varepsilon), \quad \varepsilon\to 0$$

where the latter asymptotic follows from (3.31), since $H(p,q)$ is differentiable in q.

Now (3.32) and (3.33) imply

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

and (3.22) follows from (3.17), (3.29) and (3.30). □

Remark 3.14. This Lyapunov exponents approach does not actually require neither ergodicity of the signal nor compactness of the state space. With some sophistication and under certain structural constraints both cases can be treated - [3], [16], [36].

3.3. **Conditional time reversal.** As was already mentioned above, the assumption $\nu \ll \bar\nu$ implies $\mathsf{P}\ll\bar{\mathsf{P}}$ with (recall that we work with coordinate process on the canonical space)

$$\frac{d\mathsf{P}}{d\bar{\mathsf{P}}}(x,y) = \frac{d\nu}{d\bar\nu}(x_0), \quad \mathsf{P}-a.s.$$

Consequently $\mathsf{P}^Y \ll \bar{\mathsf{P}}^Y$ and $\mathsf{P}^Y_n \ll \bar{\mathsf{P}}^Y_n$ and

$$\frac{d\mathsf{P}^Y_n}{d\bar{\mathsf{P}}^Y_n}(Y) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)|\mathscr{F}^Y_n\Big), \quad \text{and} \quad \frac{d\mathsf{P}^Y}{d\bar{\mathsf{P}}^Y}(Y) = \lim_{n\to\infty}\frac{d\mathsf{P}^Y_n}{d\bar{\mathsf{P}}^Y_n}(Y) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_{[1,\infty)}\Big),$$

where $\mathscr{F}^Y_{[1,\infty)} := \bigvee_{n\geq 1}\mathscr{F}^Y_n$. If in addition, $\nu\sim\bar\nu$, then the above measures are absolutely continuous as well and the Radon-Nikodym derivatives are positive P-a.s. and $\bar{\mathsf{P}}$-a.s. We accept the latter assumption below for simplicity, though the weaker $\nu\ll\bar\nu$ is essentially needed (the reader is referred to [24] for details). When $\nu\sim\bar\nu$, both filtering processes π and $\bar\pi$ are well defined both on $(\Omega,\mathscr{F},\mathsf{P})$ and $(\Omega,\mathscr{F},\bar{\mathsf{P}})$ as the solutions of (2.1) subject to ν and $\bar\nu$ respectively. Obviously $\bar\pi$ is the conditional distribution of X_n given $\mathscr{F}^Y_n$ under $\bar{\mathsf{P}}$ and the "wrong" filtering under P (obtained by starting (2.1) from $\bar\nu$, while (X,Y) corresponds to ν). Analogously π is the conditional distribution of X_n given $\mathscr{F}^Y_n$ under P and is the "wrong" filtering under $\bar{\mathsf{P}}$. The formula for transformation of the conditional expectations under a.c. change of measure from Lemma 2.1 implies

$$\pi_n(f) = \mathsf{E}\big(f(X_n)|\mathscr{F}^Y_n\big) = \frac{\bar{\mathsf{E}}\Big(f(X_n)\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)}{\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)} \tag{3.34}$$

for any measurable bounded f. Then

$$\begin{aligned}&\mathsf{E}\big|\pi_n(f)-\bar\pi_n(f)\big| = \bar{\mathsf{E}}\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\big|\pi_n(f)-\bar\pi_n(f)\big| = \\ &\bar{\mathsf{E}}\Big|\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\pi_n(f)-\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\bar{\mathsf{E}}\big(f(X_n)|\mathscr{F}^Y_n\big)\Big| = \\ &\bar{\mathsf{E}}\Big|\bar{\mathsf{E}}\Big(f(X_n)\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)-\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\bar{\mathsf{E}}\big(f(X_n)|\mathscr{F}^Y_n\big)\Big|,\end{aligned}$$

where the latter equality is due to (3.34). Let $|f|\le C$ for definiteness, then the latter implies [22]

$$\begin{aligned}&\mathsf{E}\big|\pi_n(f)-\bar\pi_n(f)\big| = \\ &\bar{\mathsf{E}}\Big|\bar{\mathsf{E}}\Big(f(X_n)\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\vee X_n\Big)\Big|\mathscr{F}^Y_n\Big)-\bar{\mathsf{E}}\Big(f(X_n)\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\Big|\mathscr{F}^Y_n\Big)\Big| \le \\ &\bar{\mathsf{E}}\Big|f(X_n)\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\vee X_n\Big)-f(X_n)\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\Big| \le \\ &C\bar{\mathsf{E}}\Big|\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\vee X_n\Big)-\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big)\Big|.\end{aligned} \tag{3.35}$$

