Submitted to the Annals of Applied Probability

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

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-type convergence of the total variation distance of the filter started from two different initial distributions, and a convergence in expectation are considered. The results are illustrated using generic models of non-ergodic HMM.

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. In this case, the state estimation problem is therefore a particular instance of the discrete time filtering problem. More specifically, let X and Y be 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 kernel defined for any $(x,y) \in \mathsf{X}\times\mathsf{Y}$

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

and $C \in \mathcal{X} \otimes \mathcal{Y}$ by

$$T\left[(x,y),C\right] \stackrel{\text{def}}{=} Q \otimes G[(x,y),C] = \iint Q(x,dx')\, g(x',dy') \mathbb{1}_C(x',y') \; . \tag{1}$$

We consider $\{X_k, Y_k\}_{k\geq 0}$ the Markov chain with transition kernel T and initial distribution $\nu \otimes G(C) \stackrel{\text{def}}{=} \iint \nu(dx) g(x,y) \mathbb{1}_C(x,y)$, where ν is a probability measure on $(\mathsf{X}, \mathcal{X})$. With a slight abuse in the terminology, ν is referred to the initial distribution of $\{(X_k, Y_k)\}_{k\geq 0}$. We denote by $\mathbb{P}_\nu$ the distribution of the Markov chain $\{(X_k, Y_k)\}_{k\geq 0}$ over canonical sequence space $\Omega = \mathsf{X}^{\mathbb{N}} \times \mathsf{Y}^{\mathbb{N}}$ and by $\mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k\geq 0})}$ the trace of this probability on $\mathsf{Y}^{\mathbb{N}}$. We assume that the chain $\{X_k\}_{k\geq 0}$ is not observed (*hidden*). The distribution of the hidden state X_n conditionally on the observations $Y_{0:n} \stackrel{\text{def}}{=} [Y_0, \ldots, Y_n]$. $\phi_{\nu,n}[Y_{0:n}]$ is referred to as the *filtering* distribution. These distributions can be computed recursively.

A typical question consists in finding conditions under which the filtering distribution is *stable*, which means that the 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. More precisely, assuming that $\{Y_k\}_{k\geq 0}$ is a is Y-valued stochastic process defined on some space $(\Omega, \mathcal{F}, \mathbb{P}_\star)$, we want to establish either pathwise filter stability

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

or filter stability in the mean

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

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. The filter is said to be exponentially stable in the pathwise or the mean sense if

$$\limsup_{n\to\infty} n^{-1} \log \left(\left\| \phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}] \right\|_{\mathrm{TV}} \right) < 0 \; , \tag{4}$$

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

We stress that, in contrast with most contributions on this subject, $\mathbb{P}_\star$ need not be equal to $\mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k\geq 0})}$ which means that we are interested in studying the filter stability even for mis-specified models. As stressed by [5], the most important motivation for studying the stability of the filter is the time-uniform convergence of the estimators of the filtering distribution. Because of the recursive nature of the filter, 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 filter is stable enough, then the effect of these local error does not accumulate. Another important application of the stability is for the inference (or calibration) of the transition kernel Q or the likelihood g, when these quantities depend upon a finite or an infinite-dimensional parameters. As shown in [10], the convergence of the likelihood of the observation and the consistency of the maximum likelihood estimator ultimately relies on the stability of the filter for mis-specified observations (several examples of this type will be given later).

The forgetting property of the initial condition of the optimal filter in nonlinear state space models has attracted many research efforts; see for example the in-depth tutorial of [**?**]. 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 forgetting property can be investigated using tools from the theory of positive operators, namely the Birkhoff contraction inequality for the Hilbert projective metric (see [1], [15], [14]). The results obtained using this approach require stringent *mixing* conditions for the transition kernels; these conditions state that there exist positive constants ϵ_- and ϵ_+ and a probability measure λ on $(\mathsf{X}, \mathcal{X})$ such that for $f \in \mathbb{B}^+(\mathsf{X})$,

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

where $Qf(x) = Q(x,f) \stackrel{\text{def}}{=} \int Q(x, dx') f(x')$, for any function $f \in \mathbb{B}_+(\mathsf{X})$ the set of non-negative functions $f : \mathsf{X} \to \mathbb{R}$, such that f is $\mathcal{X}/\mathcal{B}(\mathbb{R})$ measurable, with $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra. This condition in particular implies that the chain is uniformly geometrically ergodic. Similar results were obtained independently by [7] using the Dobrushin ergodicity coefficient (see [8] for further refinements under this assumption). The mixing condition has later been weakened by [4], under the assumption that the kernel Q is positive recurrent and is dominated by some reference measure λ:

$$\sup_{(x,x') \in \mathsf{X} \times \mathsf{X}} q(x, x') < \infty \quad \text{and} \quad \int \operatorname{essinf} q(x, x') \pi(x) \lambda(dx) > 0 \;,$$

where $q(x, \cdot) = \frac{dQ(x,\cdot)}{d\lambda}$, essinf is the essential infimum with respect to λ and $\pi d\lambda$ is the stationary distribution of the chain Q . If the upper bound is reasonable, the lower bound is restrictive in many applications and fails to be satisfied e.g. for the linear state space Gaussian model. In [15], 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

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

This condition implies that for any $(x', x'') \in \mathsf{C} \times \mathsf{C}$,

$$\frac{\epsilon_-(d)}{\epsilon_+(d)} < \mathrm{essinf}_{x\in\mathsf{X}} \frac{q(x', x)}{q(x'', x)} \le \mathrm{esssup}_{x\in\mathsf{X}} \frac{q(x', x)}{q(x'', x)} \le \frac{\epsilon_+(d)}{\epsilon_-(d)} \;,$$

where $q(x, \cdot) \stackrel{\text{def}}{=} dQ(x, \cdot)/d\lambda_\mathsf{C}$, and esssup and essinf denote the essential supremum and infimum with respect to λ_C. This condition is obviously more general than (6), but still it is not satisfied in the linear Gaussian case (see [15, Example 4.3]). All the above mentioned conditions are strictly stronger than the ergodicity. Perhaps surprisingly, [4, Section 5] show that the ergodicity of the signal is not sufficient to guarantee stability.

Several attempts have been made to establish the stability conditions under the so-called *small* noise condition. The first result in this direction has been obtained by [1] (in continuous time) who considered an ergodic diffusion process with constant diffusion coefficient and linear observations: when the variance of the observation noise is sufficiently small, [1] established that the filter is exponentially stable. Small noise conditions also appeared (in a discrete time setting) in [3] and [17]. These results do not allow to consider the linear gaussian state space model with arbitrary noise variance.

A significant improvement has been achieved by [13], who considered the filtering problem of a signal $\{X_k\}_{k\ge0}$ taking values in $\mathsf{X} = \mathbb{R}^d$ filtered from observations $\{Y_k\}_{k\ge0}$ 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 $\{(\zeta_k, \varepsilon_k)\}_{k\ge0}$ is a i.i.d. sequence of random vectors in $\mathbb{R}^{d+\ell}$ with density $q_\zeta(x)q_\varepsilon(y)$, $b(\cdot)$ 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 in the mean (**??**) stability of the filter under appropriate conditions on the functions b, 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 $\mathbb{P}_\star = \mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k\geq 0})}$

A new approach for ergodic HMM using the so-called *Local Doeblin property* is proposed in [9]. Both almost sure convergence and convergence in expectation for the distance in total variation norm for two filters with different initial distributions are proven. The results hold under weaker conditions than those appearing under other mixing assumptions and, in particular, cover the linear Gaussian state-space model. Moreover, assumptions on observations are relaxed and convergence theorems apply for sequences which are not necessarily HMM.

The works mentioned above mainly deal with ergodic HMM, *i.e.* the situations in which the signal process is ergodic. Results for non-ergodic signals in the linear Gaussian case are now classical; see [16]. More recently, extensions of these results to the non-linear filtering problem have been considered, triggered by the observations many models used for example in engineering or econometrics are non-ergodic (see [11] and [18] and the references therein). Non-linear non-ergodic HMM have been considered much less frequently in the literature. In [3], the observation process is the signal (state) corrupted by an additive white noise of sufficiently small noise variance. In [17], the authors assume the model (8)-(9) (under some technical conditions). The authors propose to truncate the Markov kernels on compact sets depending on the observation sequences, which are chosen in such a way that the truncated kernels satisfy strong mixing conditions. 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 $\mathbb{P}_\star = \mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k\geq 0})}$. Similar conditions have been studied in [5], which established the pathwise stability, again under $\mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k\geq 0})}$. 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 [17] allow a form of weak dependence in the signal noise.

