Elect. Comm. in Probab. **13** (2008), 562–575

ELECTRONIC COMMUNICATIONS in PROBABILITY

DISCRETE TIME NONLINEAR FILTERS WITH INFORMATIVE OBSERVATIONS ARE STABLE

RAMON VAN HANDEL
ORFE Department, Princeton University, Princeton, NJ 08544
email: `rvan@princeton.edu`

Submitted July 21, 2008, *accepted in final form* October 21, 2008

AMS 2000 Subject classification: 60J05; 62M20; 93E11; 93E15
Keywords: nonlinear filtering; prediction; asymptotic stability; hidden Markov models

Abstract

The nonlinear filter associated with the discrete time signal-observation model (X_k, Y_k) is known to forget its initial condition as $k \to \infty$ regardless of the observation structure when the signal possesses sufficiently strong ergodic properties. Conversely, it stands to reason that if the observations are sufficiently informative, then the nonlinear filter should forget its initial condition regardless of any properties of the signal. We show that for observations of additive type $Y_k = h(X_k) + \xi_k$ with invertible observation function h (under mild regularity assumptions on h and on the distribution of the noise ξ_k), the filter is indeed stable in a weak sense without any assumptions at all on the signal process. If the signal satisfies a uniform continuity assumption, weak stability can be strengthened to stability in total variation.

1 Introduction

Let $(E, \mathscr{B}(E))$ and $(F, \mathscr{B}(F))$ be Polish spaces endowed with their Borel σ-fields, and let $P : E \times \mathscr{B}(E) \to [0,1]$ be a given transition probability kernel. On the sequence space $\Omega = E^{\mathbb{Z}_+} \times F^{\mathbb{Z}_+}$ with the canonical coordinate projections $X_n(x,y) = x(n)$, $Y_n(x,y) = y(n)$, we define the family of probability measures $\mathbf{P}^\mu$ (for any probability measure μ on E) such that $(X_n)_{n\ge0}$ is a Markov chain with initial measure $X_0 \sim \mu$ and transition probability P, and such that $Y_n = H(X_n, \xi_n)$ for every $n \ge 0$ where ξ_n is an i.i.d. sequence independent of $(X_n)_{n\ge0}$. A time series model of this type, called a *hidden Markov model*, has a wide variety of applications in science, engineering, statistics and finance; see, e.g., [4]. The process $(X_n)_{n\ge0}$ is known as the *signal process* (and E is the signal state space), while $(Y_n)_{n\ge0}$ is called the *observation process* (and F is the observation state space). As E is Polish, we may define for every μ the regular conditional probabilities

$$\pi^\mu_{n-}(\,\cdot\,) := \mathbf{P}^\mu(X_n \in \cdot\,|Y_0, \dots, Y_{n-1}), \qquad n \ge 1,$$

and

$$\pi^\mu_n(\,\cdot\,) := \mathbf{P}^\mu(X_n \in \cdot\,|Y_0, \dots, Y_n), \qquad n \ge 0.$$

Here π^μ_{n-} is called the *one step predictor* of the signal given the observations, while π^μ_n is known as the *nonlinear filter*. These objects play a central role in the statistical theory of hidden Markov models. A question which has generated considerable interest in recent years is whether, as $n \to \infty$, the filter π_n^μ becomes insensitive to the choice of the initial measure μ. Broadly speaking, the filter is said to be *stable* if π_n^μ and π_n^ν converge towards one another in a suitably chosen manner as $n \to \infty$ (e.g., $\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \to 0$ $\mathbf{P}^\mu$-a.s.) for a large class of initial measures μ, ν (e.g., for all $\mu \ll \nu$).

The filter stability property is of significant practical interest, as the initial measure (a Bayesian prior) may be difficult to characterize. When the filter is stable, we can guarantee that it will nonetheless generate optimal estimates of the signal process after an initial transient. Moreover, the stability property also plays a key role in various important auxiliary problems, such as proving consistency of maximum likelihood estimates and proving uniform convergence of approximate filtering algorithms. On the other hand, the filter stability problem poses a set of interesting mathematical questions in the theory of nonlinear estimation, many of which have yet to be fully resolved. An overview of the state-of-the-art can be found in [8].

Intuitively one expects filter stability to be caused by two separate mechanisms:

1. If the signal process itself becomes insensitive to its initial condition after a long time interval (i.e., the signal is ergodic) one would expect the filter to inherit this property.

2. If the observations are informative, one would expect that the information in the observations will eventually obsolete the prior information contained in the initial measure.

In the two special cases where a detailed characterization of filter stability is available—for linear Gaussian models [12] and for finite state signals [17]—the notion of *detectability* embodies precisely this intuition. It thus seems reasonable to conjecture that it is true in great generality that these two mechanisms conspire to bring about the stability of the filter. To date, however, detectability conditions for filter stability are only known in the abovementioned special cases. To gain further insight, it is therefore instructive to study each of the mechanisms separately in a general setting. In particular, the two extreme cases lead to the following fundamental problems: (i) can we find conditions on the signal process such that the filter is stable regardless of the observation structure? and (ii) can we find conditions on the observation structure such that the filter is stable regardless of any properties of the signal?

Various solutions to Problem (i) can be found in the literature. It was shown by Atar and Zeitouni [1] and by Del Moral and Guionnet [9] that the filter is stable whenever the signal possesses a certain strong *mixing condition*, regardless of the observation structure. The mixing condition was weakened to some extent by Chigansky and Liptser [5]. Somewhat surprisingly, assuming only ergodicity of the signal is not sufficient to guarantee stability (see [5, section 5]); both the mixing condition and the condition of Chigansky and Liptser are strictly stronger than ergodicity. Under a mild nondegeneracy assumption on the observations, however, ergodicity of the signal is already sufficient to ensure stability of the filter [15].

