Appendix A.

This appendix consists of three subsections. In Section A.1 we gather together some of the key definitions from earlier in the paper for the reader's convenience and introduce some further objects and lemmas needed for the main proofs. Section A.2 and Section A.3 are dedicated to proving respectively Theorems 2 and 3. Upon first reading, one may skip straight to Section A.2, where an outline of the main steps in the proof of Theorem 2 is given.

Before commencing, a comment on generality and notation. In the main part of the paper we work with a state-space $\mathbb{X}$ which is a set of finite cardinality and a probability mass function on $\mathbb{X}^U$ for some $U \subseteq V$ is denoted, for example, $\mu(x^U)$. In order to make notation visually compact, throughout the appendix we write expectations with respect to such mass functions as measure theoretic integrals, e.g. $\int \mathbb{I}_A(x)\mu(dx)$. Here it is to be understood that $\mu(dx) = \mu(x)\psi(dx)$ where ψ denotes counting measure, and hence $\int \mathbb{I}_A(x)\mu(dx) \equiv \mu(A) \equiv \sum_{x\in A}\mu(x)$. This integral notation reflects the fact that many of our results do not rely on $\mathbb{X}$ being a set of finite cardinality and could be replaced by a polish space, although we do not pursue this further here.

A.1 Definitions and Preliminary results

Before starting we recap the main notation and definitions. Throughout the appendix we assume:

$\mathbb{X}, V, F$ are finite sets

$L := \mathbf{card}(\mathbb{X}), M := \mathbf{card}(V)$

$\mathbb{Y}$ can be discrete, $\mathbb{R}^d$ with $d \in \mathbb{N}$ or a subset thereof

$\mathbb{X}^V$ is the product space $\bigotimes_{v\in V} \mathbb{X}$ and denotes the state-space of the Markov chain$(X_t)_{t\in\{0,\ldots,T\}}$

$A \in \sigma(\mathbb{S})$ with $\sigma(\mathbb{S})$ power set of $\mathbb{S}$ ($\mathbb{S}$ depends on the analysis)

μ_0 is a probability measure over $\mathbb{X}^V$ and describe the distribution of X_0

$p : \mathbb{X}^V \times \mathbb{X}^V \to \mathbb{R}_+$ is the transition kernel of the Markov Chain $(X_t)_{t\in\{0,\ldots,T\}}$

$g : \mathbb{X}^V \times \mathbb{Y} \to \mathbb{R}_+$ is the emission density of the HMM $(X_t, Y_t)_{t\in\{0,\ldots,T\}}$ i.e. $Y_t|X_t \sim g(X_t, \cdot)$

δ_x is the Dirac measure with mass on $x \in \mathbb{X}^V$

as stated in the main paper we assume a factorization of both the transition kernel and emission distribution. In particular the factorization of the emission distribution can be represented through a factor graph $\mathcal{G} = (V, F, E)$ where V, F are the vertex sets and E is the edge set ($v \in V, f \in F$ are connected if x^v appears in factor f). Mathematically:

$$p(x, z) = \prod_{v\in V} p^v(x^v, z^v) \quad x, z \in \mathbb{X}^V$$

$$g(x, y) = \prod_{f\in F} g^f(x^{N(f)}, y) \quad x \in \mathbb{X}^V, y \in \mathbb{Y}$$

where $p^v : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$, $g^f : \mathbb{X}^{N(f)} \times \mathbb{Y} \to \mathbb{R}_+$ and $N(\cdot)$ is the neighbourhood function of $\mathcal{G}$.

Moreover, we shall need the following definitions (some of them were already introduced in the paper, but we recall them here) associated with the factor graph $\mathcal{G}$:

$$\begin{aligned}
d(J, J') &\coloneqq \min_{e \in J} \min_{e' \in J'} \{d(e, e')\}, \quad J, J' \subseteq V \times F, \\
N(J) &\coloneqq \{f \in F : \exists v \in J \text{ with } d(v, f) \leqslant 1\}, \quad J \subseteq V, \\
N^2(J) &\coloneqq \{v' \in V : \exists v \in J \text{ with } d(v', v) \leqslant 2\}, \quad J \subseteq V, \\
N_v^r(J) &\coloneqq \{v' \in V \text{ such that } \exists v \in J \text{ with } d(v, v') \leqslant 2r + 2\}, \quad J \subseteq V, \\
N_f^r(J) &\coloneqq \{f \in F \text{ such that } \exists v \in J \text{ with } d(v, f) \leqslant 2r + 1\}, \quad J \subseteq V, \\
n_J &\coloneqq \frac{1}{2} \max_{v \in V} d(J, v), \\
\Upsilon &\coloneqq \max_{v \in V} \{\mathbf{card}(N(v))\}, \\
\Upsilon^{(2)} &\coloneqq \max_{v \in V} \{\mathbf{card}(N_v^0(v))\}, \\
\tilde{\Upsilon} &\coloneqq \max_{v, v' \in V} \{\mathbf{card}(N(v) \cap N(v'))\}, \\
\widetilde{J} &\coloneqq \{v \in J : \forall f \in N(v), \, N(f) \subseteq J\}, \quad J \subseteq V, \\
\partial J &\coloneqq J \backslash \widetilde{J}, \quad J \subseteq V.
\end{aligned}$$

Remark that $N_v^0(J) = N^2(J)$ and the following inclusion relation holds:

$$J \subseteq N^2(J) = N_v^0(J) \subseteq N_v^1(J) \subseteq \cdots \subseteq N_v^m(J), \quad m \geqslant 1,$$

and similarly on the set F:

$$N(J) = N_f^0(J) \subseteq N_f^1(J) \subseteq \cdots \subseteq N_f^m(J), \quad m \geqslant 1.$$

Let $\mathbb{S}$ be a product of Polish spaces, i.e.: $\mathbb{S} = \bigotimes_{k \in I} \mathbb{S}_k$, where I is a finite index set.

Definition 1 *Given two probability distribution μ, ν on $\mathbb{S}$ we can define a total variation distance (TV) and a local total variation distance (LTV):*

- $\|\mu - \nu\| \coloneqq \sup_{A \in \sigma(\mathbb{S})} |\mu(A) - \nu(A)|.$
- $\|\mu - \nu\|_J \coloneqq \sup_{A \in \sigma(\bigotimes_{k \in J} \mathbb{S}_k)} |\mu^J(A) - \nu^J(A)|, \quad J \subset I.$

Definition 2 *Let μ be a probability distribution on $\mathbb{S}$ and let $X \sim \mu$. The conditional distribution over the component $i \in I$ is defined as:*

$$\mu_x^i(A) \coloneqq \mathbb{P}\left(X^i \in A | X^{I \backslash i} = x^{I \backslash i}\right), \quad x \in \mathbb{S}.$$

Definition 3 *Let μ be a probability distribution on $\mathbb{X}^V$ with $X \sim \mu$ and let $P(x, dz) \coloneqq p(x, z)\psi(dz)$ with $x, z \in \mathbb{X}^V$ be the transition kernel of the considered FHMM with $Z|X \sim P(X, \cdot)$, then we define for $v \in V$:*

$$\mu_{x,z}^v(A) \coloneqq \mathbb{P}\left(X^v \in A | X^{V \backslash v} = x^{V \backslash v}, Z = z\right), \quad x, z \in \mathbb{X}^V.$$