A significant weakening of these assumptions has been achieved in [21] and [22]. These papers 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. These papers require however that the initial the conditional distribution of the process $\nu \ll \nu'$ and the pathwise and the mean filter stability are obtained provided that the distribution of the observation process is $\mathbb{P}_\star = \mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k \geq 0})}$.

Using a rather different perspective, [**?**] have considered a model in which 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 the

In this contribution, we propose a new set of conditions to establish the stability of the filter under model mis-specification. Compared to the conditions in [21] and [22], 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 by $\mathbb{P}_\nu \Big|_{\sigma(\{Y_k\}_{k \geq 0})}$ and may, on the contrary, be fairly general. On the other hand, stronger conditions on the transition kernel Q are required, which are reminiscent from the Local Doeblin condition introduced in [9]. 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) and we provide an explicit bound of the deviation. Theorem 6, the average stability of the filter (**??**) is obtained, once again with computable bound. In section 3, different nonlinear state-space models are analyzed. For these models, we establish conditions for pathwise and average stability at geometric rate for model. The proofs of the main results are given in Section 4. Several technical Lemmas required to study the examples are given in Sections 5, 6.

2. Main results.

Our results require the choice of a set-valued function, referred to as *Local Doeblin set function*, which extends the so-called local Doeblin sets introduced in [23] and later exploited in [13]. The difference between LD-sets of [23] 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* $(0, \infty)^2$

- *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* $A \in \mathcal{X}$*,*

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

Some general conditions on the Local Doeblin set function involving the distributions of the observations ensure the forgetting property of the optimal filter. The case of nonlinear state-space models is studied in Section 3. Roughly speaking, inequality (10) means that the transition of the hidden chain, when the state is in a given subset $C(y)$ does not depend too much on the current state.

Consider the following assumptions on the likelihood of the observations:

H 1. For all $(x,y) \in \mathsf{X} \times \mathsf{Y}$, $g(x,y) > 0$.

For a set $A \in \mathcal{X}$ and an observation $y \in \mathsf{Y}$, the supremum of the likelihood over A is denoted

$$\Upsilon_A(y) \stackrel{\text{def}}{=} \sup_{x\in A} g(x,y) \ . \tag{11}$$

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

$$\Upsilon_{\mathsf{C}^c_\eta(y)}(y) \leq \eta \Upsilon_{\mathsf{X}}(y) \ . \tag{12}$$

Condition (H1) states that the likelihood is everywhere positive. This excludes the case of additive noise with bounded support; see for example [2]. When $\mathsf{X} = \mathbb{R}^d$, the second 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 assumption is satisfied in many models of practical interest, and roughly implies that the observation effectively provides information on the state range of values.

For a given LD-set function C , we set

$$\Phi_{\nu,\mathsf{C}}(y,y') \stackrel{\text{def}}{=} \nu\left[g(\cdot,y)Qg(\cdot,y')\mathbb{1}_{\mathsf{C}(y')}(\cdot)\right] \ , \tag{13}$$

$$\Psi_{\mathsf{C}}(y,y') \stackrel{\text{def}}{=} \lambda\left(y,y';g(\cdot,y')\mathbb{1}_{\mathsf{C}(y')}\right) \ . \tag{14}$$

Assume moreover that

[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

H 3. The functions $(y,y') \mapsto \Phi_{\nu,\mathsf{C}}(y,y')$ and $(y,y') \mapsto \Psi_{\mathsf{C}}(y,y')$ are measurable.

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 jumps to another LD set, the forgetting mechanism comes into play. From now on, for all $(x,x') \in \mathsf{X}^2$, denote by $\bar{x} = (x,x')$ 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 distribution 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, for any $A \in \mathcal{X}$ and ν and ν' two probability distributions on $(\mathsf{X},\mathcal{X})$, the difference $\phi_{\nu,n}[y_{0:n}](A) - \phi_{\nu',n}[y_{0:n}](A)$ may be expressed as

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

We compute bounds for the numerator and the denominator of the previous expression. Such bounds are given in the two following Propositions (proofs are postponed to Section 4). For an LD-set function C denote:

$$
\rho_{\mathsf{C}}(y,y') \stackrel{\text{def}}{=} 1 - (\varepsilon^{-}_{\mathsf{C}}/\varepsilon^{+}_{\mathsf{C}})^2(y,y')\;. \tag{16}
$$

PROPOSITION 2. *Let C be an LD-set function and ν and ν' be two probability measures on $(\mathsf{X},\mathcal{X})$. For any integer n and any sequence $\{y_i\}_{i=0}^{n}$ in Y, let us define*

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

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

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

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

The proofs of these two propositions are adapted from [9] and are postponed to Section 4. By combining these two Propositions, we obtain an explicit bound for the total variation distance $\|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}}$. 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'. To state the result, some additional notations are required. Under assumption (H2), for a fixed $\eta > 0$ and a corresponding LD-set function C_η, let us define, for $\alpha \in (0, 1)$ and a sequence $y_{0:n} = \{y_i\}_{i=0}^n$ in Y,

(18)

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

where ρ_η is a shorthand notation for ρ_{C_η} (see (16))

THEOREM 4. *Assume (H1)-(H2)-(H3). 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* $\eta > 0$,

(19)

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

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

PROOF. Eq. (15) implies

EM donner une indication sur la preuve qui explique que l'on travaille avec deux LD-set functions

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

where $\Delta_n(\nu, \nu', y_{0:n})$ is defined by (17). Set

$$N_{\eta,n} \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbb{1}\{(\bar{X}_{i-1}, \bar{X}_i) \in \bar{\mathsf{C}}_\eta(y_{i-1}) \times \bar{\mathsf{C}}_\eta(y_i)\} , \quad M_{\eta,n} \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}_\eta^c(y_i)}(\bar{X}_i) .$$

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

$$\begin{aligned}(20)\quad \Delta_n(\nu,\nu',y_{0:n}) \leq \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\bar{g}(\bar{X}_0,y_0)\prod_{i=1}^n \bar{g}(\bar{X}_i,y_i)\rho_\eta^{\delta_i}(y_{i-1},y_i)\mathbb{1}\{N_{\eta,n} \geq \alpha n\}\right] \\ + \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\bar{g}(\bar{X}_0,y_0)\prod_{i=1}^n \bar{g}(\bar{X}_i,y_i)\rho_\eta^{\delta_i}(y_{i-1},y_i)\mathbb{1}\{N_{\eta,n} < \alpha n\}\right] \; ,\end{aligned}$$

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 last term in the right-hand side of (20) satisfies

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

By splitting this last product, we obtain

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

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

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

The first term in the right-hand side expression of (20) satisfies

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

By combining the above relations, the result follows. □

The last step consists in finding conditions upon which the bound in the right hand side of (19) 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 $(\mathsf{Y}, \mathcal{Y})$, which is not necessarily related to the model under which the filter is computed. 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.

THEOREM 5. *Assume (H1)-(H2)-(H3). Assume moreover that there exists some LD-set function* C *such that*

$$\liminf_{n\to\infty} n^{-1} \sum_{k=2}^{n} \log \varepsilon_{\mathsf{C}}^{-}(Y_{k-1}, Y_k) > -\infty \,, \qquad \mathbb{P}_\star - \text{a.s.} \tag{21}$$

$$\limsup_{n\to\infty} n^{-1} \sum_{k=0}^{n} \log \Upsilon_{\mathsf{X}}(Y_k) < \infty \,, \qquad \mathbb{P}_\star - \text{a.s.} \tag{22}$$

$$\liminf_{n\to\infty} n^{-1} \sum_{k=2}^{n} \log \Psi_{\mathsf{C}}(Y_{k-1}, Y_k) > -\infty \,. \qquad \mathbb{P}_\star - \text{a.s.} \tag{23}$$

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

$$\limsup_{n\to\infty} n^{-1} \log \Lambda_\eta(Y_{0:n}, \alpha) < 0, \qquad \mathbb{P}_\star - \text{a.s.} \tag{24}$$

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

$$\nu Q \mathbb{1}_{\mathsf{C}(Y_1)} > 0 \,, \quad \mathbb{P}_\star - \text{a.s.} \qquad \nu' Q \mathbb{1}_{\mathsf{C}(Y_1)} > 0 \,, \quad \mathbb{P}_\star - \text{a.s.}$$

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 OF THEOREM 5. For any positive sequences $\{u_n\}$ and $\{v_n\}$,

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

Under the stated assumptions, there exists a LD-set function C and some constant $M > 0$ such that for any $\eta > 0$,

$$\limsup n^{-1} \log \left\| \phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}] \right\|_{\mathrm{TV}} \leq \frac{1-\alpha}{2} \log(\eta) + M \,, \quad \mathbb{P}_\star - \text{a.s.}$$

The proof is concluded by choosing η small enough. □

The assumptions linking the LD-set function and the observations make this theorem quite abstract. With a filtering model defined by specific equations, assumptions can be directly formulated on the model and on the observations. Such situations will be described through examples presented in Section 3.