In contrast to the first problem, a solution to Problem (ii) has hitherto been elusive. Unlike results based on ergodicity or mixing, stability results based on the structure of the observations have appeared only recently in [6, 17, 16]. It appears, however, that it is more natural in this context to study stability of the predictor than stability of the filter. In particular, the following general result was established in [16, proposition 3.11] for *additive* observations of the form $Y_n = h(X_n) + \xi_n$, where $E = F = \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^n$ is a given observation function, and ξ_n is a sequence of i.i.d. $\mathbb{R}^n$-valued random variables independent of $(X_n)_{n\geq 0}$.

Proposition 1.1 ([16]). *Suppose that the following hold:*

1. *h possesses a uniformly continuous inverse; and*

2. *the characteristic function of ξ_0 vanishes nowhere.*

Then $\|\pi^\mu_{n-} - \pi^\nu_{n-}\|_{\mathrm{BL}} \xrightarrow{n\to\infty} 0$ $\mathbf{P}^\mu$*-a.s. whenever* $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge 0}\}}$.

Here $\|\cdot\|_{\mathrm{BL}}$ denotes the dual bounded-Lipschitz distance (to be defined below). The assumptions of this result certainly conform to the idea that the observations are 'informative': Y_k is simply a noisy and distorted version of X_k. Note also that this result places no conditions whatsoever on the signal process X_n except the Markov property. However, the result is a statement about the one step predictor and not about the filter.
The main purpose of this note is to point out that under the mild additional assumption that the law of the noise variables ξ_k has a density, a slightly weaker version of proposition 1.1 holds also when the predictor is replaced by the filter. We therefore provide an affirmative answer to the conjecture that there exists a solution to Problem (ii) above. The proof of this result adapts a coupling argument due to Ocone and Pardoux [12, lemma 3.6].

Remark 1.2. In the continuous time setting, it is known that a result along the lines of proposition 1.1 holds for the filter when the signal state space E is assumed to be compact (under the mild assumption that the signal is Feller), see [17]. This is not a satisfactory solution to Problem (ii), however, as unstable signals are ruled out in a compact state space.

Even if one is willing to make assumptions on the signal process, the case of non-ergodic signals has received comparatively little attention in the literature. Previous results in the non-ergodic setting show that the filter is stable in the total variation distance, but only under strong assumptions on both the signal and the observation process [3, 13, 7]. In particular, these results only hold when the signal to noise ratio of the observations is sufficiently large (this appears to be an inherent restriction of the method of proof used in these papers). In addition to our main result, we will show that the filter is stable in the total variation distance under significantly weaker assumptions than have been required in previous work. In particular, our result holds for an arbitrary signal to noise ratio. To this end, we must investigate when the convergence in the dual bounded-Lipschitz distance in our main result can be strengthened to convergence in the total variation distance. For this purpose we will introduce a suitable uniform continuity assumption on the signal transition kernel.

2 Notation and Main Results

2.1 Notation

For any Polish space S endowed with a complete metric d (when $S = \mathbb{R}^n$ we will always choose the Euclidean metric $d(x,y) = \|x-y\|$), define

$$\|f\|_\infty = \sup_x |f(x)|, \qquad \|f\|_L = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)} \qquad \text{for any } f : S \to \mathbb{R}.$$

If $\|f\|_L < \infty$, the function f is Lipschitz continuous. Denote by $\mathrm{Lip} = \{f : S \to \mathbb{R} : \|f\|_\infty \le 1 \text{ and } \|f\|_L \le 1\}$ the unit ball in the space of 1-Lipschitz functions. Then for any two probability measures μ, ν on S, the dual bounded-Lipschitz norm is defined as

$$\|\mu - \nu\|_{\mathrm{BL}} = \sup_{f\in\mathrm{Lip}} \left| \int f(x)\,\mu(dx) - \int f(x)\,\nu(dx) \right|.$$

The supremum can equivalently be taken over a countable subfamily $\mathrm{Lip}_0 \subset \mathrm{Lip}$ (Lip_0 does not depend on μ, ν) [16, lemma A.1]. As usual, the total variation norm is defined as

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{\|f\|_\infty \le 1} \left| \int f(x)\,\mu(dx) - \int f(x)\,\nu(dx) \right|.$$

Also in this case the supremum can be replaced by the supremum over a countable subfamily $B_0 \subset \{f : \|f\|_\infty \le 1\}$ (along the lines of [16, lemma 4.1]).

2.2 Main Results

In the following, we will work with a hidden Markov model where the signal state space E is a Polish space with complete metric d, and the observation state space is Euclidean $F = \mathbb{R}^n$. We consider additive observations of the form $Y_k = h(X_k) + \xi_k$ for all $k \ge 0$, where $h : E \to \mathbb{R}^n$ is the observation function and ξ_k is a sequence of i.i.d. $\mathbb{R}^n$-valued random variables which are independent of the signal. The signal transition kernel P is fixed at the outset and is not presumed to satisfy any assumptions until further notice.
In our main result, we will impose the following assumption.

Assumption 2.1. The following hold:

1. The observation function h possesses a uniformly continuous inverse.
2. The law of ξ_k has a density q_ξ with respect to the Lebesgue measure on $\mathbb{R}^n$.
3. The Fourier transform of q_ξ vanishes nowhere.

Note that this is an assumption on the observations only: nothing at all is assumed about the signal at this point. Our main result thus holds regardless of any properties of the signal.

Theorem 2.2. *Suppose that assumption 2.1 holds. Then*

$$\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{BL}}) \xrightarrow{n\to\infty} 0 \qquad \textit{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}.$$

In order to strengthen the convergence to the total variation distance, we do need to impose an assumption on the signal. The following is essentially a uniform strong Feller assumption.

Assumption 2.3. The signal transition kernel P satisfies

$$\|P(x_n, \cdot) - P(y_n, \cdot)\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \qquad \text{whenever} \qquad d(x_n, y_n) \xrightarrow{n\to\infty} 0.$$

In other words, the measure-valued map $x \mapsto P(x, \cdot)$ is uniformly continuous for the total variation distance on the space of probability measures.

We now obtain the following result.

Theorem 2.4. *Suppose that assumptions 2.1 and 2.3 hold. Then*

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\textit{-a.s.} \qquad \textit{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}.$$

A typical example where assumption 2.3 holds is the following.

