



**HAL**  
open science

# Consistency of the maximum likelihood estimator in hidden Markov models with trends

Luc Lehéricy, Augustin Touron

► **To cite this version:**

Luc Lehéricy, Augustin Touron. Consistency of the maximum likelihood estimator in hidden Markov models with trends. 2021. hal-03140231

**HAL Id: hal-03140231**

**<https://hal.science/hal-03140231>**

Preprint submitted on 15 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Consistency of the maximum likelihood estimator in hidden Markov models with trends

L. Lehéricy\*, A. Touron\*<sup>†</sup>

## Abstract

A hidden Markov model with trends is a hidden Markov model whose emission distributions are translated by a trend that depends on the current hidden state and on the current time. Unlike standard hidden Markov models, such processes are not homogeneous and cannot be made homogeneous by a simple de-trending step. We show that when the trends are polynomial, the maximum likelihood estimator is able to recover the trends together with the other parameters and is strongly consistent. More precisely, the supremum norm of the difference between the true trends and the estimated ones tends to zero. The convergence of the maximum likelihood estimator is illustrated on simulated data.

**Keywords:** hidden Markov model, maximum likelihood, inhomogeneous process, consistency

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Model and assumptions</b>	<b>5</b>
<b>3</b>	<b>Consistency of the MLE</b>	<b>6</b>
<b>4</b>	<b>Simulations</b>	<b>12</b>
4.1	First experiment: rate of convergence . . . . .	12
4.2	Second experiment: the trends have not diverged yet . . . . .	14
<b>A</b>	<b>Block approximation</b>	<b>15</b>
A.1	Step 1: introduction of the trend block $B_t$ in the log-likelihood . . . . .	16
A.2	Step 2: conditioning on the blocks $B_1^{t-1}$ . . . . .	17
A.3	Application: existence and finiteness of the relative entropy rate . . . . .	20
A.4	Proofs . . . . .	20

---

\*Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

<sup>†</sup>EDF R&D

<b>B</b>	<b>Localization of the MLE</b>	<b>24</b>
B.1	Preliminary: compactness results . . . . .	24
B.2	The MLE is in $\mathcal{T}(\alpha, n, D)$ . . . . .	27
B.3	The MLE is in $\mathcal{U}(\beta, n, B)$ . . . . .	30
B.4	Proofs . . . . .	35
<b>C</b>	<b>Integrated log-likelihood</b>	<b>40</b>
C.1	Convergence of the log-likelihood to the integrated log-likelihood . . . . .	40
C.2	Maximizers of the integrated log-likelihood and identifiability . . . . .	44
C.3	Uniform convergence of the homogeneous log-likelihood . . . . .	45
<b>D</b>	<b>Miscellaneous proofs</b>	<b>47</b>
D.1	Proof of Lemma 24 . . . . .	47
D.2	Proof of Lemma 25 . . . . .	48

# 1 Introduction

A hidden Markov model (shortened HMM in the following) is a joint process  $(X_t, Y_t)_{t \geq 1}$  where  $(X_t)_{t \geq 1}$  is an (unobserved) Markov chain and conditionally on  $(X_t)_{t \geq 1}$ , the observations  $(Y_t)_{t \geq 1}$  are independent and the distribution of  $Y_s$  (the *emission distribution*) depends only on  $X_s$ . In most generalizations of HMM, the joint process  $(X_t, Y_t)_{t \geq 1}$  is still a Markov chain. We say that the process is *inhomogeneous* when this chain is inhomogeneous, that is when the distribution of  $(X_t, Y_t)$  conditionally to  $(X_{t-1}, Y_{t-1})$  depends on  $t$ . In particular, in inhomogeneous HMM, the transition matrix and emission distributions may vary over time.

As far as we know, few inhomogeneous generalizations of HMM have been studied theoretically. Let us review them before stating our contribution. They can be divided in four categories.

[Diehn et al. \(2019\)](#) focus on the case where a rapidly fading phenomenon affects the distribution of the observations. Their model is a trivariate process  $(X_t, Y_t, Z_t)_{t \geq 1}$  where only  $(Z_t)_{t \geq 1}$  is observed, such that  $(X_t, Y_t)_{t \geq 1}$  is an homogeneous HMM and  $(X_t, Z_t)_{t \geq 1}$  is an inhomogeneous HMM. Their key assumption is that the distance between  $Z_t$  and  $Y_t$  tends to zero fast enough when  $t$  tends to infinity. In this sense, the process  $(Z_t)_{t \geq 1}$  is a perturbation of the process  $(Y_t)_{t \geq 1}$  by a rapidly fading inhomogeneous noise. The main theoretical result of their article is that the estimators based on maximizing either the true likelihood  $\theta \mapsto p_{Z_1^n | \theta}(Z_1^n)$  or a quasi-likelihood  $\theta \mapsto p_{Y_1^n | \theta}(Z_1^n)$  are consistent (here  $p_{Y_1^n | \theta}$  is the density of the vector  $(Y_1, \dots, Y_n)$  under the parameter  $\theta$ ). This can be seen as a proof that the maximum likelihood estimator for homogeneous HMM is robust to a temporary perturbation of the data.

A second generalization has been considered by [Touron \(2018\)](#) to handle periodic phenomenon. The author introduces periodic hidden Markov models, where the transition matrix and the emission densities vary periodically over time. He shows that such models are identifiable under general assumptions and that the maximum likelihood estimator is consistent. The consistency proof relies on a transformation of the process into an homogeneous HMM. Models with periodic parameters are of particular interest

in the study of meteorological phenomena. For instance, [Ailliot and Monbet \(2012\)](#) use Markov-Switching Autoregressive (MS-AR) models with periodic emission distributions and transition matrices in order to account for both the daily and yearly seasonalities of wind speed.

Inhomogeneity can also be introduced in the process by considering that the transition probabilities of the hidden Markov chain depend on an exogeneous variable. [Bellone et al. \(2000\)](#) and [Hughes et al. \(1999\)](#) chose this approach to model precipitations, taking a vector of atmospheric variables as the exogeneous variable. However, no theoretical results on the consistency of the MLE are given in these papers.

Finally, we may assume that the transition probabilities or the emission densities of the HMM change over time in a way not encompassed by the previous cases. For example, this allows to model phenomena with slow or long-term evolution, such as economic or meteorological data recordings spanning over several decades, as in ([Touron, 2019, Part II](#)), where the author is interested in generating realistic temperature recordings over the last half century. To account for the global warming, he uses HMM with trends similar to the ones studied here. We could not find any theoretical result on this type of generalization outside of our paper.

Note that some models that are presented as time-inhomogeneous in the literature are actually homogeneous according to our definition. This is the case of the Markov-switching models considered since the work of [Hamilton \(1989\)](#) and more recently by [Pouzo et al. \(2016\)](#) and [Ailliot and Pene \(2015\)](#) for instance. These models are a generalization of HMM where the hidden state  $X_t$  depends both of the previous hidden state  $X_{t-1}$  and on previous observations, let's say  $Y_{t-1}$  for an order one model, and where the observation  $Y_t$  depends both on the corresponding hidden state  $X_t$  and on previous observations. The authors state that this model is “non-homogeneous” because the transition kernel of the hidden Markov chain depends on previous observations. Our motivation for not calling these models inhomogeneous is that the joint process  $(X_t, Y_t)_{t \geq 1}$  is a homogeneous Markov chain. As such, their proofs are based on the same approach as for HMM and other homogeneous generalizations such as the autoregressive models with Markov regime of [Douc et al. \(2004\)](#).

In this paper, we introduce a new inhomogeneous generalization: hidden Markov models with trends. A HMM with trends is a trivariate process  $(X_t, Y_t, Z_t)_{t \geq 1}$  taking values in  $\mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d$  for some finite set  $\mathcal{X}$  and some integer  $d$  (which we assume to be 1 for the sake of simplicity) where only the process  $(Y_t)_{t \geq 1}$  is observed, such that  $(X_t, Z_t)_{t \geq 1}$  is a homogeneous HMM and that there exists a vector of functions  $(T_x)_{x \in \mathcal{X}}$  from  $\mathbb{N}^*$  to  $\mathbb{R}^d$ , the *trends*, such that

$$Y_t = T_{X_t}(t) + Z_t.$$

Here, the trends are polynomial functions of time. As a consequence, they may diverge.

Adding trends to a HMM allows to deal with non periodic and non vanishing inhomogeneities. However, this makes existing approaches such as [Douc et al. \(2004\)](#) inapplicable since they rely heavily on the homogeneity of the process. Our main contribution is a new method of proof that shows that the maximum likelihood estimator recovers the parameter of the homogeneous HMM  $(X_t, Z_t)_{t \geq 1}$  as well as the trends with respect to the supremum norm, see Theorem 1.

Let us give an overview of the proof of this theorem.

First, we introduce the notion of “blocks”. This notion relies solely on the true parameters. A block is a set of hidden states whose trends are equal up to translation. In particular, since the trends are polynomials, if two trends are not in the same block, they will eventually diverge from one another.

For each time  $t$ , let  $B_t$  be the block of  $X_t$ . The block variables  $(B_t)_{t \geq 1}$  are unobserved, yet the log-likelihood of the process  $(Y_t)_{t \geq 1}$  is asymptotically the same as the log-likelihood of the process  $(Y_t, B_t)_{t \geq 1}$ , see Theorem 3: since trends of different blocks diverge, it eventually becomes apparent which block a given observation comes from.

Given a constant  $M > 0$ , we say that a function belongs to the tube of a block at time  $n$  if its distance to the trends of this block remains bounded by  $M$  during the first  $n$  time steps. In the following, the bound  $M$  on the distance must not depend on  $n$ .

The second step of the proof shows that the trends of the maximum likelihood estimator end up in close proximity to the true trends: Theorem 5 shows that eventually, the trends of the maximum estimator are all in the tube of a block and each tube contains at least one trend of the maximum likelihood estimator.

Let us give an idea of the proof of Theorem 5. When a tube contains no estimated trend, each observation from its block is far from the estimated trends, which inflicts a heavy loss on the likelihood. On the other hand, if an estimated trend is in no tube, then eventually it is far from all observations. We assume that the transition matrix is lower bounded, so each state should be visited a non-zero proportion of the time. However, when a trend is far from all observations, the posterior proportion of observations generated by the corresponding state tends to zero: each time this superfluous state isn't seen worsens the likelihood a little more, to the point that eventually, removing the trend improves the likelihood of the estimator. Thus, asymptotically, each estimated trend is in a tube and no tube is empty.

The third step is to de-trend the observations by subtracting a representative  $\mathbb{T}_{B_t}^*$  of the trends of the block:

$$Z'_t = Y_t - \mathbb{T}_{B_t}^*(t).$$

The process  $(X_t, (Z'_t, B_t))_{t \geq 1}$  obtained this way is also a HMM with trends, yet it is still not homogeneous under the parameter  $\theta$  since the residual trends  $T_x^\theta - \mathbb{T}_b^*$  of  $(Z'_t)_{t \geq 1}$  (where  $b$  is the block that contains  $T_x^\theta$ ), while bounded, are not constant. Fortunately, as the number of observations grow, these residual trends become “flatter”, because they are polynomials with bounded degree that are bounded on  $[0, n]$ . Thus, during small time intervals, the de-trended process behaves like a homogeneous HMM under the parameter  $\theta$ , which allows to explicitly compute the limit of the log-likelihood, see Theorem 10. The identifiability of the model and the consistency of the maximum likelihood estimator follow from the properties of this limit.

Simulations based on two parametric models illustrate the consistency of the maximum likelihood estimator. Moreover, an empirical study of the rates of convergence suggests that it converges with parametric rates.

Section 2 introduces the model and assumptions used in the following. Section 3 presents our main result: the consistency of the maximum likelihood estimator (Theorem 1), as well as an outline of its proof. The technical proofs are postponed to the appendices. Section 4 contains the numerical study. Finally, Appendices A to D contain the proofs of the different steps of the consistency proof.

**Notation** For each positive integer  $K$ ,  $[K]$  denotes the set  $\{1, \dots, K\}$ . For  $a \leq b$  integers, write  $Y_a^b$  instead of  $(Y_a, \dots, Y_b)$ .

## 2 Model and assumptions

Let  $K^*$  be a positive integer. Let  $\gamma^* = (\gamma_x^*)_{x \in [K^*]}$  be a vector of probability densities on  $\mathbb{R}$  with respect to the Lebesgue measure. Let  $(X_t)_{t \geq 1}$  be a Markov chain on  $[K^*]$  with transition matrix  $Q^*$  and initial distribution  $\pi^*$ . For all  $x \in [K^*]$ , let  $(Z_t^x)_{t \geq 1}$  be a sequence of i.i.d. random variables in  $\mathbb{R}$  such that these sequences are mutually independent and independent on  $(X_t)_{t \geq 1}$  and such that for all  $x \in [K^*]$ ,  $Z_1^x$  has density  $\gamma_x^*$  with respect to the Lebesgue measure. Let  $Z_t^{\max} = \max_{x \in [K^*]} |Z_t^x|$  and  $Z_t = Z_t^{X_t}$ . Finally, let  $T^* = (T_x^*)_{x \in [K^*]}$  be a family of functions  $\mathbb{R}_+ \rightarrow \mathbb{R}$  and let  $Y_t = Z_t + T_{X_t}^*(t)$  for all integers  $t \geq 1$ . The  $(T_x^*)_{x \in [K^*]}$  are called *trends*.

The process  $(X_t, Z_t)_{t \geq 1}$  is a homogeneous hidden Markov model with parameter  $(K^*, \pi^*, Q^*, \gamma^*)$  and  $(X_t, Y_t)_{t \geq 1}$  is a hidden Markov model with trends with parameter  $(K^*, \pi^*, Q^*, \gamma^*, T^*)$ .

**Remark.** The random variables  $Z_t^{\max}$  are i.i.d. and independent of  $(X_t)_{t \geq 1}$ . They allow to bound  $Z_t$  uniformly for all possible values of  $X_t$ .

Consider a sample  $(Y_1, \dots, Y_n)$  generated by a hidden Markov model with trends  $(X_t, Y_t)_{t \geq 1}$  with parameter  $\theta^* := (K^*, \pi^*, Q^*, \gamma^*, T^*)$ , which we call the *true parameter*. The goal is to recover this parameter. In the following, we write  $\mathbb{P}^*$  the distribution of the process  $(X_t, Y_t)_{t \geq 1}$  and  $\mathbb{E}^*$  the corresponding expectation.

Let  $\sigma_- \in (0, 1)$ .

**(Aerg)** A stochastic matrix  $Q$  satisfies **(Aerg)** when all its coefficients are lower bounded by  $\sigma_-$ :

$$\forall x, x', \quad Q(x, x') \geq \sigma_-.$$

For each  $K \in \mathbb{N}^*$ , let  $\Sigma_K^{\sigma_-}$  be the set of stochastic matrices of size  $K$  which satisfy **(Aerg)** and  $\Delta_K$  be the set of probability vectors of size  $K$ . Let  $\Gamma$  be a set of density functions on  $\mathbb{R}$  and for each  $d \in \mathbb{N}$ , write  $\mathbb{R}_d[X]$  the set of polynomials whose degree is at most  $d$ .

Let  $K$  and  $d$  be positive integers. The model considered in this paper is

$$\Theta = \bigcup_{K'=1}^K \left( \{[K']\} \times \Delta_{K'} \times \Sigma_{K'}^{\sigma_-} \times \Gamma^{K'} \times (\mathbb{R}_d[X])^{K'} \right)$$

Each  $\theta = (K^\theta, \pi^\theta, Q^\theta, \gamma^\theta, T^\theta) \in \Theta$  is a parameter of a HMM with trends. Write  $\mathbb{P}^\theta$  the distribution under the parameter  $\theta$ . Assume that the true parameter belong to the model:  $\theta^* \in \Theta$ .

Endow  $\Gamma$  with the pointwise convergence topology (i.e.  $g_n \xrightarrow[n \rightarrow \infty]{} g$  if and only if for all  $z \in \mathbb{R}$ ,  $g_n(z) \xrightarrow[n \rightarrow \infty]{} g(z)$ ). Assume that  $\Gamma$  is compact.

We will use the following assumptions.

**(Amax)** *Envelope function.* There exists a nonincreasing function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $g \xrightarrow{+\infty} 0$  and

$$\forall \gamma \in \Gamma \quad \forall z \in \mathbb{R} \quad \gamma(z) \leq g(|z|).$$

**(Amin)** *Lower bound function.* There exists a nonincreasing function  $m : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that

$$\forall \gamma \in \Gamma \quad \forall z \in \mathbb{R} \quad \gamma(z) \geq m(|z|) > 0.$$

**(Aint)** *Integrability of the lower bound function.* The function  $m$  defined in **(Amin)** satisfies

$$\forall M > 0 \quad \mathbb{E}^* |\log m(M + Z_t^{\max})| < \infty.$$

**(Acentering)** *Centering of the emission densities.* 0 is a median of the emission densities, that is:

$$\forall \gamma \in \Gamma \quad \int_{z \leq 0} \gamma(z) dz = \frac{1}{2}.$$

**(Aid)** *Identifiability.*  $Q^*$  is invertible and there exists  $t \in \mathbb{N}$  such that the densities  $(\gamma_x^*(\cdot - T_x^*(t)))_{x \in [K^*]}$  are pairwise distinct.

Equivalently,  $Q^*$  is invertible and the couples  $(\gamma_x^*(\cdot - \Delta(x)), \mathbf{b}^*(x))_{x \in [K^*]}$  are pairwise distinct, where the functions  $\mathbf{b}^*$  and  $\Delta$  are defined in Definition 2.

**(Areg)** *Regularity of the emission densities.* There exists a modulus of continuity  $\omega$  (that is a nondecreasing function  $\mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  that is continuous at 0 and such that  $\omega(0) = 0$ ) and a nondecreasing function  $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that

$$\forall (z, \eta) \in \mathbb{R}^2 \quad \forall \gamma \in \Gamma \quad \left| \log \frac{\gamma(z + \eta)}{\gamma(z)} \right| \leq L(|z|) \omega(|\eta|)$$

and such that

$$\forall M > 0, \quad \mathbb{E}^*[L(M + Z_1^{\max})] < \infty.$$

As an example, consider  $\Gamma$  as the set of densities of the distributions  $\mathcal{N}(0, s^2)$  where  $s \in [s_-, s_+]$  for  $0 < s_- \leq s_+ < \infty$ . Then **(Amax)**, **(Amin)** and **(Acentering)** are immediate, **(Aint)** is straightforward since  $|\log m|$  can be chosen as growing at most quadratically to infinity, the part of **(Aid)** on the densities holds when all states with equal trends have different variances, and the functions  $L$  and  $\omega$  in **(Areg)** can be computed explicitly.

### 3 Consistency of the MLE

Let  $n \in \mathbb{N}^*$ . The maximum likelihood estimator  $\hat{\theta}_n = (K^{\hat{\theta}_n}, \pi^{\hat{\theta}_n}, Q^{\hat{\theta}_n}, \gamma^{\hat{\theta}_n}, T^{\hat{\theta}_n})$  is an element of

$$\arg \max_{\theta \in \Theta} \frac{1}{n} \log p_{Y_1^n}^\theta(Y_1^n).$$

where  $p_{Y_1^n}^\theta$  is the density of the distribution of  $Y_1^n$  with respect to the Lebesgue measure under the parameter  $\theta$ .

The main result of this paper is the following theorem.

**Theorem 1.** Assume that  $K^*$  is known and that **(Amax)**, **(Amin)**, **(Aint)**, **(Acentering)**, **(Aid)** and **(Areg)** hold. Then almost surely, there exists a sequence of (random) permutations  $\tau_n$  of  $[K^*]$  such that

$$\begin{cases} \forall x, x' \in [K^*], & Q^{\hat{\theta}_n}(\tau_n(x), \tau_n(x')) \xrightarrow[n \rightarrow \infty]{} Q^*(x, x'), \\ \forall z \in \mathbb{R}, \quad \forall x \in [K^*], & \gamma_{\tau_n(x)}^{\hat{\theta}_n}(z) \xrightarrow[n \rightarrow \infty]{} \gamma_x^*(z), \\ \forall x \in [K^*], & \|T_{\tau_n(x)}^{\hat{\theta}_n} - T_x^*\|_{\infty, [0, n]} \xrightarrow[n \rightarrow \infty]{} 0. \end{cases}$$

We say that the maximum likelihood estimator converges *up to permutation of the hidden states*. Without convention on the ordering of the hidden states, it is not possible to get rid of this permutation.

Note that trends converge in supremum norm. This is considerably stronger than assuming that the coefficients of the polynomial converge. It is also intrinsic in the sense that it shows the convergence of the trends when seen as continuous functions and not just as polynomials. Therefore, it could be extended to other types of continuous trends.

Since the trends are polynomial, two trends are either equal up to translation (we say that they are *in the same block*) or diverge from one another. The first part of the proof, resulting in Theorem 3, makes use of this fact to show that the block from which an observation comes can be assumed observed once the number of observations is large enough.

**Definition 2** (Blocks of trends). Let  $\mathcal{R}^*$  be the equivalence relation defined on  $[K^*]$  by  $x \mathcal{R}^* x'$  if and only if  $T_x^* - T_{x'}^*$  is constant. Let

$$\mathcal{B}^* := [K^*]/\mathcal{R}^*$$

be the set of “blocks” of true trends.

Let us denote by  $\mathbf{b}^* : [K^*] \rightarrow \mathcal{B}^*$  the quotient mapping and let

$$B_t := \mathbf{b}^*(X_t)$$

be the block from which the observation  $Y_t$  is generated. For each  $b \in \mathcal{B}^*$ , write

$$\mathbb{T}_b^* = T_{\min\{x|x \in b\}}^*$$

the reference trend of block  $b$  and let

$$\Delta : x^* \in [K^*] \mapsto T_{x^*}^*(1) - \mathbb{T}_{\mathbf{b}^*(x^*)}^*(1)$$

be the function which maps the index of a trend to the difference between the corresponding trend and its reference trend.