Lemma 4 *Fix any collection of observations $\{y_1, \dots, y_T\}$ and consider a probability distribution μ on $\mathbb{X}^V$. Given the optimal correction operator as in* (5)*, the conditional distribution of $\mathsf{C}_t\mu$ over the component $v \in V$ is given by:*

$$(\mathsf{C}_t\mu)^v_x(A) = \frac{\int \mathbb{I}_A(x^v) \prod_{f\in N(v)} g^f(x^{N(f)}, y_t)\mu^v_x(dx^v)}{\int \prod_{f\in N(v)} g^f(x^{N(f)}, y_t)\mu^v_x(dx^v)}, \quad x \in \mathbb{X}^V.$$

Similarly, the one step forward conditional distribution over the component $v \in V$ is:

$$(\mathsf{C}_t\mu)^v_{x,z}(A) = \frac{\int \mathbb{I}_A(x^v) \prod_{f\in N(v)} g^f(x^{N(f)}, y_t)p^v(x^v, z^v)\mu^v_x(dx^v)}{\int \prod_{f\in N(v)} g^f(x^{N(f)}, y_t)p^v(x^v, z^v)\mu^v_x(dx^v)}, \quad x, z \in \mathbb{X}^V.$$

Proof The proof is a trivial consequence of the form of the correction operator as in (5) and Definition 2. The procedure is the following:

$$\begin{aligned}
(\mathsf{C}_t\mu)^v_x(A) &= \frac{(\mathsf{C}_t\mu)(A \times x^{V\setminus v})}{(\mathsf{C}_t\mu)(\mathbb{X} \times x^{V\setminus v})} \\
&= \frac{\int \mathbb{I}_{A\times x^{V\setminus v}}(\tilde{x}^v \times \tilde{x}^{V\setminus v}) \prod_{f\in F\setminus N(v)} g^f(\tilde{x}^{N(f)}, y_t) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\tilde{x}^v \times \tilde{x}^{V\setminus v}) \prod_{f\in F\setminus N(v)} g^f(\tilde{x}^{N(f)}, y_t) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu(d\tilde{x})} \\
&= \frac{\prod_{f\in F\setminus N(v)} g^f(x^{N(f)}, y_t)}{\prod_{f\in F\setminus N(v)} g^f(x^{N(f)}, y_t)} \frac{\int \mathbb{I}_A(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu(d\tilde{x})} \\
&= \frac{\int \mathbb{I}_A(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t) \frac{\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\tilde{x}^v\times\tilde{x}^{V\setminus v})\mu(d\tilde{x})}}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t) \frac{\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\tilde{x}^v\times\tilde{x}^{V\setminus v})\mu(d\tilde{x})}} \\
&= \frac{\int \mathbb{I}_A(\tilde{x}^v) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu^v_x(d\tilde{x}^v)}{\int \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_t)\mu^v_x(d\tilde{x}^v)}.
\end{aligned}$$

The form of $(\mathsf{C}_t\mu)^v_{x,z}(A)$ follow the same procedure with the addition of the kernel $p^v(x^v, z^v)$ at the end, where the isolation of the component v is a consequence of the factorization of the transition kernel.

■

Remark *We use the name "one step forward" to refer to the distribution of $X^v_t|X^{V\setminus v}_t = x^{V\setminus v}, X_{t+1} = z$*

Definition 5 *Fix any collection of observations $\{y_1, \dots, y_T\}$ and consider a probability distribution μ on $\mathbb{X}^V$. For $K \in \mathcal{K}$ and $m = \{0, \dots, n\}$ define:*

$$(\tilde{\mathsf{C}}^{m,K}_t\mu)(A) := \frac{\int \mathbb{I}_A(x^K) \prod_{f\in N^m_f(K)} g^f(x^{N(f)}, y_t)\mu(dx)}{\int \prod_{f\in N^m_f(K)} g^f(x^{N(f)}, y_t)\mu(dx)},$$

and define:

$$\tilde{\mathsf{C}}^m_t\mu := \bigotimes_{K\in\mathcal{K}} \tilde{\mathsf{C}}^{m,K}_t\mu. \tag{A.1}$$

Given that the approximated correction operator is applied to probability distributions that factorize over the partition $\mathcal{K}$, we are interested in $\mu = \bigotimes_{K \in \mathcal{K}} \mu^K$ hence:

$$(\tilde{\mathsf{C}}_t^{m,K}\mu)(A) = \frac{\int \mathbb{I}_A(x^K) \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu^K(dx^K)}{\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu^K(dx^K)},$$

meaning that $\tilde{\mathsf{C}}_t^m \mu$ can be written as:

$$(\tilde{\mathsf{C}}_t^m \mu)(A) = \frac{\int \mathbb{I}_A(x) \prod_{K \in \mathcal{K}} \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu(dx)}{\int \prod_{K \in \mathcal{K}} \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu(dx)},$$

where x_K is a collection of auxiliary variables:

$$x_K^v \coloneqq \begin{cases} x^v & \text{if } v \in K \\ \tilde{x}^v & \text{if } v \notin K \end{cases} \quad \text{with } \tilde{x} \in \mathbb{X}^{V \setminus K} \text{ and } x \in \mathbb{X}^V.$$

The above definition is nothing more than a trick to distinguish the components that are integrated out from the ones that are not.

Lemma 6 *Fix any collection of observations $\{y_1, \ldots, y_T\}$ and let μ be a probability distribution on $\mathbb{X}^V$ such that $\mu = \bigotimes_{K' \in \mathcal{K}} \mu^{K'}$. Given the new correction operator as in* (A.1) *the conditional distribution of $\tilde{\mathsf{C}}_t^m \mu$ over the component $v \in K$ with $K \in \mathcal{K}$ is given by:*

$$(\tilde{\mathsf{C}}_t^m \mu)_x^v(A) = \frac{\int \mathbb{I}_A(x^v) \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu_x^v(dx^v)}{\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) \mu_x^v(dx^v)}, \quad x \in \mathbb{X}^{\mathbb{V}}.$$

Similarly, the one step forward conditional distribution over the component $v \in K$ with $K \in \mathcal{K}$ is given by:

$$(\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v(A) = \frac{\int \mathbb{I}_A(x^v) \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)}, \quad x, z \in \mathbb{X}^{\mathbb{V}}.$$