Proposition 2.5. *Let $E = \mathbb{R}^m$. Suppose that the signal is defined by the recursion*

$$X_{k+1} = b(X_k) + \sigma(X_k)\,\eta_k,$$

where b, σ, η_k satisfy the following assumptions:

1. *$b : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ are uniformly continuous.*
2. *σ is uniformly bounded from below: $\|\sigma(x)v\| \geq \alpha\|v\|$ for all $x, v \in \mathbb{R}^m$ and some $\alpha > 0$.*
3. *η_k are i.i.d. $\mathbb{R}^m$-valued random variables, whose law possesses a density q_η with respect to the Lebesgue measure on $\mathbb{R}^m$.*

Then assumption 2.3 holds. If q_η and q_ξ are strictly positive and assumption 2.1 holds, then

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\gamma\text{-a.s.} \qquad \text{for any } \mu, \nu, \gamma.$$

This result should be compared to the main results in [3, 13, 7], where a very similar model is investigated. However, in these references total variation stability is proved only when the signal to noise ratio is sufficiently high. This is an artefact of the quantitative method of proof where two rates of expansion are compared: the filter is stable if one of the rates 'wins', which leads to a requirement on the signal to noise ratio. Our qualitative approach does not depend on the signal to noise ratio, however, so that evidently the assumptions required for the balancing of rates are stronger than is needed for filter stability (see [16] for further discussion). On the other hand, our approach can not provide an estimate of the rate of stability.

Remark 2.6. Our results require that the observation state space is Euclidean $F = \mathbb{R}^n$, as it relies on properties of convolutions in $\mathbb{R}^n$ (an extension to the case where F is a locally compact abelian group may be feasible). In contrast, we have only assumed that the signal state space is Polish. Note, however, that assumption 2.1 requires the existence of a uniformly continuous map $h^{-1} : \mathbb{R}^n \to E$ such that $h^{-1}(h(x)) = x$ for all $x \in E$. In particular, h is an embedding of E into $\mathbb{R}^n$, so that E can not be larger (e.g., of higher dimension) than $\mathbb{R}^n$. This is to be expected, of course, as filter stability in the case where h is not invertible must depend on specific properties of the signal process such as observability or ergodicity.

Remark 2.7. Theorems 2.2 and 2.4 provide stability of the filter whenever the initial measures satisfy $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\geq 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\geq 0}\}}$. Absolute continuity of the initial measures $\mu \ll \nu$ is sufficient for this to hold, but is not always necessary. For example, if $\mu P^k \ll \nu P^k$ for some $k > 0$ and the observation density q_ξ is strictly positive, then it is not difficult to prove that $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\geq 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\geq 0}\}}$ also. In particular, if the signal possesses a strictly positive transition density, then $\mu P \sim \nu P$ for every μ, ν and we obtain stability for arbitrary initial measures provided $q_\xi > 0$. This is the case, for example, in the setting of proposition 2.5.

Remark 2.8. Our results do not give a rate of stability, while most previous work on filter stability gives exponential convergence rates. The following simple example demonstrates that exponential stability can not be expected in the general setting of this paper.

Let $E = F = \mathbb{R}$ and $Y_k = X_k + \xi_k$, where ξ_k are i.i.d. $N(0,1)$ and $X_k = X_0$ for all k. This setting certainly satisfies the requirements of theorem 2.2. We choose $\mu = N(\alpha, \sigma^2)$ and $\nu = N(\beta, \sigma^2)$ for some $\alpha, \beta, \sigma \in \mathbb{R}$ (so $\mathbf{P}^\mu \sim \mathbf{P}^\nu$). Linear filtering theory shows that π_k^μ is a random Gaussian measure with mean Z_k^μ and variance V_k^μ given by

$$Z_k^\mu = \frac{\alpha}{1+\sigma^2(k+1)} + \frac{\sigma^2(k+1)}{1+\sigma^2(k+1)} \cdot \frac{1}{k+1}\sum_{\ell=0}^{k} Y_\ell, \qquad V_k^\mu = \frac{\sigma^2}{1+\sigma^2(k+1)},$$

and similarly for $\pi_k^\nu, Z_k^\nu, V_k^\nu$ where α is replaced by β. Note that by the law of large numbers, the second term in the expression for Z_k^μ (and Z_k^ν) converges $\mathbf{P}^\mu$-a.s. to X_0. But

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{BL}} \geq \left|\int \cos(x)\,\pi_n^\mu(dx) - \int \cos(x)\,\pi_n^\nu(dx)\right| = e^{-V_n^\mu/2}|\cos(Z_n^\mu) - \cos(Z_n^\nu)|.$$

Noting that $\cos(Z_n^\nu) = \cos(Z_n^\mu) - \sin(Z_n^\mu)(\beta-\alpha)/(1+\sigma^2(n+1)) + o(n^{-1})$, we find that

$$\liminf_{n\to\infty} n\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{BL}} \geq \liminf_{n\to\infty} n\, e^{-V_n^\mu/2}|\cos(Z_n^\mu) - \cos(Z_n^\nu)| = \frac{|\beta-\alpha|}{\sigma^2}\,|\sin(X_0)| > 0 \quad \mathbf{P}^\mu\text{-a.s.}$$

By Fatou's lemma $\liminf_{n\to\infty} n\,\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{BL}}) > 0$, so that evidently the stability rate of the filter is at best of order $O(n^{-1})$ and is certainly not exponential.

It is interesting to note that by [14, p. 528, theorem 4] and by the equivalence of the Hellinger and total variation distances, $\sum_{n=0}^\infty \|\pi_{n-}^\mu h^{-1} * \xi - \pi_{n-}^\nu h^{-1} * \xi\|_{\mathrm{TV}}^2 < \infty$ $\mathbf{P}^\mu$-a.s. The convergence of the expression $\|\pi_{n-}^\mu h^{-1} * \xi - \pi_{n-}^\nu h^{-1} * \xi\|_{\mathrm{TV}}$ which appears in the proof of lemma 3.1 below is therefore generally, in a sense, not much worse than $o(n^{-1/2})$. It is unclear, however, whether this property survives the subsequent manipulations that lead to stability of the filter.