For all  $\theta \in \Theta$ , write

$$\begin{aligned} \ell_n(\theta) &:= \log p_{Y_1^n}^\theta(Y_1^n) \\ \ell_n^{(Y, B)}(\theta) &:= \log p_{(Y, B)_1^n}^\theta((Y, B)_1^n) \end{aligned}$$

the log-likelihoods of  $\theta$  with respect to the processes  $(Y_t)_{t \geq 1}$  and  $(Y_t, B_t)_{t \geq 1}$ .

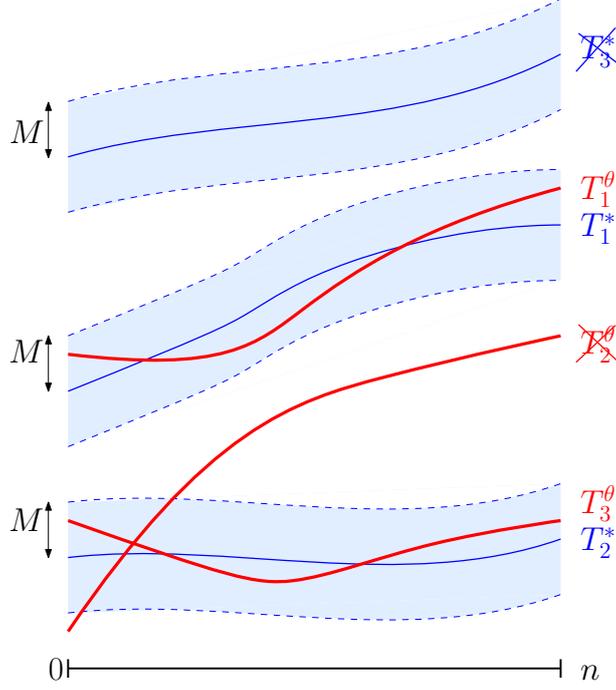


Figure 1: Illustration of the tubes of trends and of a  $\theta$  that is not in  $\Theta_n^{\text{OK}}(M)$ .

Define the tube of size  $M$  of a block at time  $n$  as the set of functions  $\mathbb{R}_+ \rightarrow \mathbb{R}$  which are at a distance smaller than  $M$  of the reference trend of this block on  $[0, n]$  (and as a consequence is at a distance smaller than  $M$  during the first  $n$  time steps). Since the observations  $(Y_t)_{t \geq 1}$  are clustered around the true trends (see Figure 3 for instance), the estimated trends and the tubes will eventually match. This behaviour is formalized in Theorem 5 below by proving that there exists a large enough  $M$  such that for  $n$  large enough,  $\hat{\theta}_n \in \Theta_n^{\text{OK}}(M)$ , where for all  $M > 0$  and  $n \in \mathbb{N}$ ,  $\Theta_n^{\text{OK}}(M)$  is the subset of  $\Theta$  defined by

$$\forall \theta \in \Theta_n^{\text{OK}}(M), \quad \forall x^* \in [K^*], \quad \exists x \in [K^\theta], \quad \|T_{x^*}^* - T_x^\theta\|_{\infty, [0, n]} \leq M \quad (1)$$

$$\text{and } \forall x \in [K^\theta], \quad \exists x^* \in [K^*], \quad \|T_{x^*}^* - T_x^\theta\|_{\infty, [0, n]} \leq M. \quad (2)$$

If  $\theta \in \Theta_n^{\text{OK}}(M)$ , each tube of size  $M + \|\Delta\|_\infty$  contains at least one trend  $T_x^\theta$ ,  $x \in [K^\theta]$ , and each trend  $T_x^\theta$ ,  $x \in [K^\theta]$  is in a tube of size  $M + \|\Delta\|_\infty$ , see Figure 1 for an illustration.

The following theorem states that when  $\theta \in \Theta_n^{\text{OK}}(M)$  (that is, informally, when the trends are loosely identified) and provided that the number of observations is large enough, the log-likelihood associated to the observed process  $(Y_t)_{t \geq 1}$  can be approximated by the log-likelihood of the process  $(Y_t, B_t)_{t \geq 1}$  where we know which block of trend each observation comes from.

Let us first specify what we mean by the log-likelihood of the process  $(Y_t, B_t)_{t \geq 1}$ . Since the true trends are either in the same block or diverge from one another, the tubes eventually have empty intersection. Once this is the case, given  $\theta \in \Theta_n^{\text{OK}}(M)$  and  $x \in [K^\theta]$ , we may define  $\mathbf{b}^\theta(x)$  as the block whose tube contains the trend  $T_x^\theta$ . The log-likelihood  $\ell_n^{(Y, B)}(\theta)$  is the log-likelihood of a inhomogeneous HMM whose emission densities at time  $t$  are  $(y, b) \in \mathbb{R} \times \mathcal{B}^* \mapsto \gamma_x^\theta(y - T_x^\theta(t)) \mathbf{1}(b = \mathbf{b}^\theta(x))$  for each  $x \in [K^\theta]$ .

**Theorem 3** (Adding block information). *Assume **(Amax)**, **(Amin)** and **(Aint)**. Then for all  $M > 0$ , almost surely,*

$$\sup_{\theta \in \Theta_n^{OK}(M)} \left| \frac{1}{n} \ell_n(\theta) - \frac{1}{n} \ell_n^{(Y,B)}(\theta) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Theorem 3 is proved in Appendix A, together with the following corollary.

**Corollary 4.** *Assume **(Amax)**, **(Amin)** and **(Aint)**. Then there exists a finite  $\ell(\theta^*)$  such that*

$$\frac{1}{n} \ell_n(\theta^*) \xrightarrow{n \rightarrow \infty} \ell(\theta^*).$$

Corollary 4 is a key argument in the proof of the following theorem, proved in Appendix B.

**Theorem 5** (Localization of the maximum likelihood estimator). *Assume **(Amax)**, **(Amin)** and **(Aint)**. Then there exists  $M > 0$  such that almost surely, there exists  $n_{loc} \in \mathbb{N}$  such that for all  $n \geq n_{loc}$ ,*

$$\hat{\theta}_n \in \Theta_n^{OK}(M).$$

Let us give an intuition on the last steps of the proof. We know by Theorem 3 that  $\frac{1}{n} \ell_n(\theta) \simeq \frac{1}{n} \ell_n^{(Y,B)}(\theta)$ . That means we may de-trend the observations by letting  $Z'_t = Y_t - \mathbb{T}_{B_t}^*(t)$ , and  $\frac{1}{n} \ell_n(\theta) \simeq \frac{1}{n} \ell_n^{(Y,B)}(\theta) = \frac{1}{n} \ell_n^{(Z',B)}(\theta)$ . Under a parameter  $\theta \in \Theta_n^{OK}(M)$ , this new HMM with trends  $(X_t, (Z'_t, B_t))_{t \geq 1}$  is “locally homogeneous”. Let us explain what this means.

Rescale the trends of  $Z'_t$  so that the time index lives in  $[0, 1]$  by defining

$$D_x^{\theta,n} : u \in [0, 1] \mapsto T_x^\theta(nu) - \mathbb{T}_{\mathbf{b}^\theta(x)}^*(nu). \quad (3)$$

Since the functions  $D_x^{\theta,n}$  are bounded and polynomial, they live in a relatively compact subset of the continuous functions on  $[0, 1]$  for all  $n, x$  and  $\theta \in \Theta_n^{OK}(M)$ . In particular, we control how fast they vary, and thus how well they can be approximated by piecewise constant functions (see Figure 2), which correspond to a HMM with a succession of homogeneous regimes.

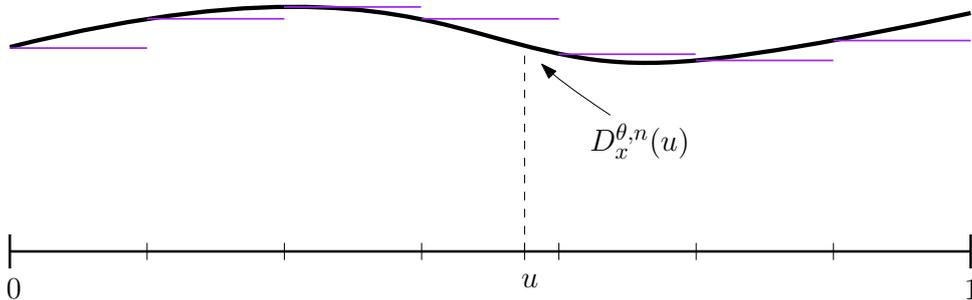


Figure 2: Approximation of one trend of the de-trended HMM (black) by the corresponding trend in a succession of homogeneous HMM (purple).

Let  $\frac{1}{n} \ell_n^{(Y,B)} [N] (\theta)$  be the log-likelihood corresponding to such an approximation when the sample (i.e. the time interval  $[0, n]$  before rescaling) is cut into  $N$  pieces, and let

$\ell^{\text{hom}}(\theta, n, u)$  be the limit of the normalized log-likelihood of a homogeneous HMM with parameters given by  $\theta$  frozen at time  $nu$ , that is of a homogeneous HMM  $(X_t, (Z'_t, B_t))_{t \geq 1}$  with transition matrix  $Q^\theta$  and emission densities  $(z', b) \mapsto \gamma_x^\theta(z' - D_x^{\theta, n}(u)) \mathbf{1}(b = \mathbf{b}^\theta(x))$ ,  $x \in [K^\theta]$ . The previous paragraph means that the following sequence of approximations holds:

$$\begin{aligned} \frac{1}{n} \ell_n(\theta) &\simeq \frac{1}{n} \ell_n^{(Y, B)}(\theta) \\ &\simeq \frac{1}{n} \ell_n^{(Y, B)}[N](\theta) \\ &\simeq \frac{1}{N} \sum_{i=1}^N \ell^{\text{hom}}\left(\theta, n, \frac{i-1}{N}\right) \quad \text{when } n \rightarrow \infty \\ &\simeq \int_{[0,1]} \ell^{\text{hom}}(\theta, n, u) du \quad \text{by Riemann integration.} \end{aligned}$$

To conclude, we use that the latter quantity, the *integrated log-likelihood*, is maximized only at the true parameters, as shown in Proposition 11.

Let us now state these intermediate results more rigorously. For all  $K' \in \mathbb{N}^*$ , for all  $K'$ -uple  $\gamma = (\gamma_x)_{x \in [K']}$  of measurable functions and for all  $\mathbf{D} = (D_x)_{x \in [K']} \in \mathbb{R}^{K'}$ , let

$$\tau(\gamma, \mathbf{D}) := (z' \mapsto \gamma_x(z' - D_x))_{x \in [K']}$$

be the vector of functions  $\gamma$  translated by the vector  $\mathbf{D}$ .

**Definition 6** (Log-likelihood of homogeneous HMM). *Let  $K' \in \mathbb{N}^*$  be a positive integer,  $\pi$  be a probability measure on  $[K']$ ,  $Q$  be a transition matrix of size  $K'$ ,  $\gamma$  be a vector of  $K'$  emission densities on  $\mathbb{R}$  and  $\mathbf{b}$  be a function  $[K'] \rightarrow \mathcal{B}^*$ . Let  $(X_t, (\tilde{Z}_t, \tilde{B}_t))_{t \geq 1}$  be a homogeneous HMM taking values in  $[K'] \times (\mathbb{R} \times \mathcal{B}^*)$  with parameter  $(\pi, Q, (\gamma_x \otimes \mathbf{1}_{\mathbf{b}(x)})_{x \in [K']})$ .*

*Denote by  $\frac{1}{n} \ell_n^{\text{hom}}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{z}, \tilde{b})_1^n\}$  (resp.  $\ell^{\text{hom}}(Q, \gamma, \mathbf{b})$ ) its normalized log-likelihood associated with the observations  $(\tilde{z}, \tilde{b})_1^n$  (resp. the limit of the log-likelihood, if it exists), that is*

$$\frac{1}{n} \ell_n^{\text{hom}}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{z}, \tilde{b})_1^n\} = \frac{1}{n} \log \sum_{x_1^n \in [K']^n} \pi(x_1) Q(x_1, x_2) \dots Q(x_{n-1}, x_n) \prod_{t=1}^n \gamma_{x_t}(\tilde{z}_t) \mathbf{1}_{\mathbf{b}(x_t) = \tilde{b}_t}$$

and

$$\ell^{\text{hom}}(Q, \gamma, \mathbf{b}) = \lim_{n \rightarrow \infty} \frac{1}{n} \ell_n^{\text{hom}}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{Z}, \tilde{B})_1^n\}. \quad (4)$$

The following Lemma ensures the existence of the limit of the normalized log-likelihood in Definition 6. It relies on results from Douc et al. (2004) on homogeneous HMM.

**Lemma 7.** *Assume **(Amax)**, **(Amin)** and **(Aint)**. Let  $K' \in \mathbb{N}^*$ . Then almost surely, for all  $Q \in \Sigma_{K'}^{\sigma^-}$ ,  $\gamma \in \Gamma^{K'}$ ,  $\mathbf{D} \in \mathbb{R}^{K'}$  and  $\mathbf{b} : [K'] \rightarrow \mathcal{B}^*$ , the quantity*

$$\ell^{\text{hom}}(Q, \tau(\gamma, \mathbf{D}), \mathbf{b})$$

*from Equation (4) exists, does not depend on the choice of the initial measure  $\pi$  and is finite when  $(\tilde{Z}_t, \tilde{B}_t)_t = (Y_t - \mathbb{T}_{B_t}^*(t), B_t)_t$ .*

**Definition 8** (Integrated log-likelihood). We call integrated log-likelihood the following mapping:

$$\ell^{int} : (Q, \gamma, \mathfrak{D}, \mathbf{b}) \in \bigcup_{K'=1}^K \Sigma_{K'}^{\sigma^-} \times \Gamma^{K'} \times \mathbf{L}^\infty([0, 1])^{K'} \times (\mathcal{B}^*)^{K'} \mapsto \int_0^1 \ell^{hom}(Q, \tau(\gamma, \mathfrak{D}(u)), \mathbf{b}) du$$

For all  $M > 0$ , let

$$\mathcal{D}(M) := \bigcup_{n \geq 4K(d+1)} \{D_x^{\theta, n} \mid \theta \in \Theta_n^{\text{OK}}(M), x \in [K^\theta]\} \quad (5)$$

where  $D_x^{\theta, n}$  is defined in Equation (3). By definition of  $\Theta_n^{\text{OK}}(M)$ ,  $\mathcal{D}(M)$  is a subset of  $\mathbf{L}^\infty([0, 1])$  uniformly bounded by  $M + \|\Delta\|_\infty$ . Let  $\text{Cl}(\mathcal{D}(M))$  be its closure in  $\mathbf{L}^\infty([0, 1])$ .

**Proposition 9.** Let  $M > 0$ . Assume **(Amax)** and **(Amin)**. Then  $\text{Cl}(\mathcal{D}(M))$  is a compact subset of  $\mathcal{C}^0([0, 1])$ .

**Theorem 10.** Assume **(Amax)**, **(Amin)**, **(Aint)** and **(Areg)**. For all  $M > 0$ , the function  $\ell^{int}$  is continuous on  $\bigcup_{K'=1}^K \Sigma_{K'}^{\sigma^-} \times \Gamma^{K'} \times \text{Cl}(\mathcal{D}(M))^{K'} \times (\mathcal{B}^*)^{K'}$  and almost surely,

$$\sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n(\theta) - \ell^{int}(Q^\theta, \gamma^\theta, (D_x^{\theta, n})_x, \mathbf{b}^\theta) \right| \xrightarrow{n \rightarrow \infty} 0. \quad (6)$$

Proposition 9 and Theorem 10 are proved in the first part of Appendix C.

Now, assume that  $K^*$  is known and take  $K = K^*$ . The following proposition ensures that the only maximizer of the integrated log-likelihood is the true parameter.

**Proposition 11.** Assume that **(Amax)**, **(Amin)**, **(Aint)**, **(Acentering)**, **(Aid)** and **(Areg)** hold. Let  $(Q, \gamma, \mathfrak{D}, \mathbf{b}) \in \Sigma_K^{\sigma^-} \times \Gamma^K \times \text{Cl}(\mathcal{D})^K \times (\mathcal{B}^*)^K$  be a maximizer of  $\ell^{int}$ , then  $\mathfrak{D}$  is constant and  $(Q, \gamma, \mathfrak{D}, \mathbf{b}) = (Q^*, \gamma^*, \Delta, \mathbf{b}^*)$  up to permutation of the hidden states.

The proof of Proposition 11 is in the second part of Appendix C.

We may now prove the consistency of the maximum likelihood estimator. By Theorem 5, there exists  $M > 0$  such that almost surely, there exists a (random) integer  $n_{\text{loc}}$  such that for all  $n \geq n_{\text{loc}}$ ,  $\hat{\theta}_n \in \Theta_n^{\text{OK}}(M)$ . For  $n \geq n_{\text{loc}}$ , let

$$\begin{cases} (Q_n, \gamma_n) := (Q^{\hat{\theta}_n}, \gamma^{\hat{\theta}_n}), \\ \mathfrak{D}_n := (D_x^{\hat{\theta}_n, n})_{x \in [K]}, \\ \mathbf{b}_n := \mathbf{b}^{\hat{\theta}_n}. \end{cases}$$

For all  $n \geq n_{\text{loc}}$ ,  $(Q_n, \gamma_n, \mathfrak{D}_n, \mathbf{b}_n) \in \Sigma_K^{\sigma^-} \times \Gamma^K \times \text{Cl}(\mathcal{D})^K \times (\mathcal{B}^*)^K$ , and this set is compact by compactness of  $\Gamma$  and Proposition 9. Let  $(Q, \gamma, \mathfrak{D}, \mathbf{b})$  be the limit of a convergent subsequence  $(Q_{\varphi(n)}, \gamma_{\varphi(n)}, \mathfrak{D}_{\varphi(n)}, \mathbf{b}_{\varphi(n)})_{n \geq 1}$ , then by continuity of  $\ell^{int}$  and by the uniform convergence of equation (6),

$$\frac{1}{\varphi(n)} \ell_{\varphi(n)}(\hat{\theta}_{\varphi(n)}) \xrightarrow{n \rightarrow \infty} \ell^{int}(Q, \gamma, \mathfrak{D}, \mathbf{b}) \leq \ell^{int}(Q^*, \gamma^*, \Delta, \mathbf{b}^*) = \ell(\theta^*),$$

by Proposition 11, and by definition of the maximum likelihood estimator

$$\frac{1}{\varphi(n)} \ell_{\varphi(n)}(\theta^*) \leq \frac{1}{\varphi(n)} \ell_{\varphi(n)}(\hat{\theta}_{\varphi(n)}).$$

Hence  $\ell^{\text{int}}(Q, \gamma, \mathfrak{D}, \mathbf{b}) = \ell(\theta^*) = \ell^{\text{int}}(Q^*, \gamma^*, \Delta, \mathbf{b}^*)$ , which means that  $(Q, \gamma, \mathfrak{D}, \mathbf{b}) = (Q^*, \gamma^*, \Delta, \mathbf{b}^*)$  up to permutation of the hidden states by Proposition 11. Thus, the MLE sequence  $(Q_n, \gamma_n, \mathfrak{D}_n, \mathbf{b}_n)_{n \geq 1}$  has only one possible limit: the true parameter. Theorem 1 follows.

## 4 Simulations

### 4.1 First experiment: rate of convergence

In this first example, we consider the following HMM with trends  $(X_t, Y_t)_{t \geq 1}$  with  $K^* = 3$  states. The emission distributions are centered Gaussian distributions with respective variances  $(\sigma_1^*)^2 = 5$ ,  $(\sigma_2^*)^2 = 10$  and  $(\sigma_3^*)^2 = 15$ . The trends are given by

$$T_1^*(t) = \alpha(t + 10^4)^2, \quad T_2^*(t) = T_1^*(t) - 5, \quad T_3^*(t) = 3T_1^*(t),$$

with  $\alpha = 10^{-8}$ . Thus  $T_1^*$  and  $T_2^*$  belong to the same block while  $T_3^*$  diverges from the two other trends. Finally, the transition matrix is

$$Q^* = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}.$$

The model is as follows. The number of states  $K^* = 3$  is assumed known. The set  $\Gamma$  of possible emission distributions is taken as the set of centered Gaussian distributions, without constraint on the variance. The lower bound on the transition probabilities is chosen as  $\sigma_- = 0$ . Even if  $\Gamma$  is not compact and  $\sigma_-$  is not positive, contrary to the theoretical result, the maximum likelihood estimator is still able to recover the parameters. Finally, the maximum degree of the trends is  $d = 4$ . Note that the model is over-parametrized, as it contains all the trends in  $\mathbb{R}_4[X]$  while the degree of the true trends is only 2. This reflects the fact that in practice, we may not know the degree of the true trends.

We simulate 10 independent realizations  $(X_t, Y_t)_{1 \leq t \leq n_{\max}}$  with  $n_{\max} = 10^5$ . Figure 3 shows the data points corresponding to one of these 10 trajectories. Figure 4 illustrates the fact that for each  $x \in [K^*]$ ,  $\|T_x^* - T_x^{\hat{\theta}_n}\|_{\infty, [0, n]} \xrightarrow[n \rightarrow \infty]{} 0$ , where  $T_x^{\hat{\theta}_n}$  is the maximum likelihood estimator of  $T_x^*$  computed from the first  $n$  observations.

For each simulated trajectory and for several  $n \in \{1, \dots, n_{\max}\}$ , we compute:

- The errors on the trends  $\|T_x^* - T_x^{\hat{\theta}_n}\|_{\infty, [0, n]}$ ,  $1 \leq x \leq K^*$ ,
- The error on the transition matrix  $\|Q^* - Q^{\hat{\theta}_n}\|_F$ , where  $\|\cdot\|_F$  denotes the Frobenius norm,
- The error on the variances  $\max_x |(\sigma_x^*)^2 - \hat{\sigma}_x^2|$ .

