FORGETTING OF THE INITIAL DISTRIBUTION FOR NON-ERGODIC HIDDEN MARKOV CHAINS

By Randal Douc

Institut Télécom / Télécom SudParis, UMR CNRS 5157

By Elisabeth Gassiat

Université Paris-Sud 11, CNRS UMR 8628

By Benoit Landelle

Thales Optronique

Université Paris-Sud 11, CNRS UMR 8628

By Eric Moulines

Institut Télécom / Télécom ParisTech, UMR CNRS 5181

Abstract In this paper, the forgetting of the initial distribution for a non-ergodic Hidden Markov Models (HMM) is studied. A new set of conditions is proposed to establish the forgetting property of the filter, which significantly extends all the existing results. Both a pathwise and mean convergence of the total variation distance of the filter started from two different initial distributions are considered. The results are illustrated using a generic non-ergodic state-space models for which both pathwise and mean exponential stability is established.

1. Introduction and notations. There are many applications where the current state of a dynamical system need to be estimated from observations up to the current time. In this paper, it is assumed that the underlying state process $\{X_k\}_{k\geq 0}$ (often referred to as the *signal* process) is a general state space discrete time Markov chain and the *observation process* $\{Y_k\}_{k\geq 0}$ is independent conditionally to the state sequence. More specifically, let X and Y be separable Polish spaces endowed with their Borel σ-fields $\mathcal{X}$ and $\mathcal{Y}$. We denote by Q the transition kernel on $(\mathsf{X}, \mathcal{X})$, μ a measure on $(\mathsf{Y}, \mathcal{Y})$ and a transition density g from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$. Consider the Markov transition

AMS 2000 subject classifications: Primary, 93E11,60G35; Secondary, 62C10

Keywords and phrases: Non-linear filtering, forgetting of the initial distribution, non-ergodic Hidden Markov Chains, Feynman-Kac semigroup

kernel defined for any $(x, y) \in \mathsf{X} \times \mathsf{Y}$ and $C \in \mathcal{X} \otimes \mathcal{Y}$ by

$$T[(x,y), C] \stackrel{\text{def}}{=} \iint Q(x, dx')\, g(x', y') \mathbb{1}_C(x', y') \mu(\mathrm{d}y') \ . \tag{1}$$

We consider $\{(X_k, Y_k)\}_{k \geq 0}$ the Markov chain with transition kernel T and initial distribution $C \mapsto \iint g(x, y) \mathbb{1}_C(x, y) \nu(dx) \mu(\mathrm{d}y)$, where ν is a probability measure on $(\mathsf{X}, \mathcal{X})$. With a slight abuse in the terminology, ν is referred to as the initial distribution of $\{(X_k, Y_k)\}_{k \geq 0}$ and we denote by $\mathbb{P}_\nu$ the distribution of this process over a suitably defined measurable space $(\Omega, \mathcal{F})$. We assume that the chain $\{X_k\}_{k \geq 0}$ is not observed. The distribution of the hidden state X_n conditionally on the observations $Y_{0:n} \stackrel{\text{def}}{=} [Y_0, \ldots, Y_n]$, denoted $\phi_{\nu,n}[Y_{0:n}]$, is referred to as the *filtering* distribution.

A typical question consists in finding conditions under which the filtering distribution is *stable*, *i.e.* that an appropriately chosen distance between the filtering distributions $\phi_{\nu,n}[Y_{0:n}]$ and $\phi_{\nu',n}[Y_{0:n}]$ for two different choices of the initial distribution ν and ν' vanishes as n goes to infinity. In this paper, assuming that $\{Y_k\}_{k \geq 0}$ is a Y-valued stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P}_\star)$, our objective is to establish either pathwise or mean filter stability in the total variation distance

$$\limsup_{n \to \infty} \|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} = 0 \ , \quad \mathbb{P}_\star - \text{a.s.} \ , \tag{2}$$

$$\limsup_{n \to \infty} \mathbb{E}_\star \left[\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right] = 0 \ , \tag{3}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. In contrast with most contributions on this subject, $\mathbb{P}_\star$ need not be equal to $\mathbb{P}_\nu$ which means that our results apply even if the filtering model is mis-specified. Under more stringent conditions, we may strengthen (2) or (3) by specifying rates of convergence. Of particular importance are the exponential rates (or exponential stability), which amounts to requiring that

$$\limsup_{n \to \infty} n^{-1} \log \left(\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right) < 0 \ , \quad \mathbb{P}_\star - \text{a.s.} \ , \tag{4}$$

$$\limsup_{n \to \infty} n^{-1} \log \mathbb{E}_\star \left[\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right] < 0 \ . \tag{5}$$

As stressed by [5], the most important motivation for studying the stability of the filter is the time-uniform convergence of estimators of the filtering distribution. Since these estimators are most often defined recursively, the approximation error at a given time instant has an impact at all subsequent time instants. As shown in [6], the propagation of error can be considered as an incorrect initialization at the time when the error was made. If the rate of convergence of the filter is fast enough (e.g. if the filter is exponentially stable), then the effects of these local errors do not build up. Another important application of the stability is for the inference of the transition kernel Q or the likelihood g, when these quantities belong either to parametric or non-parametric family of distributions. As shown in [8], the convergence of the likelihood of the observation and the consistency of the maximum likelihood estimator rely on the stability of the filter for a mis-specified model of the observations (several examples of this type will be given later).

The stability of the filter in nonlinear state space models has attracted many research efforts; see for example the in-depth tutorial of [4]. The brief overview below is mainly intended to allow comparison of assumptions and results presented in this contribution with respect to those previously reported in the literature.

The filtering equation can be seen as a positive random non-linear operator acting on the space of probability measures; the stability can be investigated using tools from the theory of positive operators, namely the Birkhoff contraction inequality for the Hilbert projective metric (see [1], [13], [12]). The results obtained using this approach require strong *mixing* conditions for the transition kernels: there exist positive constants ϵ_- and ϵ_+ and a probability measure λ on $(\mathsf{X}, \mathcal{X})$ such that

$$\epsilon_- \lambda(f) \leq Q(x, f) \leq \epsilon_+ \lambda(f) \ , \quad \text{for any } x \in \mathsf{X}, A \in \mathcal{X}. \tag{6}$$

This condition in particular implies that the chain is uniformly geometrically ergodic. Under weak additional assumptions on the likelihood, (6) allows to establish both pathwise and mean exponential stability of the filter, with bounds which are uniform with respect to the observations $Y_{0:n}$.

In [13], the stability of the optimal filter is studied for a class of kernels referred to as *pseudo-mixing*. The definition of pseudo-mixing kernel is adapted to the case where the state space is $\mathsf{X} = \mathbb{R}^d$, equipped with the Borel sigma-field $\mathcal{X}$. A kernel Q on $(\mathsf{X}, \mathcal{X})$ is *pseudo-mixing* if for any compact set C with a diameter d large enough, there exist positive constants $\epsilon_-(d) > 0$ and $\epsilon_+(d) > 0$ and a measure λ_{C} (which may be chosen to be finite without loss of generality) such that

$$\epsilon_-(d) \lambda_{\mathsf{C}}(A) \leq Q(x, A) \leq \epsilon_+(d) \lambda_{\mathsf{C}}(A) \ , \quad \text{for any } x \in \mathsf{C},\ A \in \mathcal{X}. \tag{7}$$

This condition is more general than (6), but still it is not satisfied in the linear Gaussian case (see [13, Example 4.3]).

A significant improvement has been achieved by [11], who considered the filtering problem of a signal $\{X_k\}_{k \geq 0}$ taking values in $\mathsf{X} = \mathbb{R}^d$ filtered from observations $\{Y_k\}_{k\geq 0}$ in $\mathsf{Y} = \mathbb{R}^\ell$,

$$X_{k+1} = f(X_k) + \sigma(X_k)\zeta_k \; , \tag{8}$$

$$Y_k = h(X_k) + \beta\varepsilon_k \; . \tag{9}$$

Here $\{(\zeta_k, \varepsilon_k)\}_{k\geq 0}$ is a i.i.d. sequence of random vectors in $\mathbb{R}^{d+\ell}$ with density $q_\zeta(x)q_\varepsilon(y)$, $f(\cdot)$ is a d-dimensional vector function, $\sigma(\cdot)$ a $d \times d$-matrix function, $h(\cdot)$ is a ℓ-dimensional vector-function and $\beta > 0$. The authors established both pathwise (2) and mean (3) stability of the filter under appropriate conditions on the functions f, h and σ and on the signal and measurement noise $\{(\zeta_k, \varepsilon_k)\}_{k\geq 0}$. These conditions cover (with some restrictions) the linear gaussian state space model. Note however that these results hold only if $\mathbb{P}_\star = \mathbb{P}_\nu$ and $\nu' \ll \nu$. These results were later extended in [7]. Both pathwise and mean stability are established for initial distributions ν and ν' that are not necessarily comparable and a distribution $\mathbb{P}_\star$ which might be different from $\mathbb{P}_\nu$. The results hold under weaker conditions than those mentioned above; in particular, these results cover the linear Gaussian state-space model without restriction on the measurement and noise variance.

The works mentioned above mainly are obtained under the assumption that the signal process is ergodic. Results for non-ergodic signals in the linear Gaussian case have been obtained in [14]. Non-linear non-ergodic state-space models have been considered much less frequently in the literature. These extensions are important because many models in engineering or econometrics are non-ergodic (see [9] and [16] and the references therein). In [3], the model (8)-(9) is considered: f is assumed to be Lipshitz and $h(x) = x$, for all $x \in \mathsf{X} = \mathbb{R}^d$. The pathwise exponential stability is established [3, Theorem 2.1] under the assumptions that the state and the observation noise are both gaussian and that the variance of the observation noise is small enough. More general distributions for the state and the observation noises are considered in [3, Theorem 2.4], but still the exponential stability is obtained only under the condition that the scale of the observation noise is small enough.

These results were later extended in [15] to allow more general functions h. The authors establish stability in the mean of the filter, under conditions essentially stating that the tails of the observation noise $\{\varepsilon_k\}_{k\geq 0}$ are sufficiently light compared to the tails of the signal noise $\{\zeta_k\}_{k\geq 0}$. These results are derived under the additional assumption that the two initial conditions ν and ν' are comparable (*i.e.* $\nu \ll \nu'$ and $\nu' \ll \nu$) and that the distribution of the observation $\mathbb{P}_\star = \mathbb{P}_\nu$. Similar conditions have been studied in [5], which established the pathwise stability, again under $\mathbb{P}_\nu$. The conditions in these two publications are not equivalent; in particular [5] assume that $\sigma \equiv 1$ in (8) and that the signal and observation noises are i.i.d. whereas [15] allow a form of weak dependence in the signal noise (see Section 4 for further discussion).