Proof The proof follows from Definition 2, Definition 5 and the form of the operator when μ factorizes. The form of these conditional distributions can be obtained with the same procedure as in the proof of Lemma 4. Consider $v \in K$:

$$
\begin{aligned}
(\tilde{\mathsf{C}}_t^m \mu)_x^v(A) &= \frac{(\tilde{\mathsf{C}}_t^m \mu)(A \times x^{V\setminus v})}{(\tilde{\mathsf{C}}_t^m \mu)(\mathbb{X} \times x^{V\setminus v})} \\
&= \frac{\int \mathbb{I}_{A\times x^{V\setminus v}}(\hat{x}) \prod_{K'\in\mathcal{K}} \int \prod_{f\in N_f^m(K')} g^f(\hat{x}_{K'}^{N(f)}, y_t)\mu^{V\setminus K'}(d\tilde{x})\mu(d\hat{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\hat{x}) \prod_{K'\in\mathcal{K}} \int \prod_{f\in N_f^m(K')} g^f(\hat{x}_{K'}^{N(f)}, y_t)\mu^{V\setminus K'}(d\tilde{x})\mu(d\hat{x})} \\
&= \frac{\int \mathbb{I}_{A\times x^{V\setminus v}}(\hat{x}) \int \prod_{f\in N_f^m(K)} g^f(\hat{x}_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\mu(d\hat{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\hat{x}) \int \prod_{f\in N_f^m(K)} g^f(\hat{x}_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\mu(d\hat{x})} \\
&= \frac{\int \mathbb{I}_{A\times x^{V\setminus v}}(\hat{x}) \int \prod_{f\in N_f^m(K)} g^f(\hat{x}_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\frac{\mu(d\hat{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\hat{x}^v\times \hat{x}^{V\setminus v})\mu(d\hat{x})}}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\hat{x}) \int \prod_{f\in N_f^m(K)} g^f(\hat{x}_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\frac{\mu(d\hat{x})}{\int \mathbb{I}_{\mathbb{X}\times x^{V\setminus v}}(\hat{x}^v\times \hat{x}^{V\setminus v})\mu(d\hat{x})}} \\
&= \frac{\int \mathbb{I}_{A}(x) \int \prod_{f\in N_f^m(K)} g^f(x_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\mu_x^v(dx)}{\int \prod_{f\in N_f^m(K)} g^f(x_{K}^{N(f)}, y_t)\mu^{V\setminus K}(d\tilde{x})\mu_x^v(dx)}.
\end{aligned}
$$

Moreover given that for $v \in K$ and $\mu = \bigotimes_{K'\in\mathcal{K}} \mu^{K'}$ we have $\mu_x^v = (\mu^K)_x^v$ then $(\tilde{\mathsf{C}}_t^m \mu)_x^v(A) = (\tilde{\mathsf{C}}_t^{m,K} \mu)_x^v(A)$, which can be easily checked by substituting μ_x^v with $(\mu^K)_x^v$ in the previous computations. ■

Lemma 7 *Let μ, μ' and ν, ν' probability distributions on $\mathbb{S}$. Assume that there exists $\epsilon \in (0,1)$ such that:*

$$\nu(A) \geqslant \epsilon\mu(A) \quad \textit{and} \quad \nu'(A) \geqslant \epsilon\mu'(A).$$

Then:

$$\left\|\nu - \nu'\right\| \leqslant 2(1-\epsilon) + \epsilon \left\|\mu - \mu'\right\|.$$

Proof The proof is available in Rebeschini and Van Handel (2015, Lemma 4.1, pag. 32). ■

Remark *The application of Lemma 7 involves finding a lower bound for each of the measure of interest, once the lower bounds are available we have an upper bound for the total variation norm. Note that if $\mu = \mu'$ the upper bound simplify to $2(1-\epsilon)$.*

Lemma 8 *Let μ, ν probability distributions on $\mathbb{S}$ and Λ a bounded-strictly positive measurable function on the same space. Consider:*

$$\mu_\Lambda(A) \coloneqq \frac{\int \mathbb{I}_A(x)\Lambda(x)\mu(dx)}{\int \Lambda(x)\mu(dx)}, \qquad \nu_\Lambda(A) \coloneqq \frac{\int \mathbb{I}_A(x)\Lambda(x)\nu(dx)}{\int \Lambda(x)\nu(dx)}.$$

Then:

$$\left\|\mu_\Lambda - \nu_\Lambda\right\| \leqslant 2\frac{\sup_x \Lambda(x)}{\inf_x \Lambda(x)} \left\|\mu - \nu\right\|.$$

Proof The complete proof is available in Rebeschini and Van Handel (2015, Lemma 4.2,pag. 32). ∎

Remark *To apply lemma 8 we need to reformulate the probability measures μ_Λ, ν_Λ in a fractional form where the same function Λ appears. If this is possible the lemma gives an upper bound for the total variation norm in terms of* $\sup$ *and* $\inf$ *of* Λ.

Theorem 9 (Dobrushin Comparison Theorem) *Let μ, ν be probability distributions on $\mathbb{S}$. For $i, j \in I$ define the quantities:*

$$C_{i,j} = \frac{1}{2} \sup_{\substack{x,\hat{x}\in\mathbb{S} \\ x^{I\setminus j} = \hat{x}^{I\setminus j}}} \left\| \mu_x^i - \mu_{\hat{x}}^i \right\| \quad \text{and} \quad b_j = \sup_{x\in\mathbb{S}} \left\| \mu_x^j - \nu_x^j \right\|. \tag{A.2}$$

Assume that:

$$\max_{i\in I} \sum_{j\in I} C_{i,j} < 1 \quad \text{(Dobrushin condition)},$$

then the matrix $D \coloneqq \sum_{t\geqslant 0} C^t$, where C^0 is the identity matrix, converges and for an arbitrary $J \subseteq I$ it holds:

$$\left\| \mu - \nu \right\|_J \leqslant \sum_{i\in J} \sum_{j\in I} D_{i,j} b_j.$$

Proof See for example Georgii (2011, Theorem 8.20). ∎

Lemma 10 *Let I be a finite index set with $m(\cdot,\cdot)$ a pseudometric on it. Let C be a non-negative matrix with rows and columns indexed by I. Assume that there exists $\lambda \in (0,1)$ such that:*

$$\max_{i\in I} \sum_{j\in I} e^{m(i,j)} C_{i,j} \leqslant \lambda.$$

Then the matrix $D \coloneqq \sum_{t\geqslant 0} C^t$ satisfies:

$$\max_{i\in I} \sum_{j\in I} e^{m(i,j)} D_{i,j} \leqslant \frac{1}{1-\lambda}.$$

Moreover:

$$\sum_{j\in J} D_{i,j} \leqslant \frac{e^{-m(i,J)}}{1-\lambda}.$$

Proof The proof is available in Rebeschini and Van Handel (2015) (Lemma 4.3, pag.33). Remark that a pseudometric $m(\cdot,\cdot)$ on I is a metric on I where it is allowed that $m(i,j) = 0$ even if $i \neq j$. ∎

A.2 Filtering

Theorem 2 *Fix any collection of observations* $\{y_1, \dots, y_T\}$ *and any partition* $\mathcal{K}$ *of* V. *There exists a region* $\mathcal{R}_0 \subseteq (0,1)^3$ *depending only on* $\tilde{\Upsilon}, \Upsilon$ *and* $\Upsilon^{(2)}$, *such that if, for given* $(\epsilon_-, \epsilon_+, \kappa) \in \mathcal{R}_0$,