By the Markov property of (X,Y),

$$\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\vee X_n\Big) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_{[1,n]}\vee\mathscr{F}^X_{[n,\infty)}\Big) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_{[1,\infty)}\vee\mathscr{F}^X_{[n,\infty)}\Big)$$

where $\mathscr{F}^X_{[n,\infty)} := \sigma\{X_n, X_{n+1},...\}$. By the martingale convergence theorem[23] $\bar{\mathsf{P}}$-a.s.

$$\begin{aligned}&\lim_{n\to\infty}\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_{[1,\infty)}\vee\mathscr{F}^X_{[n,\infty)}\Big) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\bigcap_{n\ge0}\mathscr{F}^Y_{[1,\infty)}\vee\mathscr{F}^X_{[n,\infty)}\Big), \\ &\lim_{n\to\infty}\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_n\Big) = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar\nu}(X_0)\big|\mathscr{F}^Y_{[1,\infty)}\Big).\end{aligned} \tag{3.36}$$

Thus (3.35) implies stability of (2.1) (in the sense $\lim_{n\to\infty}\mathsf{E}|\pi_n(f)-\bar\pi_n(f)|=0$ for any measurable and bounded function f) if

$$\bigcap_{n\ge0}\mathscr{F}^Y_{[1,\infty)}\vee\mathscr{F}^X_{[n,\infty)} = \mathscr{F}^Y_{[1,\infty)}, \quad \bar{\mathsf{P}}-a.s. \tag{3.37}$$

If the tail σ-algebra of X is $\bar{\mathsf{P}}$-a.s. empty: $\bigcap_{n\ge0}\mathscr{F}^X_{[n,\infty)}=\{\emptyset,\Omega\}$ $\bar{\mathsf{P}}$-a.s. then the latter is a particular case of the following question. Let $\mathscr{G}_n$ be a decreasing sequence of σ-algebras and $\mathscr{F}$ be a fixed σ-algebra. Is the following true (per se or P-a.s.)

$$\bigcap_{n\ge0}\mathscr{F}\vee\mathscr{G}_n \overset{?}{=} \mathscr{F}\vee\bigcap_{n\ge0}\mathscr{G}_n. \tag{3.38}$$

Little is known about the conditions, under which this relation holds, and in fact, according to D.Williams [76], it "...trapped up even Kolmogorov and Wiener" (see Sinai [66, p. 837]

[22] $\mathscr{F}^Y_n\vee X_n$ is short for the more proper notation $\mathscr{F}^Y_n\vee\sigma\{X_n\}$

[23] note that the filtration $\mathscr{F}^Y_{[1,\infty)}\vee\mathscr{F}^X_{[n,\infty)}$ is decreasing with n, while $\mathscr{F}^Y_n$ is increasing, so actually the direct and reverse martingale convergence theorems are used here

for some details). The reader can find a discussion concerning (3.38) in von Weizsäcker [75] (note, however, that the counterexample there is incorrect). Other works related to this issue are [77],[39], [34]. Different counterexamples to (3.38) appeared in Exercise 4.12 in Williams [76] and in [62]. In fact Example 1.1 is nothing but another situation when (3.38) fails. $\mathscr{F}^Y_{[1,\infty)}$ determines all the transitions $\{1,3\} \leftrightarrow \{2,4\}$ of X but does not tell where X_n resides for each $n \geq 0$. Specifying the value of X_n at any $n \geq 0$, "pins" this uncertainty and thus reveals all the information about X. In other words,

$$\mathscr{F}^Y_{[1,\infty)} \vee \mathscr{F}^X_{[n,\infty)} = \mathscr{F}^Y_{[1,\infty)} \vee \mathscr{F}^X_{[1,\infty)} \not\supseteq \mathscr{F}^Y_{[1,\infty)}$$

with strict inclusion. The signal X in this example is an ergodic finite state Markov chain and hence it is geometrically ergodic and its tail σ-algebra is empty, i.e. $\bigcap_{n\geq 0} \mathscr{F}^X_{[n,\infty)} = \{\emptyset, \Omega\}$. Thus stability of the filter is not implied merely by ergodicity of the signal! The validity of the relation similar to (3.37) was implicitly claimed by H.Kunita in [45] (under certain additional technical conditions) and this gap in the proof currently lacks justification (see [8], [14]).