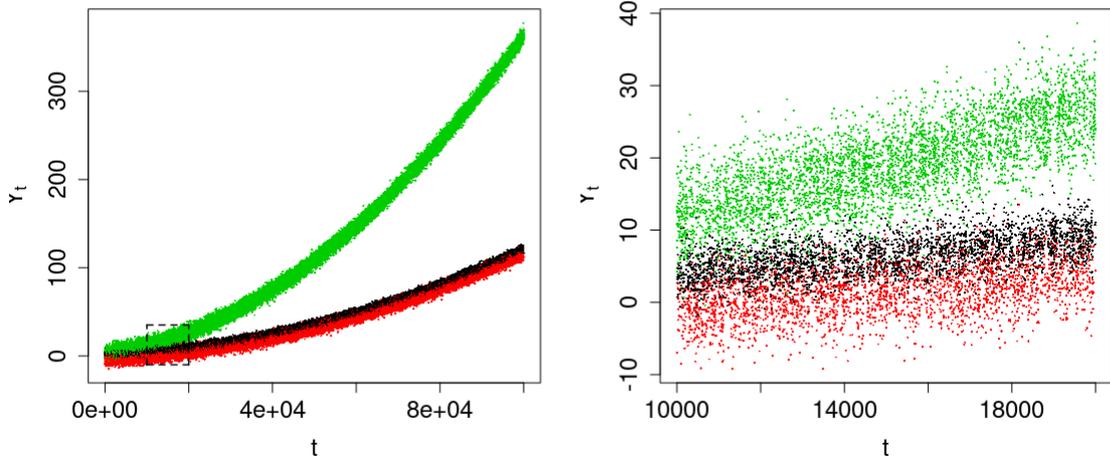


Figure 3: Left panel: a simulated trajectory of length 100000 of the observations of the HMM with trends  $(X_t, Y_t)_{t \geq 1}$ . Each color corresponds to a different state (state 1: black, state 2: red, state 3: green). Right panel: a focus on  $10000 \leq t \leq 20000$ .

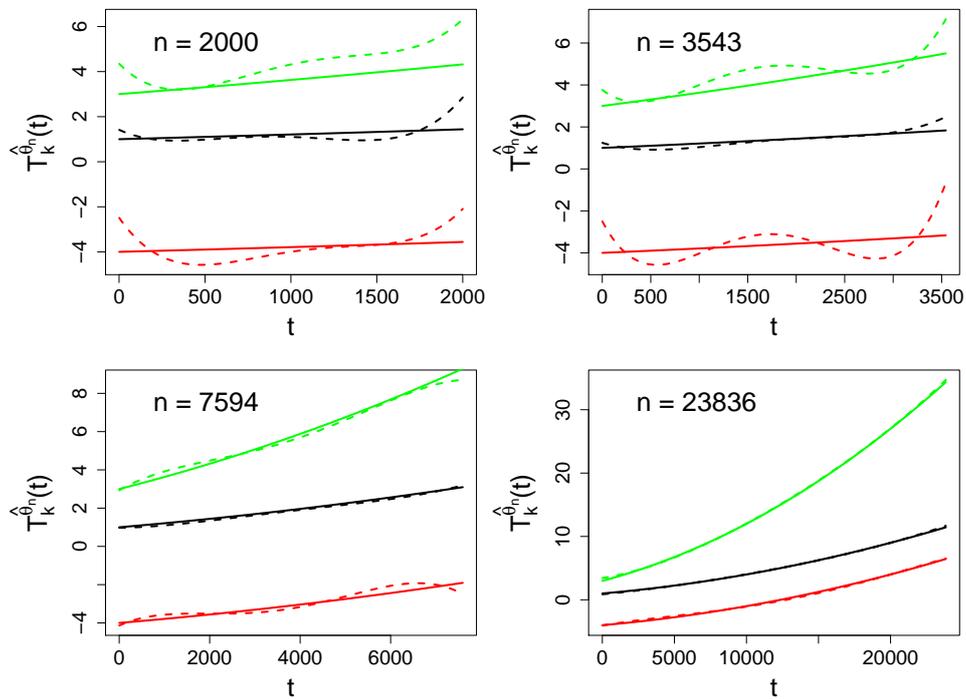


Figure 4: Convergence of the estimated trends  $T_x^{\hat{\theta}_n}$  (dashed lines) to the true trends  $T_x^*$  (solid lines).

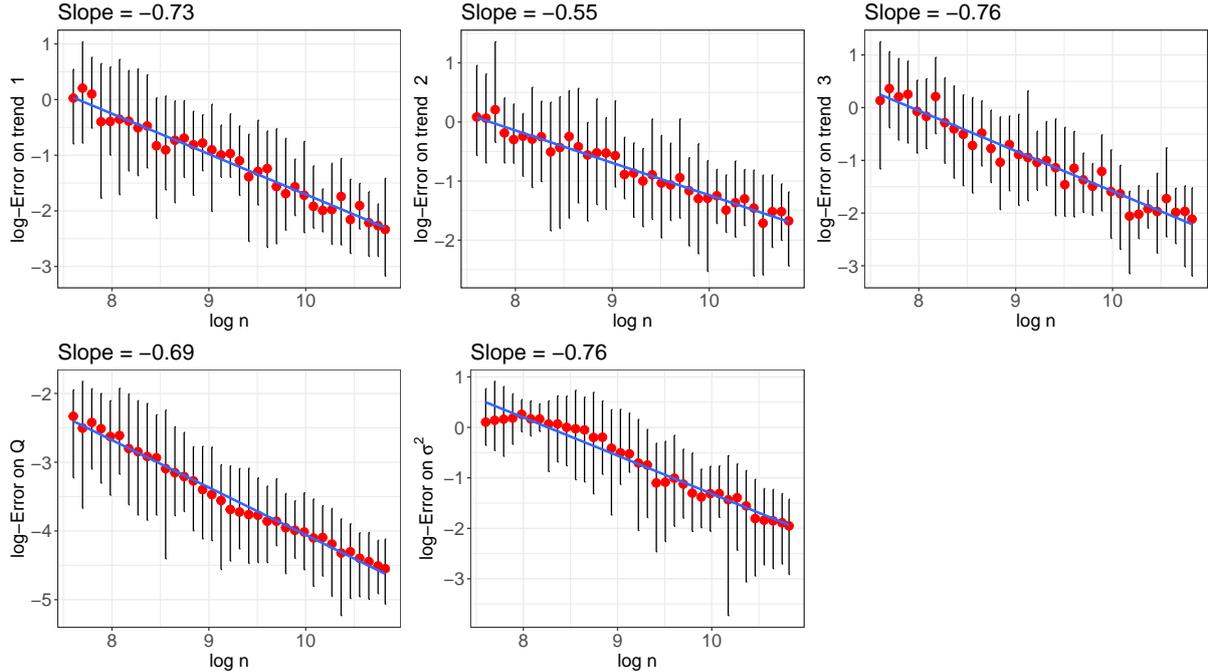


Figure 5: Rate of convergence of the maximum likelihood estimator. The red dots are the means of the log-errors across the 10 simulations, the blue line is the linear fit to these means and the black bars represent the range of the log-errors across the simulations.

We plotted the logarithm of these errors against  $\log n$  (see Figure 5). These graphs suggest a linear decrease of the logarithm of the errors with respect to  $\log n$ . Therefore, we can conjecture that as  $n$  tends to infinity,  $n^\alpha \|Q^* - Q^{\hat{\theta}_n}\|_F$  is bounded in probability for some  $\alpha > 0$ , and that the same property holds for the other parameters (possibly with a different  $\alpha$ ). Based on this experiment, it seems reasonable to conjecture that the maximum likelihood estimator achieves the parametric rate of convergence  $\alpha = 0.5$ . However, proving this claim would require further investigation that go beyond the scope of this paper. It is also worth noting that in this example, we chose simple parametric emission distributions, whereas the model described in Section 2 for which our main result holds only requires the set of emission densities to be compact, not necessarily finite-dimensional.

## 4.2 Second experiment: the trends have not diverged yet

In this section, we consider a HMM with trends whose trends have not diverged during the experiment, which is an assumption on which the proofs rely heavily. We show that the MLE is still able to recover the trends and the homogeneous parameter accurately. This is especially relevant for practical applications where one may not have enough time to see the trends diverge.

Fix the maximum number of observations to  $n = 10000$  and consider the following HMM with trends  $(X_t, Y_t)_{t \geq 1}$  with  $K^* = 2$  states. The trends are defined by  $T_1^*(t) = 0$  and  $T_2^*(t) = 3 \left( \frac{t - \frac{n}{2}}{2} \right)^2 - 1$ . The emission distributions are centered gaussian distributions

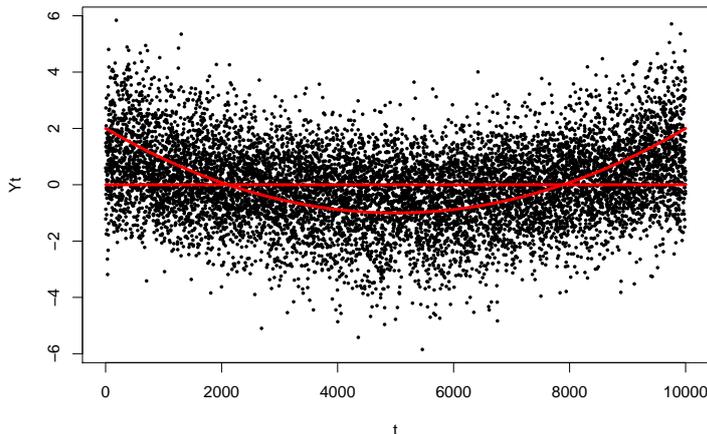


Figure 6: Simulated data points. The red lines are the true trends.

with respective variances  $(\sigma_1^*)^2 = 1$  and  $(\sigma_2^*)^2 = 2$  and the transition matrix is

$$Q^* = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

Figure 6 shows the simulated observations  $(Y_t)_{1 \leq t \leq n}$  as well as the true trends  $T_1^*$  and  $T_2^*$ . The two states are not clearly separated: the trends will eventually diverge, but we don't have enough observations to make use of this. However, the maximum likelihood estimator is able to recover the trends even in this situation.

The model is as follows. The number of states  $K^* = 2$  is assumed known. As in the previous section, we take  $\sigma_- = 0$ ,  $\Gamma$  as the set of centered Gaussian distributions and  $d = 4$ .

Figure 7 shows the estimated trends obtained using the EM algorithm (Dempster et al., 1977) together with the true trends. The estimated transition matrix and variances are

$$\hat{Q} = \begin{pmatrix} 0.74 & 0.26 \\ 0.22 & 0.78 \end{pmatrix}, \quad (\hat{\sigma}_1^2, \hat{\sigma}_2^2) = (1.13, 2.11).$$

The precision of these estimations can be further improved by increasing the number of data points. The trends and the homogeneous parameter are already well estimated. An intuition to explain this convergence even when the trends have not diverged is that the trends vary slowly enough for the homogeneous approximation of Theorem 10 to hold.

## A Block approximation

In this appendix, we shall prove Theorem 3 and Corollary 4.

Let us begin with a few definitions. Let

$$E : t \in \mathbb{N}^* \longmapsto \inf_{b, b' \in \mathcal{B}^* \text{ s.t. } b \neq b'} |\mathbb{T}_b^*(t) - \mathbb{T}_{b'}^*(t)|$$

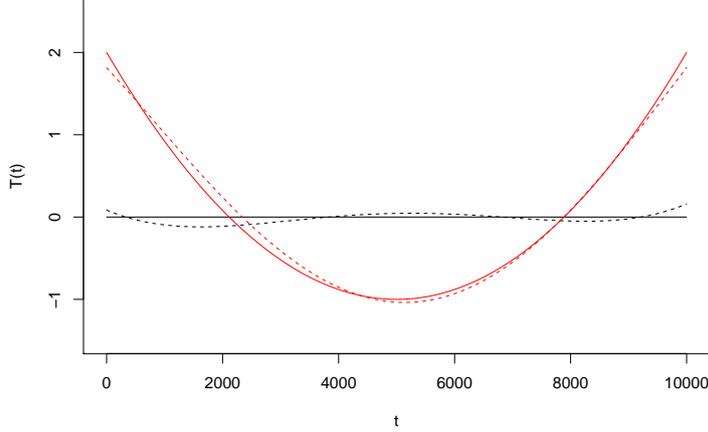


Figure 7: True (full lines) and estimated (dashed lines) trends.

be the minimum difference between the reference trends of two distinct blocks at time  $t$ . Note that  $E(t)$  diverges to  $+\infty$  since the true trends are polynomials.

Let  $M > 0$  and

$$n_1(M) := \inf\{n \in \mathbb{N}^* \mid \forall t \geq n, E(t) > 2M\}.$$

Let  $n \geq n_1(M)$ , so that the tubes of size  $M$  at time  $n$  have no intersection. Let  $\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)$  and  $x \in [K^\theta]$ , where  $\Delta$  is as in Definition 2. Then, by equation (2), there exists a block  $\mathbf{b}^\theta(x) \in \mathcal{B}^*$  (which is unique since  $n \geq n_1(M)$ ) such that  $T_x^\theta$  is in the tube of size  $M$  of  $\mathbf{b}^\theta(x)$  a time  $n$ . In particular,

$$\sup_{t \in \{1, \dots, n\}} |\mathbb{T}_{\mathbf{b}^\theta(x)}^*(t) - T_x^\theta(t)| \leq M. \quad (7)$$

The proof aims to make the following approximations rigorous.

$$\begin{aligned} \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) &\approx \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \\ &\approx \log p_{Y_t, B_t|Y_1^{t-1}, B_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}, B_1^{t-1}), \end{aligned}$$

hence the two following steps.

## A.1 Step 1: introduction of the trend block $B_t$ in the log-likelihood

Assume **(Amax)**, **(Amin)** and **(Aint)**. Let us show that the following quantity tends to 0 uniformly in  $\theta$ .

$$\begin{aligned} &\log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \\ &= \log \left( \frac{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Y_t - T_{x_t}^\theta(t))}{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Y_t - T_{x_t}^\theta(t)) \mathbf{1}_{\mathbf{b}^\theta(x_t) = B_t}} \right). \end{aligned}$$

This can be rewritten as

$$\begin{aligned}
& \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \\
&= \log(p_{B_t|Y_1^t}^\theta(B_t|Y_1^t)^{-1}) \\
&= \log\left(\left\{1 - p_{B_t|Y_1^t}^\theta(\{b \in \mathcal{B}^* \text{ s.t. } b \neq B_t\}|Y_1^t)\right\}^{-1}\right). \tag{8}
\end{aligned}$$

Intuitively, when  $t$  is large, since the trends get further from one another, the probability to get the wrong block converges to zero.

**Lemma 12.** *Assume **(Amax)** and **(Amin)**. Then for all  $t \in \mathbb{N}^*$ ,*

$$\begin{aligned}
& \sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} \left| \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \\
& \leq \log\left(\left\{1 - \frac{g(\{E(t) - M - Z_t^{\max} - \|\Delta\|_\infty\}_+)}{\sigma_- m(M + Z_t^{\max} + \|\Delta\|_\infty)}\right\}_+^{-1} \wedge \frac{g(0)}{\sigma_- m(M + Z_t^{\max} + \|\Delta\|_\infty)}\right) \\
& =: h(E(t), Z_t^{\max})
\end{aligned}$$

with the convention  $\{z\}_+^{-1} = +\infty$  if  $z \leq 0$ .

*Proof.* Proof in Section A.4.1. □

Summing in  $t$  results in

$$\sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} \left| \frac{1}{n} \ell_n(\theta) - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \leq \frac{1}{n} \sum_{t=1}^n h(E(t), Z_t^{\max}).$$

The function  $e \in \mathbb{R}_+ \mapsto h(e, z)$  is non-negative, non-increasing for all  $z \in \mathbb{R}$  and tends to 0 as  $e$  tends to  $+\infty$ . Moreover, under Assumption **(Aint)**,  $h(0, Z_1^{\max})$  is integrable by definition of  $h$ . Thus, under Assumption **(Aint)**, the law of large numbers implies that for all  $E > 0$

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} \left| \frac{1}{n} \ell_n(\theta) - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \leq \mathbb{E}^*[h(E, Z_1^{\max})].$$

The dominated convergence theorem ensures that  $\mathbb{E}^*[h(E, Z_1^{\max})] \rightarrow 0$  as  $E \rightarrow +\infty$ . Thus, we obtain the following uniform approximation of the normalized log-likelihood:

$$\sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} \left| \frac{1}{n} \ell_n(\theta) - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \xrightarrow[n \rightarrow \infty]{} 0. \tag{9}$$

## A.2 Step 2: conditioning on the blocks $B_1^{t-1}$

Assume **(Amax)** and **(Amin)**. The following lemma is a consequence of the lower bound on the transition matrices, see for instance Lemma 1 and Corollary 1 of [Douc et al. \(2004\)](#).

**Lemma 13** (Exponential forgetting). *There exists  $C > 0$  such that for all  $n \in \mathbb{N}^*$ ,  $y_1^n \in \mathbb{R}^n$ ,  $\theta \in \Theta$  and for all probability measures  $\pi, \pi'$  on  $[K^\theta]$ :*

$$\sum_{x \in [K^\theta]} |p_{X_n|Y_1^{n-1}}^\theta(x|y_1^{n-1}, X_0 \sim \pi) - p_{X_n|Y_1^{n-1}}^\theta(x|y_1^{n-1}, X_0 \sim \pi')| \leq C\rho^n$$

with  $\rho = 1 - \frac{\sigma_-}{1-\sigma_-} \in (0, 1)$ .

Besides, under **(Aerg)**, for all  $\theta \in \Theta$ ,  $x \in [K^\theta]$ ,  $y_1^{n-1} \in \mathbb{R}^n$  and for all probability measure  $\pi$  on  $[K^\theta]$ :

$$p_{X_n|Y_1^{n-1}}^\theta(x|y_1^{n-1}, X_0 \sim \pi) \geq \sigma_-.$$

Hence, using the inequality  $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$  for all  $x, y > 0$ : for all  $n \in \mathbb{N}^*$ ,  $\theta \in \Theta$ ,  $y_1^n \in \mathbb{R}^n$ ,  $b \in \mathcal{B}^*$  and for all probability measures  $\pi, \pi'$  on  $[K^\theta]$ :

$$\begin{aligned} & |\log p_{Y_n, B_n|Y_1^{n-1}}^\theta(y_n, b|y_1^{n-1}, X_0 \sim \pi) - \log p_{Y_n, B_n|Y_1^{n-1}}^\theta(y_n, b|y_1^{n-1}, X_0 \sim \pi')| \\ & \leq \frac{\sum_{x \in [K^\theta]} |p_{X_n|Y_1^{n-1}}^\theta(x|y_1^{n-1}, X_0 \sim \pi) - p_{X_n|Y_1^{n-1}}^\theta(x|y_1^{n-1}, X_0 \sim \pi')| p_{Y_n, B_n|X_n}^\theta(y_n, b|x)}{\sigma_- \sum_{x \in [K^\theta]} p_{Y_n, B_n|X_n}^\theta(y_n, b|x)} \\ & \leq \frac{C}{\sigma_-} \rho^n. \end{aligned}$$

Changing the constant  $C$  if necessary, for all  $a \in \mathbb{N}^*$ :

$$\begin{aligned} \sup_{\theta \in \Theta_n^{OK}(M-\|\Delta\|_\infty)} & \left| \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right. \\ & \left. - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}, B_1^{t-a}}^\theta(Y_t, B_t|Y_1^{t-1}, B_1^{t-a}) \right| \leq C\rho^a. \end{aligned}$$

It remains to condition on  $B_{t-a+1}^{t-1}$ .

**Lemma 14.** *Assume **(Amax)** and **(Amin)**. Then for all  $a \in \mathbb{N}^*$ ,*

$$\begin{aligned} \sup_{\theta \in \Theta_n^{OK}(M-\|\Delta\|_\infty)} & \left| \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}, B_1^{t-a}}^\theta(Y_t, B_t|Y_1^{t-1}, B_1^{t-a}) \right. \\ & \left. - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t|Y_1^{t-1}, B_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}, B_1^{t-1}) \right| \\ & \leq \frac{2aK^2}{\sigma_-^3} \frac{1}{n} \sum_{i=1}^{n-1} \left( 1 \wedge \frac{g(\{E(i) - M - Z_i^{\max} - \|\Delta\|_\infty\}_+)}{m(M + Z_i^{\max} + \|\Delta\|_\infty)} \right) \\ & =: \frac{2aK^2}{\sigma_-^3} \frac{1}{n} \sum_{i=1}^{n-1} h'(E(i), Z_i^{\max}). \end{aligned}$$

*Proof.* We show that when **(Aerg)** holds, for all  $\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)$ ,  $t \leq n$ ,  $a \in \mathbb{N}^*$ ,  $y_1^t \in \mathcal{Y}^t$  and  $b_1^t \in (\mathcal{B}^*)^t$ ,

$$\begin{aligned} & \left| \log p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-a}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-a}) - \log p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1}) \right| \\ & \leq \frac{2}{\sigma_-^\theta} p_{B_{(t-a+1)\vee 1}^{t-1} | Y_1^{t-1}, B_1^{t-a}}^\theta \left( (\mathcal{B}^*)^{(a \wedge t) - 1} \setminus \{b_{(t-a+1)\vee 1}^{t-1}\} | y_1^{t-1}, b_1^{t-a} \right) \end{aligned} \quad (10)$$

$$\leq \frac{2K^2}{\sigma_-^3} \sum_{i=(t-a+1)\vee 1}^{t-1} \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq b_i} \gamma_{x_i}^\theta(y_i - T_{x_i}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))}. \quad (11)$$

Then, we show that under **(Amax)** and **(Amin)**, for all  $\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)$  and  $i \leq n$ :

$$\begin{aligned} & \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq B_i} \gamma_{x_i}^\theta(Y_i - T_{x_i}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} \leq 1 \wedge \frac{g(\{E(i) - M - Z_i^{\max} - \|\Delta\|_\infty\}_+)}{m(M + Z_i^{\max} + \|\Delta\|_\infty)} \\ & =: h'(E(i), Z_i^{\max}), \end{aligned} \quad (12)$$

and the lemma follows by summing over  $t$  and  $i$ . The details of the proof can be found in Section A.4.2.  $\square$

Therefore, almost surely,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} & \left| \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t | Y_1^{t-1}}^\theta(Y_t, B_t | Y_1^{t-1}) \right. \\ & \left. - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(Y_t, B_t | Y_1^{t-1}, B_1^{t-1}) \right| \\ & \leq \limsup_{n \rightarrow +\infty} \left( C\rho^a + \frac{2aK^2}{\sigma_-^3} \frac{1}{n} \sum_{i=1}^{n-1} h'(E(i), Z_i^{\max}) \right) \\ & \leq C\rho^a + \frac{2aK^2}{\sigma_-^3} \mathbb{E}^*[h'(E, Z_1^{\max})] \\ & \leq C' (-\mathbb{E}^*[h'(E, Z_1^{\max})] \log \mathbb{E}^*[h'(E, Z_1^{\max})]) \end{aligned}$$

for some explicit constant  $C'$  and for all  $E$  sufficiently large to have  $\mathbb{E}^*[h'(E, Z_1^{\max})] < 1/2$  using Assumption **(Adiv)**, the law of large numbers, the fact that the mapping  $e \mapsto h'(e, z)$  is non-negative, bounded by 0 and 1 and non-increasing for all  $z$ , and by taking  $a = \lceil \frac{\log \mathbb{E}^*[h'(E, Z_1^{\max})]}{\log \rho} \rceil$ . Since the function  $e \mapsto h'(e, z)$  tends to 0 as  $e$  tends to  $+\infty$  for all  $z$ , the dominated convergence theorem ensures that almost surely,

$$\begin{aligned} \sup_{\theta \in \Theta_n^{\text{OK}}(M - \|\Delta\|_\infty)} & \left| \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t | Y_1^{t-1}}^\theta(Y_t, B_t | Y_1^{t-1}) \right. \\ & \left. - \frac{1}{n} \sum_{t=1}^n \log p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(Y_t, B_t | Y_1^{t-1}, B_1^{t-1}) \right| \xrightarrow[n \rightarrow \infty]{} 0. \end{aligned}$$

Let  $\frac{1}{n}\ell_n^{(Y,B)}(\theta) = \frac{1}{n}\log p_{(Y,B)_1^n}^\theta((Y,B)_1^n)$ . Combining the above equation with equation (9) yields Theorem 3.