In a related work, [10] have considered the stability of the filter for denumerable Markov chains. In this work, the observation equation (9) holds with $\sigma \equiv 1$ and $\{X_k\}_{k \geq 0}$ is a finite or denumerable Markov chain. The authors establish exponential pathwise stability when the noise variance is sufficiently small and h is one-to-one. Here again, $\nu \ll \nu'$ and the distribution of the observation process is $\mathbb{P}_\star = \mathbb{P}_\nu$.

A significant weakening of these assumptions has been achieved in [17] and [18]. These contributions establish the stability of the filter (in bounded Lipshitz norm) for an observation model (9) under the conditions that h possesses a uniformly continuous inverse and the noise $\{\varepsilon_k\}_{k \geq 0}$ has a density with respect to the Lebesgue measure whose Fourier transform vanishes nowhere but without imposing any assumption on the transition kernel Q of the signal. Stability in total variation distance can be obtained under the uniform strong Feller assumption, *i.e.* that $x \mapsto Q(x, \cdot)$ is uniformly continuous for the total variation distance on the space of probability measures. The pathwise and the mean filter stability are obtained are obtained under the conditions that the initial distributions of the process ν and ν' satisfy $\nu \ll \nu'$ and the distribution of the observation process is $\mathbb{P}_\star = \mathbb{P}_\nu$.

In this contribution, we propose a new set of conditions to establish pathwise and mean filter stability under possible model mis-specification. We assume an observation model that can be more general than (9) and do not assume that $\nu \ll \nu'$; in addition, the distribution of the observation process $\mathbb{P}_\star$ is not constrained to be $\mathbb{P}_\nu$ and may, on the contrary, be fairly general. Compared to the very weak conditions introduced in [17] and [18], the price to pay are stronger conditions on the transition kernel Q, which are reminiscent from the Local Doeblin condition introduced in [7].

The paper is organized as follows. In section 2, the assumptions are introduced and the main results are stated. In Theorem 5, the pathwise stability of the filter (2) is established and an explicit bound of the deviation is given. In Theorem 6, the average stability of the filter (3) is established together with a computable bound. In section 3, different nonlinear state-space models are considered. For these models, we provide conditions upon which the exponential pathwise and mean stability hold. Several technical Lemmas required to study the examples are given in Sections 4 and 5.

2. Main results. Our results require the existence of a set-valued function, referred to as *Local Doeblin (LD) set function*, which extends the so-called LD-sets introduced in [19] and later exploited in [11]. The difference between LD-sets of [19] and LD-set functions lies in the dependence on the successive observations.

DEFINITION 1 (LD-set function). *A set-valued function* $\mathsf{C} : y \longmapsto \mathsf{C}(y)$ *from* Y *to* $\mathcal{X}$ *is called a* Local Doeblin set function (LD-set function) *if there exist a measurable function* $(y, y') \longmapsto \left(\varepsilon^-_{\mathsf{C}}(y, y'), \varepsilon^+_{\mathsf{C}}(y, y')\right)$ *from* $\mathsf{Y} \times \mathsf{Y}$ *to* $(0, \infty)^2$ *and a transition kernel* $\lambda : \mathsf{Y} \times \mathsf{Y} \times \mathcal{X} \to [0, 1]$ [1] *such that, for all* $x \in \mathsf{C}(y)$ *and* $A \in \mathcal{X}$,

$$\varepsilon^-_{\mathsf{C}}(y, y')\lambda\left(y, y'; A \cap \mathsf{C}(y')\right) \le Q[x, A \cap \mathsf{C}(y')] \le \varepsilon^+_{\mathsf{C}}(y, y')\lambda\left(y, y'; A \cap \mathsf{C}(y')\right) \ . \tag{10}$$

Consider the following assumptions on the likelihood of the observations:

H 1. g is continuous and positive.

This excludes the case of additive noise with bounded support; see for example [2]. Stability of the filter may hold in such context, but the fact that the likelihood might vanish creates additional technical difficulties which will obscure the main points of the paper. In particular, under this assumption, for any distribution ν on $(\mathsf{X}, \mathcal{X})$, $n \ge 0$ and sequence $y_{0:n} \in \mathsf{Y}^{n+1}$,

$$\mathbb{E}^Q_\nu\left[\prod_{k=0}^n g(X_k, y_k)\right] \stackrel{\text{def}}{=} \int \cdots \int \nu(dx_0) \prod_{k=1}^n Q(x_{k-1}, dx_k) \prod_{k=0}^n g(x_k, y_k) > 0 \ . \tag{11}$$

The filtering distribution can thus be expressed, for $A \in \mathcal{X}$,

$$\phi_{\nu,n}[y_{0:n}] \stackrel{\text{def}}{=} \frac{\int \cdots \int \nu(dx_0) \prod_{k=1}^n Q(x_{k-1}, dx_k) \prod_{k=0}^n g(x_k, y_k) \mathbb{1}_A(x_n)}{\int \cdots \int \nu(dx_0) \prod_{k=1}^n Q(x_{k-1}, dx_k) \prod_{k=0}^n g(x_k, y_k)} \ . \tag{12}$$

The continuity can also be relaxed, also at the expense of some minor technical adaptations. The main idea of the proof is that the states belong very often to the LD-sets. Every time the state is in a LD set and

[1] for any $(y, y') \in \mathsf{Y} \to \mathsf{Y}$, $\lambda(y, y'; \cdot)$ is a σ-finite measure on $(\mathsf{X}, \mathcal{X})$ and for any $A \in \mathcal{X}$, the function $(y, y') \mapsto \lambda(y, y'; A)$ is measurable from $(\mathsf{Y} \times \mathsf{Y}, \mathcal{Y} \otimes \mathcal{Y})$ for $[0, 1]$ equipped with its Borel σ-field

jumps to another LD set, the forgetting mechanism comes into play. From now on, for all $(x, x') \in \mathsf{X}^2$, denote by $\bar{x} = (x, x')$ and by $\bar{g}$ the product $\bar{g}(\bar{x}, y) = g(x, y)g(x', y)$. Similarly, for all $A \in \mathcal{X}$, denote $\bar{A} = A \times A$, for all LD-set function C, $\bar{\mathsf{C}}$ the set-valued function $\bar{\mathsf{C}}(y) = \mathsf{C}(y) \times \mathsf{C}(y)$. For all $(x, x') \in \mathsf{X}^2$, and A, $B \in \mathcal{X}$, set $\bar{Q}(x, x', A \times B) = Q(x, A)Q(x', B)$. Finally, for ν,ν' two probability distributions on $(\mathsf{X}, \mathcal{X})$, we denote by $\mathbb{E}^{Q}_{\nu}$ and $\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}$ the expectation with respect to the distribution of a Markov chain on X (resp. on $\mathsf{X} \times \mathsf{X}$) with initial distribution ν (resp. $\nu \otimes \nu'$) and transition kernel Q (resp. $\bar{Q}$). Then, under the stated assumptions, for any $A \in \mathcal{X}$, any ν and ν' two probability distributions on $(\mathsf{X}, \mathcal{X})$, any integer n and any sequence $y_{0:n} \in \mathsf{Y}^{n+1}$, the difference $\phi_{\nu,n}[y_{0:n}](A) - \phi_{\nu',n}[y_{0:n}](A)$ may be expressed as

$$
\begin{aligned}
(13)\quad & \phi_{\nu,n}[y_{0:n}](A) - \phi_{\nu',n}[y_{0:n}](A) \\
&= \frac{\mathbb{E}^{Q}_{\nu}\left[\prod_{i=0}^{n} g(X_i, y_i)\mathbb{1}_A(X_n)\right]}{\mathbb{E}^{Q}_{\nu}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]} - \frac{\mathbb{E}^{Q}_{\nu'}\left[\prod_{i=0}^{n} g(X_i, y_i)\mathbb{1}_A(X_n)\right]}{\mathbb{E}^{Q}_{\nu'}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]}, \\
&= \frac{\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i)\mathbb{1}_A(X_n)\right] - \mathbb{E}^{\bar{Q}}_{\nu'\otimes\nu}\left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i)\mathbb{1}_A(X_n)\right]}{\mathbb{E}^{Q}_{\nu}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]\mathbb{E}^{Q}_{\nu'}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]}, \\
&= \frac{\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i)\{\mathbb{1}_A(X_n) - \mathbb{1}_A(X'_n)\}\right]}{\mathbb{E}^{Q}_{\nu}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]\mathbb{E}^{Q}_{\nu'}\left[\prod_{i=0}^{n} g(X_i, y_i)\right]}.
\end{aligned}
$$

We compute bounds for the numerator and the denominator of the previous expression. Such bounds are given in the two following Propositions. For an LD-set function C denote:

$$
(14)\qquad \rho_{\mathsf{C}}(y, y') \stackrel{\text{def}}{=} 1 - (\varepsilon^{-}_{\mathsf{C}}/\varepsilon^{+}_{\mathsf{C}})^2(y, y').
$$

For any integer n and any sequence $\{y_i\}_{i=0}^{n}$ in Y, let us define

$$
\begin{aligned}
(15)\quad & \Delta_n(\nu, \nu', y_{0:n}) \\
&= \sup_{A\in\mathcal{X}} \left| \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i)\mathbb{1}_A(X_n)\right] - \mathbb{E}^{\bar{Q}}_{\nu'\otimes\nu}\left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i)\mathbb{1}_A(X_n)\right] \right|.
\end{aligned}
$$

PROPOSITION 2. *Let C be an LD-set function and ν and ν' be two probability measures on $(\mathsf{X}, \mathcal{X})$. Then,*

$$
\Delta_n(\nu, \nu', y_{0:n}) \le \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left\{\bar{g}(\bar{X}_0, y_0)\prod_{i=1}^{n} \bar{g}(\bar{X}_i, y_i)\rho^{\delta_i}_{\mathsf{C}}(y_{i-1}, y_i)\right\},
$$

where $\delta_i = \mathbb{1}_{\bar{\mathsf{C}}(y_{i-1})\times\bar{\mathsf{C}}(y_i)}(\bar{X}_{i-1}, \bar{X}_i)$.