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+, \quad \text{and} \quad \kappa \leqslant g^f\left(x^{N(f)}, y_t\right) \leqslant \frac{1}{\kappa},$$

for all $x, z \in \mathbb{X}^V, v \in V, f \in F, t \in \{1, \dots, T\}$, *then for* $\beta > 0$ *small enough depending only on* $\tilde{\Upsilon}, \Upsilon, \Upsilon^{(2)}, \epsilon_-, \epsilon_+$ *and* κ, *we have that for any* μ_0 *satisfying:*

$$\widetilde{Corr}(\mu_0, \beta) \leqslant 2e^{-\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right)$$

and for any $K \in \mathcal{K}$, $J \subseteq K$ *and* $m \in \{0, \dots, n\}$:

$$\|\pi_t - \tilde{\pi}_t\|_J \leqslant \alpha_1(\beta)\left(1 - \kappa^{a(\mathcal{K})}\right)\mathbf{card}(J) + \gamma_1(\beta)\left(1 - \kappa^{b(\mathcal{K},m)}\right)\mathbf{card}(J)e^{-\beta m}, \quad \forall t \in \{1, \dots, T\},$$

where $\pi_t, \tilde{\pi}_t$ *are given by* (4) *and* (15) *with initial condition* μ_0; $\alpha_1(\beta), \gamma_1(\beta)$ *are constants depending only on* β, *and*

$$\begin{aligned} a(\mathcal{K}) &:= 2\max_{K\in\mathcal{K}}\max_{v\in\partial K}\mathbf{card}(N(v)\cap\partial N(K)), \\ b(m,\mathcal{K}) &:= 2\max_{K\in\mathcal{K}}\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v)), \end{aligned}$$

with the convention that the maximum over an empty set is zero.

The proof of Theorem 2 consists of three main steps:

- establish a bound on LTV between filtering distributions associated with different initial distributions, under an assumption on the decay of correlations associated with the initial distributions. This is the subject of Proposition 14 in Section A.2.1;
- control the approximation error between the Bayes update operator C_t and $\tilde{\mathsf{C}}_t^m$, under decay of correlation assumptions, this is the subject of the propositions in Section A.2.2;
- prove that the required decay of correlation assumptions hold uniformly in time, this is the subject of Section A.2.3.

These steps are then brought together to complete the proof of Theorem 2 in Section A.2.4.

Remark 11 *Note that if the aim of the proof would be to provide some time uniform bounds on the proposed approximation without considering the initial distribution, we could set* $\mu_0 = \delta_x$. *However, for the sake of completeness, we include a full proof which consider a range of* μ_0 *that satisfy a decay of correlation property.*

A.2.1 Stability of the filter with respect to initial distributions

Definition 12 *Given $v, v' \in V$ and μ probability distribution on $\mathbb{X}^V$ define the quantity:*

$$C^{\mu}_{v,v'} := \frac{1}{2} \sup_{x,\hat{x} \in \mathbb{X}^V : x^{V \setminus v'} = \hat{x}^{V \setminus v'}} \left\| \mu^v_x - \mu^v_{\hat{x}} \right\| .$$

Then for a fixed $\beta > 0$:

$$Corr(\mu, \beta) := \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v')} C^{\mu}_{v,v'}.$$

Definition 13 *Given $v, v' \in V$ and μ probability distribution on $\mathbb{X}^V$ define the quantity:*

$$\tilde{C}^{\mu}_{v,v'} := \frac{1}{2} \sup_{z \in \mathbb{X}^V} \sup_{\substack{x,\hat{x} \in \mathbb{X}^V : \\ x^{V \setminus v'} = \hat{x}^{V \setminus v'}}} \left\| \mu^v_{x,z} - \mu^v_{\hat{x},z} \right\| .$$

Then for a fixed $\beta > 0$:

$$\widetilde{Corr}(\mu, \beta) := \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v')} C^{\mu}_{v,v'}.$$

Proposition 14 *Fix any collection of observations $\{y_1, \ldots, y_T\}$ and any partition $\mathcal{K}$ of V. Suppose that there exists $(\epsilon_-, \epsilon_+) \in (0,1)^2$ and $\kappa \in (0,1)$ such that:*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+ \quad \textit{and} \quad \kappa \leqslant g^f(x^{N(f)}, y_t) \leqslant \frac{1}{\kappa},$$

for all $x \in \mathbb{X}^V, z \in \mathbb{X}^V, v \in V, f \in F, t \in \{1, \ldots, T\}$. Let μ, ν two probability distributions on $\mathbb{X}^V$, and assume there exists $\beta > 0$ such that:

$$\widetilde{Corr}(\mu, \beta) + 2e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta} \Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2},$$

then for all $t \in \{0, \ldots, T\}, K \in \mathcal{K}$ and $J \subseteq K$:

$$\left\| \mathsf{F}_T \ldots \mathsf{F}_{t+1} \mu - \mathsf{F}_T \ldots \mathsf{F}_{t+1} \nu \right\|_J \leqslant 2e^{-\beta(T-t)} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \sup_{x,z \in \mathbb{X}^V} \left\| \mu^{v'}_{x,z} - \nu^{v'}_{x,z} \right\| \right\}.$$

Proof Define the probability distributions:

$$\rho(A) \propto \int \mathbb{I}_A(x_{0:T}) \left[\prod_{k=1}^{T} p(x_{k-1}, x_k) g(x_k, y_k) \right] \bigotimes_{k=1}^{T} \psi(dx_k) \mu(dx_0),$$

$$\tilde{\rho}(A) \propto \int \mathbb{I}_A(x_{0:T}) \left[\prod_{k=1}^{T} p(x_{k-1}, x_k) g(x_k, y_k) \right] \bigotimes_{k=1}^{T} \psi(dx_k) \nu(dx_0).$$

It can be observed that:

$$\left\| \rho - \tilde{\rho} \right\|_{(T,J)} = \left\| \mathsf{F}_T \ldots \mathsf{F}_1 \mu - \mathsf{F}_T \ldots \mathsf{F}_1 \nu \right\|_J ,$$

and the proof proceeds by applying the Dobrushin theorem (Theorem 9) to $\rho, \tilde{\rho}$ where the index set is given by $I = \bigcup_{t=0}^{T} (t, V)$ and the subset is (T, J).

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9, i.e.:

$$C_{i,j} = \frac{1}{2} \sup_{\substack{x,\hat{x} \in \mathbb{X}^I: \\ x^{I \setminus j} = \hat{x}^{I \setminus j}}} \left\| \rho_x^i - \rho_{\hat{x}}^i \right\| .$$

In the following passages we consider $x = (x_0, \dots, x_T)$, where $x_t \in \mathbb{X}^V$.