### A.3 Application: existence and finiteness of the relative entropy rate

Theorem 3 implies that Corollary 4 holds: since  $\theta^* \in \Theta_n^{\text{OK}}(M)$  for all  $n \in \mathbb{N}^*$  and  $M > 0$ ,

$$\left| \frac{1}{n}\ell_n(\theta^*) - \frac{1}{n}\ell_n^{(Y,B)}(\theta^*) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Let  $Z'_t = Y_t - \mathbb{T}_{B_t}^*(t)$ . Then  $\frac{1}{n}\ell_n^{(Y,B)}(\theta^*) = \frac{1}{n}\ell_n^{(Z',B)}(\theta^*)$ . Moreover, under  $\theta^*$ , the process  $(X_t, (Z'_t, B_t))_{t \geq 1}$  is a homogeneous and ergodic HMM with emission densities  $(z', b) \mapsto \gamma_{x^*}^*(z' - \Delta(x^*))\mathbf{1}(b = \mathbf{b}^*(x^*))$  for  $x^* \in [K^*]$  with respect to the measure  $\text{Leb} \otimes \mu_{\mathcal{B}^*}$ , where  $\mu_{\mathcal{B}^*}$  is the counting measure on  $\mathcal{B}^*$ .

Since it is homogeneous and ergodic, [Barron \(1985\)](#) shows that there exists  $\ell(\theta^*) > -\infty$  such that

$$\frac{1}{n}\ell_n^{(Z',B)}(\theta^*) \longrightarrow \ell(\theta^*).$$

Then, all emission densities are upper bounded by  $g(0)$  under **(Amax)**, so that the positive part of their logarithm is integrable. Therefore, [Leroux \(1992\)](#) implies that  $\ell(\theta^*) < +\infty$  and Corollary 4 is proved.

## A.4 Proofs

### A.4.1 Proof of Lemma 12 (current block)

First note that this quantity is non-negative: the denominator contains less terms, and all of them are non-negative. Hence it is enough to find an upper bound. To this aim we will use Assumptions **(Amax)**, **(Amin)** and **(Aerg)**:

$$\begin{aligned} & \left| \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \\ & \leq \log \left( \frac{g(0)}{\sigma_- \sup_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) = B_t} m(|Y_t - T_{x_t}^\theta(t)|)} \right) \\ & \leq \log \frac{g(0)}{\sigma_-} + \sum_{x_t^* \in [K^*]} \mathbf{1}_{X_t = x_t^*} \left( -\log m \left( \inf_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) = \mathbf{b}^*(x_t^*)} |Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| \right) \right). \end{aligned}$$

When  $X_t = x_t^*$ ,

$$\begin{aligned} & \inf_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) = \mathbf{b}^*(x_t^*)} |Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| \\ & \leq |Z_t| + \inf_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) = \mathbf{b}^*(x_t^*)} |\mathbb{T}_{\mathbf{b}^*(x_t^*)}^*(t) - T_{x_t}^\theta(t)| + \Delta(x_t^*) \\ & \leq Z_t^{\max} + M + \|\Delta\|_\infty \end{aligned}$$

using equation (7), hence

$$-\log m \left( \inf_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) = \mathbf{b}^*(x_t^*)} |Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| \right) \leq -\log m(M + Z_t^{\max} + \|\Delta\|_\infty).$$

This yields

$$\left| \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \right| \leq \log \frac{g(0)}{\sigma_- m(M + Z_t^{\max} + \|\Delta\|_\infty)}.$$

Let us show the second bound. We can rewrite it as

$$\begin{aligned} & \log p_{Y_t|Y_1^{t-1}}^\theta(Y_t|Y_1^{t-1}) - \log p_{Y_t, B_t|Y_1^{t-1}}^\theta(Y_t, B_t|Y_1^{t-1}) \\ &= -\log \left( 1 - \frac{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Y_t - T_{x_t}^\theta(t)) \mathbf{1}_{\mathbf{b}^\theta(x_t) \neq B_t}}{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Y_t - T_{x_t}^\theta(t))} \right) \\ &= -\log \left( 1 - \sum_{x_t^* \in [K^*]} \mathbf{1}_{X_t = x_t^*} \underbrace{\frac{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)) \mathbf{1}_{\mathbf{b}^\theta(x_t) \neq B_t}}{\sum_{x_t \in [K^\theta]} p^\theta(X_t = x_t|Y_1^{t-1}) \gamma_{x_t}(Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t))}}_{(*)} \right). \end{aligned}$$

Using **(Amax)**, **(Amin)** and **(Aerg)**,

$$\begin{aligned} (*) & \leq \frac{\sup_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) \neq B_t} g(|Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)|)}{\sigma_- \sum_{x_t \in [K^\theta]} m(|Z_t + T_{x_t^*}^*(t) - T_{x_t}^\theta(t)|)} \\ & \leq \frac{\sup_{x_t \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_t) \neq B_t} g(\{|T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| - Z_t^{\max}\}_+)}{\sigma_- \sup_{x_t \in [K^\theta]} m(|T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| + Z_t^{\max})}. \end{aligned}$$

Let  $x \in [K^\theta]$  such that  $\mathbf{b}^\theta(x) \neq B_t$ . When  $X_t = x_t^*$ ,

$$\begin{aligned} |Y_t - T_x^\theta(t)| &= |Z_t + T_{x_t^*}^*(t) - T_x^\theta(t)| \\ &\geq |\mathbb{T}_{B_t}^*(t) - T_x^\theta(t)| - Z_t^{\max} - \Delta(x_t^*) \\ &\geq |\mathbb{T}_{B_t}^*(t) - \mathbb{T}_{\mathbf{b}^\theta(x)}^*(t)| - M - Z_t^{\max} - \|\Delta\|_\infty \\ &\geq E(t) - M - Z_t^{\max} - \|\Delta\|_\infty \end{aligned}$$

and

$$\begin{aligned} \sup_{x_t \in [K^\theta]} m(|T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| + Z_t^{\max}) &\leq m \left( \inf_{x_t \in [K^\theta]} |T_{x_t^*}^*(t) - T_{x_t}^\theta(t)| + Z_t^{\max} \right) \\ &\leq m(M + Z_t^{\max} + \|\Delta\|_\infty) \end{aligned}$$

using equation (7). Therefore,

$$(*) \leq \frac{g(\{E(t) - M - Z_t^{\max} - \|\Delta\|_\infty\}_+)}{\sigma_- m(M + Z_t^{\max} + \|\Delta\|_\infty)}. \quad (13)$$

#### A.4.2 Proof of Lemma 14 (recent blocks)

**Proof of equation (10)** Without loss of generality, one may assume  $a \leq t$  (otherwise, the proof holds by replacing  $a$  by  $a \wedge t$ ).

$$\begin{aligned} p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-a}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-a}) &= p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1}) \\ &\quad \times p_{B_{t-a+1}^{t-1} | Y_1^{t-1}, B_1^{t-a}}^\theta(b_{t-a+1}^{t-1} | y_1^{t-1}, b_1^{t-a}) \\ &\quad + p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-a}, B_{t-a+1}^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-a}, (\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\}) \\ &\quad \times p_{B_{t-a+1}^{t-1} | Y_1^{t-1}, B_1^{t-a}}^\theta((\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\} | y_1^{t-1}, b_1^{t-a}), \end{aligned}$$

hence

$$\begin{aligned} &|p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-a}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-a}) - p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1})| \\ &\leq p_{B_{t-a+1}^{t-1} | Y_1^{t-1}, B_1^{t-a}}^\theta((\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\} | y_1^{t-1}, b_1^{t-a}) \left[ p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1}) \right. \\ &\quad \left. + p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1}, (\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\}) \right] \\ &\leq 2p_{B_{t-a+1}^{t-1} | Y_1^{t-1}, B_1^{t-a}}^\theta((\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\} | y_1^{t-1}, b_1^{t-a}) \sum_{x \in [K^\theta]} p_{Y_t, B_t | X_t}^\theta(y_t, b_t | x). \end{aligned}$$

Finally, since under **(Aerg)**

$$\begin{aligned} p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-a}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-a}) &= \sum_{x \in \mathcal{X}} p_{Y_t, B_t | X_t}^\theta(y_t, b_t | x) p_{X_t | Y_1^{t-1}, B_1^{t-a}}^\theta(x | y_1^{t-1}, b_1^{t-a}) \\ &\geq \sigma_- \sum_{x \in \mathcal{X}} p_{Y_t, B_t | X_t}^\theta(y_t, b_t | x) \end{aligned}$$

and the same inequality holds for  $p_{Y_t, B_t | Y_1^{t-1}, B_1^{t-1}}^\theta(y_t, b_t | y_1^{t-1}, b_1^{t-1})$ , we obtain equation (10) using  $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$  for all  $x, y > 0$ .

**Proof of equation (11)** Since

$$(\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\} = \bigcup_{i=t-a+1}^{t-1} ((\mathcal{B}^*)^{i-(t-a+1)} \times (\mathcal{B}^* \setminus \{b_i\}) \times (\mathcal{B}^*)^{t-1-i}),$$

by union bound,

$$\begin{aligned}
& p_{B_1^{t-a+1}|Y_1^{t-1}, B_1^{t-a}}^\theta((\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\}|y_1^{t-1}, b_1^{t-a}) \\
& \leq \sum_{i=t-a+1}^{t-1} p_{B_i|Y_1^{t-1}, B_1^{t-a}}^\theta(\mathcal{B}^* \setminus \{b_i\}|y_1^{t-1}, b_1^{t-a}) \\
& = \sum_{i=t-a+1}^{t-1} \sum_{x_i \in [K^\theta]} p_{B_i|X_i}^\theta(\mathcal{B}^* \setminus \{b_i\}|x_i) p_{X_i|Y_1^{t-1}, B_1^{t-a}}^\theta(x_i|y_1^{t-1}, b_1^{t-a}) \\
& = \sum_{i=t-a+1}^{t-1} \sum_{x_i \in [K^\theta]} \mathbf{1}_{\mathbf{b}^\theta(x_i) \neq b_i} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} p_{X_i|Y_i, X_{i-1}, X_{i+1}}^\theta(x_i|y_i, x_{i-1}, x_{i+1}) \\
& \quad \times p_{X_{i-1}, X_{i+1}|Y_1^{t-1}, B_1^{t-a}}^\theta(x_{i-1}, x_{i+1}|y_1^{t-1}, b_1^{t-a}).
\end{aligned}$$

Then, use that for all  $x_{i-1}, x_{i+1} \in [K^\theta]$ ,

$$p_{X_{i-1}, X_{i+1}|Y_1^{t-1}, B_1^{t-a}}^\theta(x_{i-1}, x_{i+1}|y_1^{t-1}, b_1^{t-a}) \leq 1$$

and that by the Markov property and **(Aerg)** for all  $x_{i-1}, x, x_{i+1} \in [K^\theta]$

$$\begin{aligned}
p_{X_i|X_{i-1}, X_{i+1}}^\theta(x|x_{i-1}, x_{i+1}) &= \frac{p_{X_{i+1}|X_i}^\theta(x_{i+1}|x) p_{X_i|X_{i-1}}^\theta(x|x_{i-1})}{p_{X_{i+1}|X_{i-1}}^\theta(x_{i+1}|x_{i-1})} \\
&\geq Q^\theta(x, x_{i+1}) Q^\theta(x_{i-1}, x) \\
&\geq \sigma_-^2,
\end{aligned}$$

so that

$$\begin{aligned}
& p_{X_i|Y_i, X_{i-1}, X_{i+1}}^\theta(x_i|y_i, x_{i-1}, x_{i+1}) \\
&= \frac{p_{X_i|X_{i-1}, X_{i+1}}^\theta(x_i|x_{i-1}, x_{i+1}) \gamma_{x_i}^\theta(y_i - T_{x_i}^\theta(i))}{\sum_{x \in [K^\theta]} p_{X_i|X_{i-1}, X_{i+1}}^\theta(x|x_{i-1}, x_{i+1}) \gamma_x^\theta(y_i - T_x^\theta(i))} \\
&\leq \frac{p_{X_i|X_{i-1}, X_{i+1}}^\theta(x_i|x_{i-1}, x_{i+1}) \gamma_{x_i}^\theta(y_i - T_{x_i}^\theta(i))}{\sigma_-^2 \sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))}.
\end{aligned}$$

Thus,

$$\begin{aligned}
& p_{B_1^{t-a+1}|Y_1^{t-1}, B_1^{t-a}}^\theta((\mathcal{B}^*)^{a-1} \setminus \{b_{t-a+1}^{t-1}\}|y_1^{t-1}, b_1^{t-a}) \\
& \leq \sum_{i=t-a+1}^{t-1} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} \frac{\sum_{x_i \in [K^\theta]} \mathbf{1}_{\mathbf{b}^\theta(x_i) \neq b_i} p_{X_i|X_{i-1}, X_{i+1}}^\theta(x_i|x_{i-1}, x_{i+1}) \gamma_{x_i}^\theta(y_i - T_{x_i}^\theta(i))}{\sigma_-^2 \sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))} \\
& \leq \frac{K^2}{\sigma_-^2} \sum_{i=t-a+1}^{t-1} \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq b_i} \gamma_{x_i}^\theta(y_i - T_{x_i}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))}.
\end{aligned}$$

**Proof of equation (12)** Using **(Amax)** and **(Amin)**,

$$\begin{aligned}
& \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq B_i} \gamma_{x_i}^\theta(Y_i - T_{x_i}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} \leq \left( 1 \wedge \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq B_i} g(|Y_i - T_{x_i}^\theta(i)|)}{\sup_{x \in [K^\theta]} m(|Y_i - T_x^\theta(i)|)} \right) \\
& \leq \sum_{x_i^* \in [K^*]} \mathbf{1}_{X_i = x_i^*} \left( 1 \wedge \frac{\sup_{x_i \in [K^\theta] \text{ s.t. } \mathbf{b}^\theta(x_i) \neq \mathbf{b}^*(x_i^*)} g(|Z_i + T_{x_i^*}^*(i) - T_{x_i}^\theta(i)|)}{\sup_{x \in [K^\theta]} m(|Z_i + T_{x_i^*}^*(i) - T_x^\theta(i)|)} \right) \\
& \leq 1 \wedge \frac{g(\{E(t) - M - Z_i^{\max} - \|\Delta\|_\infty\}_+)}{m(M + Z_i^{\max} + \|\Delta\|_\infty)}
\end{aligned}$$

by the same arguments as in the control of **(\*)** in equation (13).

## B Localization of the MLE

In this section we shall prove Theorem 5. Assume **(Aerg)**, **(Amin)**, **(Amax)** and **(Aint)**.

### B.1 Preliminary: compactness results

Recall that for all  $M > 0$  and  $n \in \mathbb{N}$ ,  $\Theta_n^{\text{OK}}(M)$  is the subset of  $\Theta$  defined by

$$\forall \theta \in \Theta_n^{\text{OK}}(M), \quad \forall x^* \in [K^*], \quad \exists x \in [K^\theta], \quad \|T_{x^*}^* - T_x^\theta\|_{\infty, [0, n]} \leq M \quad (14)$$

$$\text{and } \forall x \in [K^\theta], \quad \exists x^* \in [K^*], \quad \|T_{x^*}^* - T_x^\theta\|_{\infty, [0, n]} \leq M. \quad (15)$$

To prove that the maximum likelihood estimator belongs to such a set, we consider relaxed versions of (14) and (15). Let

$$\begin{aligned}
I_{n,D}(x, x^*, \theta) &:= \{t \in \{1, \dots, n\} : |T_{x^*}^*(t) - T_x^\theta(t)| \leq D\}, \\
\mathcal{T}(\alpha, n, D) &:= \left\{ \theta \in \Theta : \left| \bigcap_{x^* \in [K^*]} \bigcup_{x \in [K^\theta]} I_{n,D}(x, x^*, \theta) \right| \geq n\alpha \right\}, \\
\mathcal{U}(\alpha, n, D) &:= \left\{ \theta \in \Theta : \forall x \in [K^\theta], \left| \bigcup_{x^* \in [K^*]} I_{n,D}(x, x^*, \theta) \right| \geq n\alpha \right\}.
\end{aligned}$$

$\mathcal{T}$  corresponds to a relaxation of (14) and contains the parameters  $\theta$  such that all true trends are close to at least one parameter trend during most of the first time steps. Likewise,  $\mathcal{U}$  corresponds to a relaxation of (15) and contains the parameters  $\theta$  whose trends are close to at least one true trend during most of the first time steps.

By construction of these sets, for each  $\theta \in \mathcal{T}(\alpha, n, D)$  (resp.  $\theta \in \mathcal{U}(\alpha, n, D)$ ) and for each  $x^* \in [K^*]$  (resp.  $x \in [K^\theta]$ ), there exists at least one  $x \in [K^\theta]$  (resp.  $x^* \in [K^*]$ ) such that  $I_{n,D}(x, x^*, \theta) \geq n\alpha/K$ .

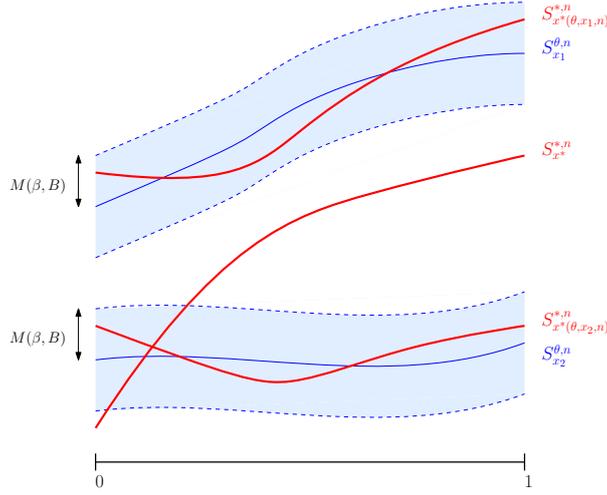


Figure 8: Rescaled trends of a parameter in  $\mathcal{U}(\beta, n, B)$ . Every parameter trend is at bounded distance of at least one true trend. However, some true trends may be far from all parameter trends.

**Proposition 15.** *For all  $\alpha \in (0, 1]$  and  $D > 0$ , there exists  $M(\alpha, D) > 0$  and  $n_0 := 4K(d+1)/\alpha$  such that*

$$\forall n \geq n_0, \forall \theta \in \Theta, \forall x \in [K^\theta], \forall x^* \in [K^*],$$

$$|I_{n,D}(x, x^*, \theta)| \geq n\alpha/K \Rightarrow \|T_{x^*}^* - T_x^\theta\|_{\infty, [0, n]} \leq M(\alpha, D).$$

It follows that  $\theta \in \mathcal{T}(\alpha, n, D)$  (resp.  $\theta \in \mathcal{U}(\alpha, n, D)$ ) is equivalent to (14) (resp. (15)) up to changing  $M$ :

**Definition 16.** *Let  $\alpha \in (0, 1]$ ,  $n \geq 1$  and  $D > 0$ .*

- *For all  $\theta \in \mathcal{T}(\alpha, n, D)$  and  $x^* \in [K^*]$ , let  $x(\theta, x^*, n, \alpha, D)$  be the smallest  $x \in [K^\theta]$  such that  $I_{n,D}(x, x^*, \theta) \geq n\alpha/K$ .*
- *For all  $\theta \in \mathcal{U}(\alpha, n, D)$  and  $x \in [K^\theta]$ , let  $x^*(\theta, x, n, \alpha, D)$  be the smallest  $x^* \in [K^*]$  such that  $I_{n,D}(x, x^*, \theta) \geq n\alpha/K$ .*

*In the following, we omit the dependency in  $\alpha$  and  $D$  and write  $x^*(\theta, x, n)$  and  $x(\theta, x^*, n)$ .*

**Corollary 17.** *For all  $\alpha \in (0, 1]$  and  $D > 0$ , there exists  $M(\alpha, D) > 0$  and  $n_0 := 4K(d+1)/\alpha$  such that*

$$\forall n \geq n_0, \begin{cases} \theta \in \mathcal{T}(\alpha, n, D) \Rightarrow \forall x^* \in [K^*], \|T_{x^*}^* - T_{x(\theta, x^*, n)}^\theta\|_{\infty, [0, n]} \leq M(\alpha, D), \\ \theta \in \mathcal{U}(\alpha, n, D) \Rightarrow \forall x \in [K^\theta], \|T_{x^*(\theta, x, n)}^* - T_x^\theta\|_{\infty, [0, n]} \leq M(\alpha, D). \end{cases}$$

Hence, for all  $\alpha, \beta \in (0, 1]$ , for all  $M > 0$  and  $D, B \geq M$  and for all  $n$ ,

$$\Theta_n^{\text{OK}}(M) \subset \mathcal{T}(\alpha, n, D) \cap \mathcal{U}(\beta, n, B) \subset \Theta_n^{\text{OK}}(M(\alpha, D) \vee M(\beta, B)). \quad (16)$$

Proposition 15 is a direct consequence of Arzelà–Ascoli’s theorem and of the following result.