However (3.35) hints that instead of studying the stability of the conditional distribution of X_n given $\mathscr{F}^Y_n$, one may study the *time reversed* conditional distribution of X_0, given $\mathscr{F}^Y_n \vee X_n$. It turns out that the latter has interesting dynamics, somewhat more amenable to stability analysis. The following theorem is taken from [24].

Theorem 3.15. *Assume that X is an ergodic chain and $\nu \sim \bar{\nu}$, then*

$$\lim_{n\to\infty} \frac{1}{n} \log \|\pi_n - \bar{\pi}_n\| \leq -\frac{\lambda_\diamond}{\lambda^*}, \tag{3.39}$$

where $\lambda_\diamond := \sum_{i=1}^d \mu_i \min_j \lambda_{ij}$.

Remark 3.16. This theorem states that the filter is stable, if Λ is m-primitive and at least one of its rows has all nonzero entries. The assertion is independent of the observation densities structure, just like (3.11). Though (3.39) is weaker than (3.11), both are stronger than just ergodicity of the chain X. This raises the following question: what is the necessary and sufficient condition for the filtering stability only in terms of the ergodic properties of the chain, or in other words, what is the weakest ergodic property to be inherited by the filter ?

Proof. Define $q_n(i,j) := \mathsf{P}(X_0 = a_i | \mathscr{F}^Y_n, X_n = a_j)$. These backward probabilities satisfy the following recursions (derived via Bayes formulae - see Lemma 3.1, [24]): for each $i = 1, ..., d$

$$q_n(i,j) = \frac{\sum_{\ell=1}^d \lambda_{\ell j} \pi_{n-1}(\ell) q_{n-1}(i,\ell)}{\sum_{\ell=1}^d \lambda_{\ell j} \pi_{n-1}(\ell)}, \quad n \geq 1, \tag{3.40}$$

subject to $q_0(i,j) := \mathbf{1}_{\{i=j\}}$. These recursions are linear in q with the time inhomogeneous coefficients depending on π_n. Namely let $q_n(i)$ denote the vector with entries $q_n(i,j)$, $j = 1, ..., d$, then

$$q_n(i) = Q_{n-1} q_{n-1}(i), \quad n \geq 1$$

where Q_n is the matrix with entries

$$Q_n(j,k) = \frac{\lambda_{kj} \pi_n(k)}{\sum_{\ell=1}^d \lambda_{\ell j} \pi_n(\ell)}.$$

Each row of Q_n sums up to 1 and hence Q_n is a (random) transition probability matrix. For a fixed i, introduce the upper and lower envelopes of $q_n(i,j)$:

$$q_n^{\max}(i) := \max_j q_n(i,j), \quad \text{and} \quad q_n^{\min}(i) := \min_j q_n(i,j).$$

Then we have

$$\begin{aligned} q_n(i,j) - q_n(i,j') = \sum_{\ell=1}^d Q_{n-1}(j,\ell) q_{n-1}(i,\ell) - \sum_{\ell=1}^d Q_{n-1}(j',\ell) q_{n-1}(i,\ell) = \\ q_n^{\max}(i) - q_n^{\min}(i) - \sum_{\ell=1}^d Q_{n-1}(j,\ell)\big(q_{n-1}^{\max}(i) - q_{n-1}(i,\ell)\big) - \\ \sum_{\ell=1}^d Q_{n-1}(j',\ell)\big(q_{n-1}(i,\ell) - q_{n-1}^{\min}(i)\big) \end{aligned}$$

Define $\Delta_n(i) := q_n^{\max}(i) - q_n^{\min}(i) > 0$ and $\alpha_n(i,\ell) := \dfrac{q_n(i,\ell) - q_n^{\min}(i)}{\Delta_n(i)}$, so that

$$\begin{aligned} q_n(i,j) - q_n(i,j') = \Delta_{n-1}(i)\Big(1 - \sum_{\ell=1}^d \Big(Q_{n-1}(j,\ell)\big(1-\alpha_{n-1}(i,\ell)\big) + Q_{n-1}(j',\ell)\alpha_{n-1}(i,\ell)\Big)\Big) \\ \le \Delta_{n-1}(i)\Big(1 - \sum_{\ell=1}^d Q_{n-1}(j,\ell) \wedge Q_{n-1}(j',\ell)\Big), \end{aligned}$$

where the latter inequality holds via minimization of the convex sum (recall that $\alpha_n \in [0,1]$). Since the latter holds for any j and j', in particular we have

$$\Delta_n(i) \le \Delta_{n-1}(i)\Big(1 - \sum_{\ell=1}^d Q_{n-1}(j,\ell) \wedge Q_{n-1}(j',\ell)\Big). \tag{3.41}$$