- Consider $i = (0, v)$ and $v \in V$ then:

$$\begin{aligned}
\rho_x^{(0,v)}(A) &= \frac{\int \mathbb{I}_A(\tilde{x}_0^v) \mathbb{I}_{\{x_0^{V\setminus v}, x_{1:T}\}}(\tilde{x}_0^{V\setminus v}, \tilde{x}_{1:T}) \prod_{k=1}^T p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^T \psi(d\tilde{x}_k) \mu(d\tilde{x}_0)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_0^v) \mathbb{I}_{\{x_0^{V\setminus v}, x_{1:T}\}}(\tilde{x}_0^{V\setminus v}, \tilde{x}_{1:T}) \prod_{k=1}^T p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^T \psi(d\tilde{x}_k) \mu(d\tilde{x}_0)} \\
&= \frac{g(x_1, y_1) \prod_{k=2}^T p(x_{k-1}, x_k) g(x_k, y_k)}{g(x_1, y_1) \prod_{k=2}^T p(x_{k-1}, x_k) g(x_k, y_k)} \frac{\int \mathbb{I}_A(\tilde{x}_0^v) \mathbb{I}_{\{x_0^{V\setminus v}, x_1\}}(\tilde{x}_0^{V\setminus v}, \tilde{x}_1) p(\tilde{x}_0, \tilde{x}_1) \mu(d\tilde{x}_0)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_0^v) \mathbb{I}_{\{x_0^{V\setminus v}, x_1\}}(\tilde{x}_0^{V\setminus v}, \tilde{x}_1) p(\tilde{x}_0, \tilde{x}_1) \mu(d\tilde{x}_0)} \\
&= \frac{\int \mathbb{I}_A(\tilde{x}_0^v) p^v(\tilde{x}_0^v, x_1^v) \mu_{x_0}^v(d\tilde{x}_0^v)}{\int p^v(\tilde{x}_0^v, x_1^v) \mu_{x_0}^v(d\tilde{x}_0^v)} = \mu_{x_0, x_1}^v(A),
\end{aligned}$$

  where the last passage follows from the factorization of the kernel and the definition of μ_x^v. Now we have to distinguish the different cases in which ρ_x^i can differ from $\rho_{\tilde{x}}^i$, where $x^{I\setminus j} = \tilde{x}^{I\setminus j}$.

  - If $j = (0, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \tilde{C}^{\mu}_{v,v'}$.
  - If $j = (1, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
    where the result follows from Lemma 7, obtained by a majorization of the kernel part.
  - If $j = (k, v')$ with $k > 1$ and $v' \in V$ then: $\quad C_{i,j} = 0$,
    because in ρ_x^i there is no dependence on x_t with $t > 1$.

- Consider $i = (t, v)$ with $0 < t < T$ and $v \in V$ define:

$$x_t^{N(f)\setminus v} := (\tilde{x}_t^v, x_t^{N(f)\setminus v}), \quad \tilde{x}_{0:T\setminus t} := (\tilde{x}_{0:t-1}, \tilde{x}_{t+1:T}) \quad \text{and} \quad x_{0:T\setminus t} := (x_{0:t-1}, x_{t+1:T})$$

then:

$$
\begin{aligned}
\rho_x^{(t,v)}(A) &= \frac{\int \mathbb{I}_A(\tilde{x}_t^v)\mathbb{I}_{\{x_t^{V\backslash v}, x_{0:T\backslash t}\}}(\tilde{x}_t^{V\backslash v}, \tilde{x}_{0:T\backslash t}) \prod_{k=1}^{T} p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_t^v)\mathbb{I}_{\{x_t^{V\backslash v}, x_{0:T\backslash t}\}}(\tilde{x}_t^{V\backslash v}, \tilde{x}_{0:T\backslash t}) \prod_{k=1}^{T} p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)} \\
&= \frac{\prod_{k\neq t} g(x_k, y_k) \prod_{k\neq t,t+1} p(x_{k-1}, x_k)}{\prod_{k\neq t} g(x_k, y_k) \prod_{k\neq t,t+1} p(x_{k-1}, x_k)} \\
&\quad \frac{\int \mathbb{I}_A(\tilde{x}_t^v)\mathbb{I}_{\{x_t^{V\backslash v}, x_{0:T\backslash t}\}}(\tilde{x}_t^{V\backslash v}, \tilde{x}_{0:T\backslash t}) p(\tilde{x}_{t-1}, \tilde{x}_t) p(\tilde{x}_t, \tilde{x}_{t+1}) g(\tilde{x}_t, y_t) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_t^v)\mathbb{I}_{\{x_t^{V\backslash v}, x_{0:T\backslash t}\}}(\tilde{x}_t^{V\backslash v}, \tilde{x}_{0:T\backslash t}) p(\tilde{x}_{t-1}, \tilde{x}_t) p(\tilde{x}_t, \tilde{x}_{t+1}) g(\tilde{x}_t, y_t) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)} \\
&= \frac{\int \mathbb{I}_A(\tilde{x}_t^v) p^v(x_{t-1}^v, \tilde{x}_t^v) p^v(\tilde{x}_t^v, x_{t+1}^v) \prod_{f\in N(v)} g^f(x_t^{N(f)\backslash v}, y_t)\psi^v(d\tilde{x}_t^v)}{\int p^v(x_{t-1}^v, \tilde{x}_t^v) p^v(\tilde{x}_t^v, x_{t+1}^v) \prod_{f\in N(v)} g^f(x_t^{N(f)\backslash v}, y_t)\psi^v(d\tilde{x}_t^v)},
\end{aligned}
$$

where the last passage follows from the factorization of the kernel and the factorial representation of the observation density. Now we have to distinguish the different cases in which ρ_x^i can differ from $\rho_{\tilde{x}}^i$, where $x^{I\backslash j} = \tilde{x}^{I\backslash j}$.

- If $j = (k, v')$ with $k \leqslant t-2$ and $v' \in V$ then: $\quad C_{i,j} = 0$.
- If $j = (t-1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
  where the result follow from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (t, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \kappa^{\mathbf{2card}(N(v)\cap N(v'))}\right) & v' \in N^2(v)\backslash v \\ 0 & \text{otherwise} \end{cases}$,
  where the result follows from Lemma 7, obtained by a majorization of the observation density part. Recall that the only factors that contains v are the ones in $N(v)$ so the components that are connected to these factors are the ones in $N^2(v)$.
- If $j = (t+1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
  where the result follows from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (k, v')$ with $k \geqslant t+2$ and $v' \in V$ then: $\quad C_{i,j} = 0$.