Compared to [9, Theorem 1] in the ergodic case , the conditions (21) and (24) 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 (21) and (24) are trivially satisfied.

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

$$\mathbb{E}_\star\Big[\big\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\big\|_{\mathrm{TV}}\Big] \leq 2\beta^n + r_0(\nu, n) + r_0(\nu', n) + \sum_{i=1}^{4} r_i(n) \tag{25}$$

where the sequences in the right-hand side of (25) *are defined by*

$$r_0(\nu, n) \stackrel{\text{def}}{=} \mathbb{P}_\star\left(\log \Phi_{\nu,\mathsf{C}}(Y_0, Y_1) \leq -M_0 n\right), \tag{26}$$

$$r_1(n) \stackrel{\text{def}}{=} \mathbb{P}_\star\left(\sum_{k=2}^{n} \log \varepsilon^-_{\mathsf{C}}(Y_{k-1}, Y_k) \leq -M_1 n\right), \tag{27}$$

$$r_2(n) \stackrel{\text{def}}{=} \mathbb{P}_\star\left(\sum_{k=0}^{n} \log \Upsilon_{\mathsf{X}}(Y_k) \geq M_2 n\right), \tag{28}$$

$$r_3(n) \stackrel{\text{def}}{=} \mathbb{P}_\star\left(\sum_{k=2}^{n} \log \Psi_{\mathsf{C}}(Y_{k-1}, Y_k) \leq -M_3 n\right), \tag{29}$$

$$r_4(n) \stackrel{\text{def}}{=} \mathbb{P}_\star\left(\log \Lambda_\eta(Y_{0:n}, \alpha) \geq -\delta n\right). \tag{30}$$

PROOF OF THEOREM 6. Denote by Ω_n the event

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

Under the stated assumptions, $\mathbb{P}_\star(\Omega_n^{\mathrm{c}}) \leq r_0(\nu, n) + r_0(\nu', n) + \sum_{i=1}^{3} r_i(n)$.

On the event Ω_n, we have

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

We can choose η small enough and $\gamma\in(0,1)$ such that $\eta^{a_n}e^{2n\sum_{i=0}^{3}M_i}\leq\gamma^n$. Then, by Theorem 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 $\beta=\max(\gamma,e^{-\delta})$. Thus,

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

and 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 (25) gives a geometric rate. In Section 3, for given filtering equations, explicit majorations of the other terms will be obtained with geometric rates. Like for the pathwise convergence, the terms r_1 and r_4 which involve the functions $\varepsilon_{\mathsf{C}}^{-}$ and $\varepsilon_{\mathsf{C}}^{+}$ are specific to the non-ergodic case.

3. Nonlinear state-space models.

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

$$\begin{cases} X_k = f(X_{k-1})+\zeta_k\ , \\ Y_k = h(X_k)+\varepsilon_k\ , \end{cases} \tag{31}$$

where f and h denote some measurable functions.

We denote by ν_0 the initial distribution of X_0. We consider the following assumptions:

E 1. f is a-Lipschitz, i.e. $|f(x)-f(x')|\leq a|x-x'|$ and h is uniformly continuous and surjective and there exist constant b_0 and b such that for all $x_1,\ x_2\in\mathsf{X}$,

$$|x_1-x_2|\leq b_0+b|h(x_1)-h(x_2)|\ .$$

E 2. The observation noise $\{\varepsilon_k\}_{k\geq 0}$ is a sequence of i.i.d. random variables with positive density υ with respect to the Lebesgue measure Leb on Y. Moreover, the density υ is bounded, $\lim_{|u|\to\infty}\upsilon(u)=0$, and, for all compact set $\mathsf{K}\subset\mathsf{Y}$, $\inf_{y\in\mathsf{K}}\upsilon(y)>0$.

Notice that f is not necessarily contracting so that the model is possibly non-ergodic. The assumption (E1) has been first considered in [17]. A function h satisfying (E1) 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 two successive observations $y_1,\ y_2 \in \mathsf{Y}$, the maximal distance between any two elements in the preimages $h^{-1}(\{y_1\})$ and $h^{-1}(\{y_2\})$ can not be arbitrarily large. Even if h is not bijective, the distance $|y_1 - y_2|$ gives information on the distance of two successive preimage states. The assumption (E2) is more classical and is satisfied, for example, by Gaussian densities.

We first consider the simple situation where the state noise $\{\zeta_k\}_{k\geq 0}$ is a sequence of i.i.d. random variables independent of the observation noise $\{\varepsilon_k\}_{k\geq 0}$ and the observations are distributed according to the model. Then, we study more general dependence structure of the state noise distribution and the case where the observations do not necessarily follow the model.

3.1. *Nonlinear state-space model with i.i.d. state noise.* In this section, we assume that

E 3. The state noise $\{\zeta_k\}_{k\geq 0}$ is a sequence of i.i.d. random variables, independent of the observation noise $\{\varepsilon_k\}_{k\geq 0}$, with density γ with respect to the Lebesgue measure Leb on X. . In addition, γ is locally bounded and for any compact subset K, $\inf_{x\in K} \gamma(x) > 0$.

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

$$Q(x, A) = \int_A \gamma[x' - f(x)]\, \mathrm{Leb}(dx')\ . \tag{32}$$

For any $\Delta \in (0, \infty)$, let us define the following set-valued function from Y to $\mathcal{X}$ by

$$y \longmapsto \mathsf{C}(y, \Delta) \stackrel{\mathrm{def}}{=} \{x \in \mathsf{X} : |h(x) - y| \leq \Delta\}\ . \tag{33}$$

For any $y \in \mathsf{Y}$, $\mathsf{C}(y, \Delta)$ is included in a neighborhood of $h^{-1}(\{y\})$, the preimage of y with respect to h. Indeed, under assumption (E1), for any $z \in \mathsf{X}$ in $h^{-1}(\{y\})$, and any $x \in \mathsf{C}(y, \Delta)$,

$$|x - z| \leq b_0 + b\Delta\ .$$

Let $(y, y') \in \mathsf{Y}^2$. By the condition (E1), h is surjective so $h^{-1}(\{y\}) \neq \emptyset$ and $h^{-1}(\{y'\}) \neq \emptyset$. We choose arbitrarily $z \in h^{-1}(\{y\})$ and $z' \in h^{-1}(\{y'\})$.

By the triangle inequality and the condition (E1), it follows that, for all $(x,x') \in \mathsf{C}(y,\Delta) \times \mathsf{C}(y',\Delta)$,
(34)
$$|f(x)-x'| \le |f(x)-f(z)|+|f(z)-z'|+|z'-x'| \le a(b_0+b\Delta)+D(y,y')+b_0+b\Delta \;,$$

where the function $(y,y') \mapsto D(y,y')$ is defined by

$$D(y,y') \stackrel{\text{def}}{=} \sup\left\{|f(z)-z'| \; : z \in h^{-1}(\{y\}), z' \in h^{-1}(\{y'\})\right\} \;. \tag{35}$$

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

$$\gamma^-(r) \stackrel{\text{def}}{=} \inf_{|s|\le r} \gamma(s) \;, \qquad \gamma^+(r) \stackrel{\text{def}}{=} \sup_{|s|\le r} \gamma(s) \;, \tag{36}$$

It follows from (32) and (34) that, for all $A \in \mathcal{X}$ and $x \in \mathsf{C}(y,\Delta)$,

$$\varepsilon^-_\Delta(y,y')\mathrm{Leb}[A \cap \mathsf{C}(y',\Delta)] \le Q[x, A \cap \mathsf{C}(y',\Delta)] \le \varepsilon^+_\Delta(y,y')\mathrm{Leb}[A \cap \mathsf{C}(y',\Delta)] \;, \tag{37}$$

where,

$$\varepsilon^-_\Delta(y,y') \stackrel{\text{def}}{=} \gamma^-[(a+1)b_0 + (a+1)b\Delta + D(y,y')] \;,$$
$$\varepsilon^+_\Delta(y,y') \stackrel{\text{def}}{=} \gamma^+[(a+1)b_0 + (a+1)b\Delta + D(y,y')] \;.$$

Under (E3), it follows by (37) that the application defined by (33) is a LD-set function. By assumption (E2), for all $\eta > 0$, we may choose Δ large enough so that $\sup_{|s|>\Delta} \upsilon(s) \le \eta \sup_{s\in\mathsf{X}} \upsilon(s)$. Since $g(x,y) = \upsilon\left(y - h(x)\right)$, the previous inequality implies assumption (H2)

$$\Upsilon_{\mathsf{C}^c(y,\Delta)}(y) \le \eta \Upsilon_{\mathsf{X}}(y) \;, \tag{38}$$

is satisfied. The positiveness of υ implies assumption (H1).