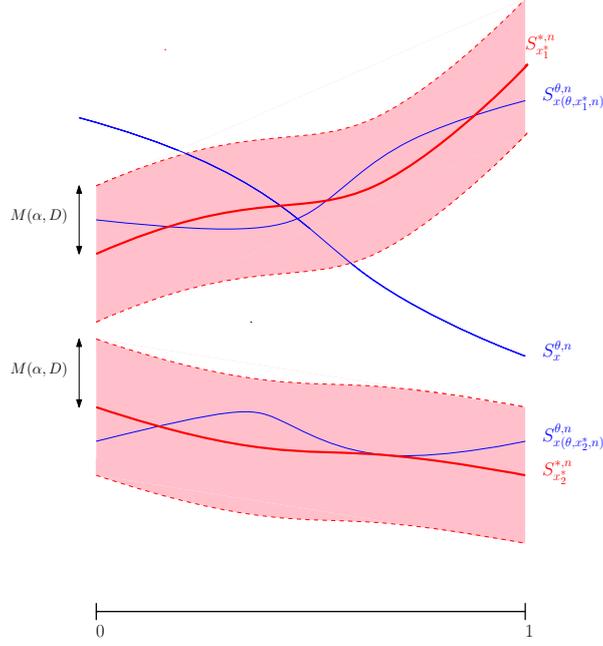


Figure 9: Rescaled trends of a parameter in  $\mathcal{T}(\alpha, n, D)$ . Every true trend is at bounded distance of at least one parameter trend. However, some parameter trends may be far from all true trends.

**Definition 18.** For all  $\theta \in \Theta$ ,  $n \in \mathbb{N}^*$  and  $x \in [K^\theta]$ , let

$$S_x^{\theta, n} : u \in [0, 1] \longmapsto T_x^\theta(nu)$$

be the trend  $T_x^\theta$  whose time variable has been rescaled from  $[0, n]$  to  $[0, 1]$ . Likewise, define  $S_{x^*}^{*, n}$  the rescaled true trend corresponding to the state  $x^*$ .

**Theorem 19.** Let  $\varepsilon \in (0, 1]$  and  $D > 0$ . Then the set

$$\mathcal{S} := \bigcup_{n \geq n_0} \bigcup_{\theta \in \Theta} \bigcup_{(x, x^*) \in [K^\theta] \times [K^*] \text{ s.t. } |I_{n, D}(x, x^*, \theta)| \geq n\varepsilon} \{S_x^{\theta, n} - S_{x^*}^{*, n}\}$$

is relatively compact in the set of continuous functions  $(\mathcal{C}^0([0, 1]), \|\cdot\|_\infty)$ .

**Remark.** This result, together with Arzelà–Ascoli’s theorem, entails that  $\mathcal{S}$  is uniformly equicontinuous in addition to being a bounded subset of  $\mathbf{L}^\infty([0, 1])$ . This will be used in Section C.

*Proof.* Let  $\varepsilon \in (0, 1]$  and  $D > 0$ . Let us first give another representation of the elements of  $\mathcal{S}$ .

**Lemma 20.** For any  $S \in \mathcal{S}$ , there exists  $(u_1^S, \dots, u_{d+1}^S) \in [0, 1]^{d+1}$  such that

$$\begin{cases} \forall i \neq j, & |u_i^S - u_j^S| \geq \frac{\varepsilon}{4K(d+1)}, \\ \forall i, & |S(u_i^S)| \leq D. \end{cases}$$

*Proof.* For  $S \in \mathcal{S}$ , let  $n(S) \geq n_0$ ,  $\theta, x \in [K^\theta]$  and  $x^* \in [K^*]$  be such that  $S = S_x^{\theta, n} - S_{x^*}^{*, n}$  and  $|I_{n, D}(x, x^*, \theta)| \geq n\varepsilon$ . Let us define  $(u_1^S, \dots, u_{d+1}^S)$  iteratively. Let  $\mathcal{A}_0 = I_{n, D}(x, x^*, D)$  and, for all  $i \geq 1$ ,

- $t_i^S \in \mathcal{A}_{i-1}$  ;
- $\mathcal{A}_i = \mathcal{A}_{i-1} \setminus \bar{B}\left(t_i^S, \frac{\varepsilon}{4K(d+1)}n\right)$ .

The closed ball  $\bar{B}\left(t_i^S, \frac{\varepsilon}{4K(d+1)}n\right)$  contains at most  $1 + \left\lfloor 2\frac{\varepsilon}{4K(d+1)}n \right\rfloor \leq 3\frac{\varepsilon}{4K(d+1)}n$  elements since  $\frac{\varepsilon}{4K(d+1)}n \geq 1$ . Thus, for all  $i \geq 0$ ,

$$|\mathcal{A}_i| \geq n\frac{\varepsilon}{K} \left(1 - \frac{3}{4(d+1)}i\right).$$

In particular,  $\mathcal{A}_i \neq \emptyset$  for all  $i \in \{0, \dots, d+1\}$ , which makes it possible to define  $(t_i^S)_{1 \leq i \leq d+1}$ . Taking  $u_i^S = \frac{t_i^S}{n(S)}$  for all  $i \in \{1, \dots, d+1\}$  concludes the proof.  $\square$

The next lemma is a straightforward consequence of the Lagrange form of the interpolation polynomial.

**Lemma 21.** *The mapping*

$$\begin{aligned} \left\{ (u_i)_i \in [0, 1]^{d+1} \text{ s.t. } \inf_{i \neq j} |u_i - u_j| > 0 \right\} \times \mathbb{R}^{d+1} &\longmapsto (\mathcal{C}^0([0, 1]), \|\cdot\|_\infty) \\ (u_1, \dots, u_{d+1}, s_1, \dots, s_{d+1}) &\longmapsto P_{u, s} \end{aligned} \quad (17)$$

is continuous, where  $P_{u, s}$  is the only polynomial with degree at most  $d$  such that  $P(u_i) = s_i$  for all  $i \in \{1, \dots, d+1\}$ .

To conclude, note that  $\mathcal{S}$  is a subset of the image of the compact set  $\{(u_i)_i \in [0, 1]^{d+1} \text{ s.t. } \inf_{i \neq j} |u_i - u_j| \geq \varepsilon/(4K(d+1))\} \times [-D, D]^{d+1}$  by the mapping (17).  $\square$

## B.2 The MLE is in $\mathcal{T}(\alpha, n, D)$

The key idea of this section is that if one of the true trends is far from all parameter trends, then the observations coming from this true trend will significantly reduce the likelihood.

Let  $\alpha \in (0, 1)$ ,  $n \geq 1$ ,  $D > 0$  and  $\theta \notin \mathcal{T}(\alpha, n, D)$ , then

$$\left| \bigcup_{x^* \in [K^*]} \bigcap_{x \in [K^\theta]} I_{n, D}(x, x^*, \theta)^c \right| > n(1 - \alpha)$$

with the notations of Section B.1. In particular, there exists  $x_\mathcal{T}^*(\theta) \in [K^*]$  such that

$$\left| \bigcap_{x \in [K^\theta]} I_{n, D}(x, x_\mathcal{T}^*(\theta), \theta)^c \right| \geq n\frac{1 - \alpha}{K^*}. \quad (18)$$

Write  $I_n^{\text{far}}(\theta) := \bigcap_{x \in [K^\theta]} I_{n,D}(x, x_{\mathcal{T}}^*(\theta), \theta)^{\complement}$ , then

$$\begin{aligned} \frac{1}{n} \ell_n(\theta) &= \frac{1}{n} \sum_{t=1}^n \log p^\theta(Y_t | Y_1^{t-1}) \\ &\leq \log g(0) + \frac{1}{n} \sum_{t \in I_n^{\text{far}}(\theta)} \mathbf{1}_{X_t = x_{\mathcal{T}}^*(\theta)} \log \frac{g(\{D - Z_t^{\max}\}_+)}{g(0)}. \end{aligned} \quad (19)$$

We used the fact that under Assumption **(Amax)**,

$$\begin{aligned} p^\theta(Y_t | Y_1^{t-1}) &= \sum_{x \in [K^\theta]} p^\theta(X_t = x | Y_1^{t-1}) \gamma_x^\theta(Y_t - T_x^\theta(t)) \\ &\leq \sum_{x^* \in [K^*]} \mathbf{1}_{X_t = x^*} \sup_{x \in [K^\theta]} \gamma_x^\theta(Z_t + T_{x^*}^*(t) - T_x^\theta(t)) \\ &\leq \sum_{x^* \in [K^*]} \mathbf{1}_{X_t = x^*} \sup_{x \in [K^\theta]} g(\{|T_{x^*}^*(t) - T_x^\theta(t)| - Z_t^{\max}\}_+) \\ &\leq \sum_{x^* \in [K^*]} \mathbf{1}_{X_t = x^*} g(\inf_{x \in [K^\theta]} \{|T_{x^*}^*(t) - T_x^\theta(t)| - Z_t^{\max}\}_+). \end{aligned}$$

**Lemma 22.** For all  $\theta \in \Theta$ ,  $D > 0$  and  $x^* \in [K^*]$ , the set

$$\bigcap_{x \in [K^\theta]} \{t \in [0, n] : |T_{x^*}^*(t) - T_x^\theta(t)| > D\} \quad (20)$$

has at most  $A := (d+1)^K$  connected components.

Note that  $I_n^{\text{far}}(\theta) = (20) \cap \{1, \dots, n\}$ .

*Proof.* The functions  $(t \mapsto T_{x^*}^*(t) - T_x^\theta(t))_{x \in [K^\theta]}$  are polynomials whose degree is at most  $d$ . Their derivatives vanish at most  $d-1$  times, and the set of times  $t$  where they are larger than  $D$  in absolute value is a union of segments containing either, a zero of their derivative,  $+\infty$  or  $-\infty$ . Hence there are at most  $d+1$  such segments. Thus,  $\tilde{I}_{n,D,x^*}(\theta)$  is an intersection of at most  $K$  sets, each of them having at most  $d+1$  connected components. Therefore, one may take  $A = (d+1)^K$ .  $\square$

**Definition 23.** For all  $n \in \mathbb{N}^*$ ,  $D > 0$ ,  $x^* \in [K^*]$  and  $\theta \in \Theta$ , write  $J(n, D, x^*, \theta)$  the largest connected component of (20). In case of tie, choose the first one for the usual order in  $\mathbb{R}$ .

Thus, by the pigeonhole principle and equation (18),

$$|J(n, D, x_{\mathcal{T}}^*(\theta), \theta) \cap \{1, \dots, n\}| \geq n \frac{1 - \alpha}{AK^*}.$$

Write  $J_n^{\text{far}}(\theta) := J(n, D, x_{\mathcal{T}}^*(\theta), \theta) \cap \{1, \dots, n\}$ , then using equation (19):

$$\frac{1}{n} \ell_n(\theta) \leq \log g(0) + \frac{1}{n} \sum_{t \in J_n^{\text{far}}(\theta)} \mathbf{1}_{X_t = x_{\mathcal{T}}^*(\theta)} \log \frac{g(\{D - Z_t^{\max}\}_+)}{g(0)}. \quad (21)$$

**Lemma 24.** Let  $\delta > 0$  and assume **(Aerg)**. Then, almost surely,

$$\liminf_{n \rightarrow \infty} \inf_{\substack{S \subset \{1, \dots, n\} \\ S \text{ segment} \\ |S| \geq \delta n}} \inf_{x^* \in [K^*]} \frac{1}{n} \sum_{t \in S} \mathbf{1}_{X_t = x^*} \geq \frac{\delta \sigma_-}{4}.$$

By “segment”, we mean a set of the form  $[a, b] \cap \mathbb{Z}$  for some  $(a, b) \in \mathbb{R}^2$ .

*Proof.* The idea is to split  $\{1, \dots, n\}$  into segments of size  $\frac{\delta}{2}n$  and to control the infimum of the empirical mean over each segment. Each segment of size larger than  $\delta n$  contains at least one of those segments. The proof is detailed in Section D.1.  $\square$

Apply Lemma 24 to  $S = J_n^{\text{far}}(\theta)$ : almost surely,

$$\liminf_{n \rightarrow \infty} \inf_{\theta \notin \mathcal{T}(\alpha, n, D)} \frac{1}{n} \sum_{t \in J_n^{\text{far}}(\theta)} \mathbf{1}_{X_t = x_T^*(\theta)} \geq \frac{(1 - \alpha)\sigma_-}{4AK^*}. \quad (22)$$

**Lemma 25.** Let  $\delta \in (0, 1)$ ,  $(U_t)_{t \geq 1}$  a sequence of i.i.d. non-positive integrable random variables and  $(\delta_n)_{n \geq 1}$  a non-decreasing sequence of  $[0, 1]$ -valued random variables such that  $\liminf_{n \rightarrow \infty} \delta_n \geq \delta$  a.s. For all  $\beta \in [0, 1]$ , let us denote by  $q_U(\beta)$  the  $\beta$ -quantile of  $U_1$ , i.e.

$$q_U(\beta) = \inf\{u \text{ s.t. } \mathbb{P}(U_1 \leq u) \geq \beta\}.$$

Then, almost surely,

$$\limsup_{n \rightarrow \infty} \sup_{\substack{S \subset \{1, \dots, n\} \\ |S| \geq \delta n n}} \frac{1}{n} \sum_{t \in S} U_t \leq \mathbb{E}[U_1 \mathbf{1}_{U_1 > q_U(1-\delta)}].$$

Equivalently, if  $(V_t)_{t \geq 1}$  is a sequence of non-negative i.i.d. integrable random variables and  $(\delta_n)_{n \geq 1}$  a non-increasing sequence of  $[0, 1]$ -valued random variables such that  $\limsup_{n \rightarrow \infty} \delta_n \leq \delta$  a.s., almost surely

$$\limsup_{n \rightarrow \infty} \sup_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq \delta n n}} \frac{1}{n} \sum_{t \in S} V_t \leq \mathbb{E}[V_1 \mathbf{1}_{V_1 \geq q_V(1-\delta)}].$$

**Remark.** The supremum is taken over all subsets  $S$ , not only segments.

*Proof.* Proof in Section D.2.  $\square$

For all  $t \geq 1$  and  $D > 0$ , let  $U_t^D = \log \frac{g(\{D - Z_t^{\max}\}_+)}{g(0)}$ .  $U_t^D$  is non-positive by definition. Then, taking

$$\begin{cases} \delta = \frac{(1-\alpha)\sigma_-}{4AK^*}, \\ \delta_n = \inf_{m \geq n} \inf_{\theta \notin \mathcal{T}(\alpha, m, D)} \frac{1}{m} \sum_{t \in J_m^{\text{far}}(\theta)} \mathbf{1}_{X_t = x_T^*(\theta)}, \end{cases}$$

one has  $\liminf_{n \rightarrow \infty} \delta_n \geq \delta$  by equation (22). Therefore, Lemma 25 combined with equation (21) implies that almost surely,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \notin \mathcal{T}(\alpha, n, D)} \frac{1}{n} \ell_n(\theta) \leq \log g(0) + \mathbb{E}[U_1^D \mathbf{1}_{U_1^D \geq q_{UD}(1 - \frac{(1-\alpha)\sigma_-}{4AK^*})}].$$

Note that  $U_t^D = f^D(Z_t^{\max}) \leq 0$  where  $f^D : z \in \mathbb{R}_+ \mapsto \log \frac{g(\{D-z\}_+)}{g(0)}$ .  $f^D$  is non-decreasing, so  $q_{f^D(Z^{\max})}(1 - \delta) \leq f^D(q_{Z^{\max}}(1 - \delta))$  for all  $\delta > 0$ . Thus, for all  $z \geq 0$ ,

$$\mathbf{1}_{z \geq q_{Z^{\max}}(1 - \delta)} \leq \mathbf{1}_{f^D(z) \geq f^D(q_{Z^{\max}}(1 - \delta))} \leq \mathbf{1}_{f^D(z) \geq q_{UD}(1 - \delta)},$$

and since  $U_1^D \leq 0$ ,

$$\mathbb{E}[U_1^D \mathbf{1}_{U_1^D \geq q_{UD}(1 - \frac{(1-\alpha)\sigma_-}{4AK^*})}] \leq \mathbb{E}[U_1^D \mathbf{1}_{Z_1^{\max} \geq q_{Z^{\max}}(1 - \frac{(1-\alpha)\sigma_-}{4AK^*})}].$$

Then, for all  $\delta > 0$ , the monotone convergence theorem applied to the right-hand side entails

$$\mathbb{E}[U_1^D \mathbf{1}_{U_1^D \geq q_{UD}(1 - \delta)}] \xrightarrow{D \rightarrow +\infty} -\infty.$$

Thus, under the assumptions of Corollary 4, there exists  $D(\alpha) < \infty$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \notin \mathcal{T}(\alpha, n, D(\alpha))} \frac{1}{n} \ell_n(\theta) \leq \ell(\theta^*) - 1,$$

so that almost surely, for  $n$  large enough,

$$\hat{\theta}_n \in \mathcal{T}(\alpha, n, D(\alpha)).$$

### B.3 The MLE is in $\mathcal{U}(\beta, n, B)$

Let  $\alpha, \beta \in (0, 1)$ ,  $n \geq \frac{4K(d+1)}{1-\alpha}$ ,  $D > 0$ ,  $B > M(\alpha, D)$  and  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ . Since  $\theta \notin \mathcal{U}(\beta, n, B)$ , one of its trends is far from all true trends. We show that removing the state of this trend increases the likelihood of the observations. An interpretation is that because of **(Aerg)**, a proportion at least  $\sigma_-$  of the observations is supposed to come from the (superfluous) state, but no observations appears in the vicinity of its trend, which penalizes the likelihood.

By definition of  $\mathcal{U}(\beta, n, B)$ , there exists  $x_{\mathcal{U}}(\theta) \in [K^\theta]$  such that

$$\left| \bigcap_{x^* \in [K^*]} I_{n,B}(x_{\mathcal{U}}(\theta), x^*, \theta)^{\mathbb{C}} \right| > n(1 - \beta). \quad (23)$$

Write  $I_n^U(\theta) := \bigcap_{x^* \in [K^*]} I_{n,B}(x_{\mathcal{U}}(\theta), x^*, \theta)^{\mathbb{C}}$  and let  $\theta^U$  be the parameter of the HMM with  $K^\theta - 1$  hidden states, with hidden state space  $\mathcal{X}^{\theta^U} = [K^\theta] \setminus \{x_{\mathcal{U}}(\theta)\}$  and other parameters defined by

$$\begin{cases} \forall x \in \mathcal{X}^{\theta^U}, & \pi^{\theta^U}(x) = \frac{\pi^\theta(x)}{1 - \pi^\theta(x_{\mathcal{U}}(\theta))}, \\ \forall x, x' \in \mathcal{X}^{\theta^U}, & Q^{\theta^U}(x, x') = \frac{Q^\theta(x, x')}{1 - Q^\theta(x, x_{\mathcal{U}}(\theta))}, \\ \forall x \in \mathcal{X}^{\theta^U}, & \gamma_x^{\theta^U} = \gamma_x^\theta, \\ \forall x \in \mathcal{X}^{\theta^U}, & T_x^{\theta^U} = T_x^\theta. \end{cases}$$

By construction of  $\Theta$ ,  $\theta^U \in \Theta$ . Note that for all  $x, x' \in \mathcal{X}^{\theta^U}$ ,

$$\begin{aligned}\pi^{\theta^U}(x) &= \mathbb{P}^\theta(X_1 = x \mid X_1 \neq x_U(\theta)) \\ &= \mathbb{P}^\theta(X_1 = x \mid \forall t \geq 1, X_t \neq x_U(\theta))\end{aligned}$$

and

$$\begin{aligned}\forall s \geq 1, \quad Q^{\theta^U}(x, x') &= \mathbb{P}^\theta(X_{s+1} = x' \mid X_s = x, X_{s+1} \neq x_U(\theta)) \\ &= \mathbb{P}^\theta(X_{s+1} = x' \mid X_s = x, \forall t \geq 1, X_t \neq x_U(\theta)),\end{aligned}$$

so that

$$\forall \mathcal{A} \in \sigma(Y_t \mid t \geq 1), \quad \mathbb{P}^{\theta^U}(\mathcal{A}) = \mathbb{P}^\theta(\mathcal{A} \mid \forall t \geq 1, X_t \neq x_U(\theta)).$$

Then

$$\frac{1}{n} \ell_n(\theta^U) = \frac{1}{n} \sum_{t=1}^n \log p^\theta(Y_t \mid Y_1^{t-1}, x_U(\theta) \notin X_1^t),$$

with the abuse of notation  $x \in X_1^t \Leftrightarrow (\exists s \in \{1, \dots, t\} X_s = x)$ . By Assumption **(Aerg)**,

$$\begin{aligned}p^\theta(Y_t \mid Y_1^{t-1}) &= p^\theta(Y_t \mid X_t = x_U(\theta)) p^\theta(X_t = x_U(\theta) \mid Y_1^{t-1}) \\ &\quad + p^\theta(Y_t \mid X_t \neq x_U(\theta), Y_1^{t-1}) p^\theta(X_t \neq x_U(\theta) \mid Y_1^{t-1}) \\ &\leq (1 - \sigma_-) p^\theta(Y_t \mid X_t = x_U(\theta)) + (1 - \sigma_-) p^\theta(Y_t \mid X_t \neq x_U(\theta), Y_1^{t-1}),\end{aligned}$$

hence

$$\begin{aligned}\frac{1}{n} \ell_n(\theta) - \underbrace{\frac{1}{n} \sum_{t=1}^n \log p^\theta(Y_t \mid X_t \neq x_U(\theta), Y_1^{t-1})}_{(i)} &\leq \log(1 - \sigma_-) \\ &\quad + \underbrace{\frac{1}{n} \sum_{t=1}^n \log \left( 1 + \frac{p^\theta(Y_t \mid X_t = x_U(\theta))}{p^\theta(Y_t \mid X_t \neq x_U(\theta), Y_1^{t-1})} \right)}_{(ii)}.\end{aligned}$$

The next steps are:

- Prove that (i) is close to  $\frac{1}{n} \ell_n(\theta^U)$  for large enough  $n$ .
- Prove that (ii) goes to zero uniformly in  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ .

**First step: controlling (i)** We shall prove that for an adequate choice of  $\beta$  and  $B$ , almost surely

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \left| (i) - \frac{1}{n} \ell_n(\theta^U) \right| \leq \frac{-\log(1 - \sigma_-)}{3}.$$

The following forgetting property allows to control what happens in the distant past.

**Lemma 26.** For all  $t \geq 1$ ,  $\theta \in \Theta$ , for any probability measures  $\mu$  and  $\nu$  on  $[K^\theta]$ , for all  $x \in [K^\theta]$  and  $Y_0^t$ ,

$$|\log p^\theta(Y_t | Y_0^{t-1}, X_t \neq x, X_0 \sim \mu) - \log p^\theta(Y_t | Y_0^{t-1}, X_t \neq x, X_0 \sim \nu)| \leq C\rho^t$$

where  $\rho = 1 - \frac{\sigma_-}{1-\sigma_-}$  and  $C = \frac{2}{\rho(1-\rho)^3}$ .

*Proof.* Proof in Section B.4.1. □

Let  $a \in \mathbb{N}^*$ . It follows from Lemma 26 that for all  $t$ , almost surely,

$$|\log p^\theta(Y_t | Y_1^{t-1}, X_t \neq x_U(\theta)) - \log p^\theta(Y_t | Y_1^{t-1}, X_t \neq x_U(\theta), x_U(\theta) \notin X_1^{t-a})| \leq C\rho^a. \quad (24)$$

It remains to add  $X_{t-a+1}^{t-1}$  to the conditioning. This is the goal of the following lemma.

**Lemma 27.** Assume **(Amax)** and **(Amin)**. Then for all  $a \in \mathbb{N}^*$ ,

$$\begin{aligned} & \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \left| \frac{1}{n} \sum_{t=1}^n \log p^\theta(Y_t | Y_1^{t-1}, X_t \neq x_U(\theta), x_U(\theta) \notin X_1^{t-a}) \right. \\ & \qquad \qquad \qquad \left. - \frac{1}{n} \sum_{t=1}^n \log p^\theta(Y_t | Y_1^{t-1}, x_U(\theta) \notin X_1^t) \right| \\ & \leq \frac{2aK^2}{\sigma_-^3} \left( \beta + \frac{1}{n} \sum_{i=1}^{n-1} \left( 1 \wedge \frac{g(\{B - Z_i^{\max}\}_+)}{m(Z_i^{\max} + M(\alpha, D))} \right) \right) \\ & =: \frac{2aK^2}{\sigma_-^3} \left( \beta + \frac{1}{n} \sum_{i=1}^{n-1} h_U(B, Z_i^{\max}) \right). \end{aligned}$$

*Overview of the proof.* First, show that for all  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ ,  $t \leq n$ ,  $a \in \mathbb{N}^*$  and  $y_1^t \in \mathcal{Y}^t$ ,

$$\begin{aligned} & \left| \log p^\theta(y_t | y_1^{t-1}, X_t \neq x_U(\theta), x_U(\theta) \notin X_1^{t-a}) - \log p^\theta(y_t | y_1^{t-1}, x_U(\theta) \notin X_1^t) \right| \\ & \leq \frac{2}{\sigma_-} p^\theta \left( x_U(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, X_t \neq x_U(\theta), x_U(\theta) \notin X_1^{t-a} \right) \end{aligned} \quad (25)$$

$$\leq \frac{2K^2}{\sigma_-^3} \sum_{i=(t-a+1)\vee 1}^{t-1} \frac{\gamma_{x_U(\theta)}^\theta(y_i - T_{x_U(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))}. \quad (26)$$

Then, under **(Amax)** and **(Amin)**, for all  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$  and  $i \leq n$

$$\frac{\gamma_{x_U(\theta)}^\theta(Y_i - T_{x_U(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} \leq \begin{cases} 1 \wedge \frac{g(\{B - Z_i^{\max}\}_+)}{m(Z_i^{\max} + M(\alpha, D))} =: h_U(B, Z_i^{\max}) & \text{if } i \in I_n^U(\theta), \\ 1 & \text{if } i \notin I_n^U(\theta). \end{cases} \quad (27)$$

so that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\gamma_{x_{\mathcal{U}}(\theta)}^\theta(Y_i - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} &\leq \frac{1}{n} \sum_{i \notin I_n^U(\theta)} 1 + \frac{1}{n} \sum_{i \in I_n^U(\theta)} h_U(B, Z_i^{\max}) \\ &\leq \beta + \frac{1}{n} \sum_{i=1}^n h_U(B, Z_i^{\max}) \end{aligned}$$

since  $|I_n^U(\theta)| > n(1 - \beta)$  by Equation (23) and  $h_U(b, z) \geq 0$  for all  $b, z \geq 0$ . The lemma follows by summing equation (26) over  $t$ . The details of the proof can be found in Section B.4.2.  $\square$

Thus, equation (24) and Lemma 27 imply

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \left| (i) - \frac{1}{n} \ell_n(\theta^U) \right| \leq C\rho^a + \frac{2aK^2}{\sigma_-^3} (\beta + \mathbb{E}^* [h_U(B, Z_i^{\max})]).$$

Now choose  $a$  large enough that

$$C\rho^a \leq \frac{-\log(1 - \sigma_-)}{9},$$

then  $\beta$  such that

$$\frac{2aK^2}{\sigma_-^3} \beta \leq \frac{-\log(1 - \sigma_-)}{9}.$$

Finally, note that  $0 \leq h_U(b, z) \leq 1$  for all  $b, z \geq 0$  and that  $h_U(b, z) \rightarrow 0$  when  $b \rightarrow \infty$  for all  $z$ , so that by the dominated convergence theorem, there exists  $B$  such that

$$\frac{2aK^2}{\sigma_-^3} \mathbb{E}^* [h_U(B, Z_i^{\max})] \leq \frac{-\log(1 - \sigma_-)}{9},$$

which ensures that for all  $\alpha \in (0, 1)$  and  $D > 0$ , there exists  $\beta \in (0, 1)$  and  $B > 0$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \left| (i) - \frac{1}{n} \ell_n(\theta^U) \right| \leq \frac{-\log(1 - \sigma_-)}{3}.$$

This concludes the proof of the first step.

**Second step: controlling (ii)**

**Lemma 28.** Assume **(Amax)** and **(Amin)**. Then

$$\begin{aligned}
& \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \quad (ii) \\
& \leq \frac{1}{n} \sum_{t=1}^n \log \left( 1 + \frac{g(\{B - Z_t^{\max}\}_+)}{\sigma_- m(Z_t^{\max} + M(\alpha, D))} \right) \\
& \quad + \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{t \notin I_n^U(\theta)} \log \left( \frac{g(0) + \sigma_- m(Z_t^{\max} + M(\alpha, D))}{g(\{B - Z_t^{\max}\}_+) + \sigma_- m(Z_t^{\max} + M(\alpha, D))} \right) \\
& =: \frac{1}{n} \sum_{t=1}^n h'_U(B, Z_t^{\max}) + \frac{1}{n} \sum_{t \notin I_n^U(\theta)} V_t^B.
\end{aligned}$$

*Proof.* We show that

$$\frac{p^\theta(Y_t | X_t = x_{\mathcal{U}}(\theta))}{p^\theta(Y_t | X_t \neq x_{\mathcal{U}}(\theta), Y_1^{t-1})} \leq \begin{cases} \frac{g(\{B - Z_t^{\max}\}_+)}{\sigma_- m(Z_t^{\max} + M(\alpha, D))} & \text{if } t \in I_n^U(\theta), \\ \frac{g(0)}{\sigma_- m(Z_t^{\max} + M(\alpha, D))} & \text{if } t \notin I_n^U(\theta). \end{cases}$$

The lemma follows by summing over  $t$ . The details of the proof can be found in Section B.4.3.  $\square$

Note that under Assumption **(Aint)**,

$$\mathbb{E}^*[-\log m(Z_t^{\max} + M(\alpha, D))] < \infty.$$

Hence,

$$\begin{aligned}
\mathbb{E}^* |h'_U(0, Z_t^{\max})| & \leq \mathbb{E}^* [\{\log(\sigma_- m(Z_t^{\max} + M(\alpha, D)) + g(0))\}_+] \\
& \quad + \mathbb{E}^*[-\log(\sigma_- m(Z_t^{\max} + M(\alpha, D)))] \\
& \leq \log 2 + |\log g(0)| + \mathbb{E}^*[-\log m(Z_t^{\max} + M(\alpha, D))] - \log \sigma_- \\
& < \infty.
\end{aligned}$$

Thus, since  $b \mapsto h'_U(b, z)$  is nonincreasing and converges to zero when  $b \rightarrow \infty$  for all  $z$ , the dominated convergence theorem together with the law of large numbers imply that there exists  $B$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n h'_U(B, Z_t^{\max}) \leq \frac{-\log(1 - \sigma_-)}{6}.$$

Then, apply Lemma 25 to the i.i.d. non-negative random variables  $(V_t^B)_{t \geq 1}$  using the fact that  $|I_n^U(\theta)| > n(1 - \beta)$ , which yields

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \frac{1}{n} \sum_{t \notin I_n^U(\theta)} V_t^B \leq \mathbb{E}^*[V_1^B \mathbf{1}_{V_1^B \geq q_{V^B}(1-\beta)}].$$

Note that

$$\mathbb{E}^* V_1^B \leq \log((1 + \sigma_-)g(0)) - \log \sigma_- + \mathbb{E}^*[-\log m(Z_t^{\max} + M(\alpha, D))],$$

which is finite thanks to **(Aint)**. Thus,

$$\mathbb{E}^*[V_1^B \mathbf{1}_{V_1^B \geq q_{VB}(1-\beta)}] \xrightarrow{\beta \rightarrow 0} 0,$$

so that there exists  $\beta$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \frac{1}{n} \sum_{t \notin I_n^U(\theta)} V_t^B \leq \frac{-\log(1 - \sigma_-)}{6}.$$

Hence, we proved that there exists  $\beta(\alpha, D) \in (0, 1)$  and  $B(\alpha, D) > 0$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} (ii) \leq \frac{-\log(1 - \sigma_-)}{3},$$

which ends the second step.

Putting together the results of the two steps, one gets that for all  $\alpha \in (0, 1)$  and  $D > 0$ , there exists  $\beta \in (0, 1)$  and  $B > 0$  such that almost surely,

$$\limsup_{n \rightarrow \infty} \left( \sup_{\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)} \frac{1}{n} \ell_n(\theta) - \sup_{\theta \in \Theta} \frac{1}{n} \ell_n(\theta) \right) \leq \frac{\log(1 - \sigma_-)}{3} < 0,$$

so that for  $n$  large enough,  $\hat{\theta}_n \notin (\mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B))$ .

Together with Section B.2, this implies that for all  $\alpha \in (0, 1)$ , there exists  $\beta \in (0, 1)$  and  $D, B > 0$  such that  $\hat{\theta}_n \in \mathcal{T}(\alpha, n, D) \cap \mathcal{U}(\beta, n, B)$  for  $n$  large enough, which entails Theorem 5 by equation (16).

## B.4 Proofs

### B.4.1 Proof of Lemma 26

We shall use the inequality

$$|\log C_\mu - \log C_\nu| \leq \frac{|C_\mu - C_\nu|}{C_\mu \wedge C_\nu}$$

with  $C_\mu = p^\theta(Y_t | Y_0^{t-1}, X_t \neq x, X_0 \sim \mu)$  and  $C_\nu = p^\theta(Y_t | Y_0^{t-1}, X_t \neq x, X_0 \sim \nu)$ .

$$\begin{aligned} |C_\mu - C_\nu| &= \left| \frac{p^\theta(Y_t, X_t \neq x | Y_0^{t-1}, X_0 \sim \mu)}{p^\theta(X_t \neq x | Y_0^{t-1}, X_0 \sim \mu)} - \frac{p^\theta(Y_t, X_t \neq x | Y_0^{t-1}, X_0 \sim \nu)}{p^\theta(X_t \neq x | Y_0^{t-1}, X_0 \sim \nu)} \right| \\ &=: \left| \frac{B_\mu}{A_\mu} - \frac{B_\nu}{A_\nu} \right|. \end{aligned}$$

One has

$$\begin{aligned} B_\mu &= \sum_{x' \neq x} p^\theta(Y_t | X_t = x') p^\theta(X_t = x' | Y_0^{t-1}, X_0 \sim \mu) \\ &= \sum_{x' \neq x} p^\theta(Y_t | X_t = x') \sum_{x'' \in [K^\theta]} Q_{x''x'}^\theta p^\theta(X_{t-1} = x'' | Y_0^{t-1}, X_0 \sim \mu), \end{aligned}$$

which yields

$$\sigma_- \sum_{x' \neq x} p^\theta(Y_t | X_t = x') \leq B_\mu \leq (1 - \sigma_-) \sum_{x' \neq x} p^\theta(Y_t | X_t = x')$$

and the same result holds for  $B_\nu$ . Besides,

$$\begin{aligned} A_\mu &= \sum_{x \neq x'} p^\theta(X_t = x' | Y_0^{t-1}, X_0 \sim \mu) \\ &= \sum_{x' \neq x} \sum_{x'' \in [K^\theta]} Q_{x''x'}^\theta p^\theta(X_{t-1} = x'' | Y_0^{t-1}, X_0 \sim \mu). \end{aligned}$$

Hence,

$$\sigma_- \leq A_\mu \leq 1 - \sigma_-$$

and the same result holds for  $A_\nu$ . Then, letting  $\phi_\mu(x') = p^\theta(X_{t-1} = x' | Y_0^{t-1}, X_0 \sim \mu)$ , we get, using the above expressions:

$$\begin{aligned} |A_\mu - A_\nu| &\leq (1 - \sigma_-) \|\phi_\mu - \phi_\nu\|_1 \\ |B_\mu - B_\nu| &\leq (1 - \sigma_-) \sum_{x' \neq x} p^\theta(Y_t | X_t = x') \|\phi_\mu - \phi_\nu\|_1. \end{aligned}$$

Thus,

$$\begin{aligned} |C_\mu - C_\nu| &= \left| \frac{B_\mu}{A_\mu} - \frac{B_\nu}{A_\nu} \right| \\ &\leq \frac{1}{A_\mu A_\nu} (B_\mu |A_\mu - A_\nu| + A_\mu |B_\mu - B_\nu|) \\ &\leq \frac{2(1 - \sigma_-)^2}{\sigma_-^2} \sum_{x' \neq x} p^\theta(Y_t | X_t = x') \|\phi_\mu - \phi_\nu\|_1. \end{aligned}$$

Furthermore,

$$\frac{1}{C_\mu \wedge C_\nu} \leq \frac{(1 - \sigma_-)}{\sigma_- \sum_{x' \neq x} p^\theta(Y_t | X_t = x')}$$

Finally,

$$|\log C_\mu - \log C_\nu| \leq \frac{2}{(1 - \rho)^3} \|\phi_\mu - \phi_\nu\|_1.$$

It remains to prove that  $\|\phi_\mu - \phi_\nu\|_1 \leq \rho^{t-1}$ , which follows from the geometric ergodicity of the HMM. See for instance Corollary 1 of [Douc et al. \(2004\)](#) or Proposition 2.1 of [De Castro et al. \(2017\)](#).

### B.4.2 Proof of Lemma 27

**Proof of equation (25)** For all  $t \geq 1$  and  $y_1^t \in \mathbb{R}^t$ ,

$$\begin{aligned} & p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\ &= p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^t) p^\theta(x_{\mathcal{U}}(\theta) \notin X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\ & \quad + p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta), x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1}) \\ & \quad \times p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)), \end{aligned}$$

so that

$$\begin{aligned} & |p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) - p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^t)| \\ & \leq p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\ & \quad \times \left( p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^t) + p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta), x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1}) \right) \\ & \leq 2p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \sum_{x \in [K^\theta] \setminus \{x_{\mathcal{U}}(\theta)\}} \gamma_x^\theta(y_t - T_x^\theta(t)). \end{aligned}$$

In addition, under **(Aerg)**,

$$\begin{aligned} & p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\ &= \sum_{x \in [K^\theta] \setminus \{x_{\mathcal{U}}(\theta)\}} p^\theta(y_t | X_t = x) p^\theta(X_t = x | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}) \\ & \geq \sigma_- \sum_{x \in [K^\theta] \setminus \{x_{\mathcal{U}}(\theta)\}} \gamma_x^\theta(y_t - T_x^\theta(t)) \end{aligned}$$

and the same holds for  $p^\theta(y_t | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^t)$ , so that using that  $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$  for all  $x, y > 0$ , we obtain that for all  $t \geq 1$  and  $y_1^t \in \mathbb{R}^t$

$$\begin{aligned} & |\log p^\theta(Y_t | X_1^{t-a} \neq x_{\mathcal{U}}(\theta), X_t \neq x_{\mathcal{U}}(\theta), Y_1^{t-1}) - \log p^\theta(Y_t | X_1^t \neq x_{\mathcal{U}}(\theta), Y_1^{t-1})| \\ & \leq \frac{2}{\sigma_-} p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)). \end{aligned}$$

**Proof of equation (26)** By union bound,

$$\begin{aligned}
& p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\
& \leq \sum_{i=(t-a+1)\vee 1}^{t-1} p^\theta(X_i = x_{\mathcal{U}}(\theta) | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\
& = \sum_{i=(t-a+1)\vee 1}^{t-1} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} p^\theta(X_i = x_{\mathcal{U}}(\theta) | y_i, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}) \\
& \quad \times p^\theta(X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\
& \leq \sum_{i=(t-a+1)\vee 1}^{t-1} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} p^\theta(X_i = x_{\mathcal{U}}(\theta) | y_i, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}) \\
& = \sum_{i=(t-a+1)\vee 1}^{t-1} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} \frac{p^\theta(X_i = x_{\mathcal{U}}(\theta) | X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}) \gamma_{x_{\mathcal{U}}(\theta)}^\theta(y_i - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} p^\theta(X_i = x | X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}) \gamma_x^\theta(y_i - T_x^\theta(i))}.
\end{aligned}$$

Using the Markov property and **(Aerg)**, for all  $x_{i-1}, x_{i+1} \in [K^\theta]$ ,

$$p^\theta(X_i = x | X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}) \in [\sigma_-^2, 1].$$

Hence,

$$\begin{aligned}
& p^\theta(x_{\mathcal{U}}(\theta) \in X_{(t-a+1)\vee 1}^{t-1} | y_1^{t-1}, x_{\mathcal{U}}(\theta) \notin X_1^{t-a}, X_t \neq x_{\mathcal{U}}(\theta)) \\
& \leq \sum_{i=(t-a+1)\vee 1}^{t-1} \sum_{x_{i-1}, x_{i+1} \in [K^\theta]} \frac{\gamma_{x_{\mathcal{U}}(\theta)}^\theta(y_i - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sigma_-^2 \sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))} \\
& \leq \frac{K^2}{\sigma_-^2} \sum_{i=(t-a+1)\vee 1}^{t-1} \frac{\gamma_{x_{\mathcal{U}}(\theta)}^\theta(y_i - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(y_i - T_x^\theta(i))}
\end{aligned}$$

which concludes the proof.

**Proof of equation (27)** This quantity is always bounded by 1 since all terms are nonnegative. In addition, under Assumptions **(Amax)** and **(Amin)**, for all  $i \in I_n^U(\theta)$ ,

$$\begin{aligned}
& \frac{\gamma_{x_{\mathcal{U}}(\theta)}^\theta(Y_i - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} \leq \sum_{x \in [K^*]} \mathbf{1}_{X_i = x^*} \left( 1 \wedge \frac{\gamma_{x_{\mathcal{U}}(\theta)}^\theta(Z_i + T_{x^*}^*(i) - T_{x_{\mathcal{U}}(\theta)}^\theta(i))}{\sup_{x \in [K^\theta]} \gamma_x^\theta(Z_i + T_x^*(i) - T_x^\theta(i))} \right) \\
& \leq 1 \wedge \frac{g \left( \left\{ \inf_{x^* \in [K^*]} |T_{x^*}^*(i) - T_{x_{\mathcal{U}}(\theta)}^\theta(i)| - Z_i^{\max} \right\}_+ \right)}{m \left( Z_i^{\max} + \sup_{x^* \in [K^*]} \inf_{x \in [K^\theta]} |T_{x^*}^*(i) - T_x^\theta(i)| \right)}.
\end{aligned}$$

Since  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ , Corollary 17 ensures that  $\inf_{x \in [K^\theta]} |T_{x^*}^*(i) - T_x^\theta(i)| \leq M(\alpha, D)$  for all  $x^* \in [K^*]$  and  $i \in \{1, \dots, n\}$ . Moreover, by definition of  $I_n^U(\theta)$ ,  $\inf_{x^* \in [K^*]} |T_{x^*}^*(i) - T_{x_{\mathcal{U}}^\theta}^\theta(i)| \geq B$  for all  $i \in I_n^U(\theta)$ , so that

$$\frac{\gamma_{x_{\mathcal{U}}^\theta}^\theta(Y_i - T_{x_{\mathcal{U}}^\theta}^\theta(i))}{\sum_{x \in [K^\theta]} \gamma_x^\theta(Y_i - T_x^\theta(i))} \leq 1 \wedge \frac{g(\{B - Z_i^{\max}\}_+)}{m(Z_i^{\max} + M(\alpha, D))}$$

for all  $i \in I_n^U(\theta)$ , which concludes the proof.