- Consider $i = (T, v)$ and $v \in V$ then:

$$\rho_x^{(T,v)}(A) = \frac{\int \mathbb{I}_A(\tilde{x}_T^v)\mathbb{I}_{\{x_T^{V\setminus v}, x_{0:T-1}\}}(\tilde{x}_T^{V\setminus v}, \tilde{x}_{0:T-1}) \prod_{k=1}^{T} p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_T^v)\mathbb{I}_{\{x_T^{V\setminus v}, x_{0:T-1}\}}(\tilde{x}_T^{V\setminus v}, \tilde{x}_{0:T-1}) \prod_{k=1}^{T} p(\tilde{x}_{k-1}, \tilde{x}_k) g(\tilde{x}_k, y_k) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}$$

$$= \frac{\prod_{k=1}^{T-1} g(x_k, y_k) \prod_{k=0}^{T-1} p(x_{k-1}, x_k)}{\prod_{k=1}^{T-1} g(x_k, y_k) \prod_{k=0}^{T-1} p(x_{k-1}, x_k)}$$

$$\frac{\int\int \mathbb{I}_A(\tilde{x}_T^v)\mathbb{I}_{\{x_T^{V\setminus v}, x_{0:T-1}\}}(\tilde{x}_T^{V\setminus v}, \tilde{x}_{0:T-1}) p(\tilde{x}_{T-1}, \tilde{x}_T) g(\tilde{x}_T, y_T) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}{\int\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_T^v)\mathbb{I}_{\{x_T^{V\setminus v}, x_{0:T-1}\}}(\tilde{x}_T^{V\setminus v}, \tilde{x}_{0:T-1}) p(\tilde{x}_{T-1}, \tilde{x}_T) g(\tilde{x}_T, y_T) \bigotimes_{k=1}^{T} \psi(d\tilde{x}_k)\mu(d\tilde{x}_0)}$$

$$= \frac{\int \mathbb{I}_A(\tilde{x}_T^v) p(x_{T-1}^v, \tilde{x}_T^v) \prod_{f \in N(v)} g^f(x_T^{N(f)\setminus v}, y_T)\psi^v(dx_T^v)}{\int p(x_{T-1}^v, \tilde{x}_T^v) \prod_{f \in N(v)} g^f(x_T^{N(f)\setminus v}, y_T)\psi^v(d\tilde{x}_T^v)}.$$

  - If $j = (k, v')$ with $k \leqslant T-2$ and $v' \in V$ then: $\quad C_{i,j} = 0$.
  - If $j = (T-1, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
    where the result follows from Lemma 7, obtained by a majorization of the kernel part.
  - If $j = (T, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}\right) & v' \in N^2(v)\setminus v \\ 0 & \text{otherwise} \end{cases}$,
    where the result follows from Lemma 7, obtained by a majorization of the observation density part.

Given the previous results, for any $v \in K$:

$$\sum_{j \in I} e^{m(i,j)} C_{i,j} \leqslant \begin{cases} \sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}^{\mu}_{v,v'} + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & i = (0, v) \\ 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \sum_{v' \in N^2(v)} (1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}) e^{\beta d(v,v')} & i = (t, v) \\ e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \sum_{v' \in N^2(v)} (1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}) e^{\beta d(v,v')} & i = (T, v) \end{cases}$$

where $m(i,j) = \beta|k - k'| + \beta d(v, v')$ for $i = (k, v)$ and $j = (k', v')$ with $k, k' \in \{0, \ldots, T\}$ and $v, v' \in V$ is the pseudometric of interest on the index set I. But then by combining the above calculation with the assumption:

$$\begin{aligned}\max_{i\in I}\sum_{j\in I} C_{i,j} &\leqslant \max_{i\in I}\sum_{j\in I} e^{m(i,j)}C_{i,j}\\ &\leqslant \widetilde{\mathrm{Corr}}(\mu,\beta)+2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+\left(1-\kappa^{2\tilde{\Upsilon}}\right)\max_{v\in V}\sum_{v'\in N^2(v)} e^{\beta d(v,v')}\\ &= \widetilde{\mathrm{Corr}}(\mu,\beta)+2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\leqslant\frac{1}{2}.\end{aligned}$$

Given that $\sum_{j\in I} C_{i,j} \leqslant \sum_{j\in I} e^{m(i,j)}C_{i,j}$ then the Dobrushin theorem (Theorem 9) can be applied, meaning that:

$$\left\|\mathsf{F}_T\dots\mathsf{F}_1\mu-\mathsf{F}_T\dots\mathsf{F}_1\nu\right\|_J=\|\rho-\tilde{\rho}\|_{(T,J)}\leqslant\sum_{v\in J}\sum_{j\in I} D_{(T,v),j}b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9:

$$b_j=\sup_{x\in\mathbb{X}^I}\left\|\rho_x^j-\tilde{\rho}_x^j\right\|. \tag{A.3}$$

Remark that the form of ρ_x^j is already known from the study on $C_{i,j}$, hence we can compute just $\tilde{\rho}_x^j$ and then compare it.

- If $j=(0,v')$ and $v'\in V$ then:

$$\tilde{\rho}_x^j(A)=\frac{\int \mathbb{I}_A(x_0^{v'})p^{v'}(x_0^{v'},x_1^{v'})\nu_{x_0}^{v'}(dx_0^{v'})}{\int p^{v'}(x_0^{v'},x_1^{v'})\nu_{x_0}^{v'}(dx_0^{v'})}=\nu_{x_0,x_1}^{v}(A),$$

  where the procedure is exactly the same as for $\tilde{\rho}_x^{(0,v')}(A)$ with ν rather than μ, hence:

$$b_j=\sup_{x_0,x_1\in\mathbb{X}^V}\left\|\mu_{x_0,x_1}^{v'}-\nu_{x_0,x_1}^{v'}\right\|.$$

- If $j=(k',v')$ with $k'\geqslant 1$ and $v'\in V$ then:

$$\rho_x^j(A)=\tilde{\rho}_x^j(A),$$

  because the difference is only on the initial distribution which disappear as consequence of the Markov property, hence:

$$b_j=0.$$

Moreover, given that $\max_{i\in I}\sum\limits_{j\in I} e^{m(i,j)}C_{i,j}\leqslant\frac{1}{2}$ then Lemma 10 can be applied and so:

$$\max_{i\in I}\sum_{j\in J} e^{m(i,J)}D_{i,j}\leqslant 2.$$

The first step allows us to apply the Dobrushin theorem to the quantity of interest, while step two gives us a bound for the quantity of interest, hence by using the control over (A.3) we get:

$$\begin{aligned}
&\|\mathsf{F}_T \dots \mathsf{F}_1\mu - \mathsf{F}_T \dots \mathsf{F}_1\nu\|_J \\
&\quad \leqslant \sum_{v\in J}\sum_{j\in I} D_{(T,v),j} b_j \\
&\quad \leqslant \sum_{v\in J}\sum_{v'\in V} D_{(T,v),(0,v')} b_{(0,v')} \\
&\quad \leqslant \sum_{v\in J}\sum_{v'\in V} e^{\beta|T|+\beta d(v,v')} D_{(T,v),(0,v')} e^{-\beta|T|-\beta d(v,v')} \sup_{x_0,x_1\in\mathbb{X}^V} \left\| \mu^{v'}_{x_0,x_1} - \nu_{x_0,x_1} \right\| \\
&\quad \leqslant 2e^{-\beta T} \sum_{v\in J} \max_{v'\in V} \left\{ e^{-\beta d(v,v')} \sup_{x_0,x_1\in\mathbb{X}^V} \left\| \mu^{v'}_{x_0,x_1} - \nu_{x_0,x_1} \right\| \right\}.
\end{aligned}$$

Given that the above bound depends only on how many times the operator F_t is applied then:

$$\|\mathsf{F}_T \dots \mathsf{F}_{t+1}\mu - \mathsf{F}_T \dots \mathsf{F}_{t+1}\nu\|_J \leqslant 2e^{-\beta(T-t)} \sum_{v\in J} \max_{v'\in V} \left\{ e^{-\beta d(v,v')} \sup_{x,z\in\mathbb{X}^V} \left\| \mu^{v'}_{x,z} - \nu^{v'}_{x,z} \right\| \right\}.$$