Taking a closer look at the expressions in the sum, we obtain

$$\begin{aligned} Q_{n-1}(j,\ell) \wedge Q_{n-1}(j',\ell) = \frac{\lambda_{\ell j}\pi_n(\ell)}{\sum_{\ell=1}^d \lambda_{\ell j}\pi_n(\ell)} \bigwedge \frac{\lambda_{\ell j'}\pi_n(\ell)}{\sum_{\ell=1}^d \lambda_{\ell j'}\pi_n(\ell)} \ge \\ \frac{\big(\lambda_{\ell j} \wedge \lambda_{\ell j'}\big)}{\lambda^*}\pi_n(\ell) \ge \frac{\min_j \lambda_{\ell j}}{\lambda^*}\pi_n(\ell) \end{aligned}$$

Iterating the latter inequality one gets

$$\max_{i,j,k} \big|q_n(i,j) - q_n(i,k)\big| \le \max_i \Delta_n(i) \le \prod_{k=1}^n \Big(1 - \sum_{\ell=1}^d \frac{\min_j \lambda_{\ell j}}{\lambda^*}\pi_{k-1}(\ell)\Big). \tag{3.42}$$

Finally, using the formula (3.34), one gets ($\bar{q}_n(i,j) := \bar{\mathsf{P}}(X_0 = a_i|\mathscr{F}^Y_n, X_n = a_j)$ and $0/0 = 0$ is agreed here)

$$\|\pi_n - \bar{\pi}_n\| = \sum_{j=1}^d |\pi_n(j) - \bar{\pi}_n(j)| =$$

$$\frac{1}{\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_n\Big)} \sum_{j=1}^d \bigg|\bar{\mathsf{E}}\Big(\mathbf{1}_{\{X_n=a_j\}}\frac{d\nu}{d\bar{\nu}}(X_0)|\mathscr{F}^Y_n\Big) - \bar{\pi}_n(j)\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_n\Big)\bigg| =$$

$$\frac{1}{\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_n\Big)} \sum_{j=1}^d \bigg|\bar{\pi}_n(j)\sum_{i=1}^d \frac{d\nu}{d\bar{\nu}}(a_i)\bar{q}_n(i,j) - \bar{\pi}_n(j)\sum_{i=1}^d\sum_{k=1}^d \frac{d\nu}{d\bar{\nu}}(a_i)\bar{q}_n(i,k)\bar{\pi}_n(k)\bigg| \le$$

$$\frac{1}{\bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_n\Big)} \sum_{j=1}^d \bar{\pi}_n(j)\sum_{k=1}^d \bar{\pi}_n(k)\sum_{i=1}^d \frac{d\nu}{d\bar{\nu}}(a_i)\big|\bar{q}_n(i,j) - \bar{q}_n(i,k)\big| \le$$

$$d\max_i(\bar{\nu}_i/\nu_i)\max_i(\nu_i/\bar{\nu}_i)\prod_{k=1}^n\Big(1 - \sum_{\ell=1}^d \frac{\min_j \lambda_{\ell j}}{\lambda^*}\bar{\pi}_{k-1}(\ell)\Big).$$

This implies (3.39):

$$\lim_{n\to\infty}\frac{1}{n}\log\|\pi_n - \bar{\pi}_n\| = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \log\Big(1 - \sum_{\ell=1}^d \frac{\min_j \lambda_{\ell j}}{\lambda^*}\bar{\pi}_{k-1}(\ell)\Big) \le$$

$$-\sum_{\ell=1}^d \frac{\min_j \lambda_{\ell j}}{\lambda^*}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \bar{\pi}_{k-1}(\ell)$$

where the latter inequality holds $\bar{\mathsf{P}}$-a.s. (and thus also P-a.s.) by the law of large numbers for the filtering process $\bar{\pi}_n$, which is the exact conditional distribution under $\bar{\mathsf{P}}$. Indeed