### B.4.3 Proof of Lemma 28

Under Assumption **(Amax)**,

$$\begin{aligned} p^\theta(Y_t | X_t = x_{\mathcal{U}}(\theta)) &= \gamma_{x_{\mathcal{U}}(\theta)}^\theta(Y_t - T_{x_{\mathcal{U}}(\theta)}^\theta(t)) \\ &= \sum_{x^* \in [K^*]} \mathbf{1}_{X_t = x^*} \gamma_{x_{\mathcal{U}}(\theta)}^\theta(Z_t + T_{x^*}^*(t) - T_{x_{\mathcal{U}}(\theta)}^\theta(t)) \\ &\leq \sup_{x^* \in [K^*]} g\left(\{|T_{x^*}^*(t) - T_{x_{\mathcal{U}}(\theta)}^\theta(t)| - Z_t^{\max}\}_+\right) \\ &\leq g\left(\left\{\inf_{x^* \in [K^*]} |T_{x^*}^*(t) - T_{x_{\mathcal{U}}(\theta)}^\theta(t)| - Z_t^{\max}\right\}_+\right), \end{aligned}$$

hence, for all  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ ,

$$p^\theta(Y_t | X_t = x_{\mathcal{U}}(\theta)) \leq \begin{cases} g(\{B - Z_t^{\max}\}_+) & \text{if } t \in I_n^U(\theta), \\ g(0) & \text{otherwise.} \end{cases} \quad (28)$$

On the other hand, under Assumptions **(Amin)** and **(Aerg)**,

$$\begin{aligned} p^\theta(Y_t | X_t \neq x_{\mathcal{U}}(\theta), Y_1^{t-1}) &= \frac{p^\theta(Y_t, X_t \neq x_{\mathcal{U}}(\theta) | Y_1^{t-1})}{p^\theta(X_t \neq x_{\mathcal{U}}(\theta) | Y_1^{t-1})} \\ &\geq \sum_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} p^\theta(Y_t | X_t = x, Y_1^{t-1}) p^\theta(X_t = x | Y_1^{t-1}) \\ &\geq \sigma_- \sum_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} p^\theta(Y_t | X_t = x) \\ &= \sigma_- \sum_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} \gamma_x^\theta(Y_t - T_x^\theta(t)) \\ &= \sigma_- \sum_{x^* \in [K^*]} \mathbf{1}_{X_t = x^*} \sum_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} \gamma_x^\theta(Z_t + T_{x^*}^*(t) - T_x^\theta(t)) \\ &\geq \sigma_- \inf_{x^* \in [K^*]} \sup_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} m(Z_t^{\max} + |T_{x^*}^*(t) - T_x^\theta(t)|) \\ &\geq \sigma_- m \left( Z_t^{\max} + \sup_{x^* \in [K^*]} \inf_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} |T_{x^*}^*(t) - T_x^\theta(t)| \right). \end{aligned}$$

Using  $\theta \notin \mathcal{T}(\alpha, n, D)$  and Corollary 17, for all  $x^* \in [K^*]$ ,

$$\inf_{x \in [K^\theta], x \neq x_{\mathcal{U}}(\theta)} |T_{x^*}^*(t) - T_x^\theta(t)| \leq M(\alpha, D).$$

For example, we can choose  $x = x(\theta, x^*, n)$  (defined in Definition 16). We know that  $x_{\mathcal{U}}(\theta) \neq x(\theta, x^*, n)$  because we chose  $B > M(\alpha, D)$ , so that  $|T_{x^*}^*(i) - T_{x_{\mathcal{U}}(\theta)}^\theta(i)| > M(\alpha, D)$  for at least one  $i \in \{1, \dots, n\}$ , and because  $|T_{x^*}^*(i) - T_{x(\theta, x^*, n)}^\theta(i)| \leq M(\alpha, D)$  for all  $i \in \{1, \dots, n\}$ . Therefore, for all  $\theta \in \mathcal{T}(\alpha, n, D) \setminus \mathcal{U}(\beta, n, B)$ ,

$$p^\theta(Y_t | X_t \neq x_{\mathcal{U}}(\theta), Y_1^{t-1}) \geq \sigma_{-m}(Z_t^{\max} + M(\alpha, D)),$$

which concludes the proof together with equation (28).

## C Integrated log-likelihood

In this section, we use the fact that the observed process  $(Y_t)_{t \geq 1}$  may be replaced by the process  $(Y_t - \mathbb{T}_{B_t}^*(t), B_t)_{t \geq 1}$ . While this process is not homogeneous under the parameter  $\theta$ , its distribution varies slowly over time. We take advantage of this property to show the uniform convergence of the log-likelihood by approximating  $(Y_t - \mathbb{T}_{B_t}^*(t), B_t)_{t \geq 1}$  by a process that is locally (in time) homogeneous under the parameter  $\theta$ . The limit can be written as an integral of limits of log-likelihoods of homogeneous HMM, hence the name *integrated log-likelihood*.

### C.1 Convergence of the log-likelihood to the integrated log-likelihood

Assume **(Aerg)**, **(Amax)**, **(Amin)**, **(Aint)** and **(Areg)**. In this section we shall prove Theorem 10. Let  $M > 0$ .

The normalized log-likelihood associated with the HMM  $(Y_t, B_t)_{t \geq 1}$  can be written as

$$\begin{aligned} \frac{1}{n} \ell_n^{(Y, B)}(\theta) &= \frac{1}{n} \log \sum_{\substack{x_1^n \text{ s.t.} \\ \forall t, \mathbf{b}^\theta(x_t) = B_t}} \pi^\theta(x_1) Q^\theta(x_1, x_2) \dots Q^\theta(x_{n-1}, x_n) \prod_{t=1}^n \gamma_{x_t}^\theta(Y_t - T_{x_t}^\theta(t)) \\ &= \frac{1}{n} \log \sum_{\substack{x_1^n \text{ s.t.} \\ \forall t, \mathbf{b}^\theta(x_t) = B_t}} \pi^\theta(x_1) Q^\theta(x_1, x_2) \dots Q^\theta(x_{n-1}, x_n) \prod_{t=1}^n \gamma_{x_t}^\theta \left( Z'_t - D_{x_t}^{\theta, n} \left( \frac{t}{n} \right) \right), \end{aligned} \tag{29}$$

with  $D_x^{\theta, n} : u \in [0, 1] \mapsto T_x^\theta(nu) - \mathbb{T}_{\mathbf{b}^\theta(x)}^*(nu)$  and  $Z'_t := Y_t - \mathbb{T}_{B_t}^*(t)$ . Note that  $Z'_t = Z_t + \Delta(X_t)$ . Recall that  $\mathcal{D}(M)$  is defined by Equation (5).

Theorem 19 implies that  $\text{Cl}(\mathcal{D}(M))$  is compact, where  $\text{Cl}(\cdot)$  denotes the closure with respect to the supremum norm topology, hence Proposition 9 holds. Together with Arzelà–Ascoli’s theorem, this entails that  $\mathcal{D}(M)$  is uniformly equicontinuous and uniformly bounded by  $M$ . Hence there exists a continuity modulus  $\nu$  such that for all  $\delta > 0$  and all  $(s, u) \in [0, 1]^2$ ,

$$|s - u| \leq \delta \Rightarrow \sup_{n \geq 4K(d+1)} \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \sup_{x \in [K^\theta]} |D_x^{\theta, n}(s) - D_x^{\theta, n}(u)| \leq \nu(\delta). \tag{30}$$

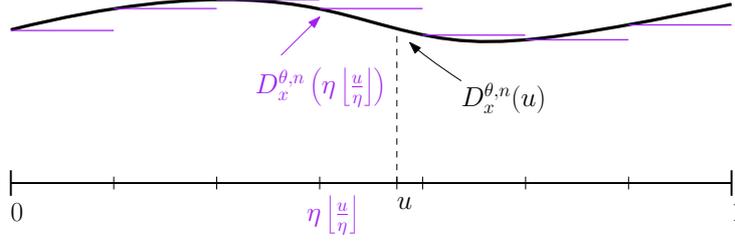


Figure 10: Construction of the trends of the homogenized process.

**Definition 29** (Log-likelihood of the homogenized process). *For each  $\eta > 0$  and  $\theta \in \Theta_n^{\text{OK}}(M)$ , let*

$$\begin{aligned} \frac{1}{n} \ell_n^{(Y,B)}[\eta](\theta) &:= \frac{1}{n} \log \sum_{x_1^n \text{ s.t. } \forall t, \mathbf{b}^\theta(x_t) = B_t} \pi^\theta(x_1) Q^\theta(x_1, x_2) \dots Q^\theta(x_{n-1}, x_n) \\ &\quad \times \prod_{t=1}^n \gamma_{x_t}^\theta \left( Z'_t - D_{x_t}^{\theta, n} \left( \eta \left\lfloor \frac{t}{\eta n} \right\rfloor \right) \right), \end{aligned}$$

be the normalized log-likelihood of the process where each residual trend is made constant over segments of length  $\eta$ .

**Remark.** *This quantity is indeed a log-likelihood:  $\frac{1}{n} \ell_n^{(Y,B)}[\eta](\theta) = \frac{1}{n} \ell_n^{(Y,B)}(\theta[N, n])$ , where the parameter  $\theta[N, n]$  is defined by  $\pi^{\theta[N, n]} = \pi^\theta$ ,  $Q^{\theta[N, n]} = Q^\theta$ ,  $\gamma^{\theta[N, n]} = \gamma^\theta$ ,  $\mathbf{b}^{\theta[N, n]} = \mathbf{b}^\theta$  and*

$$\forall x \in [K^\theta], \quad T_x^{\theta[N, n]}(t) = \mathbb{T}_{\mathbf{b}^\theta(x)}^*(t) + D_x^{\theta, n}(\lfloor N \frac{t}{n} \rfloor). \quad (31)$$

However,  $\theta[N, n]$  has piecewise polynomial trends instead of polynomial trends, so that it does not belong to  $\Theta$ .

$\frac{1}{n} \ell_n^{(Y,B)}[\eta](\theta)$  is an approximation of the log-likelihood of equation (29). Assumption **(Areg)** together with Equation (30) ensure that for all  $\delta > 0$ ,  $n \geq 1$ ,  $\theta \in \Theta_n^{\text{OK}}(M)$ ,  $x \in [K^\theta]$  and  $t \in \{1, \dots, n\}$ ,

$$\gamma_{x_t}^\theta \left( Z'_t - D_{x_t}^{\theta, n} \left( \frac{t}{n} \right) \right) \in \left[ e^{-L(|Z'_t| + M)\omega(\nu(\delta))}, e^{L(|Z'_t| + M)\omega(\nu(\delta))} \right] \gamma_{x_t}^\theta \left( Z'_t - D_{x_t}^{\theta, n} \left( \delta \left\lfloor \frac{t}{\delta n} \right\rfloor \right) \right),$$

hence, for all  $\delta > 0$  and  $n \geq 1$ ,

$$\begin{aligned} \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n^{(Y,B)}(\theta) - \frac{1}{n} \ell_n^{(Y,B)}[\delta](\theta) \right| &\leq \omega(\nu(\delta)) \times \frac{1}{n} \sum_{t=1}^n L(|Z'_t| + M) \\ &\leq \omega(\nu(\delta)) \times \frac{1}{n} \sum_{t=1}^n L(\|\Delta\|_\infty + M + Z_t^{\max}). \quad (32) \end{aligned}$$

**Remark.** *Under **(Areg)**, the law of large numbers entails that almost surely, for all  $N \geq 1$ ,*

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n^{(Y,B)}(\theta) - \frac{1}{n} \ell_n^{(Y,B)} \left[ \frac{1}{N} \right] (\theta) \right| \\ \leq \omega \left( \nu \left( \frac{1}{N} \right) \right) \mathbb{E}^* [L(\|\Delta\|_\infty + M + Z_1^{\max})]. \end{aligned}$$

Recall the following notation. For all  $K' \in \mathbb{N}^*$ , for all  $K'$ -uple  $\gamma = (\gamma_x)_{x \in [K']}$  of measurable functions and for all  $\mathbf{D} = (D_x)_{x \in [K']} \in \mathbb{R}^{K'}$ , let

$$\tau(\gamma, \mathbf{D}) := (z' \mapsto \gamma_x(z' - D_x))_{x \in [K']}$$

the vector of functions  $\gamma$  translated by the vector  $\mathbf{D}$ .

**Definition 30.** Let  $\pi$  be a probability measure on  $[K']$ ,  $Q$  be a  $K' \times K'$  transition matrix,  $\gamma$  be a vector of  $K'$  emission densities on  $\mathbb{R}$  and  $\mathbf{b}$  be a function  $[K'] \rightarrow \mathcal{B}^*$ . Let  $(X_t, (\tilde{Z}_t, \tilde{B}_t))_{t \geq 1}$  be a homogeneous HMM taking values in  $[K'] \times (\mathbb{R} \times \mathcal{B}^*)$  with parameter  $(\pi, Q, (\gamma_x \otimes \mathbf{1}_{\mathbf{b}(x)})_{x \in [K']})$ .

Denote by  $\frac{1}{n} \ell_n^{hom}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{z}, \tilde{b})_1^n\}$  (resp.  $\ell^{hom}(Q, \gamma, \mathbf{b})$ ) the normalized log-likelihood of the parameter  $(\pi, Q, \gamma, \mathbf{b})$  for the observations  $(\tilde{z}, \tilde{b})_1^n$  (resp. the limit of the log-likelihood, if it exists), that is

$$\frac{1}{n} \ell_n^{hom}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{z}, \tilde{b})_1^n\} = \frac{1}{n} \log \sum_{x_1^n \in [K']^n} \pi(x_1) Q(x_1, x_2) \dots Q(x_{n-1}, x_n) \prod_{t=1}^n \gamma_{x_t}(\tilde{z}_t) \mathbf{1}_{\mathbf{b}(x_t) = \tilde{b}_t}$$

and

$$\ell^{hom}(Q, \gamma, \mathbf{b}) = \lim_{n \rightarrow \infty} \frac{1}{n} \ell_n^{hom}(\pi, Q, \gamma, \mathbf{b})\{(\tilde{Z}, \tilde{B})_1^n\}. \quad (33)$$

The following Lemma ensures the existence of the limit of the normalized log-likelihood in Definition 30 as well as its uniform continuity with respect to the parameter. It is a consequence of a result concerning homogeneous HMM stated in Douc et al. (2004).

**Lemma 31.** Assume **(Amax)**, **(Amin)**, **(Aint)** and **(Areg)**. Let  $K' \in \mathbb{N}^*$ . The following points hold.

- Almost surely, for all  $Q \in \Sigma_{K'}^{\sigma^-}$ ,  $\gamma \in \Gamma^{K'}$ ,  $\mathbf{D} \in \mathbb{R}^{K'}$  and  $\mathbf{b} : [K'] \rightarrow \mathcal{B}^*$ , the quantity

$$\ell^{hom}(Q^\theta, \tau(\gamma^\theta, \mathbf{D}), \mathbf{b})$$

from Equation (33) exists and is finite almost surely under  $\mathbb{P}^*$  when  $(\tilde{Z}_t, \tilde{B}_t)_t = (Z'_t, B_t)_t$ .

- For all  $K' \in \mathbb{N}^*$ , the mapping

$$\begin{aligned} (Q^\theta, \gamma^\theta, \mathfrak{D}, u, \mathbf{b}) \in \Sigma_{K'}^{\sigma^-} \times \Gamma^{K'} \times Cl(\mathcal{D}(M))^{K'} \times [0, 1] \times (\mathcal{B}^*)^{K'} \\ \mapsto \ell^{hom}(Q^\theta, \tau(\gamma^\theta, \mathfrak{D}(u)), \mathbf{b}) \end{aligned}$$

is continuous and its domain is compact, so that it is uniformly continuous.

- Almost surely, for all  $N \in \mathbb{N}^*$ ,

$$\begin{aligned} \sup_{(\pi, Q, \gamma, \mathfrak{D}, u, \mathbf{b})} \sup_{s \in \{0, \dots, (N-1)n\}} \left| \frac{1}{n} \ell_n^{hom}(\pi, Q, \tau(\gamma, \mathfrak{D}(u)), \mathbf{b})\{(Z', B)_{s+1}^{s+n}\} \right. \\ \left. - \ell^{hom}(Q, \tau(\gamma, \mathfrak{D}(u)), \mathbf{b}) \right| \xrightarrow[n \rightarrow \infty]{} 0 \end{aligned}$$

where the supremum is taken for  $(\pi, Q, \gamma, \mathfrak{D}, u, \mathbf{b}) \in \Delta_{K'} \times \Sigma_{K'}^{\sigma^-} \times \Gamma^{K'} \times Cl(\mathcal{D}(M))^{K'} \times [0, 1] \times (\mathcal{B}^*)^{K'}$ .

*Proof.* Proof in Section C.3 □

As a consequence, the family of functions

$$\bigcup_{K'=1}^K \{u \in [0, 1] \mapsto \ell^{\text{hom}}(Q, \tau(\gamma, \mathfrak{D}(u)), \mathbf{b})\}_{Q \in \Sigma_{K'}^{\sigma^-}, \gamma \in \Gamma^{K'}, \mathfrak{D} \in \mathcal{D}(M)^{K'}, \mathbf{b} \in (\mathcal{B}^*)^{K'}}$$

is uniformly equicontinuous, which ensures the following result.

**Corollary 32** (Riemann approximation of the integral). *The quantity*

$$R_N := \sup_{n \geq n_1(M)} \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{N} \sum_{i=0}^{N-1} \ell^{\text{hom}} \left( Q^\theta, \tau \left( \gamma^\theta, \mathfrak{D}^{\theta, n} \left( \frac{i}{N} \right) \right), \mathbf{b}^\theta \right) - \int_0^1 \ell^{\text{hom}} \left( Q^\theta, \tau \left( \gamma^\theta, \mathfrak{D}^{\theta, n}(u) \right), \mathbf{b}^\theta \right) du \right| \quad (34)$$

satisfies

$$R_N \xrightarrow{N \rightarrow +\infty} 0.$$

The integrated log-likelihood  $\ell^{\text{int}}$  from Definition 8 is continuous by uniform continuity of  $\ell^{\text{hom}}$ . We may now prove the main result of this section, that is the convergence of the normalized log-likelihood to the integrated log-likelihood.

**Proof of Theorem 10** By the triangle inequality and using Equations (34) and (32), for all  $n \geq n_1(M)$  and  $N \in \mathbb{N}^*$ ,

$$\begin{aligned} & \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n^{(Y, B)}(\theta) - \ell^{\text{int}}(Q^\theta, \gamma^\theta, \mathfrak{D}^{\theta, n}, \mathbf{b}^\theta) \right| \\ & \leq \omega \left( \nu \left( \frac{1}{N} \right) \right) \frac{1}{n} \sum_{t=1}^n L(\|\Delta\|_\infty + M + Z_t^{\max}) + R_N \\ & + \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n^{(Y, B)} \left[ \frac{1}{N} \right] (\theta) - \frac{1}{N} \sum_{i=0}^{N-1} \ell^{\text{hom}} \left( Q^\theta, \tau \left( \gamma^\theta, \mathfrak{D}^{\theta, n} \left( \frac{i}{N} \right) \right), \mathbf{b}^\theta \right) \right|. \end{aligned}$$

For the sake of simplicity, assume that  $\frac{n}{N}$  is an integer. By Equation (31), for all  $\theta \in \Theta_n^{\text{OK}}(M)$ , there exists  $\theta[N, n]$  such that

$$\frac{1}{n} \ell_n^{(Y, B)} \left[ \frac{1}{N} \right] (\theta) = \frac{1}{n} \ell_n^{(Y, B)}(\theta[N, n]),$$

so that

$$\begin{aligned} \frac{1}{n} \ell_n^{(Y, B)} \left[ \frac{1}{N} \right] (\theta) &= \frac{1}{n} \sum_{i=0}^{N-1} \log p^{\theta[N, n]} \left( (Y, B)_{1+i\frac{n}{N}}^{\frac{n}{N}+i\frac{n}{N}} \mid (Y, B)_1^{i\frac{n}{N}} \right) \\ &= \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{n} \ell_{\frac{n}{N}}^{\text{hom}} \left( \pi_{i\frac{n}{N}}^{\theta[N, n]}, Q^\theta, \tau \left( \gamma^\theta, \mathfrak{D}^{\theta, n} \left( \frac{i}{N} \right) \right), \mathbf{b}^\theta \right) \left\{ (Z', B)_{1+i\frac{n}{N}}^{\frac{n}{N}+i\frac{n}{N}} \right\}, \end{aligned}$$

where  $\pi_{i\frac{n}{N}}^{\theta[N,n]}$  is defined as the distribution of  $X_{1+i\frac{n}{N}}$  conditionally to  $(Y, B)_{1+i\frac{n}{N}}^{\frac{n}{N}+i\frac{n}{N}}$  under the parameter  $\theta[N, n]$ . Hence, Lemma 31 implies that almost surely,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \sup_{\theta \in \Theta_n^{\text{OK}}(M)} \left| \frac{1}{n} \ell_n^{(Y, B)}(\theta) - \ell^{\text{int}}(Q^\theta, \gamma^\theta, \mathfrak{D}^{\theta, n}, \mathbf{b}^\theta) \right| \\ \leq \inf_{N \in \mathbb{N}^*} \left[ \omega \left( \nu \left( \frac{1}{N} \right) \right) \mathbb{E}^*[L(\|\Delta\|_\infty + M + Z_1^{\max})] + R_N \right] \\ = 0. \end{aligned}$$

The conclusion follows from Theorem 3.

## C.2 Maximizers of the integrated log-likelihood and identifiability

In this section we prove Proposition 11. Assume **(Amax)**, **(Amin)**, **(Areg)**, **(Aid)** and **(Acentering)** and assume that  $K^* = K$  is known.