PROOF. For convenience, we write $\mathsf{C}_i = \mathsf{C}(y_i)$, $\varepsilon_i^- = \varepsilon_{\mathsf{C}}^-(y_{i-1}, y_i)$, $\varepsilon_i^+ = \varepsilon_{\mathsf{C}}^+(y_{i-1}, y_i)$, $g_i(x) = g(x, y_i)$, $\lambda_i(\cdot) = \lambda(y_{i-1}, y_i; \cdot)$ and $\rho_i = 1 - (\varepsilon_i^-/\varepsilon_i^+)^2$. Let us define $\bar{\lambda}_i \stackrel{\text{def}}{=} \lambda_i \otimes \lambda_i$. Since C is an LD-set function, for all $i = 1, \ldots, n$, $\bar{x} \in \bar{\mathsf{C}}_{i-1}$, and $\bar{f}$ a non-negative function on $\mathsf{X} \times \mathsf{X}$,

$$(\varepsilon_i^-)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \leq \bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \leq (\varepsilon_i^+)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \; . \tag{16}$$

Define the sequence of unnormalized kernels $\bar{Q}_i^0$ and $\bar{Q}_i^1$ as follows: for all $\bar{x} \in \mathsf{X}^2$, and $\bar{f}$ a non-negative function on $\mathsf{X} \times \mathsf{X}$,

$$\begin{aligned}
\bar{Q}_i^0(\bar{x}, \bar{f}) &= (\varepsilon_i^-)^2 \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \; , \\
\bar{Q}_i^1(\bar{x}, \bar{f}) &= \bar{Q}(\bar{x}, \bar{f}) - (\varepsilon_i^-)^2 \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \; .
\end{aligned}$$

It follows from (16) that, for all $\bar{x}$ in $\bar{\mathsf{C}}_{i-1}$, $0 \leq \bar{Q}_i^1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \leq \rho_i \bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f})$ which implies that, for all $\bar{x} \in \mathsf{X}^2$,

$$\begin{aligned}
\bar{Q}_i^1(\bar{x}, \bar{f}) &= \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{Q}_i^1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{Q}_i^1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i^c} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}_{i-1}^c}(\bar{x}) \bar{Q}_i^1(\bar{x}, \bar{f}) \; , \\
&\leq \rho_i \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \bar{Q}_i^1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i^c} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}_{i-1}^c}(\bar{x}) \bar{Q}_i^1(\bar{x}, \bar{f}) \; , \\
&\leq \bar{Q}\left(\bar{x}, \rho_i^{\mathbb{1}_{\bar{\mathsf{C}}_{i-1}}(\bar{x}) \mathbb{1}_{\bar{\mathsf{C}}_i}} \bar{f}\right) \; .
\end{aligned} \tag{17}$$

We write $\Delta_n(\nu, \nu', y_{0:n}) = \sup_{A \in \mathcal{X}} |\Delta_n(A)|$, where

$$\Delta_n(A) \stackrel{\text{def}}{=} \nu \otimes \nu'(\bar{g}_0 \bar{Q} \bar{g}_1 \ldots \bar{Q} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q} \bar{g}_1 \ldots \bar{Q} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \; .$$

We decompose $\Delta_n(A)$ into $\Delta_n(A) = \sum_{t_{0:n-1} \in \{0,1\}^n} \Delta_n(A, t_{0:n-1})$, where

$$\begin{aligned}
\Delta_n(A, t_{0:n-1}) \stackrel{\text{def}}{=} \nu \otimes \nu'(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \\
- \nu' \otimes \nu(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \; .
\end{aligned}$$

Note that, for any $t_{0:n-1} \in \{0,1\}^n$ and any sets $A, B \in \mathcal{X}$,

$$\nu \otimes \nu'(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times B}) = \nu' \otimes \nu(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{B \times A}) \; .$$

If there is an index $i \in \{0, \ldots, n-1\}$ such that $t_i = 0$, then

$$\begin{aligned}
&\nu \otimes \nu'(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{i-1}^{t_{i-1}} \bar{g}_i \mathbb{1}_{\bar{\mathsf{C}}_i}) \times (\varepsilon_{i+1}^-)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_{i+1}} \bar{g}_{i+1} \bar{Q}_{i+1}^{t_{i+1}} \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \; , \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_0^{t_0} \bar{g}_1 \ldots \bar{Q}_{i-1}^{t_{i-1}} \bar{g}_i \mathbb{1}_{\bar{\mathsf{C}}_i}) \times (\varepsilon_{i+1}^-)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_{i+1}} \bar{g}_{i+1} \bar{Q}_{i+1}^{t_{i+1}} \ldots \bar{Q}_{n-1}^{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}}) \; .
\end{aligned}$$

Thus, $\Delta_n(A, t_{0:n-1}) = 0$ except if for all $i \in \{0, \ldots, n-1\}$, $t_i = 1$, and we obtain

$$\Delta_n(A) = \nu \otimes \nu' \Big[\bar{g}_0 \bar{Q}_0^1 \bar{g}_1 \ldots \bar{Q}_{n-1}^1 \bar{g}_n (\mathbb{1}_{A \times \mathsf{X}} - \mathbb{1}_{\mathsf{X} \times A})\Big] .$$

It then follows using (17)

$$\Delta_n(\nu, \nu', y_{0:n}) \leq \nu \otimes \nu' (\bar{g}_0 \bar{Q}_0^1 \bar{g}_1 \ldots \bar{Q}_{n-1}^1 \bar{g}_n) \leq \mathbb{E}^{\bar{Q}}_{\nu \otimes \nu'} \left[\bar{g}(\bar{X}_0, y_0) \prod_{i=1}^n \bar{g}(\bar{X}_i, y_i) \rho_i^{\delta_i} \right] ,$$

with $\delta_i = \mathbb{1}_{\bar{\mathsf{C}}_{i-1} \times \bar{\mathsf{C}}_i}(\bar{X}_{i-1}, \bar{X}_i)$. □

We now compute a bound for the denominator. For a given LD-set function C , we set

(18)
$$\Phi_{\nu, \mathsf{C}}(y, y') \stackrel{\text{def}}{=} \mathbb{E}^Q_\nu \left[g(\cdot, y_0) g(\cdot, y') \mathbb{1}_{\mathsf{C}(y_1)}(\cdot) \right] = \nu \left[g(\cdot, y) Q g(\cdot, y') \mathbb{1}_{\mathsf{C}(y')}(\cdot) \right] ,$$

(19)
$$\Psi_{\mathsf{C}}(y, y') \stackrel{\text{def}}{=} \lambda \left(y, y'; g(\cdot, y') \mathbb{1}_{\mathsf{C}(y')} \right) .$$

Proposition 3. *Let C be an LD-set function and $\{y_i\}_{i=0}^n$ a sequence in Y. We have for all $n \in \mathbb{N}$*

$$\mathbb{E}^Q_\nu \left[\prod_{i=0}^n g(X_i, y_i) \right] \geq \Phi_{\nu, \mathsf{C}}(y_0, y_1) \prod_{i=2}^n \left(\varepsilon^-_{\mathsf{C}}(y_{i-1}, y_i) \Psi_{\mathsf{C}}(y_{i-1}, y_i) \right) . \tag{20}$$

Proof of Proposition 3. Since C is an LD-set function, there exist some applications $\varepsilon^-_{\mathsf{C}}$, $\varepsilon^+_{\mathsf{C}}$ such that, for all $i = 1, \ldots, n$, for all $x \in \mathsf{C}(y_{i-1})$ and for all $A \in \mathcal{X}$ with $A \subset \mathsf{C}(y_i)$,

$$\varepsilon^-_{\mathsf{C}}(y_{i-1}, y_i) \lambda(y_{i-1}, y_i; A) \leq Q(x, A) \leq \varepsilon^+_{\mathsf{C}}(y_{i-1}, y_i) \lambda(y_{i-1}, y_i; A) . \tag{21}$$

Obviously,

$$\mathbb{E}^Q_\nu \left[\prod_{i=0}^n g(X_i, y_i) \right] \geq \mathbb{E}^Q_\nu \left[g(X_0, y_0) \prod_{i=1}^n g(X_i, y_i) \mathbb{1}_{\mathsf{C}(y_i)}(X_i) \right] .$$

Then, the right-hand side of this expression may be bounded using (21) by

$$\begin{aligned}
&\mathbb{E}^Q_\nu \left[g(X_0, y_0) \prod_{i=1}^n g(X_i, y_i) \mathbb{1}_{\mathsf{C}(y_i)}(X_i) \right] \\
&\quad = \mathbb{E}^Q_\nu \left[g(X_0, y_0) g(X_1, y_1) \mathbb{1}_{\mathsf{C}(y_1)}(X_1) \prod_{i=2}^n g(X_i, y_i) \mathbb{1}_{\mathsf{C}(y_{i-1}) \times \mathsf{C}(y_i)}(X_{i-1}, X_i) \right] , \\
&\quad \geq \nu [g(\cdot, y_0) Q g(\cdot, y_1) \mathbb{1}_{\mathsf{C}(y_1)}(\cdot)] \prod_{i=2}^n \varepsilon^-_{\mathsf{C}}(y_{i-1}, y_i) \lambda \left(y_{i-1}, y_i; g(\cdot, y_i) \mathbb{1}_{\mathsf{C}(y_i)} \right) .
\end{aligned}$$

□

Under (H1), $\Phi_{\nu,\mathsf{C}}(y_0, y_1) > 0$ for any initial distribution ν such that $\nu Q[\mathsf{C}(y_1)] > 0$. In the examples considered in Section 3, this condition is satisfied for the choices of Local Doeblin sets by any initial distributions. Following the same lines as above, it is easily seen that the lower bound (20) can be more generally written as

$$\mathbb{E}_\nu^Q\left[\prod_{i=0}^n g(X_i, y_i)\right] \geq \mathbb{E}_\nu^Q\left[\prod_{i=0}^k g(\cdot, y_k)\mathbb{1}_{\mathsf{C}(y_k)}(\cdot)\right] \times \prod_{i=k+1}^n \left(\varepsilon_{\mathsf{C}}^-(y_{i-1}, y_i)\Psi_{\mathsf{C}}(y_{i-1}, y_i)\right) .$$

This lower bound is positive as soon as $\nu Q^k[\mathsf{C}(y_k)] > 0$. The statements of the results below can be directly extended to handle this more general condition.

By combining these two Propositions, we obtain an explicit bound for the total variation distance $\left\|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\right\|_{\mathrm{TV}}$. For a set $A \in \mathcal{X}$ and an observation $y \in \mathsf{Y}$, the supremum of the likelihood over A is denoted

$$\Upsilon_A(y) \stackrel{\text{def}}{=} \sup_{x \in A} g(x, y) . \tag{22}$$

Consider the following assumption:

H 2. For any $\eta \in (0,1)$, there exists an LD-set function C_η such that $y \mapsto \Upsilon_{\mathsf{C}_\eta^c(y)}(y)$ is measurable and for all $y \in \mathsf{Y}$,

$$\Upsilon_{\mathsf{C}_\eta^c(y)}(y) \leq \eta \Upsilon_{\mathsf{X}}(y) . \tag{23}$$

When $\mathsf{X} = \mathbb{R}^d$, this assumption is typically satisfied when, for any given y, the likelihood goes to zero as the state $|x|$ goes to infinity: $\lim_{|x|\to\infty} g(x, y) = 0$. This condition is satisfied in many models of practical interest, and roughly implies that the observation effectively provides information on the state range of values.