$$\bar{\pi}_n = \Lambda^*\bar{\pi}_{n-1} + M_n,$$

where $M_n := \bar{\pi}_n - \Lambda^*\bar{\pi}_{n-1}$ are bounded martingale differences under $\bar{\mathsf{P}}$. Hence

$$\frac{1}{n}\sum_{m=1}^n M_m = 0$$

(Theorem 4. Ch. VII, Section 5 in [64]). Since $\frac{1}{n}\sum_{k=1}^n \pi_k \in \mathcal{S}^{d-1}$ and by ergodicity of X, the equation $x = \Lambda^* x$ has a unique root in $\mathcal{S}^{d-1}$, i.e. $x := \mu$, the limit $\frac{1}{n}\sum_{k=1}^n \pi_k$ exists and equals μ. $\square$

Similar result holds in the general setting, as in Theorem 3.6

Theorem 3.17 (Theorem 1.1 in [24])**.** *Let $\mu(x)$ be the unique invariant density of X and assume that $\lambda_\diamond := \int_{\mathbb{S}} \operatorname{ess\,inf}_{u\in\mathbb{S}} \lambda(x,u)\mu(x)\psi(dx) > 0$ and $\lambda(x,u) \le \lambda^* < \infty$. Then for any initial densities $\nu(x)$ and $\bar{\nu}(x) \ge \nu_* > 0$*

$$\varlimsup_{n\to\infty}\frac{1}{n}\log\|\pi_n - \bar{\pi}_n\| \le -\lambda_\diamond/\lambda^*, \quad \mathsf{P}-a.s.$$

References

1. *Oseledec's multiplicative ergodic theorem: a proof*, Random matrices and their applications (Brunswick, Maine, 1984), Contemp. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1986, pp. 23–30. MR MR841079 (87k:60029)
2. Ludwig Arnold, *Random dynamical systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998. MR MR1723992 (2000m:37087)
3. Rami Atar, *Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain*, Ann. Probab. **26** (1998), no. 4, 1552–1574. MR MR1675039 (99k:93088)
4. Rami Atar, Frederi Viens, and Ofer Zeitouni, *Robustness of Zakai's equation via Feynman-Kac representations*, Stochastic analysis, control, optimization and applications, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 1999, pp. 339–352. MR MR1702969 (2000f:93086)
5. Rami Atar and Ofer Zeitouni, *Exponential stability for nonlinear filtering*, Ann. Inst. H. Poincaré Probab. Statist. **33** (1997), no. 6, 697–725. MR MR1484538 (98i:60070)
6. ______, *Lyapunov exponents for finite state nonlinear filtering*, SIAM J. Control Optim. **35** (1997), no. 1, 36–55. MR MR1430282 (97k:93065)
7. ______, *A note on the memory length of optimal nonlinear filters*, Systems Control Lett. **35** (1998), no. 2, 131–135. MR MR1654897 (99j:93100)
8. Peter Baxendale, Pavel Chigansky, and Robert Liptser, *Asymptotic stability of the Wonham filter: ergodic and nonergodic signals*, SIAM J. Control Optim. **43** (2004), no. 2, 643–669 (electronic). MR 2086177 (2005e:93152)
9. A. G. Bhatt, A. Budhiraja, and R. L. Karandikar, *Markov property and ergodicity of the nonlinear filter*, SIAM J. Control Optim. **39** (2000), no. 3, 928–949 (electronic). MR MR1786337 (2001m:93122)
10. Garrett Birkhoff, *Lattice theory*, Third edition. American Mathematical Society Colloquium Publications, Vol. XXV, American Mathematical Society, Providence, R.I., 1967. MR MR0227053 (37 #2638)
11. David Blackwell, *The entropy of functions of finite-state Markov chains*, Transactions of the first Prague conference on information theory, Statistical decision functions, random processes held at Liblice near Prague from November 28 to 30, 1956, Publishing House of the Czechoslovak Academy of Sciences, Prague, 1957, pp. 13–20. MR MR0100297 (20 #6730)
12. A. Budhiraja, *Ergodic properties of the nonlinear filter*, Stochastic Process. Appl. **95** (2001), no. 1, 1–24. MR MR1847089 (2002e:60065)
13. ______, *On invariant measures of discrete time filters in the correlated signal-noise case*, Ann. Appl. Probab. **12** (2002), no. 3, 1096–1113. MR MR1925453 (2003h:60057)
14. ______, *Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter*, Ann. Inst. H. Poincaré Probab. Statist. **39** (2003), no. 6, 919–941. MR MR2010391 (2004h:60063)
15. A. Budhiraja and D. Ocone, *Exponential stability of discrete-time filters for bounded observation noise*, Systems Control Lett. **30** (1997), no. 4, 185–193. MR MR1455877 (98c:93110)
16. ______, *Exponential stability in discrete-time filtering for non-ergodic signals*, Stochastic Process. Appl. **82** (1999), no. 2, 245–257. MR MR1700008 (2000d:94010)
17. Amarjit Budhiraja and Harold J. Kushner, *Robustness of nonlinear filters over the infinite time interval*, SIAM J. Control Optim. **36** (1998), no. 5, 1618–1637 (electronic). MR MR1626884 (99f:93133)
18. ______, *Approximation and limit results for nonlinear filters over an infinite time interval*, SIAM J. Control Optim. **37** (1999), no. 6, 1946–1979 (electronic). MR MR1720146 (2000i:93080)
19. ______, *Approximation and limit results for nonlinear filters over an infinite time interval. II. Random sampling algorithms*, SIAM J. Control Optim. **38** (2000), no. 6, 1874–1908 (electronic). MR MR1776660 (2001f:93070)
20. ______, *Monte Carlo algorithms and asymptotic problems in nonlinear filtering*, Stochastics in finite and infinite dimensions, Trends Math., Birkhäuser Boston, Boston, MA, 2001, pp. 59–87. MR MR1797081 (2001i:93120)
21. Frédéric Cérou, *Long time behavior for some dynamical noise free nonlinear filtering problems*, SIAM J. Control Optim. **38** (2000), no. 4, 1086–1101 (electronic). MR MR1760061 (2001c:93096)
22. Pavel Chigansky, *An ergodic theorem for filtering with applications to stability*, to appear in System and Control Letters, 2006.