3 Proof of Theorem 2.2

Let us begin by recalling a part of the proof of proposition 1.1.

Lemma 3.1. *Suppose that the characteristic function of ξ_k vanishes nowhere, and that moreover $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\geq 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\geq 0}\}}$. Then $\|\pi_{n-}^\mu h^{-1} - \pi_{n-}^\nu h^{-1}\|_{\mathrm{BL}} \xrightarrow[n\to\infty]{} 0$ $\mathbf{P}^\mu$-a.s.*

Proof. Denote the law of ξ_k as ξ. It is easily verified that for any probability measure ρ

$$\mathbf{P}^\rho(Y_n \in \cdot\,|Y_0,\ldots,Y_{n-1}) = \pi_{n-}^\rho h^{-1} * \xi,$$

where $*$ denotes convolution. A classic result of Blackwell and Dubins [2, section 2] shows that

$$\|\pi_{n-}^\mu h^{-1} * \xi - \pi_{n-}^\nu h^{-1} * \xi\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.},$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm. The result now follows from [16, proposition C.2]. □

To proceed, recall that due to the Bayes formula (e.g., [4, section 3.2.2])

$$\int f(x)\,\pi_n^\mu(dx) = \frac{\int f(x)\,q_\xi(Y_n - h(x))\,\pi_{n-}^\mu(dx)}{\int q_\xi(Y_n - h(x))\,\pi_{n-}^\mu(dx)} \qquad \text{for all bounded } f: E \to \mathbb{R} \qquad \mathbf{P}^\mu\text{-a.s.}$$

Note that the denominator of this expression is strictly positive $\mathbf{P}^\mu$-a.s. Moreover, we have

$$\mathbf{E}^\mu(f(Y_n)|Y_0,\ldots,Y_{n-1}) = \int f(y)\,q_\xi(y - h(x))\,\pi_{n-}^\mu(dx)\,dy \qquad \mathbf{P}^\mu\text{-a.s.}$$

for any bounded function $f:\mathbb{R}^n\to\mathbb{R}$ as in the proof of lemma 3.1. Therefore, it follows from the disintegration of measures that $\mathbf{P}^\mu$-a.s.

$$\mathbf{E}^\mu(\|\pi_n^\mu-\pi_n^\nu\|_{\mathrm{BL}}|Y_0,\ldots,Y_{n-1})= \int\sup_{f\in\mathrm{Lip}_0}\left|\frac{\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\mu(dx)}{\int q_\xi(y-h(x))\,\pi_{n-}^\mu(dx)}-\frac{\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}{\int q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}\right| \times\left\{\int q_\xi(y-h(x))\,\pi_{n-}^\mu(dx)\right\}dy,$$

where it should be noted that by the assumption that $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}}\ll\mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}$ in theorem 2.2 (which we presume to be in force throughout) all quantities in this expression as $\mathbf{P}^\mu$-a.s. uniquely defined and both denominators are strictly positive $\mathbf{P}^\mu$-a.s.
It will be useful for what follows to rewrite the above expression in a more convenient form:

$$\mathbf{E}^\mu(\|\pi_n^\mu-\pi_n^\nu\|_{\mathrm{BL}}|Y_0,\ldots,Y_{n-1})=\int\sup_{f\in\mathrm{Lip}_0}\left|\int f(h^{-1}(x))\,q_\xi(y-x)\,\pi_{n-}^\mu h^{-1}(dx)\right. \left.-\frac{\int f(h^{-1}(x))\,q_\xi(y-x)\,\pi_{n-}^\nu h^{-1}(dx)}{\int q_\xi(y-x)\,\pi_{n-}^\nu h^{-1}(dx)}\int q_\xi(y-x)\,\pi_{n-}^\mu h^{-1}(dx)\right|dy.$$

Here we have fixed a uniformly continuous function $h^{-1}:\mathbb{R}^n\to E$ such that $h^{-1}(h(x))=x$ for all $x\in E$; the existence of this function is guaranteed by assumption 2.1.

Lemma 3.2. *Let ρ,ρ' be two probability measures on $\mathbb{R}^n$, and let Z,Z' be (not necessarily independent) $\mathbb{R}^n$-valued random variables such that $Z\sim\rho$ and $Z'\sim\rho'$. Then*

$$\int\sup_{f\in\mathrm{Lip}_0}\left|\int f(h^{-1}(x))\,q_\xi(y-x)\,\rho(dx)-\frac{\int f(h^{-1}(x))\,q_\xi(y-x)\,\rho'(dx)}{\int q_\xi(y-x)\,\rho'(dx)}\int q_\xi(y-x)\,\rho(dx)\right|dy \le\mathbf{E}(d(h^{-1}(Z),h^{-1}(Z'))\wedge 2)+2\int\mathbf{E}(|q_\xi(y-Z)-q_\xi(y-Z')|)\,dy,$$

where by convention $0/0=1$.

Proof. The left hand side of the expression in the statement of the lemma is

$$\Delta:=\int\sup_{f\in\mathrm{Lip}_0}\left|\mathbf{E}(f(h^{-1}(Z))\,q_\xi(y-Z))-\frac{\mathbf{E}(f(h^{-1}(Z'))\,q_\xi(y-Z'))}{\mathbf{E}(q_\xi(y-Z'))}\mathbf{E}(q_\xi(y-Z))\right|dy.$$