■

A.2.2 Control on the approximation error

Our next objective is to bound the approximation errors:

$$\begin{aligned}
&\sup_{x,z\in\mathbb{X}^V} \left\| (\tilde{\mathsf{F}}^m_t \tilde{\pi}_{t-1})^v_{x,z} - (\mathsf{F}_t \tilde{\pi}_{t-1})^v_{x,z} \right\|, \quad t < T, \\
&\left\| \tilde{\mathsf{F}}^m_T \tilde{\pi}_{t-1} - \mathsf{F}_t \tilde{\pi}_{t-1} \right\|_J, \quad t = T,
\end{aligned}$$

or equivalently:

$$\begin{aligned}
&\sup_{x,z\in\mathbb{X}^V} \left\| (\tilde{\mathsf{C}}^m_t \mu)^v_{x,z} - (\mathsf{C}_t \mu)^v_{x,z} \right\|, \quad t < T, \\
&\left\| \tilde{\mathsf{C}}^m_t \mu - \mathsf{C}_t \mu \right\|_J, \quad t = T,
\end{aligned}$$

where $\mu = \mathsf{P}\tilde{\pi}_{t-1}$.

Proposition 15 *(case: $t < T$) Fix any collection of observations $\{y_1, \dots, y_T\}$ and any partition $\mathcal{K}$ of V. Suppose that there exist $\kappa \in (0,1)$ such that:*

$$\kappa \leqslant g^f\left(x^{N(f)}, y_t\right) \leqslant \frac{1}{\kappa},$$

for all $x \in \mathbb{X}^V, f \in F, t \in \{1, \dots, T\}$ *and* $p^v(x^v, z^v) > 0$ *for all* $x, z \in \mathbb{X}^V, v \in V$. *Let* μ *be a probability distribution on* $\mathbb{X}^V$ *such that* $\mu = \bigotimes_{K \in \mathcal{K}} \mu^K$ *and assume that there exists* $\beta > 0$ *such that:*

$$2\kappa^{-2\Upsilon} Corr(\mu, \beta) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Then for a fixed $t \in \{1, \dots, T-1\}$, $K \in \mathcal{K}$, $v \in K$ *and* $m \in \{0, \dots, n\}$*:*

$$\sup_{x,z \in \mathbb{X}^V} \left\| (\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\| \leqslant 2\left(1 - \kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1 - \kappa^{b(\mathcal{K},m)}\right),$$

where

$$\begin{aligned} a(\mathcal{K}) &:= 2 \max_{K \in \mathcal{K}} \max_{v \in \partial K} \mathbf{card}(N(v) \cap \partial N(K)), \\ b(m, \mathcal{K}) &:= 2 \max_{K \in \mathcal{K}} \max_{v \notin N_v^{m-1}(K)} \mathbf{card}(N(v)). \end{aligned}$$

Proof Remark that if $\mu = \bigotimes_{K' \in \mathcal{K}} \mu^{K'}$ and we choose $v \in K$ with $K \in \mathcal{K}$ then from Lemma 6 we have:

$$\begin{aligned} (\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v(A) &= \frac{\int \mathbb{I}_A(x^v) \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)} \\ &= (\tilde{\mathsf{C}}_t^{m,K} \mu)_{x,z}^v(A), \end{aligned}$$

where $x, z \in \mathbb{X}^V$.

Given $v \in K$ with $K \in \mathcal{K}$, it must be noticed that we can distinguish two cases: v connected with factors that are connected only with elements inside K (using our notation $v \in \tilde{K}$) and its complement (exists a factor connected with v that is connected with elements outside K).

Consider the case $v \in \widetilde{K}$, then $N(v)$ are factors that depend only on components in K, then for $x, z \in \mathbb{X}^V$:

$$\begin{aligned} &(\tilde{\mathsf{C}}_t^{m,K} \mu)_{x,z}^v(A) \\ &= \frac{\int \mathbb{I}_A(x^v) \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) \int \prod_{f \in N_f^m(K) \setminus N(v)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) \int \prod_{f \in N_f^m(K) \setminus N(v)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)} \\ &= \frac{\int \prod_{f \in N_f^m(K) \setminus N(v)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x})}{\int \prod_{f \in N_f^m(K) \setminus N(v)} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x})} \frac{\int \mathbb{I}_A(x^v) \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)} \\ &= \frac{\int \mathbb{I}_A(x^v) \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in N(v)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)} = (\mathsf{C}_t \mu)_{x,z}^v(A), \end{aligned}$$

where the last passage follows because we remove all the dependencies on v inside the previous integral. From the above procedure we have proved that $(\tilde{\mathsf{C}}_t^{m,K} \mu)_{x,z}^v(A) = (\mathsf{C}_t \mu)_{x,z}^v(A)$

for $v \in \widetilde{K}$ and so:

$$\sup_{x,z \in \mathbb{X}^V} \left\| (\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\| = 0,$$

which proves the statement for $v \in \tilde{K}$.

Consider now the case $v \in \partial K$, using the triangular inequality:

$$\left\| (\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\| \leqslant \left\| (\tilde{\mathsf{C}}_t^m \mu)_{x,z}^v - (\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v \right\| + \left\| (\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\|,$$

where $n \coloneqq \max_{K \in \mathcal{K}} n_K$.

Firstly, we control $\left\| (\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\|$ by rewriting $(\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v$ as an integration of $(\mathsf{C}_t \mu)_{x,z}^v$. Indeed, given that $N_f^n(K) = F$, it is possible to rearrange $(\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v$ as follows:

$$\begin{aligned}
(\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v(A) &= \frac{\int \mathbb{I}_A(x^v) \int \prod_{f \in F} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int \prod_{f \in F} g^f(x_K^{N(f)}, y_t) \mu^{V \setminus K}(d\tilde{x}) p^v(x^v, z^v) \mu_x^v(dx^v)} \\
&= \frac{\int \mathbb{I}_A(x^v) g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \mu^{V \setminus K}(d\tilde{x})}{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \mu^{V \setminus K}(d\tilde{x})} \\
&= \frac{\int \int \mathbb{I}_A(x^v) g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \frac{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)} \mu^{V \setminus K}(d\tilde{x})}{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \mu^{V \setminus K}(d\tilde{x})} \\
&= \int \frac{(\mathsf{C}_t \mu)_{x_K,z}^v(A) \int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \mu^{V \setminus K}(d\tilde{x})} \mu^{V \setminus K}(d\tilde{x}),
\end{aligned}$$

where $x, z \in \mathbb{X}^V$. But then:

$$\begin{aligned}
&|(\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v(A) - (\mathsf{C}_t \mu)_{x,z}^v(A)| \\
&\quad = |\int \frac{[(\mathsf{C}_t \mu)_{x_K,z}^v(A) - (\mathsf{C}_t \mu)_{x,z}^v(A)] \int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v)}{\int g(x_K, y_t) p^v(x^v, z^v) \mu_x^v(dx^v) \mu^{V \setminus K}(d\tilde{x})} \mu^{V \setminus K}(d\tilde{x})| \\
&\quad \leqslant \sup_{\substack{\hat{x}, x \in \mathbb{X}^{V \setminus v}: \\ \hat{x}^K = x^K}} |(\mathsf{C}_t \mu)_{\hat{x},z}^v(A) - (\mathsf{C}_t \mu)_{x,z}^v(A)|.
\end{aligned}$$