**Remark.** Assumptions **(Areg)**, **(Amax)** and **(Amin)** can be replaced by

$$\begin{cases} \forall \theta \in \Theta, \quad \forall x \in [K^*], \quad z \in \mathbb{R} \mapsto \gamma^\theta(z) \text{ is continuous,} \\ \forall x \in [K^*], \quad \gamma_x^*(z) \xrightarrow{|z| \rightarrow +\infty} 0 \\ \forall x \in [K^*], \quad \forall z \in \mathbb{R}, \quad \gamma_x^*(z) > 0. \end{cases}$$

The maximum of  $\ell^{\text{int}}$  is reached at  $(Q, \gamma, \mathfrak{D} = (D_x)_{x \in [K^*]}, \mathbf{b})$  if and only if the integrand is maximal for almost every  $u \in [0, 1]$ , which means under **(Aid)** that

$$\left( Q, (\gamma(\cdot - D_x(u)) \otimes \mathbf{1}_{\mathbf{b}(x)})_{x \in [K^*]} \right) = \left( Q^*, (\gamma_x^*(\cdot - \Delta(x)) \otimes \mathbf{1}_{\mathbf{b}^*(x)})_{x \in [K^*]} \right)$$

up to permutation of the hidden states for all  $u \in [0, 1]$ .

Let us assume that the permutation is not constant at  $u$ . Since there are only a finite number of possible permutations of  $[K^*]$ , there exist two sequences  $(u_i)_{i \geq 1}$  and  $(v_i)_{i \geq 1}$  converging to  $u$ , one corresponding to a permutation  $p$  and the other to a permutation  $p' \neq p$ , that is

$$\forall i \geq 1, \quad \forall x \in [K^*], \quad \begin{cases} \gamma_x(\cdot - D_x(u_i)) = \gamma_{p(x)}^*(\cdot - \Delta(p(x))) & \text{and } \mathbf{b}(x) = \mathbf{b}^*(p(x)) \\ \gamma_x(\cdot - D_x(v_i)) = \gamma_{p'(x)}^*(\cdot - \Delta(p'(x))) & \text{and } \mathbf{b}(x) = \mathbf{b}^*(p'(x)) \end{cases}$$

Therefore, by continuity, for all  $x \in [K^*]$

$$(\gamma_{p(x)}^*(\cdot - \Delta(p(x))), \mathbf{b}^*(p(x))) = (\gamma_{p'(x)}^*(\cdot - \Delta(p'(x))), \mathbf{b}^*(p'(x))),$$

so that  $p = p'$  according to **(Aid)**, which contradicts the assumption that the permutation is not constant in  $u$ . Therefore, the permutation does not depend on  $u$ .

One may assume without loss of generality that the permutation is the identity, in other words  $Q = Q^*$ ,  $\mathbf{b} = \mathbf{b}^*$  and

$$\forall u \in [0, 1], \quad \forall x \in [K^*], \quad \gamma_x(\cdot - D_x(u)) = \gamma_x^*(\cdot - \Delta(x)).$$

Here, we took  $u$  in the whole segment  $[0, 1]$  instead of a subset with measure 1 because the mapping  $u \in [0, 1] \mapsto \gamma_x(\cdot - D_x(u))$  is continuous under **(Areg)**. If  $D_x$  is not constant at some  $x \in [K^*]$ , this entails that  $\gamma_x^*$  is invariant by translation, so that it is constant, which contradicts **(Amax)**. Therefore,  $\mathfrak{D}$  is constant.

Finally,

$$\forall x \in [K^*], \quad \frac{1}{2} = \int_{z \leq D_x} \gamma_x(z - D_x) dz = \int_{z \leq D_x} \gamma_x^*(z - \Delta(x)) dz$$

using **(Acentering)**, so that  $D_x$  is a median of  $\gamma_x^*$ . To conclude, note that under **(Amin)** and **(Acentering)**,  $\Delta(x)$  is the only median of  $\gamma_x^*$ .

### C.3 Uniform convergence of the homogeneous log-likelihood

Let us prove Lemma 31. The following theorem is a reformulation of Proposition 2 of Douc et al. (2004). Note that their proof also works when the space of parameters is not parametric.

**Theorem 33.** *Let  $\mathcal{V}$  be a Polish space and write  $\mathcal{D}(\mathcal{V})$  the set of nonnegative functions of  $\mathcal{V}$ . Let  $K' \in \mathbb{N}^*$ . Let  $\mathcal{Q}_{K'}$  be the set of transition matrices of size  $K'$  and  $\Delta_{K'}$  the set of probability measures on  $[K']$ . Let  $(V_t)_{t \geq 1}$  be an ergodic and stationary process taking values in  $\mathcal{V}$  with distribution  $\mathbb{P}^*$ .*

*Consider a compact metric space  $\Omega$  and mappings  $\omega \mapsto Q^\omega \in \mathcal{Q}_{K'}$  and  $\omega \mapsto \gamma^\omega \in \mathcal{D}(\mathcal{V})^{K'}$ . Assume that  $\omega \mapsto Q^\omega$  is continuous and for all  $v \in \mathcal{V}$ , the mapping  $\omega \mapsto \gamma^\omega(v) \in \mathbb{R}_+^{K'}$  is continuous. Finally, assume that there exists a constant  $\sigma_- > 0$  such that*

$$\inf_{\omega \in \Omega} \inf_{x, x' \in [K']} Q^\omega(x, x') \geq \sigma_-, \quad (35)$$

$$\sup_{\omega \in \Omega} \sup_{x \in [K']} \sup_{v \in \mathcal{V}} \gamma_x^\omega(v) < \infty, \quad (36)$$

$$\mathbb{E}^* \left[ \sup_{\omega \in \Omega} \left( \log \sum_{x \in [K']} \gamma_x^\omega(V_1) \right)_- \right] < \infty. \quad (37)$$

For all  $\pi \in \Delta_{K'}$ ,  $\omega \in \Omega$  and  $v_1^n \in \mathcal{V}^n$ , let

$$\frac{1}{n} l_n(\pi, Q^\omega, \gamma^\omega) \{v_1^n\} := \frac{1}{n} \log \sum_{x_1^n \in [K']^n} \pi(x_1) Q^\omega(x_1, x_2) \dots Q^\omega(x_{n-1}, x_n) \prod_{t=1}^n \gamma_{x_t}^\omega(v_t)$$

be the log-likelihood corresponding to the HMM with parameters  $(\pi, Q^\omega, \gamma^\omega)$  and to the observations  $v_1^n$ .

Then for all  $\pi \in \Delta_{K'}$  and  $\omega \in \Omega$ , there exists a finite  $l(Q^\omega, \gamma^\omega)$  such that almost surely,

$$\frac{1}{n} l_n(\pi, Q^\omega, \gamma^\omega) \{V_1^n\} \xrightarrow[n \rightarrow \infty]{} l(Q^\omega, \gamma^\omega).$$

In addition, for all  $N \in \mathbb{N}^*$ , the mapping  $\omega \mapsto l(Q^\omega, \gamma^\omega)$  is continuous and

$$\sup_{\omega \in \Omega} \sup_{\pi \in \Delta_{K'}} \left| \frac{1}{n} l_n(\pi, Q^\omega, \gamma^\omega) \{V_1^n\} - l(Q^\omega, \gamma^\omega) \right| \xrightarrow[n \rightarrow \infty]{} 0$$

almost surely.

Let us check these assumptions. First, let  $\mathcal{V} = \mathbb{R} \times \mathcal{B}^*$ ,  $V_t = (Z'_t, B_t)$  and

$$\Omega = \left\{ \omega = (Q, \gamma, \mathfrak{D}, u, \mathbf{b}) \in \Sigma_{K'}^{\sigma^-} \times \Gamma^{K'} \times \text{Cl}(\mathcal{D}(M))^{K'} \times [0, 1] \times (\mathcal{B}^*)^{K'} \right\}.$$

By Proposition 9,  $\Omega$  is compact. It is also metrizable under **(Areg)**: for instance, let  $(x_i)_{i \geq 1}$  be a dense sequence in  $\mathbb{R}$ , endow  $\Gamma$  with the distance

$$d_\Gamma(\gamma, \gamma') = \sum_{i \geq 1} 2^{-i} (|\gamma(x_i) - \gamma'(x_i)| \wedge 1)$$

and thus  $\Omega$  is metrizable as a product of metric spaces.

By the uniform continuity of  $\text{Cl}(\mathcal{D}(M))$ , the mappings

$$\begin{cases} \omega = (Q, \gamma, \mathfrak{D}, u, \mathbf{b}) \in \Omega \longmapsto Q^\omega := Q \\ \omega = (Q, \gamma, \mathfrak{D}, u, \mathbf{b}) \in \Omega \longmapsto \gamma^\omega(z', b) := (\gamma_x(z' - D_x(u)) \mathbf{1}_{\mathbf{b}(x)}(b))_{x \in [K']} \end{cases}$$

are continuous for all  $(z', b) \in \mathbb{R} \times \mathcal{B}^*$ . The lower bound (35) on the transition matrices is ensured by **(Aerg)** and the upper bound (36) on the densities is implied by **(Amax)**. Finally, the integrability condition (37) follows from the fact that for all  $\omega = (Q, \gamma, \mathfrak{D} = (D_x)_{x \in [K']}, u, \mathbf{b}) \in \Omega$ ,

$$\sum_{x \in [K']} \gamma_x^\omega(Z'_1, B_1) \geq \inf_{x \in [K']} \gamma_x(Z'_1 - D_x(u)) \geq m(M + Z_1^{\max})$$

by **(Amin)**, and  $\mathbb{E}^*[-\log m(M + Z_1^{\max})] < \infty$  by **(Aint)**.

Thus, the previous theorem holds, which shows that the application

$$\omega \longmapsto l(Q^\omega, \gamma^\omega) =: l^{\text{hom}}(Q, \tau(\gamma, \mathfrak{D}(u)), \mathbf{b})$$

is continuous on  $\Omega$ . For the uniform convergence, let  $\pi_U$  be the uniform distribution on  $[K']$  and let

$$S_{s,n}(\omega) = \frac{1}{n} l_n(\pi_U, Q^\omega, \gamma^\omega) \{V_{s+1}^{s+n}\}$$

for all  $s, n \in \mathbb{N}^*$  and  $\omega \in \Omega$ .

The theorem implies that almost surely,

$$\limsup_{n \rightarrow \infty} \sup_{\omega \in \Omega} \left| \frac{1}{n} S_{0,n}(\omega) - l(Q^\omega, \gamma^\omega) \right| = 0.$$

Hence, for all  $\epsilon > 0$ , there exists a (random)  $n(\epsilon)$  such that,

$$\forall n \geq n(\epsilon), \quad \sup_{\omega \in \Omega} \left| \frac{1}{n} S_{0,n}(\omega) - l(Q^\omega, \gamma^\omega) \right| \leq \epsilon. \quad (38)$$

The following Lemma is a reformulation of Lemma 2 of [Douc et al. \(2004\)](#) for compact nonparametric parameter spaces.

**Lemma 34.** *Under the same assumptions as the previous theorem, for all  $v_1^n \in \mathcal{V}^n$ ,*

$$\sup_{\omega \in \Omega} \sup_{\pi \in \Delta_K} |l_n(\pi, Q^\omega, \gamma^\omega) \{v_1^n\} - l_n(\pi_U, Q^\omega, \gamma^\omega) \{v_1^n\}| \leq \frac{1}{\sigma_-^2}.$$

Therefore,

$$|nS_{s,n}(\omega) - l_n(\pi_{X_{s+1}|V_1^s, X_1 \sim \pi_U}, Q^\omega, \gamma^\omega)\{V_{s+1}^{s+n}\}| \leq \frac{1}{\sigma_-^2}.$$

Note that

$$\begin{aligned} l_n(\pi_{X_{s+1}|V_1^s, X_1 \sim \pi_U}, Q^\omega, \gamma^\omega)\{V_{s+1}^{s+n}\} &= l_{s+n}(\pi_U, Q^\omega, \gamma^\omega)\{V_1^{s+n}\} - l_s(\pi_U, Q^\omega, \gamma^\omega)\{V_1^s\} \\ &= (s+n)S_{0,s+n}(\omega) - sS_{0,s}(\omega), \end{aligned}$$

so that

$$|nS_{s,n}(\omega) - (s+n)S_{0,s+n}(\omega) - sS_{0,s}(\omega)| \leq \frac{1}{\sigma_-^2}.$$

Thus, equation (38) entails that for all  $s \geq 1$ ,  $n \geq n(\epsilon)$  and  $\omega \in \Omega$ ,

$$|nS_{s,n}(\omega) - nl(Q^\omega, \gamma^\omega)| \leq (2s+n)\epsilon + \frac{1}{\sigma_-^2}.$$

Therefore, by Lemma 34, for all  $n \geq n(\epsilon)$  and  $s \in \{0, \dots, (N-1)n\}$ :

$$\sup_{\omega \in \Omega} \sup_{\pi \in \Delta_{K'}} \left| \frac{1}{n} l_n(\pi, Q^\omega, \gamma^\omega)\{V_{s+1}^{s+n}\} - l(Q^\omega, \gamma^\omega) \right| \leq (2N-1)\epsilon + \frac{2}{n\sigma_-^2},$$

which concludes the proof.

## D Miscellaneous proofs

### D.1 Proof of Lemma 24

Let us first state a Hoeffding inequality for uniformly ergodic Markov chains using **(Aerg)** (see e.g. [Glynn and Ormoneit \(2002\)](#)): for all  $\epsilon > 0$ ,  $x_1 \in [K^*]$  and  $n \geq \frac{1}{2\epsilon\sigma_-}$ ,

$$\mathbb{P} \left( \mathbb{P}(X_1 = x^*) - \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{X_t=x^*} \geq \epsilon \mid X_1 = x_1 \right) \leq \exp \left( -\frac{\sigma_-^2 \epsilon^2}{2} n \right).$$

The value of  $\mathbb{P}(X_1 = x^*)$  in the inequality is the one corresponding to the stationary distribution, so it is bounded below by  $\sigma_-$  using **(Aerg)**. Thus, for all  $\delta > 0$ ,  $\epsilon > 0$ ,  $n \geq \frac{1}{\delta\epsilon\sigma_-}$  and  $x_1 \in [K^*]$ ,

$$\mathbb{P} \left( \frac{2}{\delta n} \sum_{t=1}^{\delta n/2} \mathbf{1}_{X_t=x^*} \leq \sigma_- - \epsilon \mid X_1 = x_1 \right) \leq \exp \left( -\frac{\sigma_-^2 \epsilon^2 \delta}{4} n \right).$$

Assume  $n \geq \frac{2}{\delta(\sigma_-)^2}$ . Choose  $\epsilon = \sigma_-/2$  and apply a union bound on a covering  $\mathcal{R}$  of  $\{1, \dots, n\}$  in at most  $2n/\delta$  segments of size  $\delta n/2$ :

$$\mathbb{P} \left( \inf_{S \in \mathcal{R}} \frac{1}{n} \sum_{t \in S} \mathbf{1}_{X_t=x^*} \leq \frac{\delta\sigma_-}{4} \right) \leq \frac{2n}{\delta} \exp \left( -\frac{\sigma_-^4 \delta}{16} n \right).$$

Borel-Cantelli's lemma yields the result.

## D.2 Proof of Lemma 25

Without loss of generality, we assume  $\delta_n \rightarrow \delta$  almost surely (this is possible by replacing  $\delta_n$  by  $\delta_n \wedge \delta$  in the first statement and  $\delta_n \vee \delta$  in the second).

Let us first show that the two statements are equivalent. Assume that the second one holds. Let  $(U_t)_{t \geq 1}$  and  $(\delta_n)_n$  as in the first statement. Apply the second one to the i.i.d sequence of non-negative integrable random variables  $(-U_t)_{t \geq 1}$ :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq \delta_n n}} \frac{1}{n} \sum_{t \in S} (-U_t) &\leq \mathbb{E}[(-U_1) \mathbf{1}_{-U_1 \geq -q_U(\delta)}] \\ &\leq -\mathbb{E}[U_1 \mathbf{1}_{U_1 \leq q_U(\delta)}]. \end{aligned}$$

Add  $\mathbb{E}[U_1]$  on each side and use the law of large numbers:

$$\limsup_{n \rightarrow \infty} \sup_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq \delta_n n}} \frac{1}{n} \sum_{t \notin S} U_t \leq \mathbb{E}[U_1 \mathbf{1}_{U_1 > q_U(\delta)}].$$

Finally, replace  $\delta_n$  by  $1 - \delta_n$ ,  $\delta$  by  $1 - \delta$  and the sets  $S$  by their complementary to obtain the first statement.

Let us now show the second statement (regarding non-negative random variables). Given a random vector  $(V_1, \dots, V_n)$ , we write  $V_{(1)} \leq V_{(2)} \leq \dots \leq V_{(n)}$  its order statistics. Let  $\delta \in (0, 1)$  and let  $(\delta_n)_n$  be a non-increasing sequence of  $[0, 1]$ -valued random variables whose limit is  $\delta$  almost surely.

For all  $\beta \in (0, 1)$ , write  $\hat{q}_V(\beta) := V_{(\lfloor \beta n \rfloor)}$  the empirical  $\beta$ -quantile. Then

$$\sup_{\substack{S \subset \{1, \dots, n\} \\ |S| \leq \delta_n n}} \frac{1}{n} \sum_{t \in S} V_t = \frac{1}{n} \sum_{t=1}^n V_t \mathbf{1}_{V_t \geq \hat{q}_V(1 - \delta_n)}.$$

Let us show that almost surely

$$\left| \frac{1}{n} \sum_{t=1}^n V_t \mathbf{1}_{V_t \geq \hat{q}_V(1 - \delta_n)} - \frac{1}{n} \sum_{t=1}^n V_t \mathbf{1}_{V_t \geq q_V(1 - \delta)} \right| \xrightarrow[n \rightarrow \infty]{} 0$$

and the result will follow by the law of large numbers. Thus we have to show that

$$\frac{1}{n} \sum_{t=1}^n V_t (\mathbf{1}_{q_V(1 - \delta) \leq V_t \leq \hat{q}_V(1 - \delta_n)} + \mathbf{1}_{\hat{q}_V(1 - \delta_n) \leq V_t \leq q_V(1 - \delta)})$$

goes to 0 almost surely. Using Hoeffding's inequality, for all  $\beta \in (0, 1)$ ,

$$\mathbb{P} \left( \left| \frac{|\{t \in \{1, \dots, n\} \text{ s.t. } V_t \geq q_V(\beta)\}|}{n} - \mathbb{P}(V_1 \geq q_V(\beta)) \right| \geq \sqrt{\frac{\log n}{n}} \right) \leq 2n^{-2}.$$

In particular, taking  $\beta = 1 - \delta$ , Borel-Cantelli's lemma shows that almost surely, for large enough  $n$ ,

$$\hat{q}_V \left( 1 - \delta - \sqrt{\frac{\log n}{n}} \right) \leq q_V(1 - \delta) \leq \hat{q}_V \left( 1 - \delta + \sqrt{\frac{\log n}{n}} \right),$$

so that there are at most  $\sqrt{n \log n}$  terms between  $q_V(1 - \delta)$  and  $\hat{q}_V(1 - \delta)$ . Hence there are at most  $\sqrt{n \log n} + (\delta_n - \delta)n$  terms between  $q_V(1 - \delta)$  and  $\hat{q}_V(1 - \delta_n)$ . As  $(\delta_n)_n$  is non-increasing, this yields

$$\hat{q}_V(1 - \delta_n) \leq \hat{q}_V(1 - \delta) \leq q_V \left( 1 - \delta + \sqrt{\frac{\log n}{n}} \right),$$

which ensures that these terms are bounded above by  $q_V(1 - \delta + \sqrt{\frac{\log n}{n}})$ . Thus, almost surely, for  $n$  large enough,

$$\frac{1}{n} \sum_{t=1}^n V_t \left( \mathbf{1}_{q_V(1-\delta) \leq V_t \leq \hat{q}_V(1-\delta_n)} + \mathbf{1}_{\hat{q}_V(1-\delta_n) \leq V_t \leq q_V(1-\delta)} \right) \leq \left[ \sqrt{\frac{\log n}{n}} + (\delta_n - \delta) \right] q_V(1 - \delta/2),$$

which indeed converges to 0.

## References

- Pierre Ailliot and Valérie Monbet. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30:92–101, 2012.
- Pierre Ailliot and Françoise Pene. Consistency of the maximum likelihood estimate for non-homogeneous markov-switching models. *ESAIM: Probability and Statistics*, 19: 268–292, 2015.
- Andrew R Barron. The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem. *The Annals of Probability*, 13(4):1292–1303, 1985.
- Enrica Bellone, James P Hughes, and Peter Guttorp. A hidden markov model for down-scaling synoptic atmospheric patterns to precipitation amounts. *Climate research*, 15 (1):1–12, 2000.
- Yohann De Castro, Elisabeth Gassiat, and Sylvain Le Corff. Consistent estimation of the filtering and marginal smoothing distributions in nonparametric hidden Markov models. *IEEE Transactions on Information Theory*, 2017.
- Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977.
- Manuel Diehn, Axel Munk, and Daniel Rudolf. Maximum likelihood estimation in hidden Markov models with inhomogeneous noise. *ESAIM: Probability and Statistics*, 23:492–523, 2019.
- Randal Douc, Eric Moulines, and Tobias Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics*, 32(5):2254–2304, 2004.

- Peter W Glynn and Dirk Ormoneit. Hoeffding's inequality for uniformly ergodic Markov chains. *Statistics & Probability Letters*, 56(2):143–146, 2002.
- James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384, 1989.
- James P Hughes, Peter Guttorp, and Stephen P Charles. A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(1):15–30, 1999.
- Brian G Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications*, 40(1):127–143, 1992.
- Demian Pouzo, Zacharias Psaradakis, and Martin Sola. Maximum likelihood estimation in possibly misspecified dynamic models with time-inhomogeneous Markov regimes. *arXiv preprint arXiv:1612.04932*, 2016.
- Augustin Touron. Consistency of the maximum likelihood estimator in seasonal hidden Markov models. *Statistics and Computing*, pages 1–21, 2018.
- Augustin Touron. *Modélisation multivariée de variables météorologiques*. PhD thesis, Université Paris-Saclay, EDF, 2019.