23. ———, *Stability of the nonlinear filter for slowly switching markov chains*, Stochastic Process. Appl. **116** (2006), no. 8, 1185–1194.
24. Pavel Chigansky and Robert Liptser, *Stability of nonlinear filters in nonmixing case*, Ann. Appl. Probab. **14** (2004), no. 4, 2038–2056. MR MR2099662 (2005h:62265)
25. ———, *On a role of predictor in the filtering stability*, Electronic Comm. Prob. **11** (2006), 129–140.
26. Pavel Chigansky and Ramon van Handel, *Model robustness of finite state nonlinear filtering over the infinite time horizon*, submitted, 2006.
27. J. M. C. Clark, D. L. Ocone, and C. Coumarbatch, *Relative entropy and error bounds for filtering of Markov processes*, Math. Control Signals Systems **12** (1999), no. 4, 346–360. MR MR1728373 (2001m:60095)
28. Giuseppe Da Prato, Marco Fuhrman, and Paul Malliavin, *Asymptotic ergodicity of the process of conditional law in some problem of non-linear filtering*, J. Funct. Anal. **164** (1999), no. 2, 356–377. MR MR1695555 (2000e:93070)
29. Pierre Del Moral and Alice Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Ann. Inst. H. Poincaré Probab. Statist. **37** (2001), no. 2, 155–194. MR MR1819122 (2002k:60013)
30. Pierre Del Moral and Laurent Miclo, *On the stability of nonlinear Feynman-Kac semigroups*, Ann. Fac. Sci. Toulouse Math. (6) **11** (2002), no. 2, 135–175. MR MR1988460 (2004d:47081)
31. Bernard Delyon and Ofer Zeitouni, *Lyapunov exponents for filtering problems*, Applied stochastic analysis (London, 1989), Stochastics Monogr., vol. 5, Gordon and Breach, New York, 1991, pp. 511–521. MR MR1108433 (92i:93092)
32. Giovanni B. Di Masi and Łukasz Stettner, *Ergodicity of hidden Markov models*, Math. Control Signals Systems **17** (2005), no. 4, 269–296. MR MR2177319
33. Robert J. Elliott, Lakhdar Aggoun, and John B. Moore, *Hidden Markov models*, Applications of Mathematics (New York), vol. 29, Springer-Verlag, New York, 1995, Estimation and control. MR MR1323178 (96h:93001)
34. M. Émery and W. Schachermayer, *On Vershik's standardness criterion and Tsirelson's notion of cosiness*, Séminaire de Probabilités, XXXV, Lecture Notes in Math., vol. 1755, Springer, Berlin, 2001, pp. 265–305. MR MR1837293 (2002i:60071)
35. Eberhard Hopf, *An inequality for positive linear integral operators*, J. Math. Mech. **12** (1963), 683–692. MR MR0165325 (29 #2614)
36. Ze-Chun Hu and Wei Sun, *A note on exponential stability of the nonlinear filter for denumerable markov chains*, to appear in Systems & Control Letters, 2006.
37. Thomas Kaijser, *A limit theorem for partially observed Markov chains*, Ann. Probability **3** (1975), no. 4, 677–696. MR MR0383536 (52 #4417)
38. Gopinath Kallianpur, *Stochastic filtering theory*, Applications of Mathematics, vol. 13, Springer-Verlag, New York, 1980. MR MR583435 (82f:60089)
39. Brunon Kamiński, *Decreasing nets of σ-algebras and their applications to ergodic theory*, Tohoku Math. J. (2) **43** (1991), no. 2, 263–274. MR MR1104432 (92g:28034)
40. R. Z. Khasminskii, *Stochastic stability of differential equations*, Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis, vol. 7, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980, Translated from the Russian by D. Louvish. MR MR600653 (82b:60064)
41. Rafail Khasminskii and Ofer Zeitouni, *Asymptotic filtering for finite state Markov chains*, Stochastic Process. Appl. **63** (1996), no. 1, 1–10. MR MR1411186 (97f:93046)
42. Marina Kleptsyna and Alexander Veretennikov, *On continuous time ergodic filters with wrong initial data*, available at `http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna`, 2006.
43. ———, *On discrete time ergodic filters with wrong initial data*, available at `http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna`, 2006.
44. Fred Kochman and Jim Reeds, *A simple proof of kaijser's unique ergodicity result for hidden markov α-chains*, to appear in Annals of Applied Probability, 2006.
45. Hiroshi Kunita, *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal. **1** (1971), 365–393. MR MR0301812 (46 #967)