It is worthwhile to note that the bound we obtain is valid for any sequence $y_{0:n}$ and any initial distributions ν and ν'. Under assumption (H2), for any $\eta \in (0,1)$ there exists a LD-set function C_η satisfying (23). For any $\alpha \in (0,1)$ and a sequence $y_{0:n} = \{y_i\}_{i=0}^n$ in Y, define

$$\Lambda_\eta(y_{0:n}, \alpha) \stackrel{\text{def}}{=} \max\left\{\prod_{k=1}^n \rho_\eta^{\delta_k}(y_{k-1}, y_k) \ : \{\delta_k\}_{k=1}^n \in \{0,1\}^n , \sum_{k=1}^n \delta_k \geq \alpha n\right\} , \tag{24}$$

where ρ_η is a shorthand notation for ρ_{C_η} (see (14)).

PROPOSITION 4. *Assume (H1)-(H2). Let C be an LD-set function. Let α be some number in $(0,1)$, ν and ν' some probability measures on $(\mathsf{X},\mathcal{X})$ and $\{y_i\}_{i=0}^n$ a sequence in Y. Then, for any $\eta > 0$,*

$$(25) \quad \|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}} \leq 2\Lambda_\eta(y_{0:n},\alpha) + \\ 2\eta^{a_n} \prod_{i=0}^n \Upsilon_{\mathsf{X}}^2(y_i) \prod_{i=2}^n \left(\varepsilon_{\mathsf{C}}^-(y_{i-1},y_i)\Psi_{\mathsf{C}}(y_{i-1},y_i)\right)^{-2} \Phi_{\nu,\mathsf{C}}^{-1}(y_0,y_1)\Phi_{\nu',\mathsf{C}}^{-1}(y_0,y_1) ,$$

with $a_n \stackrel{\mathrm{def}}{=} \frac{(1-\alpha)n}{2} - \frac{1}{2}$.

PROOF. Eq. (13) implies

$$\|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}} = \frac{2\Delta_n(\nu,\nu',y_{0:n})}{\mathbb{E}_\nu^Q\left[\prod_{i=0}^n g(X_i,y_i)\right]\mathbb{E}_{\nu'}^Q\left[\prod_{i=0}^n g(X_i,y_i)\right]} ,$$

where $\Delta_n(\nu,\nu',y_{0:n})$ is defined by (15). We stress that we will use two different LD-set functions, C_η for the numerator and C for the denominator. Set

$$N_{\eta,n} \stackrel{\mathrm{def}}{=} \sum_{i=1}^n \mathbb{1}\{(\bar{X}_{i-1},\bar{X}_i) \in \bar{\mathsf{C}}_\eta(y_{i-1}) \times \bar{\mathsf{C}}_\eta(y_i)\} .$$

Using Proposition 2, we obtain

$$(26) \quad \Delta_n(\nu,\nu',y_{0:n}) \leq \mathbb{E}_{\nu\otimes\nu'}^{\bar{Q}}\left[\bar{g}(\bar{X}_0,y_0)\prod_{i=1}^n \bar{g}(\bar{X}_i,y_i)\rho_\eta^{\delta_i}(y_{i-1},y_i)\mathbb{1}\{N_{\eta,n} \geq \alpha n\}\right] \\ + \mathbb{E}_{\nu\otimes\nu'}^{\bar{Q}}\left[\bar{g}(\bar{X}_0,y_0)\prod_{i=1}^n \bar{g}(\bar{X}_i,y_i)\rho_\eta^{\delta_i}(y_{i-1},y_i)\mathbb{1}\{N_{\eta,n} < \alpha n\}\right] ,$$

with $\delta_i = \mathbb{1}_{\bar{\mathsf{C}}_\eta(y_{i-1})\times\bar{\mathsf{C}}_\eta(y_i)}(\bar{X}_{i-1},\bar{X}_i)$. The first term in the right-hand side expression of (26) satisfies

$$\mathbb{E}_{\nu\otimes\nu'}^{\bar{Q}}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i,y_i)\prod_{i=1}^n \rho_\eta^{\delta_i}(y_{i-1},y_i)\mathbb{1}\{N_{\eta,n} \geq \alpha n\}\right] \\ \leq \mathbb{E}_{\nu\otimes\nu'}^{\bar{Q}}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i,y_i)\right]\Lambda_\eta(y_{0:n},\alpha) .$$

Consider now the second term of the rhs of (26). Let $M_{\eta,n} \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}^c_\eta(y_i)}(\bar{X}_i)$. For any sequence $\{u_j\}$, such that $u_j \in \{0,1\}$ for $j \in \{0,\ldots,n\}$,

$$n \geq \sum_{i=0}^{n-1} u_i \vee u_{i+1} = \sum_{i=0}^{n-1} (u_i + u_{i+1} - u_i u_{i+1}) \geq 2 \sum_{i=0}^{n-1} u_i - 1 - \sum_{i=0}^{n-1} u_i u_{i+1} ,$$

which implies that $\sum_{i=0}^{n-1} u_i \leq (n+1)/2 + (1/2) \sum_{i=1}^{n} u_{i-1} u_i$. Using this inequality with $u_i = \mathbb{1}\{\bar{X}_i \in \bar{\mathsf{C}}_\eta(y_i)\}$ for $i \in \{0,\ldots,n\}$ shows that $N_{\eta,n} < \alpha n$ implies that $M_{\eta,n} \geq a_n$. Then,

$$\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'} \left[\bar{g}(\bar{X}_0, y_0) \prod_{i=1}^{n} \bar{g}(\bar{X}_i, y_i) \rho_\eta^{\delta_i}(y_{i-1}, y_i) \mathbb{1}\{N_{\eta,n} < \alpha n\} \right] \\ \leq \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'} \left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i) \mathbb{1}\{M_{\eta,n} \geq a_n\} \right] .$$

By splitting this last product, we obtain

$$\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i) \mathbb{1}\{M_{\eta,n} \geq a_n\} \\ = \prod\nolimits_1 \bar{g}(\bar{X}_i, y_i) \times \prod\nolimits_2 \bar{g}(\bar{X}_i, y_i) \mathbb{1}\{M_{\eta,n} \geq a_n\} \leq \eta^{a_n} \prod_{i=0}^{n} \Upsilon^2_{\mathsf{X}}(y_i) ,$$

where $\prod_1$ is the product over the indices $i \in \{0,\ldots,n\}$ such that $\bar{X}_i \in \bar{\mathsf{C}}^c_\eta(y_i)$ and $\prod_2$ is the product on the remaining indices. This implies that

$$\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'} \left[\prod_{i=0}^{n} \bar{g}(\bar{X}_i, y_i) \mathbb{1}\{M_{\eta,n} \geq a_n\} \right] \leq \eta^{a_n} \prod_{i=0}^{n} \Upsilon^2_{\mathsf{X}}(y_i) .$$

By combining the above relations and Proposition 3, the result follows. $\square$

The last step consists in finding conditions upon which the bound in the right hand side of (25) is small. This bound depends explicitly on the observations Y's; it is therefore not difficult to state general conditions upon which this quantity is small. Let $\{Y_k\}_{k\geq 0}$ be a stochastic process with probability distribution $\mathbb{P}_\star$ in $(\mathsf{Y}, \mathcal{Y})$. We first formulate an almost sure convergence on the total variation distance of the filter initialized with two different probability measures ν and ν' and then later establish a convergence of the expectation. To complete this program, a last condition measurability condition is required:

H 3. For any $\eta > 0$, the functions $(y, y') \mapsto \Phi_{\nu,\mathsf{C}_\eta}(y, y')$ and $(y, y') \mapsto \Psi_{\mathsf{C}_\eta}(y, y')$ are measurable.

THEOREM 5. *Assume (H1)-(H2)-(H3). Assume moreover that for some η_0, there exists some LD-set function C_{η_0} such that*

$$\limsup_{n\to\infty}\left[-n^{-1}\sum_{k=2}^{n}\log\varepsilon^{-}_{\mathsf{C}_{\eta_0}}(Y_{k-1},Y_k)\right] < \infty\ , \qquad \mathbb{P}_\star - \text{a.s.} \tag{27}$$

$$\limsup_{n\to\infty}\left[n^{-1}\sum_{k=0}^{n}\log\Upsilon_{\mathsf{X}}(Y_k)\right] < \infty\ , \qquad \mathbb{P}_\star - \text{a.s.} \tag{28}$$

$$\limsup_{n\to\infty}\left[-n^{-1}\sum_{k=2}^{n}\log\Psi_{\mathsf{C}_{\eta_0}}(Y_{k-1},Y_k)\right] < \infty\ . \qquad \mathbb{P}_\star - \text{a.s.} \tag{29}$$

Assume in addition that there exists $\alpha \in (0,1)$ such that for all $\eta > 0$,

$$\limsup_{n\to\infty} n^{-1}\log\Lambda_\eta(Y_{0:n},\alpha) < 0, \qquad \mathbb{P}_\star - \text{a.s.} \tag{30}$$

Then, for any initial probability distributions ν and ν' on $(\mathsf{X},\mathcal{X})$ such that

$$\Phi_{\nu,\mathsf{C}_{\eta_0}}(Y_0,Y_1) < \infty \quad \textit{and} \quad \Phi_{\nu',\mathsf{C}_{\eta_0}}(Y_0,Y_1) < \infty\ , \quad \mathbb{P}_\star\textit{-p.s.}$$

we have

$$\limsup_{n\to\infty} n^{-1}\log\left\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\right\|_{\mathrm{TV}} < 0\ , \quad \mathbb{P}_\star - \text{a.s.}\ .$$

PROOF. We apply (25) with $\mathsf{C} = \mathsf{C}_{\eta_0}$. Note that for any positive sequences $\{u_n\}$ and $\{v_n\}$,

$$\limsup_{n\to\infty} n^{-1}\log(u_n+v_n) \le \sup\left(\limsup n^{-1}\log u_n, \limsup n^{-1}\log v_n\right)\ .$$

Under the stated assumptions, there exists some constant $0 < M < \infty$ such that for any $\eta > 0$,

$$\begin{aligned}&\limsup n^{-1}\log\left\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\right\|_{\mathrm{TV}} \\ &\qquad \le \sup\left(\limsup_{n\to\infty} n^{-1}\log\Lambda_\eta(Y_{0:n},\alpha), \frac{1-\alpha}{2}\log(\eta) + M\right)\ , \quad \mathbb{P}_\star - \text{a.s.}\end{aligned}$$

The proof is concluded by choosing η small enough so that $\log(\eta)(1-\alpha)/2 + M < 0$. $\square$

Compared to [7, Theorem 1] in the ergodic case , the conditions (27) and (30) are specific to the non-ergodic case, since they involve the functions $\varepsilon^-_{\mathsf{C}}$ and $\varepsilon^+_{\mathsf{C}}$. In the ergodic case, these functions are constant and assumptions (27) and (30) are trivially satisfied.