Now note that for any $f\in\mathrm{Lip}$ and $x,y\in\mathbb{R}^n$ we have $|f(x)-f(y)|\le d(x,y)\wedge 2$, so

$$\int\sup_{f\in\mathrm{Lip}_0}|\mathbf{E}(f(h^{-1}(Z))\,q_\xi(y-Z))-\mathbf{E}(f(h^{-1}(Z'))\,q_\xi(y-Z))|\,dy \le\int\mathbf{E}(\{d(h^{-1}(Z),h^{-1}(Z'))\wedge 2\}\,q_\xi(y-Z))\,dy=\mathbf{E}(d(h^{-1}(Z),h^{-1}(Z'))\wedge 2),$$

where we have used the Fubini-Tonelli theorem to exchange the order of integration. Thus

$$\Delta \leq \mathbf{E}(d(h^{-1}(Z),h^{-1}(Z'))\wedge 2)+ \int \sup_{f\in\mathrm{Lip}_0}\left|\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z))-\frac{\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))}{\mathbf{E}(q_\xi(y-Z'))}\mathbf{E}(q_\xi(y-Z))\right|dy.$$

Estimating $\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z))$ by $\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))$, we similarly obtain

$$\Delta \leq \mathbf{E}(d(h^{-1}(Z),h^{-1}(Z'))\wedge 2)+\int \mathbf{E}(|q_\xi(y-Z)-q_\xi(y-Z')|)\,dy+ \int \sup_{f\in\mathrm{Lip}_0}\left|\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))-\frac{\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))}{\mathbf{E}(q_\xi(y-Z'))}\mathbf{E}(q_\xi(y-Z))\right|dy.$$

We now substitute in this expression

$$\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))=\frac{\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))}{\mathbf{E}(q_\xi(y-Z'))}\mathbf{E}(q_\xi(y-Z')),$$

and note that $|\mathbf{E}(f(h^{-1}(Z'))q_\xi(y-Z'))/\mathbf{E}(q_\xi(y-Z'))|\leq 1$ whenever $\|f\|_\infty\leq 1$. The remainder of the proof is now immediate. □

A remarkable result due to Dudley [10, theorem 11.7.1], which extends the classical Skorokhod representation theorem to the $\|\cdot\|_{\mathrm{BL}}$-uniformity, allows us to put this lemma to good use.

Lemma 3.3. *Let ρ_n and ρ_n', $n\geq 0$ be two sequences of probability measures on $\mathbb{R}^n$ such that $\|\rho_n-\rho_n'\|_{\mathrm{BL}}\to 0$ as $n\to\infty$. Then the following quantity*

$$\int \sup_{f\in\mathrm{Lip}_0}\left|\int f(h^{-1}(x))q_\xi(y-x)\rho_n(dx)-\frac{\int f(h^{-1}(x))q_\xi(y-x)\rho_n'(dx)}{\int q_\xi(y-x)\rho_n'(dx)}\int q_\xi(y-x)\rho_n(dx)\right|dy$$

converges to zero as $n\to\infty$.

Proof. By [10, theorem 11.7.1] we can construct two sequences of $\mathbb{R}^n$-valued random variables Z_n and Z_n', $n\geq 0$ on some underlying probability space such that $Z_n\sim\rho_n$ and $Z_n'\sim\rho_n'$ for every n and $\|Z_n-Z_n'\|\to 0$ a.s. as $n\to\infty$. By the previous lemma, the expression Δ_n in the statement of the present lemma is bounded by

$$\Delta_n\leq \mathbf{E}(d(h^{-1}(Z_n),h^{-1}(Z_n'))\wedge 2)+2\int \mathbf{E}(|q_\xi(y-Z_n)-q_\xi(y-Z_n')|)\,dy$$

for every n. As h^{-1} is uniformly continuous and $\|Z_n-Z_n'\|\to 0$ a.s. as $n\to\infty$, we find that $d(h^{-1}(Z_n),h^{-1}(Z_n'))\to 0$ a.s. as $n\to\infty$. Thus the first term evidently converges to zero by dominated convergence. To deal with the second term, note that

$$\int |q_\xi(y-Z_n)-q_\xi(y-Z_n')|\,dy=\int |q_\xi(y+Z_n'-Z_n)-q_\xi(y)|\,dy=\|T_{Z_n-Z_n'}q_\xi-q_\xi\|_{L^1(dy)},$$

where $(T_z f)(x) = f(x-z)$ denotes translation. But recall that translation is continuous in the L^1-topology [11, proposition 8.5], so we find that

$$\|T_{Z_n - Z_n'} q_\xi - q_\xi\|_{L^1(dy)} \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

On the other hand,

$$\|T_{Z_n - Z_n'} q_\xi - q_\xi\|_{L^1(dy)} \le 2\,\|q_\xi\|_{L^1(dy)} = 2 \qquad \text{for all } n.$$

Dominated convergence gives

$$\int \mathbf{E}(|q_\xi(y-Z_n) - q_\xi(y-Z_n')|)\,dy = \mathbf{E}(\|T_{Z_n-Z_n'} q_\xi - q_\xi\|_{L^1(dy)}) \xrightarrow{n\to\infty} 0,$$

where we have used the Fubini-Tonelli theorem to exchange the order of integration. ☐

The proof of theorem 2.2 is now easily completed. Indeed, under our assumptions we obtain $\|\pi^\mu_{n-} h^{-1} - \pi^\nu_{n-} h^{-1}\|_{\mathrm{BL}} \to 0$ $\mathbf{P}^\mu$-a.s. as $n\to\infty$ by lemma 3.1, so the previous lemma gives

$$\mathbf{E}^\mu(\|\pi^\mu_n - \pi^\nu_n\|_{\mathrm{BL}} | Y_0,\ldots,Y_{n-1}) \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.}$$

Taking the expectation with respect to $\mathbf{P}^\mu$, and noting that $\|\pi^\mu_n - \pi^\nu_n\|_{\mathrm{BL}} \le 2$ so that the dominated convergence theorem applies, yields the proof.

4 Proof of Theorem 2.4

We begin by showing that the one step predictor is stable in mean total variation.