To check assumptions (21) and (24), it is required to compute an upper bound for $\{D(Y_{k-1},Y_k)\}_{k\ge1}$. For $z,\ z' \in \mathsf{X}$ such that $h(z) = Y_{k-1},\ h(z') = Y_k$, it follows from the triangle inequality and assumption (E1) that

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

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

$$D(Y_{k-1},Y_k) \le (a+1)b_0 + ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k| \;. \tag{39}$$

Thanks to this bound, assumptions (21) and (24) are satisfied by applying the Law of Large Numbers, see Propositions 7 and 9 and their proofs. Since γ^- is a non increasing function, it follows by (39) that, for all integer $k \geq 1$, $\log \varepsilon_\Delta^-(Y_{k-1}, Y_k) \geq -Z_k^\Delta$ where for all $\Delta > 0$ and all integer $k \geq 1$,

$$Z_k^\Delta \stackrel{\text{def}}{=} -\log \gamma^- \left[2(a+1)b_0 + (a+1)b\Delta + ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|\right] . \tag{40}$$

PROPOSITION 7. *Let us consider the state-space model* (31). *Assume (E1)-(E-2)-(E3) and, for all* $\Delta > 0$,

$$\mathbb{E}|Z_1^\Delta| < \infty . \tag{41}$$

Let $\mathbb{P}_{\nu_0}$ *be the probability distribution on* $\mathsf{Y}^\mathbb{N}$ *of the process* $\{Y_k\}_{k\geq 0}$ *defined by* (31) *with initial distribution* ν_0 *. Then, for any initial probability distributions* 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 (41), is not very restrictive. For example, assume that $\{\zeta_k\}_{k\geq 0}$ and $\{\varepsilon_k\}_{k\geq 0}$ are sequences of Gaussian random variables. It follows, that $\gamma^-(r) = \gamma(r)$ for all $r \geq 0$. The condition (41) holds if $\mathbb{E}(|\varepsilon_1|^2) < \infty$ and $\mathbb{E}(|\zeta_1|^2) < \infty$ which are trivially satisfied.

With more stringent conditions on initial distributions, the convergence of the expected value of the total variation distance $\left\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\right\|_{\mathrm{TV}}$ may be shown to be geometric. Let us recall the definition of the log-moment generating function that will be used in the sequel.

DEFINITION 8. *The log-moment generating function* $\psi_Z(\alpha)$ *of the random variable* Z *is defined on the set* $\{\alpha \geq 0 : \mathbb{E}[\mathrm{e}^{\alpha Z}] < \infty\}$ *by* $\psi_Z(\alpha) \stackrel{\text{def}}{=} \log \mathbb{E}[\mathrm{e}^{\alpha Z}]$.

PROPOSITION 9. *Let us consider the state-space model defined by* (31). *Assume that (E1)-(E2)-(E3) hold and, for all* $\Delta > 0$, *there exists* $\delta > 0$ *such that*

$$\psi_{Z_1^\Delta} \text{ is finite on } [0, \delta). \tag{42}$$

Let $\mathbb{P}_{\nu_0}$ *be the probability distribution on* $\mathsf{Y}^\mathbb{N}$ *of the process* $\{Y_k\}_{k\geq 0}$ *defined by* (31) *with initial distribution* ν_0 *. Let* C *be the LD-set function defined by* (33). *Then, for* ν *and* ν' *two probability measures on* $(\mathsf{X}, \mathcal{X})$ *and* $\Delta > 0$ *such that, for some* $\alpha > 0$, EM affaiblir

$$(43)\quad \mathbb{E}_\star\left\{\exp\left(\alpha\left[\log \nu g(\cdot, Y_0) Q \mathbb{1}_{\mathsf{C}(Y_1,\Delta)}\right]_-\right)\right\} < \infty\;, \qquad \mathbb{E}_\star\left\{\exp\left(\alpha\left[\log \nu' g(\cdot, Y_0) Q \mathbb{1}_{\mathsf{C}(Y_1,\Delta)}\right]_-\right)\right\} < \infty\;,$$

we have

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

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

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

where ς denotes some positive constant. Therefore, for $\alpha > 0$ small enough, the condition (42) is satisfied.

Proofs of Propositions 7 and 9 are given in Section 5.

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 [17, 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]. This example illustrates that the forgetting property is kept even when the distributions of the observations differ from the model.

G 1. $\{\zeta_k\}_{k\geq 0}$ is a sequence of random variables such that, for all integer k, ζ_k is independent of ε_k and for all $A \in \mathcal{X}$,

$$\mathbb{P}\left[\zeta_k \in A \middle| X_{0:k-1}\right] = \int q(X_{k-1}, u) \mathbb{1}_A(u)\, \mathrm{Leb}(du)\;.$$

Moreover, there exist a probability density ψ and positive constants μ_- and μ_+ such that ψ is locally bounded and bounded away from zero on any compact subset, and for all $x,\ u \in \mathsf{X}$,

$$\mu^- \psi(u) \leq q(x, u) \leq \mu^+ \psi(u)\;.$$

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

$$X_k = f(X_{k-1}) + \sigma(X_{k-1})\xi_k\;, \tag{44}$$

where $\{\xi_k\}_{k\geq 0}$ 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:

$$\sigma^-|u|^2 \leq \langle u, \sigma(x)\sigma^{\mathrm{T}}(x)u\rangle \leq \sigma^+|u|^2 \ , \tag{45}$$

where σ^-, σ^+ are positive constants and the superscript T denotes the transposition. Another important example where assumption (G1) is satisfied is the case of certain discretely sampled diffusions. Let $(X_t)_{t\geq 0}$ 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 $\{X_k\}_{k\geq 0}$ satisfies assumption (G1) if the function σ is uniformly elliptic (condition (45)); see [17]. The assumptions (E1), (E2) and (G1) are similar to those made in [17]. 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. Let us denote by Q the transition kernel for $\{X_k\}_{k\geq 0}$. Then, for all $A \in \mathcal{X}$ and for all $x \in \mathsf{X}$,

EM il faut modifier le commentaire et en plus comparer aussi avec les conditions de Crisan et Heine