THEOREM 6. *Assume (H1)-(H2)-(H3). Then, for any $\eta_0 > 0$, $M_i > 0$, $i = 0, \ldots, 3$, $\delta > 0$ and $\alpha \in (0,1)$, there exist constants $\eta > 0$ and $\beta \in (0,1)$ such that, for all $n \in \mathbb{N}$,*

$$\mathbb{E}_\star\left[\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} \right] \leq 2 \left(\beta^n + r_0(\nu, n) + r_0(\nu', n) + \sum_{i=1}^4 r_i(n) \right) \tag{31}$$

where the sequences in the right-hand side of (31) *are defined by*

$$r_0(\nu, n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(-\log \Phi_{\nu, \mathsf{C}_{\eta_0}}(Y_0, Y_1) \geq M_0 n \right), \tag{32}$$

$$r_1(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(-\sum_{k=2}^n \log \varepsilon^-_{\mathsf{C}_{\eta_0}}(Y_{k-1}, Y_k) \geq M_1 n \right), \tag{33}$$

$$r_2(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(\sum_{k=0}^n \log \Upsilon_{\mathsf{X}}(Y_k) \geq M_2 n \right), \tag{34}$$

$$r_3(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(-\sum_{k=2}^n \log \Psi_{\mathsf{C}_{\eta_0}}(Y_{k-1}, Y_k) \geq M_3 n \right), \tag{35}$$

$$r_4(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(\log \Lambda_\eta(Y_{0:n}, \alpha) \geq -\delta n \right). \tag{36}$$

PROOF. For any $\alpha \in (0,1)$ and $\gamma \in (0,1)$, we can choose η small enough and such that for all $n \geq 0$, $\eta^{a_n} \mathrm{e}^{2n \sum_{i=0}^3 M_i} \leq \gamma^n$ where $a_n = n(1-\alpha)/2 - 1/2$. Denote by Ω_n the event

$$\begin{aligned}
\Omega_n = \Big\{ & -\log \Phi_{\nu, \mathsf{C}_{\eta_0}}(Y_0, Y_1) < M_0 n, \ -\log \Phi_{\nu', \mathsf{C}_{\eta_0}}(Y_0, Y_1) < M_0 n, \\
& -\sum_{i=2}^n \log \varepsilon^-_{\mathsf{C}_{\eta_0}}(Y_{i-1}, Y_i) < M_1 n, \sum_{i=0}^n \log \Upsilon_{\mathsf{X}}(Y_i) < M_2 n, \\
& -\sum_{i=2}^n \log \Psi_{\mathsf{C}_{\eta_0}}(Y_{i-1}, Y_i) < M_3 n, \ \log \Lambda_\eta(Y_{0:n}, \alpha) < -\delta n \Big\}.
\end{aligned}$$

Under the stated assumptions, $\mathbb{P}_\star(\Omega_n^{\mathrm{c}}) \leq r_0(\nu, n) + r_0(\nu', n) + \sum_{i=1}^4 r_i(n)$.

On the event Ω_n, we have

$$\Phi^{-1}_{\nu,\mathsf{C}_{\eta_0}}(Y_0,Y_1)\Phi^{-1}_{\nu',\mathsf{C}_{\eta_0}}(Y_0,Y_1)\prod_{i=2}^{n}\left[\varepsilon^{-}_{\mathsf{C}_{\eta_0}}(Y_{i-1},Y_i)\Psi_{\mathsf{C}_{\eta_0}}(Y_{i-1},Y_i)\right]^{-2}\prod_{i=0}^{n}\Upsilon^2_{\mathsf{X}}(Y_i) \\ \leq \mathrm{e}^{2n\sum_{i=0}^{3}M_i} .$$

Then, by Proposition 4, on the event Ω_n, we have $\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} \leq 2\beta^n$ where $\beta=\max(\gamma,e^{-\delta})$. Since

$$\mathbb{E}_\star\Big[\,\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\,\Big] \\ \leq \mathbb{E}_\star\Big[\,\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\,\mathbb{1}_{\Omega_n}\Big]+2\mathbb{P}_\star(\Omega_n^{\mathrm{c}}) ,$$

the result follows. □

Theorem 6 does not provide directly a rate of convergence. Indeed, only the first term of the right-hand side of equation (31) gives a geometric rate. In Section 3, for specific models, explicit bounds of the other terms will be obtained with geometric rates.

3. Nonlinear state-space models.

Let $\mathsf{X}=\mathbb{R}^{d_\mathsf{X}}$, $\mathsf{Y}=\mathbb{R}^{d_\mathsf{Y}}$, $\mathsf{Z}=\mathbb{R}^{\ell}$ with $d_\mathsf{Y}\leq d_\mathsf{X}$, endowed with the Borel σ-algebra $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. We consider the non-linear state-space model:

$$\begin{cases} X_k = f_*(X_{k-1})+\tau^*(X_{k-1},\zeta_k) , \quad X_0\sim\nu_0 \\ Y_k = h_*(X_k)+\varepsilon_k , \end{cases} \tag{37}$$

where $f_*:\mathsf{X}\to\mathsf{X}$, $h_*:\mathsf{X}\to\mathsf{Y}$ $\tau^*:\mathsf{X}\times\mathsf{Z}\to\mathsf{X}$ are some functions, and ν_0 the initial distribution of X_0. In the sequel, it is assumed that $\{\zeta_k\}_{k\geq0}$ and $\{\varepsilon_k\}_{k\geq0}$ are independent and $\{\varepsilon_k\}_{k\geq0}$ is an i.i.d sequence.

For $a>0$, define the set Lip_a of a-Lipshitz functions, *i.e.* $g\in\mathrm{Lip}_a$ if for all $x,x'\in\mathsf{X}^2$, $|g(x)-g(x')|\leq a|x-x'|$. For $b_0,b>0$, consider the set $\mathrm{S}_{b_0,b}$ the set of functions $g:\mathsf{X}\to\mathsf{Y}$ which are surjective and which satisfy, for all $x,x'\in\mathsf{X}^2$, $|x-x'|\leq b_0+b|g(x)-g(x')|$.

For a function $h:\mathsf{X}\to\mathsf{Y}$, a transition density kernel $t:\mathsf{X}\times\mathsf{X}\to\mathbb{R}$ and a probability density υ on Y, we compute the filtering distribution (12) $\phi_{\nu,n}[Y_{0:n}]$ using

$$Q(x,A)=\int_A t\left(x,x'-f(x)\right)\mathrm{Leb}(\mathrm{d}x') , \quad x\in\mathsf{X} , A\in\mathcal{X} , \tag{38}$$

$$g(x,y)=\upsilon\left(y-h(x)\right) . \tag{39}$$

Consider the following assumptions:

M 1. There exist constants $a, b_0, b > 0$ such that f is Lip_a and $h \in \mathrm{S}_{b_0,b}$.

M 2. The density υ is positive, continuous, and $\lim_{|u|\to\infty} \upsilon(u) = 0$.

Notice that f^* and f are not necessarily contracting so that the both the observations and the filtering model used to construct $\phi_{\nu,n}[Y_{0:n}]$ are (possibly) non-ergodic. The assumption (M1) has been first considered in [15]. A function h satisfying (M1) can be viewed as a perturbation of a bijective function whose inverse is b-Lipschtiz. The rationale for considering such assumption is the following. For $y_1,\ y_2 \in \mathsf{Y}$, the maximal distance between any two elements in the preimages $h^{-1}(\{y_1\})$ and $h^{-1}(\{y_2\})$ is controlled by $|y_1 - y_2|$. The assumption (M2) is satisfied, for example, by Gaussian densities.

3.1. *Nonlinear state-space model with i.i.d. state noise.* In this Section, we assume that the observations and the filtering model are matched, *i.e.* $f^* = f$ and $h^* = h$. In addition, we consider the following assumption:

E 1. The state noise $\{\zeta_k\}_{k\geq 0}$ is a sequence of i.i.d. random variables with a positive and continuous density γ with respect to the Lebesgue measure Leb on X. In addition, the transition kernel t in (38) is taken to be $t(x, x') = \gamma(x')$, for all $(x, x') \in \mathsf{X}^2$.

E 2. The function $\tau^*(x, \zeta)$ in (37) is given by: $\tau^*(x, \zeta) = \zeta$.

Under this assumption, for any $A \in \mathcal{X}$, the transition kernel Q may be expressed as

$$Q(x, A) = \int_A \gamma[x' - f(x)] \, \mathrm{Leb}(dx') \; . \tag{40}$$

Define by $D : \mathsf{Y} \times \mathsf{Y} \to \mathbb{R}$ the function

$$D(y, y') \stackrel{\mathrm{def}}{=} \sup \left\{ |f(z) - z'| \; : z \in h^{-1}(\{y\}), z' \in h^{-1}(\{y'\}) \right\} \; . \tag{41}$$

For any $r > 0$, we consider the minimum and the maximum of the state noise density over a ball of radius r:

$$\gamma^-(r) \stackrel{\mathrm{def}}{=} \inf_{|s|\leq r} \gamma(s) \; , \qquad \gamma^+(r) \stackrel{\mathrm{def}}{=} \sup_{|s|\leq r} \gamma(s) \; , \tag{42}$$

LEMMA 7. *. Assume (M1)-(M2)-(E1). Then, for any $\Delta \in (0, \infty)$, the set valued function $\mathsf{C}_\Delta : \mathsf{Y} \to \mathcal{X}$, defined by*

$$y \longmapsto \mathsf{C}_\Delta(y) \stackrel{\mathrm{def}}{=} \{x \in \mathsf{X} : |h(x) - y| \leq \Delta\} \; . \tag{43}$$

is a LD-set function: for all $A \in \mathcal{X}$ and $x \in \mathsf{C}_\Delta(y)$; more precisely,

$$\varepsilon^-_\Delta(y,y')\mathrm{Leb}[A \cap \mathsf{C}_\Delta(y')] \le Q[x, A \cap \mathsf{C}_\Delta(y')] \le \varepsilon^+_\Delta(y,y')\mathrm{Leb}[A \cap \mathsf{C}_\Delta(y')] \;, \tag{44}$$

where, setting $c = (a+1)b_0$ and $d = (a+1)b$,

$$\begin{aligned} \varepsilon^-_\Delta(y,y') &\stackrel{\text{def}}{=} \gamma^-[c + d\Delta + D(y,y')] \;, \\ \varepsilon^+_\Delta(y,y') &\stackrel{\text{def}}{=} \gamma^+[c + d\Delta + D(y,y')] \;. \end{aligned} \tag{45}$$

Let Z^Δ_k be defined as:

$$Z^\Delta_k \stackrel{\text{def}}{=} -\log \gamma^- \left[2c + d\Delta + ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|\right] \;. \tag{46}$$