Lemma 4.1. *Suppose that assumptions 2.1 and 2.3 hold. Then*

$$\mathbf{E}^\mu(\|\pi^\mu_{n-} - \pi^\nu_{n-}\|_{\mathrm{TV}}) \xrightarrow{n\to\infty} 0 \qquad \textit{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}.$$

Proof. Recall that

$$\int f(x)\,\pi^\mu_{n-}(dx) = \int f(x')\,P(x,dx')\,\pi^\mu_{n-1}(dx)$$

for any bounded measurable function f. Therefore

$$\|\pi^\mu_{n-} - \pi^\nu_{n-}\|_{\mathrm{TV}} = \|\pi^\mu_{n-1} - \pi^\mu_{n-1}\|_G := \sup_{f\in G}\left|\int f(x)\,\pi^\mu_{n-1}(dx) - \int f(x)\,\pi^\nu_{n-1}(dx)\right|,$$

where $G = \{Pf : f \in B_0\}$ (recall that B_0 is a countable family of functions such that $\|\mu-\nu\|_{\mathrm{TV}} = \|\mu-\nu\|_{B_0}$, see section 2.1). We now claim that the family G is uniformly bounded and uniformly equicontinuous. Indeed, it is immediate that $\|f\|_\infty \le 1$ for every $f\in G$, as this is the case for every $f \in B_0$. To prove uniform equicontinuity, note that

$$\sup_{d(x,y)\le\delta} |f(x)-f(y)| \le \sup_{d(x,y)\le\delta}\sup_{f\in B_0} |Pf(x)-Pf(y)| = \sup_{d(x,y)\le\delta} \|P(x,\cdot)-P(y,\cdot)\|_{\mathrm{TV}} := \varpi_P(\delta)$$

for every $f\in G$. By assumption 2.3, we evidently have $\varpi_P(\delta)\to 0$ as $\delta\to0$. Uniform equicontinuity of G is therefore established. To complete the proof, it remains to show that

$$\mathbf{E}^\mu(\|\pi^\mu_n - \pi^\nu_n\|_G) \xrightarrow{n\to\infty} 0 \qquad \text{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}.$$

This is established precisely as in the proof of theorem 2.2, however: the only modification that must be made in the present setting is that the term $\mathbf{E}(d(h^{-1}(Z),h^{-1}(Z')) \wedge 2)$ in lemma 3.2 is replaced by $\mathbf{E}(\varpi_P(d(h^{-1}(Z),h^{-1}(Z'))) \wedge 2)$. $\square$

We now proceed to show stability of the filter (rather than the one step predictor). Stability in the mean follows trivially from the previous lemma and the following estimate.

Lemma 4.2. *Whenever* $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge 0}\}}$, *we have*

$$\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}}|Y_0,\ldots,Y_{n-1}) \le 2\,\|\pi_{n-}^\mu - \pi_{n-}^\nu\|_{\mathrm{TV}} \quad \mathbf{P}^\mu\text{-}a.s.$$

Proof. As in the proof of theorem 2.2, we can write

$$\begin{aligned}\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}}|Y_0,\ldots,Y_{n-1}) = \int \sup_{f\in B_0}\bigg|\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\mu(dx) \\ - \frac{\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}{\int q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}\int q_\xi(y-h(x))\,\pi_{n-}^\mu(dx)\bigg|\,dy.\end{aligned}$$

It follows directly that we can estimate $\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}}|Y_0,\ldots,Y_{n-1}) \le \Delta_1 + \Delta_2$, where

$$\begin{aligned}\Delta_1 &= \int \sup_{f\in B_0}\bigg|\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\mu(dx) - \int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)\bigg|\,dy,\\ \Delta_2 &= \int \sup_{f\in B_0}\bigg|\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx) \\ &\qquad - \frac{\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}{\int q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)}\int q_\xi(y-h(x))\,\pi_{n-}^\mu(dx)\bigg|\,dy \\ &\le \int \bigg|\int q_\xi(y-h(x))\,\pi_{n-}^\mu(dx) - \int q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)\bigg|\,dy.\end{aligned}$$

To estimate Δ_1, note that

$$\begin{aligned}&\sup_{f\in B_0}\bigg|\int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\mu(dx) - \int f(x)\,q_\xi(y-h(x))\,\pi_{n-}^\nu(dx)\bigg| \\ &\qquad \le \sup_{f\in B_0}\int |f(x)|\,q_\xi(y-h(x))\,|\pi_{n-}^\mu - \pi_{n-}^\nu|(dx) \le \int q_\xi(y-h(x))\,|\pi_{n-}^\mu - \pi_{n-}^\nu|(dx).\end{aligned}$$

Therefore, the Fubini-Tonelli theorem gives

$$\Delta_1 \le \int \left\{\int q_\xi(y-h(x))\,dy\right\}|\pi_{n-}^\mu - \pi_{n-}^\nu|(dx) = \|\pi_{n-}^\mu - \pi_{n-}^\nu\|_{\mathrm{TV}}.$$

Δ_2 is estimated in the same fashion, and the proof is complete. $\square$

We have now shown that under assumptions 2.1 and 2.3,

$$\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}}) \xrightarrow{n\to\infty} 0 \qquad \text{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge 0}\}}.$$

It remains to prove that in fact

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.} \qquad \text{whenever} \qquad \mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge 0}\}}.$$

Clearly it suffices to show that $\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}}$ is $\mathbf{P}^\mu$-a.s. convergent.
To this end, let $\mu \ll \gamma$. Then (see [15, corollary 5.7])

$$\|\pi_n^\mu - \pi_n^\gamma\|_{\mathrm{TV}} = \frac{\mathbf{E}^\gamma(|\mathbf{E}^\gamma(\frac{d\mu}{d\gamma}(X_0)|\mathscr{F}_+^Y \vee \mathscr{F}_{[n,\infty[}^X) - \mathbf{E}^\gamma(\frac{d\mu}{d\gamma}(X_0)|\mathscr{F}_{[0,n]}^Y)|\ |\mathscr{F}_{[0,n]}^Y)}{\mathbf{E}^\gamma(\frac{d\mu}{d\gamma}(X_0)|\mathscr{F}_{[0,n]}^Y)} \quad \mathbf{P}^\mu\text{-a.s.},$$