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

For the same reasons as above, we consider the same set-valued function C (33) as before. Let $(y, y') \in \mathsf{Y}^2$. Like in (34), it follows by (E1) and the triangle inequality that, for all $(x, x') \in \mathsf{C}(y, \Delta) \times \mathsf{C}(y', \Delta)$,

$$|f(x) - x'| \leq c + d\Delta + D(y, y') \ ,$$

where D is defined in (35), $c = (a + 1)b_0$ and $d = (a + 1)b$. By setting

$$q^-(r) \stackrel{\mathrm{def}}{=} \mu^- \times \inf_{|v|\leq r} \psi(v) \ , \qquad q^+(r) \stackrel{\mathrm{def}}{=} \mu^+ \times \sup_{|v|\leq r} \psi(v) \ , \tag{46}$$

it follows from condition (G1) that, for all $A \in \mathcal{X}$ and $x \in \mathsf{C}(y, \Delta)$,

$$\varepsilon^-_\Delta(y, y')\mathrm{Leb}[A \cap \mathsf{C}(y', \Delta)] \leq Q[x, A \cap \mathsf{C}(y', \Delta)] \leq \varepsilon^+_\Delta(y, y')\mathrm{Leb}[A \cap \mathsf{C}(y', \Delta)] \ , \tag{47}$$

where

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

Under assumption (G1), the application defined by (33) is a LD-set function. As in Section 3.1, assumptions (H1) and (H2) are satisfied. Assume now that the process $\{Y^*_k\}_{k\geq 0}$ is generated by a state-model similar to (31) with two different functions f^* and h^*,

$$\begin{cases} X^*_k = f^*(X^*_{k-1}) + \zeta^*_k \;, \\ Y^*_k = h^*(X^*_k) + \varepsilon^*_k \;, \end{cases} \tag{48}$$

where $\{\varepsilon^*_k\}_{k\geq 0}$ is a sequence of i.i.d random variables, f^* is a^*-Lipschitz, h^* is surjective and for all $x_1,\ x_2 \in \mathsf{X}$,

$$|x_1 - x_2| \leq b^*_0 + b^* |h^*(x_1) - h^*(x_2)| \;,$$

for some positive constants $b^*_0,\ b^*$. For all integer $k \geq 1$, ζ^*_k is independent of ε^*_k and, for all $A \in \mathcal{X}$,

$$\mathbb{P}\left[\zeta^*_k \in A |\, {X^*}_{0:k-1}\right] = \int q^*(X^*_{k-1}, u) \mathbb{1}_A(u)\, \mathrm{Leb}(du) \;.$$

There exists probability density ψ^* and positive constants $\mu^*_-,\ \mu^*_+$ such that, for all $x,\ u \in \mathsf{X}$, EM Vérifier que cette condition est suffisante

$$\mu^*_- \psi^*(u) \leq q^*(x,u) \leq \mu^*_+ \psi^*(u) \;. \tag{49}$$

We assume that

O 1. f^* and h^* are such that $\|f - f^*\|_\infty < \infty$ and $\|h - h^*\|_\infty < \infty$.

Lemma 10. *Let $\{Y^*_k\}_{k\geq 0}$ be the sequence following (48). Under (O1), for all integer $k \geq 1$,*

$$D(Y^*_{k-1}, Y^*_k) \leq \kappa + 2a^* b^* + a^* b^* |\varepsilon^*_{k-1}| + b^* |\varepsilon^*_k| + |\zeta^*_k| \;,$$

where

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

Proof. 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 $h^*(u) = Y^*_{k-1},\ h^*(u') = Y^*_k$, it follows by the triangle inequality that

$$\begin{aligned} |f(z) - z'| &\leq |f(z) - f^*(z)| + |f^*(z) - f^*(u)| + |f^*(u) - u'| + |u' - z'| \;, \\ &\leq \|f - f^*\|_\infty + a^* |z - u| + |f^*(u) - u'| + |u' - z'| \;. \end{aligned} \tag{50}$$

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, for the same reasons, $|z' - u'| \leq K$. Combining these two majorations with (50) leads to

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

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

$$D(Y_{k-1}^*, Y_k^*) \leq \kappa + 2a^*b^* + a^*b^*|\varepsilon_{k-1}^*| + b^*|\varepsilon_k^*| + |\zeta_k^*|\ .$$

□

Let us define for all $\Delta > 0$

$$V_+^{*\Delta} = \log q^- \left[c + d\Delta + \kappa + 2a^*b^* + a^*b^*|\varepsilon_0^*| + b^*|\varepsilon_1^*| + |\zeta_+^*|\right]\ , \tag{51}$$

where ζ_+^* is a random variable independent of $\{\varepsilon_k^*\}_{k\geq 0}$ with density ψ^*.

Proposition 11. *Let us consider the state-space model defined by* (31) *and satisfying (E1), (E2) and (G1). Let* C *be the LD-set function defined by* (33) *and let* $\{Y_k^*\}_{k\geq 0}$ *be the sequence following* (48) *such that (O1) holds and, for all* $\Delta > 0$,

$$\mathbb{E}\left(|V_+^{*\Delta}| \log_+ |V_+^{*\Delta}|\right) < \infty\ . \tag{52}$$

Then, for any initial probability distributions ν *and* ν' *on* $(\mathsf{X}, \mathcal{X})$ *and* $\Delta > 0$, *we have*

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

Without loss of generality, let us write $\zeta_k^* = \tau^*(X_{k-1}^*, A_k^*)$ where τ^* denotes a measurable function and $\{A_k^*\}_{k\geq 0}$ a sequence of i.i.d. random variables with uniform law on $(0,1)^{d_X}$.

O 2. There exists a measurable function τ_+^* such that, for all $x \in \mathsf{X}$ and $a \in (0,1)$, $|\tau^*(x,a)| \leq \tau_+^*(a)$;

O 3. Let $\{Z_k^{*\Delta}\}_{k\geq 0}$ be the sequence defined by, for all $\Delta > 0$ and for all integer $k \geq 1$,

$$Z_k^{*\Delta} = -\log q^- \left[c + d\Delta + \kappa + 2a^*b^* + a^*b^*|\varepsilon_{k-1}^*| + b^*|\varepsilon_k^*| + \tau_+^*(A_k^*)\right]\ ,$$

For all $\Delta > 0$, there exists $\delta > 0$ such that the log-moment generating function $\Psi_{Z_1^{*\Delta}}$ is finite on $[0, \delta)$.

PROPOSITION 12. *Let us consider the state-space model defined by* (31) *and satisfying (E1), (E2) and (G1). Let* $\{Y_k^*\}_{k\geq 0}$ *be the sequence following* (48) *such that (O1), (O2) and (O3) hold and let* C *be the LD-set function defined by* (33). *Then, for* ν *and* ν' *two probability measures on* $(\mathsf{X}, \mathcal{X})$ *and* $\Delta > 0$ *such that, for some* $\alpha > 0$,

$$
\begin{aligned}
(53) \quad & \mathbb{E}_\star \left\{ \exp\left(\alpha [\log \nu g(\cdot, Y_0^*) Q \mathbb{1}_{\mathsf{C}(Y_1^*, \Delta)}]_- \right) \right\} < \infty \;, \\
& \qquad \mathbb{E}_\star \left\{ \exp\left(\alpha [\log \nu' g(\cdot, Y_0^*) Q \mathbb{1}_{\mathsf{C}(Y_1^*, \Delta)})]_- \right) \right\} < \infty \;,
\end{aligned}
$$

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 11 and 12 are given in Section 6.

EM commenter le résultat et le situer par rapport à l'existant, Oudjane & Rubenthaler et Crisan & Heine

4. Proofs of Propositions 2 and 3.

PROOF OF PROPOSITION 2. 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 $\mathsf{X} \times \mathsf{X}$,

$$
(\varepsilon_i^-)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \leq \bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \leq (\varepsilon_i^+)^2 \bar{\lambda}_i(\mathbb{1}_{\bar{\mathsf{C}}_i} \bar{f}) \;. \tag{54}
$$

Let us define the sequence of unnormalized kernels $\bar{Q}_i^0$ and $\bar{Q}_i^1$ by, 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 (54) 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 $\bar{x} \in \mathsf{X}^2$,

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

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

$$\Delta_n(A, t_{0:n-1}) \stackrel{\text{def}}{=} \nu\otimes\nu'(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times\mathsf{X}}) \\ - \nu'\otimes\nu(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times\mathsf{X}}) \; .$$

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

$$\nu\otimes\nu'(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times B}) = \nu'\otimes\nu(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{B\times A}) \; .$$

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

$$\begin{aligned}
&\nu\otimes\nu'(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times\mathsf{X}}) \\
&= \nu\otimes\nu'(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{i-1}^{t_{i-1}} \bar g_i \mathbb{1}_{\bar{\mathsf{C}}_i}) \times (\varepsilon_{i+1}^-)^2 \bar\lambda_i(\mathbb{1}_{\bar{\mathsf{C}}_{i+1}} \bar g_{i+1} \bar Q_{i+1}^{t_{i+1}} \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times\mathsf{X}}) \; , \\
&= \nu'\otimes\nu(\bar g_0 \bar Q_0^{t_0} \bar g_1 \dots \bar Q_{i-1}^{t_{i-1}} \bar g_i \mathbb{1}_{\bar{\mathsf{C}}_i}) \times (\varepsilon_{i+1}^-)^2 \bar\lambda_i(\mathbb{1}_{\bar{\mathsf{C}}_{i+1}} \bar g_{i+1} \bar Q_{i+1}^{t_{i+1}} \dots \bar Q_{n-1}^{t_{n-1}} \bar g_n \mathbb{1}_{A\times\mathsf{X}}) \; .
\end{aligned}$$