46. ———, *Ergodic properties of nonlinear filtering processes*, Spatial stochastic processes, Progr. Probab., vol. 19, Birkhäuser Boston, Boston, MA, 1991, pp. 233–256. MR MR1144099 (92m:60062)
47. Harold J. Kushner and Amarjit S. Budhiraja, *A nonlinear filtering algorithm based on an approximation of the conditional distribution*, IEEE Trans. Automat. Control **45** (2000), no. 3, 580–585. MR MR1762882 (2001c:93101)
48. Huibert Kwakernaak and Raphael Sivan, *Linear optimal control systems*, Wiley-Interscience [John Wiley & Sons], New York, 1972. MR MR0406607 (53 #10394)
49. Alain Le Breton and Marie-Christine Roubaud, *Asymptotic optimality of approximate filters in stochastic systems with colored noises*, SIAM J. Control Optim. **39** (2000), no. 3, 917–927 (electronic). MR MR1786336 (2001h:93093)
50. François Le Gland and Laurent Mevel, *Basic properties of the projective product with application to products of column-allowable nonnegative matrices*, Math. Control Signals Systems **13** (2000), no. 1, 41–62. MR MR1742139 (2001b:15033)
51. ———, *Exponential forgetting and geometric ergodicity in hidden Markov models*, Math. Control Signals Systems **13** (2000), no. 1, 63–93. MR MR1742140 (2001b:93075)
52. François Le Gland and Nadia 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. MR MR2023019 (2004k:62219)
53. François LeGland and Nadia 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. MR MR1989630 (2004i:93184)
54. Robert S. Liptser and Albert N. Shiryaev, *Statistics of random processes. I*, expanded ed., Applications of Mathematics (New York), vol. 5, Springer-Verlag, Berlin, 2001, General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. MR MR1800857 (2001k:60001a)
55. ———, *Statistics of random processes. II*, expanded ed., Applications of Mathematics (New York), vol. 6, Springer-Verlag, Berlin, 2001, Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. MR MR1800858 (2001k:60001b)
56. Armand M. Makowski, *Filtering formulae for partially observed linear systems with non-Gaussian initial conditions*, Stochastics **16** (1986), no. 1-2, 1–24. MR MR834558 (87d:93094)
57. D. L. Ocone, *Asymptotic stability of Beneš filters*, Stochastic Anal. Appl. **17** (1999), no. 6, 1053–1074. MR MR1721933 (2001a:60050)
58. Daniel Ocone, *Entropy inequalities and entropy dynamics in nonlinear filtering of diffusion processes*, Stochastic analysis, control, optimization and applications, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 1999, pp. 477–496. MR MR1702976 (2000f:93091)
59. Daniel Ocone and Etienne Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM J. Control Optim. **34** (1996), no. 1, 226–243. MR MR1372912 (97e:60073)
60. Nadia Oudjane and Sylvain Rubenthaler, *Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals*, Stoch. Anal. Appl. **23** (2005), no. 3, 421–448. MR MR2140972 (2005m:93153)
61. Anastasia Papavasiliou, *Parameter estimation and asymptotic stability in stochastic filtering*, Stochastic Process. Appl. **116** (2006), no. 7, 1048–1065.
62. L. C. G. Rogers, *Coupling and the tail σ-field of a one-dimensional diffusion*, Stochastic calculus in application (Cambridge, 1987), Pitman Res. Notes Math. Ser., vol. 197, Longman Sci. Tech., Harlow, 1988, pp. 78–88. MR MR995472 (90f:60142)
63. E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006, Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. MR MR2209438
64. A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, vol. 95, Springer-Verlag, New York, 1996, Translated from the first (1980) Russian edition by R. P. Boas. MR MR1368405 (97c:60003)
65. Louis Shue, Brian D.O. Anderson, and Subhrakanti Dey, *Exponential stability of filters and smoothers for hidden markov models*, IEEE Trans. Sig. Proc. **46** (1998), no. 8, 2180–2194.