PROPOSITION 8. *Assume (M1), (M2), (E1) and (E2). Assume in addition that for all $\Delta > 0$,*

$$\mathbb{E}|Z^\Delta_1| < \infty \;. \tag{47}$$

Then, for any probability distributions ν_0, ν, and ν' on $(\mathsf{X}, \mathcal{X})$, we have

$$\limsup_{n\to\infty} n^{-1} \log \left\| \phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}] \right\|_{\mathrm{TV}} < 0, \quad \mathbb{P}_{\nu_0} - \text{a.s.}$$

The condition (47) is not very restrictive. For example, assume that $\{\zeta_k\}_{k\ge0}$ and $\{\varepsilon_k\}_{k\ge0}$ are sequences of Gaussian random variables. It follows, that $\gamma^-(r) = \gamma(r)$ for all $r \ge 0$. The condition (47) holds if $\mathbb{E}(|\varepsilon_1|^2) < \infty$ and $\mathbb{E}(|\zeta_1|^2) < \infty$ which are trivially satisfied. This result also extends [5, Theorem 1.1]; these authors assume that the densities υ and γ are upper and lower bounded, *i.e.* that there exist positive constants m_υ, M_υ, α_υ, β_υ, and m_γ, M_γ, α_γ and β_γ such that

$$m_\upsilon \exp(-\alpha_\upsilon |x|^{\beta_\upsilon}) \le \upsilon(x) \le M_\upsilon \exp(-\alpha_\upsilon |x|^{\beta_\upsilon}) \;, \tag{48}$$

$$m_\gamma \exp(-\alpha_\gamma |x|^{\beta_\gamma}) \le \gamma(x) \le M_\gamma \exp(-\alpha_\gamma |x|^{\beta_\gamma}) \tag{49}$$

together with (M1) and a condition which is slightly more restrictive than (M2). Under these assumptions, we may set $\gamma_-(r) = m_\upsilon \exp(-\alpha_\upsilon r^{\beta_\upsilon})$ for $r > 0$ and the condition (47) simply reads,

$$\mathbb{E}\left[|\epsilon_1|^{\beta_\upsilon}\right] < \infty \quad \text{and} \quad \mathbb{E}\left[|\zeta_1|^{\beta_\upsilon}\right] < \infty \;,$$

which is of course satisfied under (48) and (49), without any conditions on the constants α_υ, β_υ, α_γ, and β_γ. The stability of the filter holds without requiring that the tails of the observation noise be *light* compared to the tails of the signal noise (this type of conditions is prevalent in many works on this topic since [3]).

With more stringent conditions on initial distributions, the convergence of the expected value of the total variation distance $\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}$ may be shown to be geometric. Define the log-moment generating function $\psi_Z(\alpha)$ of the random variable Z is defined by $\psi_Z(\alpha) \stackrel{\mathrm{def}}{=} \log \mathbb{E}[\mathrm{e}^{\alpha Z}]$.

PROPOSITION 9. *Assume that (M1), (M2), (E1), and (E2) hold. Then, for all $\Delta > 0$, there exists $\alpha_0 > 0$ such that*

$$\psi_{Z_1^\Delta}(\alpha_0) > 0 \; . \tag{50}$$

Let C be the LD-set function defined by (43). Then, for any ν_0, ν and ν' probability measures on $(\mathsf{X}, \mathcal{X})$ and $\Delta > 0$ such that,

$$\mathbb{E}_{\nu_0}\left\{\exp\left(\alpha_0 \left[\log \Phi_{\nu,\mathsf{C}}(Y_0, Y_1)\right]_-\right)\right\} < \infty \; , \tag{51}$$
$$\mathbb{E}_{\nu_0}\left\{\exp\left(\alpha_0 \left[\log \Phi_{\nu',\mathsf{C}}(Y_0, Y_1)\right]_-\right)\right\} < \infty \; ,$$

we have

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{E}_{\nu_0}\left[\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right] < 0 \; .$$

Assume that $\{\zeta_k\}_{k\geq 0}$ and $\{\varepsilon_k\}_{k\geq 0}$ are sequences of Gaussian random variables. The condition $\mathbb{E}[\mathrm{e}^{\alpha_0 \, Z_1^\Delta}] < \infty$ is equivalent to

$$\int_{\mathbb{R}^{d_{\mathsf{X}}+2d_{\mathsf{Y}}}} \exp\left[(\alpha_0 - \varsigma)|u|^2\right] \, \mathrm{d}u < \infty \; ,$$

where ς denotes some positive constant. Therefore, for $\alpha_0 > 0$ small enough, the condition (50) is satisfied. Proofs of Lemma 7, Propositions 8 and 9 are postponed to Section 4.

3.2. *Nonlinear state-space model with dependent state noise.* We now consider the case where the state noise $\{\zeta_k\}_{k\geq 0}$ can depend on previous states. This model has been introduced in [15, Section 3] and is important because it covers the case of partially observed discretely sampled diffusions, as well as partially observed stochastic volatility models [3, Section 2].

G 1. There exist ψ a positive continuous probability density function and constants $\mu_- > 0$ and $\mu_+ > 0$ such that, and for all $x, x' \in \mathsf{X}^2$,

$$\mu_- \psi(x') \leq t(x, x') \leq \mu_+ \psi(x') \; .$$

where t is the density transition kernel defined in (38).

This example also illustrates that the forgetting property is kept even when the distributions of the observations differ from the model.

G 2. 1. For some positive constants a^*, b_0^* and b^*, $f^* \in \mathrm{Lip}_{a^*}$ and $h^* \in \mathrm{S}_{b_0^*, b^*}$, where f^* and h^* are defined in (37).
2. There exists a function $\tau_+ : \mathsf{Z} \to \mathbb{R}$ such that for all $x \in \mathsf{X}$ and $\zeta \in \mathsf{Z}$, $|\tau^*(x, \zeta)| \le \tau_+(\zeta)$.
3. f^* and h^* are such that $\|f - f^*\|_\infty < \infty$ and $\|h - h^*\|_\infty < \infty$.

A first example of state equation satisfying (G1) is considered in [3]. A signal takes its values in X and follows the equation

$$X_k = f(X_{k-1}) + \sigma(X_{k-1})\xi_k \ , \tag{52}$$

where $\{\xi_k\}_{k\ge 0}$ is a sequence of i.i.d random variables and where $\sigma : \mathsf{X} \to \mathbb{R}^{d_\mathsf{X} \times d_\mathsf{X}}$ is a measurable function that satisfies, for all $x,\ u \in \mathsf{X}$, the following uniform ellipticity condition:

$$\sigma^- |u|^2 \le |\sigma^*(x)u|^2 \le \sigma^+ |u|^2 \ , \tag{53}$$

where σ^-, σ^+ are positive constants. Another important example where assumption (G1) is satisfied is the case of certain discretely sampled diffusions. Let $(X_t)_{t\ge 0}$ be the unique solution of the following stochastic differential equation

$$\mathrm{d}X_t = \rho(X_t)dt + \sigma(X_t)\mathrm{d}B_t \ ,$$

where B is the d_X-dimensional Brownian motion and the functions $\rho : \mathbb{R}^{d_\mathsf{X}} \to \mathbb{R}^{d_\mathsf{X}}$ and $\sigma : \mathbb{R}^{d_\mathsf{X}} \to \mathbb{R}^{d_\mathsf{X} \times d_\mathsf{X}}$ are respectively of class C^1 and C^3 . Then, the sequence $\{X_k\}_{k\ge 0}$ satisfies assumption (G2) if the function σ is uniformly elliptic (condition (53)); The assumptions (M1), (M2) and (G2) are similar to those made in [3] and [15]. This allows to establish the forgetting of the initial condition with probability one without restriction on the signal-to-noise ratio and for sequences of observations which are not necessarily distributed according to the model used to compute the filtering distribution.

For the same reasons as above, we consider the set-valued function C_Δ defined in (43). Denote

$$q^-(r) \stackrel{\text{def}}{=} \mu_- \times \inf_{|v|\le r} \psi(v) \ , \qquad q^+(r) \stackrel{\text{def}}{=} \mu_+ \times \sup_{|v|\le r} \psi(v) \ , \tag{54}$$

LEMMA 10. *Assume (M1)-(M2)-(G1). Then, for all $A \in \mathcal{X}$, $y \in \mathsf{Y}$, and $x \in \mathsf{C}_\Delta(y)$,*

$$\begin{aligned} \varepsilon_\Delta^-(y, y')\mathrm{Leb}[A \cap \mathsf{C}_\Delta(y')] \le Q[x, A \cap \mathsf{C}_\Delta(y')] \\ \le \varepsilon_\Delta^+(y, y')\mathrm{Leb}[A \cap \mathsf{C}_\Delta(y')] \ , \end{aligned} \tag{55}$$

where

$$\varepsilon_\Delta^-(y,y') \stackrel{\text{def}}{=} q^-[c+d\Delta+D(y,y')] \;, \qquad \varepsilon_\Delta^+(y,y') \stackrel{\text{def}}{=} q^+[c+d\Delta+D(y,y')] \;.$$

The proof is similar to the proof of Lemma 7 and is omitted for brevity.

LEMMA 11. *Under (G1)-(G2), for all integer $k \geq 1$,*

$$D(Y_{k-1},Y_k) \leq \kappa + (a^*+1)b_0^* + a^*b^*|\varepsilon_{k-1}| + b^*|\varepsilon_k| + \tau_+(\zeta_k) \;,$$

where

$$\kappa \stackrel{\text{def}}{=} \|f-f^*\|_\infty + (b_0 + b\,\|h^*-h\|_\infty)(1+a^*)$$

Now, define for all $\Delta > 0$

(56)

$$V_+^{*\Delta} \stackrel{\text{def}}{=} -\log q^-\left[c+d\Delta+\kappa+(1+a^*)b_0^*+a^*b^*|\varepsilon_0|+b^*|\varepsilon_1|+\tau_+(\zeta_0)\right] \;.$$

PROPOSITION 12. *Assume (M1), (M2),(G1), and (G2). Assume in addition that for all $\Delta > 0$,*

$$\mathbb{E}\left(|V_+^{*\Delta}|\right) < \infty \;. \tag{57}$$

Then, for any initial probability distributions ν and ν' on $(\mathsf{X},\mathcal{X})$, we have

$$\limsup_{n\to\infty} n^{-1}\log\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} < 0, \qquad \mathbb{P}_\star-\text{a.s.} \;,$$

where $\mathbb{P}_\star$ is the distribution of the process specified by (37)*.*

Let $\{Z_k^{*\Delta}\}_{k\geq 0}$ be the sequence defined for all $\Delta > 0$ and for all integer $k \geq 1$ by

$$Z_k^{*\Delta} \stackrel{\text{def}}{=} -\log q^-\left[c+d\Delta+\kappa+(a^*+1)b_0^*+a^*b^*|\varepsilon_{k-1}|+b^*|\varepsilon_k|+\tau_+^*(A_k^*)\right] \;.$$

PROPOSITION 13. *Assume (M1), (M2), (G1), and (G2). Assume in addition that for all $\Delta > 0$, there exists a neighborhood $\alpha_0 > 0$ such that $\Psi_{Z_1^{*\Delta}}(\alpha_0) < \infty$. Let C_Δ be the LD-set function defined by* (43)*. Then, for ν and ν' two probability measures on $(\mathsf{X},\mathcal{X})$ and $\Delta > 0$ such that,*

$$\mathbb{E}_\star\left\{\exp\left(\alpha_0[\log\Phi_{\nu,\mathsf{C}_\Delta}(Y_0,Y_1)]_-\right)\right\} < \infty \;, \qquad \mathbb{E}_\star\left\{\exp\left(\alpha_0[\log\Phi_{\nu',\mathsf{C}_\Delta}(Y_0,Y_1)]_-\right)\right\} < \infty \;, \tag{58}$$

we have

$$\limsup_{n\to\infty} n^{-1}\log\mathbb{E}_\star\left[\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right] < 0 \;.$$

Proofs of Propositions 12 and 13 are given in Section 5.