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

$$\Delta_n(A) = \nu\otimes\nu'\Big[\bar g_0 \bar Q_0^1 \bar g_1 \dots \bar Q_{n-1}^1 \bar g_n (\mathbb{1}_{A\times\mathsf{X}} - \mathbb{1}_{\mathsf{X}\times A})\Big] \; .$$

It then follows

$$\Delta_n(\nu, \nu', y_{0:n}) \le \nu\otimes\nu'(\bar g_0 \bar Q_0^1 \bar g_1 \dots \bar Q_{n-1}^1 \bar g_n) \le \mathbb{E}_{\nu\otimes\nu'}^{\bar Q}\left[\bar g(\bar X_0, y_0) \prod_{i=1}^n \bar g(\bar X_i, y_i) \rho_i^{\delta_i}\right] \; ,$$

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

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 $i = 1, \dots, n$, for all $x \in \mathsf{C}(y_{i-1})$ and for all $A \in \mathcal{X}$ with $A \subset \mathsf{C}(y_i)$,

$$\varepsilon_{\mathsf{C}}^-(y_{i-1}, y_i)\lambda(y_{i-1}, y_i; A) \le Q(x, A) \le \varepsilon_{\mathsf{C}}^+(y_{i-1}, y_i)\lambda(y_{i-1}, y_i; A) \; . \tag{55}$$

Let us write the obvious inequality

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

Then, for the right-hand side of this expression, by (55) we have

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

□

5. Proofs of Propositions 7 and 9 .

Proof of Proposition 7. Since, by definition (36), γ^- is a non-increasing function, the inequality (39) leads to

$$n^{-1} \sum_{k=2}^{n} \log \varepsilon_\Delta^-(Y_{k-1}, Y_k) \geq -n^{-1} \sum_{k=2}^{n} Z_k^\Delta \ , \tag{56}$$

where Z_k^Δ is defined in (40). 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 (41) yield

$$\lim_{n\to\infty} n^{-1} \sum_{k=2}^{n} Z_k^\Delta = \mathbb{E}(Z_1^\Delta) < \infty \ , \qquad \mathbb{P}_{\nu_0} - \text{a.s.} \tag{57}$$

By combining (56) and (57), the first condition (21) of Theorem 5 is satisfied. By assumption (E2), the density υ is bounded which implies that $\sup_{y\in\mathsf{Y}} \Upsilon_{\mathsf{X}}(y) \leq \sup \upsilon$. Hence, the second condition (22) of Theorem 5 is satisfied. We now consider the third condition (23). Since the measure appearing in the definition of the LD-set function does not depend on y, y', the function $(y, y') \mapsto \Psi_{\mathsf{C}(y',\Delta)}(y, y')$, defined in (14), does not depend on y and is given by

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

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 $\mathrm{Leb}[\mathsf{C}(y', \Delta)]$ is bounded below by the volume of a ball of radius δ in $\mathbb{R}^{d_\mathsf{X}}$. Thus, we have, for all y, $y' \in \mathsf{Y}$,

$$\Psi_{\mathsf{C}(y',\Delta)}(y, y') \geq \varrho_\Delta \ , \tag{58}$$

for some $\varrho_\Delta > 0$, depending only on Δ. The third condition (23) of Theorem 5 follows. Since assumption (H2) is satisfied, for any fixed $\eta > 0$, we choose $\Delta > 0$ such that inequality (38) holds. Let us write

$$R_\Delta(x) \stackrel{\text{def}}{=} \log \left[1 - (\gamma^-/\gamma^+)^2 (2(a+1)b_0 + (a+1)b\Delta + x)\right] \ , \tag{59}$$

where the constants a, b_0, and b are defined in assumption (E1). We will repeatedly use the following representation of the so-called L-statistic (see [19, Chapter 8]):

LEMMA 13. *Let $\{U_1, \ldots, U_n\}$ be a sequence and $U_{n,1} \leq U_{n,2} \leq \ldots \leq U_{n,n}$ the upper ordered statistic. Then,*

$$n^{-1} \sum_{k=j}^{n} U_{n,k} = \int_{(j-1)/n}^{1} F_{n,U}^{-1}(s)\, ds \;,$$

where $F_{n,U}^{-1}(s) \stackrel{\text{def}}{=} \inf\{t \in \mathbb{R},\; F_{n,U}(t) \geq s\}$ is the empirical quantile function, i.e. the generalized inverse of the empirical distribution function $F_{n,U}(t) \stackrel{\text{def}}{=} n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq t\}}$.

Using the definition (18) of $\Lambda_\eta(Y_{0:n}, \alpha)$,

$$n^{-1} \log \Lambda_\eta(Y_{0:n}, \alpha) = \max_{\{\{\delta_k\}_{k=1}^n \in \{0,1\}^n : \sum_{k=1}^n \delta_k \geq \alpha n\}} n^{-1} \sum_{k=1}^{n} \delta_k R_\Delta(ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|) \;.$$

Applying Lemma 13 with $U_{n,k} = R_\Delta(ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|)$ and j such that $n - j + 1 \geq \alpha n$, we obtain

$$\begin{aligned} n^{-1} \log \Lambda_\eta(Y_{0:n}, \alpha) &\leq n^{-1} \sum_{k=n-\lceil \alpha n \rceil+1}^{n} U_{n,k} \\ &= \int_0^1 \mathbb{1}\{u \geq 1 - r_n\} F_n^{-1}(u)\, du \;, \end{aligned} \tag{60}$$

where $r_n = \lceil n\alpha \rceil / n$, $F_n(t) = n^{-1} \sum_{k=1}^{n} \mathbb{1}\{R_\Delta(ab|\varepsilon_{k-1}| + |\zeta_k| + b|\varepsilon_k|) \leq t\}$ and F_n^{-1} its generalized inverse. The function R_Δ defined by (59) is negative and then, $F_n(0) = 1$ which implies that $F_n^{-1}(u) \leq 0$ for all $u \in (0, 1)$. Thus, by Fatou's lemma,

$$\begin{aligned} \limsup_{n\to\infty} \int_0^1 &\mathbb{1}\{u \geq 1 - r_n\} F_n^{-1}(u)\, du \\ &\leq \int_0^1 \limsup_{n\to\infty} \mathbb{1}\{u \geq 1 - r_n\} F_n^{-1}(u)\, du \end{aligned} \tag{61}$$

The following lemma is a generalization of [20, Lemma 21.2].

LEMMA 14. *Let $\{\Psi_n\}_{n\geq 0}$ be a sequence of nondecreasing functions and Ψ a bounded nondecreasing function defined on $\mathbb{R}$ such that for all $x \in \mathbb{R}$, $\lim_{n\to\infty} \Psi_n(x) = \Psi(x)$. Then, Ψ^{-1} has at most a countable number of discontinuity points and at any point u where Ψ^{-1} is continuous,*

$$\lim_{n\to\infty} \Psi_n^{-1}(u) = \Psi^{-1}(u) \;.$$

Let us denote $F(t) = \mathbb{P}[R_\Delta(ab|\varepsilon_0| + |\zeta_1| + b|\varepsilon_1|) \leq t]$ and notice that $F(0) = 1$. Then, combining (60), (61) and using the law of large numbers, Lemma 14 leads to

$$\limsup_{n\to\infty} n^{-1} \log \Lambda_\eta(Y_{0:n}, \alpha) \leq \int_{1-\alpha}^{1} F^{-1}(u)\, du < 0 \;, \quad \mathbb{P}_{\nu_0} - \text{a.s.} \;.$$

This shows that the fourth condition (24) is satisfied and finally, Theorem 5 applies.

□

Let us 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

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

Proof of Proposition 9. We start by giving an exponential inequality for m-dependent variables.

Lemma 15. *Let $\{Z_k\}_{k\geq 0}$ be a sequence of m-dependent stationary random variables. There exists some constant $C > 0$ such that, for all $M \geq 0$,*

$$\mathbb{P}\left(\sum_{k=1}^{n} Z_k \geq Mn\right) \leq C \exp[-n\psi_{Z_1}^*(2Mm)/(2m)] \;.$$

The proof is elementary and left to the reader. It follows by equation (56) that

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

Thanks to (42), by applying Lemma 15, 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 (58), for all $(y, y') \in \mathsf{Y}^2$, $\Psi_{\mathsf{C}(y',\Delta)}(y, y') \geq \varrho_\Delta$, for some $\varrho_\Delta > 0$. Then, by choosing M_3 large enough, we have $r_3(n) = 0$. For $r_4(n)$, we need an exponential inequality for L-statistics based on m-dependent variables.