where $\mathscr{F}_{[0,n]}^Y := \sigma\{Y_0,\ldots,Y_n\}$, $\mathscr{F}_+^Y := \sigma\{Y_k : k\ge 0\}$, and $\mathscr{F}_{[n,\infty[}^X := \sigma\{X_k : k \ge n\}$. When $d\mu/d\gamma$ is bounded, the numerator converges $\mathbf{P}^\gamma$-a.s. (hence $\mathbf{P}^\mu$-a.s.) by the martingale convergence theorem (see, e.g., [2, theorem 2]) while the denominator converges to a $\mathbf{P}^\mu$-a.s. strictly positive quantity $\mathbf{E}^\gamma(d\mu/d\gamma(X_0)|\mathscr{F}_{[0,n]}^Y) \to \mathbf{E}^\gamma(d\mu/d\gamma(X_0)|\mathscr{F}_+^Y) > 0$ $\mathbf{P}^\mu$-a.s. Evidently

$$\|\pi_n^\mu - \pi_n^\gamma\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.} \qquad \text{whenever} \qquad \mu \ll \gamma,\ \|d\mu/d\gamma\|_\infty < \infty.$$

Now set $\gamma = (\mu+\nu)/2$, and note that $\|d\mu/d\gamma\|_\infty \le 2$, $\|d\nu/d\gamma\|_\infty \le 2$. Therefore

$$\|\pi_n^\mu - \pi_n^\gamma\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.}, \qquad \|\pi_n^\nu - \pi_n^\gamma\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\nu\text{-a.s.}$$

If $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge 0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge 0}\}}$ the second statement holds also $\mathbf{P}^\mu$-a.s. The proof of theorem 2.4 is now easily completed by applying the triangle inequality.

Remark 4.3. From the above expression, it can be read off that under assumptions 2.1 and 2.3

$$\mathbf{E}^\nu\left(f(X_0)\,\middle|\,\bigcap_{n\ge 0}\mathscr{F}_+^Y \vee \mathscr{F}_{[n,\infty[}^X\right) = \mathbf{E}^\nu(f(X_0)|\mathscr{F}_+^Y) \quad \text{whenever} \quad \|f\|_\infty < \infty$$

for every ν. With a little more work, one can show that similarly for every ν

$$\mathbf{E}^\nu\left(f(X_0,\ldots,X_k)\,\middle|\,\bigcap_{n\ge 0}\mathscr{F}_+^Y \vee \mathscr{F}_{[n,\infty[}^X\right) = \mathbf{E}^\nu(f(X_0,\ldots,X_k)|\mathscr{F}_+^Y) \quad \text{whenever} \quad \|f\|_\infty < \infty,$$

which implies that for every ν

$$\bigcap_{n\ge 0}\mathscr{F}_+^Y \vee \mathscr{F}_{[n,\infty[}^X = \mathscr{F}_+^Y \qquad \mathbf{P}^\nu\text{-a.s.}$$

For the significance of this identity, we refer to [15] and the references therein.

5 Proof of Proposition 2.5

We begin by proving the following representation.

Lemma 5.1. *For fixed $x, x' \in E$, we have*

$$\|P(x,\cdot) - P(x',\cdot)\|_{\mathrm{TV}} = \int \left| \frac{q_\eta(\sigma(x)^{-1}\{\sigma(x')z - b(x) + b(x')\})}{\det(\sigma(x')^{-1}\sigma(x))} - q_\eta(z)\right| dz.$$

Proof. Note that $\sigma(x)$ is invertible for every x as it is presumed to be lower bounded. Therefore $P(x,\cdot)$ has density $p(x,z) = q_\eta(\sigma(x)^{-1}\{z-b(x)\})/\det(\sigma(x))$ with respect to the Lebesgue measure on $\mathbb{R}^m$ for every x. This implies that

$$\|P(x,\cdot)-P(x',\cdot)\|_{\mathrm{TV}} = \int \left| \frac{q_\eta(\sigma(x)^{-1}\{z-b(x)\})}{\det(\sigma(x))} - \frac{q_\eta(\sigma(x')^{-1}\{z-b(x')\})}{\det(\sigma(x'))} \right| dz.$$

The result follows through a change of variables. □

We now prove that assumption 2.3 holds in this setting.

Lemma 5.2. $\|P(x_n,\cdot)-P(y_n,\cdot)\|_{\mathrm{TV}} \to 0$ *whenever* $d(x_n,y_n)\to 0$.

Proof. Fix any sequence x_n, y_n such that $d(x_n,y_n)\to 0$. By the previous lemma, it evidently suffices to show that the following function converges to $q_\eta(z)$ in $L^1(dz)$ as $n\to\infty$:

$$g_n(z) = q_\eta(\sigma(x_n)^{-1}\{\sigma(y_n)z - b(x_n)+b(y_n)\})\det(\sigma(x_n)^{-1}\sigma(y_n)).$$

Suppose first that q_η is continuous. Note that

$$\|\sigma(x_n)^{-1}\sigma(y_n)-I\| = \|\sigma(x_n)^{-1}\{\sigma(y_n)-\sigma(x_n)\}\| \le \alpha^{-1}\|\sigma(y_n)-\sigma(x_n)\| \xrightarrow{n\to\infty} 0$$

as σ is uniformly continuous and lower bounded by $\alpha>0$, while $\|\sigma(x_n)^{-1}\{b(x_n)-b(y_n)\}\| \le \alpha^{-1}\|b(x_n)-b(y_n)\|\to 0$ as $n\to\infty$ as b is uniformly continuous. Therefore, if q_η is continuous, then $g_n(z)$ converges to $q_\eta(z)$ pointwise. By Scheffé's lemma $g_n\to q_\eta$ in $L^1(dz)$.
Now suppose that q_η is not continuous. Then there is for every $\varepsilon>0$ a nonnegative continuous function with compact support q_η^ε such that $\|q_\eta - q_\eta^\varepsilon\|_{L^1(dz)} < \varepsilon$ [11, proposition 7.9]. Using the triangle inequality, we easily estimate

$$\|P(x_n,\cdot)-P(y_n,\cdot)\|_{\mathrm{TV}} \le 2\varepsilon + \int \left| \frac{q_\eta^\varepsilon(\sigma(x_n)^{-1}\{\sigma(y_n)z-b(x_n)+b(y_n)\})}{\det(\sigma(y_n)^{-1}\sigma(x_n))} - q_\eta^\varepsilon(z)\right| dz.$$