4. Proofs of Propositions 8 and 9 .

PROOF OF LEMMA 7. Under assumption (M1), for any z in $h^{-1}(\{y\})$, and $x \in \mathsf{C}_\Delta(y)$,

$$|x - z| \leq b_0 + b\Delta \; . \tag{59}$$

Let $(y, y') \in \mathsf{Y}^2$. By (M1), h is surjective and we may pick $z \in h^{-1}(\{y\})$ and $z' \in h^{-1}(\{y'\})$. Using again (M1), it follows from (59) that, for all $(x, x') \in \mathsf{C}_\Delta(y) \times \mathsf{C}_\Delta(y')$,

$$\begin{aligned} |f(x) - x'| &\leq |f(x) - f(z)| + |f(z) - z'| + |z' - x'| \\ &\leq a(b_0 + b\Delta) + D(y, y') + b_0 + b\Delta \; , \end{aligned} \tag{60}$$

The proof follows from (40) and (60). □

PROOF OF PROPOSITION 8. We will apply Theorem 5 by successively checking the assumptions (H1–3) and (27–30).

(H1) is satisfied. By (M2), for all $\eta > 0$, we may choose Δ_η large enough so that $\sup_{|s| > \Delta_\eta} \upsilon(s) \leq \eta \sup_{s \in \mathsf{X}} \upsilon(s)$. This, combined with (44), implies (H2) with $\mathsf{C}_\eta = \mathsf{C}_{\Delta_\eta}$ and $\Upsilon_\mathsf{X} = \sup_\mathsf{X} \upsilon$. (H3) is obvious.

To check (27–30), it will be needed to bound $\{D(Y_{k-1}, Y_k)\}_{k \geq 1}$ where D is defined in (41). For $z,\ z' \in \mathsf{X}$ such that $h(z) = Y_{k-1},\ h(z') = Y_k$, it follows from (M1) that

$$\begin{aligned} |f(z) - z'| &\leq |f(z) - f(X_{k-1})| + |f(X_{k-1}) - X_k| + |X_k - z'| \; , \\ &\leq a(b_0 + b|\varepsilon_{k-1}|) + |\zeta_k| + b_0 + b|\varepsilon_k| \; . \end{aligned}$$

Therefore, for all integer $k \geq 1$,

$$D(Y_{k-1}, Y_k) \leq c + ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k| \; . \tag{61}$$

We now consider conditions (27–29). Let η_0 be fixed and set $\Delta = \Delta_{\eta_0}$. Since by definition (42), γ^- is a nonincreasing function, it follows by plugging the bound (61) into (45) that

$$-n^{-1} \sum_{k=2}^{n} \log \varepsilon_\Delta^-(Y_{k-1}, Y_k) \leq n^{-1} \sum_{k=2}^{n} Z_k^\Delta \; , \tag{62}$$

where Z_k^Δ is defined in (46). Since the process $\{ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|\}_{k \geq 1}$ is stationary 2-dependent, the strong law of large numbers for m-dependent sequences and the integrability condition (47) yield

$$\lim_{n \to \infty} n^{-1} \sum_{k=2}^{n} Z_k^\Delta = \mathbb{E}(Z_1^\Delta) < \infty \; , \qquad \mathbb{P}_{\nu_0} - \text{a.s.} \tag{63}$$

By combining (62) and (63), the first condition (27) of Theorem 5 is satisfied. By assumption (M2), the density υ is bounded which implies that $\sup_{y \in \mathsf{Y}} \Upsilon_{\mathsf{X}}(y) \leq \sup \upsilon$. Hence, the second condition (28) of Theorem 5 is satisfied. We now consider the third condition (29). Since the measure appearing in the definition of the LD-set function does not depend on $y,\ y'$, the function $(y, y') \mapsto \Psi_{\mathsf{C}_\Delta(y')}(y, y')$, defined in (19), does not depend on y and is given by

$$\Psi_{\mathsf{C}_\Delta(y')}(y, y') = \int_{\mathsf{C}_\Delta(y')} \upsilon[y' - h(x)] \,\mathrm{Leb}(dx) \geq \mathrm{Leb}[\mathsf{C}_\Delta(y')] \times \inf_{|s| \leq \Delta} \upsilon(s) \ .$$

Since the function h is uniformly continuous, for any fixed $\Delta > 0$, there exists $\delta > 0$ such that, for all $x,\ x' \in \mathsf{X}$ satisfying $|x - x'| \leq \delta$, we have $|h(x) - h(x')| \leq \Delta$. Since h is surjective, it follows that $\mathrm{Leb}[\mathsf{C}_\Delta(y')]$ is bounded below by the volume of a ball of radius δ in $\mathbb{R}^{d_{\mathsf{X}}}$. Thus, we have, for all $y,\ y' \in \mathsf{Y}$,

$$\Psi_{\mathsf{C}_\Delta(y')}(y, y') \geq \varrho_\Delta \ , \tag{64}$$

for some $\varrho_\Delta > 0$, depending only on Δ. The third condition (29) of Theorem 5 follows.

We now prove (30). We have

$$\log \Lambda_\eta(Y_{0:n}, \alpha) \leq \max \left\{ \sum_{k=1}^n \delta_k U_k : \{\delta_k\}_{k=1}^n \in \{0,1\}^n, \sum_{k=1}^n \delta_k \geq \alpha n \right\} \ .$$

where

$$\begin{aligned} R_{\Delta_\eta}(x) &\stackrel{\mathrm{def}}{=} \log \left\{ 1 - (\gamma^- / \gamma^+)^2 \left[2c + d\Delta_\eta + x \right] \right\} \ , \\ U_k &\stackrel{\mathrm{def}}{=} R_{\Delta_\eta}(ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|) \ . \end{aligned} \tag{65}$$

Then, since for any $\upsilon > 0$, $U_k \leq -\upsilon \mathbb{1}\{U_k < -\upsilon\}$, we have

$$\begin{aligned} \log \Lambda_\eta(Y_{0:n}, \alpha) &\leq \max_{\left\{\{\delta_k\}_{k=1}^n \in \{0,1\}^n : \sum_{k=1}^n \delta_k \geq \alpha n \right\}} -\upsilon \sum_{k=1}^n \delta_k \mathbb{1}\{U_k < -\upsilon\} \\ &\leq -\upsilon \lfloor (\alpha n - \sum_{k=1}^n \mathbb{1}\{U_k \geq -\upsilon\})^+ \rfloor \ . \end{aligned} \tag{66}$$

Dividing by n and letting n goes to infinity, the strong LLN for 2-dependent sequences yields that $\mathbb{P}_{\nu_0}$ − a.s.,

$$\limsup_{n \to \infty} n^{-1} \log \Lambda_\eta(Y_{0:n}, \alpha) \leq -\upsilon(\alpha - \mathbb{P}(U_1 > -\upsilon))) \ . \tag{67}$$

Note that U_1 is non positive and $\mathbb{P}(U_1 = 0) = \mathbb{P}(|\varepsilon_{k-1}| + |\zeta_k| + |\varepsilon_k| = \infty) = 0$ by the integrability condition (47). Hence, U_1 is almost surely negative and $\lim_{\upsilon \to 0} \mathbb{P}(U_1 > -\upsilon) = 0$; we may thus choose υ small enough so that $\alpha - \mathbb{P}(U_1 > -\upsilon) > 0$. The rhs is then negative by taking υ sufficiently small. □

Proof of Proposition 9. (51) implies that $r_0(\nu, n) \vee r_0(\nu', n) \leq c_0 \mathrm{e}^{-\delta_0 n}$ for some c_0, $\delta_0 > 0$. Now, recall that ψ_Z denotes the log-moment generating function of the random variable Z defined by $\psi_Z(\lambda) \stackrel{\text{def}}{=} \log \mathbb{E}[\mathrm{e}^{\lambda Z}]$ and we define its Legendre's transformation by

$$\psi_Z^*(x) = \sup_{\lambda \geq 0} \{x\lambda - \psi_Z(\lambda)\} \ .$$

We start by giving an exponential inequality for m-dependent variables whose proof is elementary.

Lemma 14. *Let $\{Z_k\}_{k \geq 0}$ be a sequence of m-dependent stationary random variables. Then, for all $M \geq 0$,*

$$\mathbb{P}\left(\sum_{k=1}^{n} Z_k \geq Mn\right) \leq m \exp[-\lfloor n/m \rfloor \psi_{Z_1}^*(M)] \ .$$

It follows by equation (62) that

$$\mathbb{P}\left(-n^{-1} \sum_{k=2}^{n} \log \varepsilon_{\Delta}^{-}(Y_{k-1}, Y_k) \geq M_1 n\right) \leq \mathbb{P}\left(\sum_{k=2}^{n} Z_k^{\Delta} \geq M_1 n\right) \ .$$

Thanks to (50), by applying Lemma 14, there exist some constant c_1, $\delta_1 > 0$ such that $r_1(n) \leq c_1 \mathrm{e}^{-\delta_1 n}$. Since υ is bounded, we can choose M_2 large enough such that $r_2(n) = 0$. By (64), for all $(y, y') \in \mathsf{Y}^2$, $\Psi_{\mathsf{C}_\Delta(y')}(y, y') \geq \varrho_\Delta$, for some $\varrho_\Delta > 0$. Then, by choosing M_3 large enough, we have $r_3(n) = 0$. Using (66), $r_4(n)$ is bounded by

$$r_4(n) \leq \mathbb{P}\left(-\upsilon \left\lfloor \left(\sum_{k=1}^{n} (\alpha - \mathbb{1}\{U_k \geq -\upsilon\})\right)^+ \right\rfloor \geq -\delta n\right)$$