Lemma 16. *Let $\{U_k\}_{k\geq 0}$ be a sequence of m-dependent stationary negative random variables. For all $\alpha \in (0, 1)$, there exists a real $r > 0$ such that*

$$\lim_{n\to\infty} n^{-1} \log \mathbb{P}\left(\sum_{k=n-\lceil \alpha n\rceil+1}^{n} U_{n,k} \geq -rn\right) < 0 \;.$$

Proof of Lemma 16. For $j \in \{1, \ldots, m\}$, define $I_j = \{j, j+m, j+2m, \ldots\} \cap \{1, \ldots, n\}$ and let $n_j = |I_j|$ the cardinal of I_j. For any $j \in \{1, \ldots, m\}$, the sequence $\{U_k, k \in I_j\}$ is i.i.d.. Denote $\{U_k^{(j)}\}_{1 \leq k \leq n_j}$ the sequence $\{U_k, k \in I_j\}$. Since $U_k < 0$ for all integer k, it then follows that

$$\sum_{k=n-\lceil \alpha n \rceil+1}^{n} U_{n,k} \leq \sum_{j=1}^{m} \sum_{k=(n_j-\lceil \alpha n \rceil+1)\vee 0}^{n_j} U_{n_j,k}^{(j)} ,$$

and

$$\mathbb{P}\left(\sum_{k=n-\lceil \alpha n \rceil+1}^{n} U_{n,k} \geq -rn \right) \leq \sum_{j=1}^{m} \mathbb{P}\left(\sum_{k=(n_j-\lceil \alpha n \rceil+1)\vee 0}^{n_j} U_{n_j,k}^{(j)} \geq -rn/m \right) ,$$

for all $n \geq N$ larger than some integer N. The sequence $\{U_k^{(j)}\}_{1 \leq k \leq n_j}$ is a sequence of i.i.d. negative random variable. Then, using [**?** , Theorem 6.1], we have

$$\lim_{n \to \infty} n_j^{-1} \log \mathbb{P}\left(n_j^{-1} \sum_{k=(n_j-\lceil \alpha n \rceil+1)\vee 0}^{n_j} U_{n_j,k}^{(j)} \geq -\delta \right) < 0 ,$$

for some positive δ and the result follows since $n_j/n = 1/m + o(1)$.

□

Then, by equation (60) and by applying Lemma 16, there exist some constants c_4, $\delta_4 > 0$ such that $r_4(n) \leq c_4 \mathrm{e}^{-\delta_4 n}$. Finally, (43) implies that $r_0 n \leq c_0 \mathrm{e}^{-\delta_0 n}$ for some c_0, $\delta_0 > 0$, so that Theorem 6 applies and provides a geometric rate.

□

6. Proofs of Propositions 11 and 12 .

Proof of Proposition 11. Let us define, for all $\Delta > 0$ and for all integer $k \geq 1$,

$$V_k^{*\Delta} = \log q^- \left[c + d\Delta + \kappa + 2a^* b^* + a^* b^* |\varepsilon_{k-1}^*| + b^* |\varepsilon_k^*| + |\zeta_k^*| \right] . \tag{62}$$

Using the definitions (46), (47) of q^- and ε_Δ^-, Lemma 10 shows that

$$n^{-1} \sum_{k=2}^{n} \varepsilon_\Delta^-(Y_{k-1}^*, Y_k^*) \geq n^{-1} \sum_{k=2}^{n} V_k^{*\Delta} . \tag{63}$$

Thus, to check (21), it suffices to control the asymptotic behavior of the right-hand side of this inequality. We use the following result [12, Chapter 2, Section 6].

LEMMA 17. *Let us denote by* $\{\mathcal{H}_k\}_{k\geq 0}$ *a filtration and consider a sequence* $\{U_k\}_{k\geq 0}$ *of random variables adapted to* $\{\mathcal{H}_k\}_{k\geq 0}$. *Let us assume that there exists a random variable* U *such that* $\mathbb{E}\left(|U|\log_+|U|\right) < \infty$ *and* $\mathbb{P}(|U_k| > x) \leq c\,\mathbb{P}(|U| > x)$ *for all* $x > 0$ *and some* $c > 0$. *Then*

$$\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n}\left[U_k - \mathbb{E}(U_k|\mathcal{H}_{k-1})\right] = 0 \;, \qquad \mathbb{P}-\text{a.s.}$$

Define the filtration $\{\mathcal{F}^*_k\}_{k\geq 0}$ where $\mathcal{F}^*_k = \sigma\left(\{X^*_j\}_{0\leq j\leq k}, \{\zeta^*_j\}_{0\leq j\leq k}, \{\varepsilon^*_j\}_{j\geq 0}\right)$. Using (49), for all $x > 0$, $\mathbb{P}(|V^{*\Delta}_k| > x) \leq \mu^*_+\mathbb{P}(|V^{*\Delta}_+| > x)$, where $V^{*\Delta}_+$ is defined in (51). Hence, we may apply Lemma 17 which yields, for any $\Delta > 0$,

$$\liminf_{n\to\infty} n^{-1}\sum_{k=2}^{n} V^{*\Delta}_k = \liminf_{n\to\infty} n^{-1}\sum_{k=2}^{n}\mathbb{E}\{V^{*\Delta}_k|\mathcal{F}^*_{k-1}\} \;, \qquad \mathbb{P}_\star - \text{a.s.} \tag{64}$$

By (49), since for all $x > 0$, $\log x \geq -\log_- x$, then, by the strong law of large numbers,

$$\liminf_{n\to\infty} n^{-1}\sum_{k=2}^{n}\mathbb{E}\{V^{*\Delta}_k|\mathcal{F}^*_{k-1}\} \geq -\mathbb{E}[H_\Delta(a^*b^*|\varepsilon^*_0| + b^*|\varepsilon^*_1|)] \;, \qquad \mathbb{P}_\star - \text{a.s.} \tag{65}$$

where $H_\Delta(x) = \mu^*_+ \times \int \log_- q^-[c + d\Delta + \kappa + 2a^*b^* + x + |w|]\psi^*(w)\,dw$. By (52), $\mathbb{E}[H_\Delta(a^*b^*|\varepsilon^*_0| + b^*|\varepsilon^*_1|)] < \infty$, it then follows by (63), (64) and (65) that

$$\liminf_{n\to\infty} n^{-1}\sum_{k=2}^{n}\log\varepsilon^-_\Delta(Y^*_{k-1}, Y^*_k) \geq \liminf_{n\to\infty} n^{-1}\sum_{k=2}^{n} V^{*\Delta}_k > -\infty \;, \qquad \mathbb{P}_\star - \text{a.s.}$$

and the condition (21) is satisfied. The proof of assumptions (22) and (23) can be checked as in Proposition 7. Since (H2) is satisfied, for a fixed $\eta > 0$, we choose $\Delta > 0$ such that $\Upsilon_{\mathsf{C}^c(y,\Delta)}(y) \leq \eta\Upsilon_{\mathsf{X}}(y)$. Applying Lemma 13 yields

$$n^{-1}\log\Lambda_\eta(Y^*_{0:n}, \alpha) \leq \int_0^1 \mathbb{1}\{1 - r_n \leq u\}F^{*-1}_n(u)\,du \;, \qquad \mathbb{P}_\star - \text{a.s.}$$

where $r_n = \lceil n\alpha\rceil/n$ and F^{*-1}_n is the generalized inverse of the distribution function:

$$F^*_n(t) = n^{-1}\sum_{k=1}^{n}\mathbb{1}\{R_\Delta[\kappa + 2a^*b^* + a^*b^*|\varepsilon^*_{k-1}| + b^*|\varepsilon^*_k| + |\zeta^*_k|] \leq t\} \;, \tag{66}$$

with R_Δ is defined in (59). For convenience, let us write $G(\varepsilon^*_{k-1}, \varepsilon^*_k, \zeta^*_k) = R_\Delta[\kappa + 2a^*b^* + a^*b^*|\varepsilon^*_{k-1}| + b^*|\varepsilon^*_k| + |\zeta^*_k|]$. Setting

$$H^*_n(t) = n^{-1} \sum_{k=1}^{n} \mathbb{P}\left\{G(\varepsilon^*_{k-1}, \varepsilon^*_k, \zeta^*_k) \leq t | \mathcal{F}^*_{k-1}\right\} \ , \tag{67}$$

it follows from Lemma 17 that, for a fixed $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \{F^*_n(t) - H^*_n(t)\} = 0 \ , \qquad \mathbb{P}_\star - \text{a.s.} \tag{68}$$