But we have already established that the second term on the right converges to zero as $n\to\infty$, and $\varepsilon>0$ is arbitrary. This completes the proof. □

Filter stability now follows from theorem 2.4 whenever $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}$. It remains to prove that when $q_\eta, q_\xi>0$ the absolute continuity requirement is in fact superfluous:

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\mu\text{-a.s.} \qquad \text{for any } \mu,\nu.$$

Indeed, if this is the case, then by the triangle inequality

$$\|\pi_n^\mu - \pi_n^\nu\|_{\mathrm{TV}} \le \|\pi_n^\gamma - \pi_n^\mu\|_{\mathrm{TV}} + \|\pi_n^\gamma - \pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0 \quad \mathbf{P}^\gamma\text{-a.s.} \qquad \text{for any } \mu,\nu,\gamma,$$

which completes the proof of proposition 2.5.
To establish the claim, note that when $q_\eta>0$ the transition kernel $P(x,\cdot)$ has a strictly positive density with respect to the Lebesgue measure for every $x\in E$. In particular, $P(x,\cdot)\sim P(z,\cdot)$ for every $x,z\in E$. As $q_\xi>0$ the filtering recursion is well defined under any initial measure, and it is immediately evident from the filtering recursion that π_{1-}^μ has a strictly positive density with

respect to the Lebesgue measure for every μ. In particular, this implies that $\pi^\mu_{1-} \sim \pi^\nu_{1-}$ for every μ, ν. But it is not difficult to establish that (see, e.g., the proof of [15, lemma 5.12])

$$\mathbf{E}^\mu\Big(\limsup_{n\to\infty} \|\pi^\mu_n - \pi^\nu_n\|_{\mathrm{TV}}\Big| Y_0 = y\Big) = \mathbf{E}^{\tilde\mu(y)}\Big(\limsup_{n\to\infty} \|\pi^{\tilde\mu(y)}_n - \pi^{\tilde\nu(y)}_n\|_{\mathrm{TV}}\Big),$$

where $\tilde\mu(y)$ and $\tilde\nu(y)$ denote the regular conditional probabilities $\mathbf{P}^\mu(X_1 \in \cdot|Y_0 = y)$ and $\mathbf{P}^\nu(X_1 \in \cdot|Y_0 = y)$ (i.e., $\tilde\mu(Y_0) = \pi^\mu_{1-}$ and $\tilde\nu(Y_0) = \pi^\nu_{1-}$). As $\tilde\mu(y) \sim \tilde\nu(y)$ for (almost) every $y \in \mathbb{R}^n$, we have $\mathbf{P}^{\tilde\mu(y)}|_{\sigma\{(Y_k)_{k\ge0}\}} \sim \mathbf{P}^{\tilde\nu(y)}|_{\sigma\{(Y_k)_{k\ge0}\}}$ and the claim follows from theorem 2.4.

Remark 5.3. An almost identical argument shows that when $q_\eta, q_\xi > 0$, absolute continuity of the observations $\mathbf{P}^\mu|_{\sigma\{(Y_k)_{k\ge0}\}} \ll \mathbf{P}^\nu|_{\sigma\{(Y_k)_{k\ge0}\}}$ holds for any pair of initial measures μ, ν. Together with theorem 2.4 this gives the desired claim. Note in particular that in this setting π^μ_n is $\mathbf{P}^\nu$-a.s. uniquely defined for any μ, ν, so that the statement of the proposition 2.5 is in fact well posed (i.e., we do not need to be careful to choose a specific version of the filter).

References

[1] Rami Atar and Ofer Zeitouni, *Exponential stability for nonlinear filtering*, Ann. Inst. H. Poincaré Probab. Statist. **33** (1997), no. 6, 697–725. MR1484538

[2] David Blackwell and Lester Dubins, *Merging of opinions with increasing information*, Ann. Math. Statist. **33** (1962), 882–886. MR0149577 (26 #7062)

[3] A. Budhiraja and D. Ocone, *Exponential stability in discrete-time filtering for non-ergodic signals*, Stochastic Process. Appl. **82** (1999), no. 2, 245–257. MR1700008

[4] Olivier Cappé, Eric Moulines, and Tobias Rydén, *Inference in hidden Markov models*, Springer Series in Statistics, Springer, New York, 2005, With Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat. MR2159833

[5] Pavel Chigansky and Robert Liptser, *Stability of nonlinear filters in nonmixing case*, Ann. Appl. Probab. **14** (2004), no. 4, 2038–2056. MR2099662

[6] ———, *On a role of predictor in the filtering stability*, Electron. Comm. Probab. **11** (2006), 129–140 (electronic). MR2240706

[7] D. Crisan and K. Heine, *Stability of the discrete time filter in terms of the tails of noise distributions*, J. London Math. Soc. **78** (2008), 441–458.

[8] D. Crisan and B. Rozovsky (eds.), *The Oxford University handbook of nonlinear filtering*, Oxford University Press, 2009, To appear.

[9] 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. MR1819122

[10] R. M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002, Revised reprint of the 1989 original. MR1932358

[11] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR1681462

[12] 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. MR1372912

[13] Nadia Oudjane and Sylvain Rubenthaler, *Stability and uniform particle approximation of non-linear filters in case of non ergodic signals*, Stoch. Anal. Appl. **23** (2005), no. 3, 421–448. MR2140972

[14] 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. MR1368405

[15] R. van Handel, *The stability of conditional Markov processes and Markov chains in random environments*, 2008, Preprint, http://arxiv.org/abs/0801.4366.

[16] ———, *Uniform observability of hidden Markov models and filter stability for unstable signals*, 2008, Preprint, http://arxiv.org/abs/0804.2885.

[17] ———, *Observability and nonlinear filtering*, Probab. Th. Rel. Fields (2009), To appear. Published electronically at `http://dx.doi.org/10.1007/s00440-008-0161-y`.