Choosing υ such that $\alpha - \mathbb{P}(U_k > -\upsilon) > 0$ and then δ such that $\upsilon(\alpha - \mathbb{P}(U_k > -\upsilon)) > \delta$ and applying Lemma 14 with $Z_k = \alpha - \mathbb{1}\{U_k \geq -\upsilon\}$ which is bounded provides the existence of constants c_4, $\delta_4 > 0$ such that $r_4(n) \leq c_4 \mathrm{e}^{-\delta_4 n}$. Thus, Theorem 6 applies and provides a geometric rate. □

5. Proofs of Propositions 12 and 13 .

Proof of Lemma 11. For all integer $k \geq 1$, for $z,\ z' \in \mathsf{X}$ such that $h(z) = Y_{k-1},\ h(z') = Y_k$ and for $u,\ u' \in \mathsf{X}$ such that $h^*(u) = Y_{k-1},\ h^*(u') = Y_k$, we have

$$
\begin{aligned}
|f(z) - z'| &\leq |f(z) - f^*(z)| + |f^*(z) - f^*(u)| + |f^*(u) - u'| + |u' - z'| \;, \\
&\leq \|f - f^*\|_\infty + a^*|z - u| + |f^*(u) - u'| + |u' - z'| \;. \qquad (68)
\end{aligned}
$$

Let us notice that

$$|z - u| \leq b_0 + b|h(z) - h(u)| \leq b_0 + b\underbrace{|h(z) - h^*(u)|}_{=0} + b|h^*(u) - h(u)| \;.$$

Then, by denoting $K = b_0 + b\,\|h^* - h\|_\infty$, it follows that $|z - u| \leq K$ and similarly, $|z' - u'| \leq K$. Combining these two upper bounds with (68) leads to

$$
\begin{aligned}
&|f(z) - z'| \leq \kappa + |f^*(u) - f^*(X_{k-1})| + |f^*(X_{k-1}) - X_k| + |X_k - u'| \;, \\
&\leq \kappa + a^*[b_0^* + b^*|h^*(z) - h^*(X_{k-1})|] + \tau_+(\zeta_k) + b_0^* + b^*|h^*(X_k) - h^*(u')| \;,
\end{aligned}
$$

where $\kappa = \|f - f^*\|_\infty + K(1 + a^*)$. Thus, it is proven that, for all integer $k \geq 1$,

$$D(Y_{k-1}, Y_k) \leq \kappa + (a^* + 1)b_0^* + a^*b^*|\varepsilon_{k-1}| + b^*|\varepsilon_k| + \tau_+(\zeta_k) \;.$$

□

Proof of Proposition 12. Define, for all $\Delta > 0$ and for all integer $k \geq 1$,

(69)

$$V_k^{*\Delta} = -\log q^- \left[c + d\Delta + \kappa + (a^* + 1)b_0^* + a^*b^*|\varepsilon_{k-1}| + b^*|\varepsilon_k| + \tau_+(\zeta_k)\right] \;.$$

Using the definitions (54), (55) of q^- and ε_Δ^-, Lemma 11 shows that

$$-n^{-1}\sum_{k=2}^{n} \log \varepsilon_\Delta^-(Y_{k-1}, Y_k) \leq n^{-1}\sum_{k=2}^{n} V_k^{*\Delta} \;. \qquad (70)$$

Thus, (27) follows from LLN for 2-dependent sequences.

The proof of assumptions (28) and (29) can be checked as in Proposition 8. It remains to check (30). Let

$$
\begin{aligned}
U_n(\upsilon) &\stackrel{\text{def}}{=} n^{-1}\sum_{k=1}^{n} \mathbb{1}\{G(\varepsilon_{k-1}, \varepsilon_k, \zeta_k) \geq -\upsilon\} \;, \\
V_n(\upsilon) &\stackrel{\text{def}}{=} n^{-1}\sum_{k=1}^{n} \mathbb{P}\left\{G(\varepsilon_{k-1}, \varepsilon_k, \zeta_k) \geq -\upsilon | \mathcal{F}_{k-1}\right\} \;,
\end{aligned}
$$

where $G(\varepsilon_{k-1}, \varepsilon_k, \zeta_k) \stackrel{\text{def}}{=} R_{\Delta_\eta}[\kappa + (a^* + 1)b_0^* + a^* b^* |\varepsilon_{k-1}| + b^* |\varepsilon_k| + \tau_+(\zeta_k)]$ and R_{Δ_η} is defined in (65). Similarly to (66), we have for any $\upsilon > 0$,

$$\log \Lambda_\eta(Y_{0:n}, \alpha) \leq -\upsilon \lfloor (\alpha n - nU_n(\upsilon)\})^+ \rfloor \ .$$

so that $\mathbb{P}_\star$-a.s.

$$\limsup_n \log \Lambda_\eta(Y_{0:n}, \alpha) \leq -\upsilon \lfloor (\alpha - \limsup_n U_n(\upsilon)\})^+ \rfloor \ . \tag{71}$$

Moreover, using the LLN for 2-dependent sequences, we have that $\mathbb{P}_\star$-a.s.

$$\lim_{n\to\infty} U_n(\upsilon) = \lim_{n\to\infty} V_n(\upsilon) \leq \mu_+^* \mathbb{E}\left[\int \mathbb{1}\{G(\varepsilon_0, \varepsilon_1, w) \geq -\upsilon\} \psi^*(w)\, dw \right] \ .$$

Since G is $\mathbb{P}_\star$-a.s. negative, the right-hand side of the above equation thus converges $\mathbb{P}_\star$-a.s. to 0 as υ tends to 0. Thus, the right-hand side of (71) is negative by choosing υ sufficiently small. □

PROOF OF PROPOSITION 13. (58) implies that $r_0(\nu, n) \vee r_0(\nu', n) \leq c_0 \mathrm{e}^{-\delta_0 n}$ for some c_0, $\delta_0 > 0$. It follows, by definition of r_1 and Lemma 11 that

$$r_1(n) = \mathbb{P}_\star \left(-n^{-1} \sum_{k=2}^{n} \log q^- [c + d\Delta + D(Y_{k-1}, Y_k)] \geq M_1 n \right) \leq \mathbb{P}_\star \left(n^{-1} \sum_{k=2}^{n} Z_k^{*\Delta} \geq M_1 n \right) \ .$$

with $c_0 = c + d\Delta + \kappa + (a^* + 1)b_0^*$. Then, applying Lemma 14, there exist some constants c_1, $\delta_1 > 0$ such that $r_1(n) \leq c_1 \mathrm{e}^{-\delta_1 n}$. By the same arguments as in proof of Proposition 9, the real numbers M_2 and M_3 can be chosen large enough such that $r_2(n) = 0$ and $r_3(n) = 0$. Let us denote by $\{U_k^+\}_{k \geq 0}$ the sequence defined by $U_k^+ \stackrel{\text{def}}{=} R_\Delta[\kappa + (a^* + 1)b_0^* + a^* b^* |\varepsilon_{k-1}| + b^* |\varepsilon_k| + \tau_+(\zeta_k)]$, for all integer $k \geq 1$. Similarly to the proof of Proposition 9, for any $\delta > 0$,

$$r_4(n) \leq \mathbb{P}\left(-\upsilon \left\lfloor \left(\sum_{k=1}^{n} (\alpha - \mathbb{1}\{U_k^+ \geq -\upsilon\}) \right)^+ \right\rfloor \geq -\delta n \right) \ .$$

We first choose υ small enough so that $\alpha - \mathbb{P}(U_k^+ > -\upsilon) > 0$ holds; then δ is chosen such that $\upsilon(\alpha - \mathbb{P}(U_k^+ > -\upsilon)) > \delta$. By applying Lemma 14 with $Z_k = \alpha - \mathbb{1}\{U_k^+ \geq -\upsilon\}$ which is a bounded random variable, there exist constants c_4, $\delta_4 > 0$ such that $r_4(n) \leq c_4 \mathrm{e}^{-\delta_4 n}$. Thus, Theorem 6 applies and provides a geometric rate. □

References.

[1] Atar, R. and Zeitouni, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):697–725.

[2] Budhiraja, A. and Ocone, D. (1997). Exponential stability of discrete-time filters for bounded observation noise. *Systems Control Lett.*, 30:185–193.

[3] Budhiraja, A. and Ocone, D. (1999). Exponential stability in discrete-time filtering for non-ergodic signals. *Stochastic Process. Appl.*, 82(2):245–257.

[4] Chigansky, P., Liptser, R., and van Handel., R. (2008). Intrinsic methods in filter stability. to appear in Handbook of Nonlinear Filtering, Oxford University Press.

[5] Crisan, D. and Heine, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *J. Lond. Math. Soc. (2)*, 78(2):441–458.

[6] Del Moral, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications.* Springer.

[7] Douc, R., Fort, G., Moulines, E., and Priouret, P. (2008). Forgetting the initial distribution for hidden markov models. *Stochastic Processes and their Applications*, In Press, Corrected Proof:–.

[8] Douc, R., Moulines, E., and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, 32(5):2254–2304.

[9] Doucet, A., De Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice.* Springer, New York.

[10] Hu, Z.-C. and Sun, W. (2006). A note on exponential stability of the nonlinear filter for denumerable Markov chains. *Systems Control Lett.*, 55(11):955–960.

[11] Kleptsyna, M. L. and Veretennikov, A. Y. (2008). On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields*, 141(3-4):411–444.

[12] Le Gland, F. and Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the hilbert metric and application to particle filters. *Ann. Appl. Probab.*, 14:144–187.

[13] LeGland, F. and Oudjane, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Process. Appl.*, 106(2):279–316.

[14] Ocone, D. and Pardoux, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control*, 34:226–243.

[15] Oudjane, N. and Rubenthaler, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.*, 23(3):421–448.

[16] Ristic, B., Arulampalam, M., and Gordon, A. (2004). *Beyond Kalman Filters: Particle Filters for Target Tracking.* Artech House.

[17] Van Handel, R. (2008a). Discrete time nonlinear filters with informative observations are stable. *Electron. Commun. Probab.*, 13:562–575.

[18] Van Handel, R. (2008b). Uniform observability of hidden markov models and filter stability for unstable signals.

[19] Veretennikov, A. Y. (2002). Coupling method for Markov chains under integral Doeblin type condition. In *Proceedings of the Conference Dedicated to the 90th Anniversary of Boris Vladimirovich Gnedenko (Kyiv, 2002)*, volume 8, pages 383–390.

Télécom SudParis,
9 rue Charles Fourier, 91011 Evry Cedex.
E-mail: randal.douc@it-sudparis.eu

Laboratoire de Mathématiques d'Orsay,
Bâtiment 425, 15, rue Georges Clemenceau,
91405 Orsay Cedex Orsay, France.
E-mail: gassiat@math.u-psud.fr

Systèmes Aéroportes,
2 av Gay Lussac,
78990 Elancourt , France
E-mail: benoit.landelle@math.u-psud.fr

Télécom ParisTech,
46 rue Barrault, 75634, Paris, France.
E-mail: eric.moulines@telecom-paristech.fr