The convergence in (68) may be shown to hold uniformly in t:

LEMMA 18. *Let us consider the stochastic functions F^*_n and H^*_n defined by* (66), (67). *Then,*

$$\lim_{n\to\infty} \|F^*_n - H^*_n\|_\infty = 0 \ , \qquad \mathbb{P}_\star - \text{a.s.} \tag{69}$$

PROOF. Let us define

$$J^*_n(t) = n^{-1} \sum_{k=1}^{n} \int \mathbb{P}\left\{G(\varepsilon^*_{k-1}, \varepsilon^*_k, w) \leq t \, | \mathcal{F}^*_{k-1}\right\} \psi^*(w) \, dw \ , \tag{70}$$

$$J^*(t) = \mathbb{E}\left[\int \mathbb{1}\{G(\varepsilon^*_0, \varepsilon^*_1, w) \leq t\} \psi^*(w) \, dw\right] . \tag{71}$$

By the Glivenko-Cantelli Theorem, $\lim_{n\to\infty} \|J^*_n - J^*\|_\infty = 0$, $\mathbb{P}_\star$-a.s. Set $\varepsilon > 0$ and a sequence $-\infty = t_0 \leq t_1 \ldots \leq t_N = \infty$ such that $J^*(t_i^-) - J^*(t_{i-1}) < \varepsilon/\mu^*_+$ for every i. By (49), for all real numbers $t < t'$, $\mathbb{P}_\star$-a.s.

$$H^*_n(t') - H^*_n(t) = n^{-1} \sum_{k=1}^{n} \mathbb{P}(t < G(\varepsilon^*_{k-1}, \varepsilon^*_k, \zeta^*_k) \leq t' | \mathcal{F}^*_{k-1}) \leq \mu^*_+[J^*_n(t') - J^*_n(t)] \ ,$$

and then

$$\limsup_{n\to\infty} |H^*_n(t') - H^*_n(t)| \leq \mu^*_+ |J^*(t') - J^*(t)| \ , \qquad \mathbb{P}_\star - \text{a.s.}$$

For all $t \in \mathbb{R}$, there exists an index i such that $t_{i-1} \leq t < t_i$. Since F^*_n and H^*_n are increasing functions, it follows that

$$F^*_n(t_{i-1}) \leq F^*_n(t) \leq F^*_n(t_i^-) \ , \quad H^*_n(t_{i-1}) \leq H^*_n(t) \leq H^*_n(t_i^-) \ .$$

These inequalities imply

$$\sup_{t\in\mathbb{R}} |F^*_n(t) - H^*_n(t)| \leq \max_{0\leq i\leq N} |F^*_n(t_i^-) - H^*_n(t_i^-)| + \max_{1\leq i\leq N} |H^*_n(t_i^-) - H^*_n(t_{i-1})| \ ,$$

and then

$$\limsup_{n\to\infty} \sup_{t\in\mathbb{R}} |F^*_n(t) - H^*_n(t)| \leq \varepsilon \ , \qquad \mathbb{P}_\star - \text{a.s.}$$

□

By (49), for all $t \in \mathbb{R}$,

$$F_n^*(t) = F_n^*(t) - H_n^*(t) + H_n^*(t) \geq F_n^*(t) - H_n^*(t) + \mu_-^* J_n^*(t) , \qquad \mathbb{P}_\star - \text{a.s.}$$

Hence, using the limit (69), for a given $\delta > 0$, there exists a random integer l such that, for all $n \geq l$ and $t \in \mathbb{R}$,

$$F_n^*(t) \geq \mu_-^* J_n^*(t) - \delta , \qquad \mathbb{P}_\star - \text{a.s.} \tag{72}$$

Let us notice that J_n^* is an increasing function with $\lim_{t \to -\infty} J_n^*(t) = 0$ and $\lim_{t \to +\infty} J_n^*(t) = 1$. Then, we can define its generalized inverse denoted by J_n^{*-1}. By (72), it follows that, for all $u \in [0, (\mu_-^* - \delta) \vee 0]$,

$$F_n^{*-1}(u) \leq J_n^{*-1}[(u+\delta)/\mu_-^*] , \qquad \mathbb{P}_\star - \text{a.s.}$$

By choosing $\delta > 0$ such that $\mu_-^* - \delta > 1 - \alpha$, there exists a random integer $i \geq l$ such that, for all $n \geq i$, we have

$$\int_0^1 \mathbb{1}\{1 - r_n \leq u\} F_n^{*-1}(u)\, du \\ \leq \int_0^1 \mathbb{1}\{1 - r_n \leq u \leq \mu_-^* - \delta\} J_n^{*-1}[(u+\delta)/\mu_-^*]\, du , \qquad \mathbb{P}_\star - \text{a.s.}$$

By Fatou's lemma,

$$\limsup_{n\to\infty} \int_0^1 \mathbb{1}\{1 - r_n \leq u\} F_n^{*-1}(u)\, du \\ \leq \int_0^1 \limsup_{n\to\infty} \mathbb{1}\{1 - r_n \leq u \leq \mu_-^* - \delta\} J_n^{*-1}[(u+\delta)/\mu_-^*]\, du , \quad \mathbb{P}_\star - \text{a.s.}$$

It follows by Lemma 14 that

$$\limsup_{n\to\infty} n^{-1} \log \Lambda_\eta(Y_{0:n}^*, \alpha) \leq \int_{1-\alpha}^{\mu_-^* - \delta} J^{*-1}[(u+\delta)/\mu_-^*]\, du < 0 , \qquad \mathbb{P}_\star - \text{a.s.}$$

Thus, condition (24) is satisfied and Theorem 5 applies.

□

Proof of Proposition 12. It follows, by definition of r_1, Lemma 10 and (O3), that

$$r_1(n) = \mathbb{P}_\star\left(n^{-1} \sum_{k=2}^n \log q^-[c + d\Delta + D(Y_{k-1}^*, Y_k^*)] \leq -M_1 n\right) \leq \\ \mathbb{P}_\star\left(n^{-1} \sum_{k=2}^n \log q^- \left[c_0 + a^* b^* |\varepsilon_{k-1}^*| + b^* |\varepsilon_k^*| + g_+^*(A_k^*)\right] \leq -M_1 n\right) .$$

with $c_0 = c + d\Delta + \kappa + 2a^*b^*$. Then, by (O4) and applying Lemma 15, 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 $U_k^+ = R_\Delta[\kappa + 2a^*b^* + a^*b^*|\varepsilon_{k-1}^*| + b^*|\varepsilon_k^*| + g_+^*(A_k^*)]$, for all integer $k \geq 1$. By definition of Λ_η,

$$n^{-1} \log \Lambda_\eta(Y_{0:n}^*, \alpha) \leq n^{-1} \sum_{k=n-\lceil \alpha n \rceil + 1}^{n} U_{n,k}^+ \,. \tag{73}$$

By applying Lemma 16, there exist some constants c_4, $\delta_4 > 0$ such that $r_4(n) \leq c_4 \mathrm{e}^{-\delta_4 n}$. Finally, (53) implies that $r_0 n \leq c_0 \mathrm{e}^{-\delta_0 n}$ for some c_0, $\delta_0 > 0$, and 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. and Lipster, R. (2004). Stability of nonlinear filters in nonmixing case. *Ann. Appl. Probab.*, 14(4):2038–2056.

[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] Del Moral, P. and Guionnet, A. (1998). Large deviations for interacting particle systems: applications to non-linear filtering. *Stoch. Proc. App.*, 78:69–95.

[8] Del Moral, P., Ledoux, M., and Miclo, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields*, 126(3):395–420.

[9] 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:–.

[10] 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.

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

[12] Hall, P. and Heyde, C. C. (1981). Rates of convergence in the martingale central limit theorem. *Ann. Probab.*, 9(3):395–404.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

[19] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[20] Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

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

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

[23] 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.

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