66. Ya. G. Sinaĭ, *Kolmogorov's work on ergodic theory*, Ann. Probab. **17** (1989), no. 3, 833–839. MR MR1009437 (91e:01017)
67. Richard B. Sowers and Armand M. Makowski, *Discrete-time filtering for linear systems in correlated noise with non-Gaussian initial conditions: formulas and asymptotics*, Robust control of linear systems and nonlinear control (Amsterdam, 1989), Progr. Systems Control Theory, vol. 4, Birkhäuser Boston, Boston, MA, 1990, pp. 407–419. MR MR1115412 (92i:93096)
68. ———, *Discrete-time filtering for linear systems with non-Gaussian initial conditions: asymptotic behavior of the difference between the MMSE and LMSE estimates*, IEEE Trans. Automat. Control **37** (1992), no. 1, 114–120. MR MR1139622 (92j:93130)
69. Wilhelm Stannat, *Stability of the filter equation for a time-dependent signal on* $\mathbb{R}^d$, Appl. Math. Optim. **52** (2005), no. 1, 39–71. MR MR2141176
70. ———, *Stability of the optimal filter via pointwise gradient estimates*, available `http://www.math.uni-bielefeld.de/~stannat/`, 2006.
71. ———, *Stability of the pathwise filter equation on* $\mathbb{R}^d$, available at `http://www.math.uni-bielefeld.de/~stannat/`, 2006.
72. Ł. Stettner, *Invariant measures of the pair: state, approximate filtering process*, Colloq. Math. **62** (1991), no. 2, 347–351. MR MR1142935 (93a:60066)
73. Łukasz Stettner, *On invariant measures of filtering processes*, Stochastic differential systems (Bad Honnef, 1988), Lecture Notes in Control and Inform. Sci., vol. 126, Springer, Berlin, 1989, pp. 279–292. MR MR1236074
74. Vladislav B. Tadić and Arnaud Doucet, *Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models*, Stochastic Process. Appl. **115** (2005), no. 8, 1408–1436. MR MR2152381 (2006f:62094)
75. Heinrich von Weizsäcker, *Exchanging the order of taking suprema and countable intersections of* σ-*algebras*, Ann. Inst. H. Poincaré Sect. B (N.S.) **19** (1983), no. 1, 91–100. MR MR699981 (85c:28001)
76. David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. MR MR1155402 (93d:60002)
77. Marc Yor, *Tsirelson's equation in discrete time*, Probab. Theory Related Fields **91** (1992), no. 2, 135–152. MR MR1147613 (93d:60104)

Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel
E-mail address: `pavel.chigansky@weizmann.ac.il`