But given that A is arbitrary it follows:

$$\left\| (\tilde{\mathsf{C}}_t^n \mu)_{x,z}^v - (\mathsf{C}_t \mu)_{x,z}^v \right\| \leqslant \sup_{\substack{\hat{x}, x \in \mathbb{X}^{V \setminus v}: \\ \hat{x}^K = x^K}} \left\| (\mathsf{C}_t \mu)_{\hat{x},z}^v(A) - (\mathsf{C}_t \mu)_{x,z}^v(A) \right\|.$$

Note that the supremum is constrained on $\hat{x}^K = x^K$ meaning that we can try to remove the factors in $N(v)$ that call elements outside K:

$$(\mathsf{C}_t\mu)^v_{\hat{x},z}(A) \geqslant \frac{\int \mathbb{I}_A(\hat{x}^v) \prod_{f\in N(v)\cap\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) \prod_{f\in N(v)\backslash\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) p^v(\hat{x}^v, z^v)\mu^v_{\hat{x}}(d\hat{x}^v)}{\int \prod_{f\in N(v)\cap\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) \prod_{f\in N(v)\backslash\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) p^v(\hat{x}^v, z^v)\mu^v_{\hat{x}}(d\hat{x}^v)}$$
$$\geqslant \kappa^{2\mathbf{card}(N(v)\cap\partial N(K))} \frac{\int \mathbb{I}_A(\hat{x}^v) \prod_{f\in N(v)\backslash\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) p^v(\hat{x}^v, z^v)\mu^v_{\hat{x}}(d\hat{x}^v)}{\int \prod_{f\in N(v)\backslash\partial N(K)} g^f(\hat{x}^{N(f)}, y_t) p^v(\hat{x}^v, z^v)\mu^v_{\hat{x}}(d\hat{x}^v)},$$

where the majorization follows from the assumption on the kernel density. The procedure can be repeated for $(\mathsf{C}_t\mu)^v_{x,z}$ and we obtain the same inequality, that because all the differences between x and $\hat{x}$ are outside K. Hence by applying Lemma 7:

$$\left\|(\tilde{\mathsf{C}}^n_t\mu)^v_{x,z} - (\mathsf{C}_t\mu)^v_{x,z}\right\| \leqslant 2\left(1-\kappa^{2\mathbf{card}(N(v)\cap\partial N(K))}\right) \leqslant 2\left(1-\kappa^{2\max_{K\in\mathcal{K}}\max_{v\in\partial K}\mathbf{card}(N(v)\cap\partial N(K))}\right)$$
$$\leqslant 2\left(1-\kappa^{a(\mathcal{K})}\right).$$

Secondly, we control $\left\|(\tilde{\mathsf{C}}^m_t\mu)^v_{x,z} - (\tilde{\mathsf{C}}^n_t\mu)^v_{x,z}\right\|$, to do so we will use the Dobrushin theorem. Define the probability distributions:

$$\rho(A) := \frac{\int \mathbb{I}_A(x_K^{V\backslash K}, x^K) g(x_K, y_t) p^v(x^v, z^v)\mu^{V\backslash K}(d{x_K}^{V\backslash K})\mu^K(dx^K)}{\int g(x_K, y_t) p^v(x^v, z^v)\mu^{V\backslash K}(d{x_K}^{V\backslash K})\mu^K(dx^K)},$$

$$\tilde{\rho}(A) := \frac{\int \mathbb{I}_A(x_K^{V\backslash K}, x^K) \prod_{f\in N^m_f(K)} g(x_K^{N(f)}, y_t) p^v(x^v, z^v)\mu^{V\backslash K}(d{x_K}^{V\backslash K})\mu^K(dx^K)}{\int \prod_{f\in N^m_f(K)} g(x_K^{N(f)}, y_t) p^v(x^v, z^v)\mu^{V\backslash K}(d{x_K}^{V\backslash K})\mu^K(dx^K)},$$

where $x, z \in \mathbb{X}^V$. It can be observed that by construction:

$$\|\rho - \tilde{\rho}\|_{(1,v)} = \left\|(\tilde{\mathsf{C}}^m_t\mu)^v_{x,z} - (\tilde{\mathsf{C}}^n_t\mu)^v_{x,z}\right\|,$$

meaning that the Dobrushin theorem can be applied to $\rho, \tilde{\rho}$ where the index set is $I = (0, V\backslash K) \cup (1, K)$ and the subset is $(1, v)$. Remark that in this case the first number has not a meaning of time, but they are just some indexes that distinguish the spaces.

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9.

- Let consider $i = (0, b)$ and $b \in V \backslash K$ with $b \in K'$ then:

$$
\begin{aligned}
\rho^{(0,b)}_{x^{V\backslash K}, x^K}(A) &= \frac{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K\backslash b}, x^K\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) g(\tilde{x}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})\mu^K(d\tilde{x}^K)}{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K\backslash b}, x^K\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) g(\tilde{x}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})\mu^K(d\tilde{x}^K)} \\
&= \frac{\prod_{f \in F\backslash N(b)} g(x^{N(b)}, y_t)}{\prod_{f \in F\backslash N(b)} g(x^{N(b)}, y_t)} \\
&\quad \frac{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K\backslash b}, x^K\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) \prod_{f \in N(b)} g(\tilde{x}^{N(b)}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K\backslash b}, x^K\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) \prod_{f \in N(b)} g(\tilde{x}^{N(b)}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})} \\
&= \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)\mu^b_x(dx^b)}{\int \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)\mu^b_x(dx^b)} \\
&= \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)(\mu^{K'})^b_x(dx^b)}{\int \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)(\mu^{K'})^b_x(dx^b)}.
\end{aligned}
$$

  - If $j = (0, b')$ and $b' \in V \backslash K$ then we have to distinguish two cases:
    * if $b' \in N^2(b)$ then by Lemma 7:

$$
C_{i,j} \leqslant \left(1 - \kappa^{2\mathbf{card}(N(b)\cap N(b'))}\right) + \kappa^{2\mathbf{card}(N(b)\cap N(b'))} C^{\mu}_{b,b'};
$$

    * if $b' \notin N^2(b)$ then we have the form $\frac{\int \mathbb{I}_A(x)\Lambda(x)\nu(dx)}{\int \Lambda(x)\nu(dx)}$ and we can apply Lemma 8:

$$
C_{i,j} \leqslant 2\kappa^{-2\mathbf{card}(N(b))} C^{\mu}_{b,b'},
$$

      where the bound follows from a multiple application of the kernel assumption $\kappa^{\mathbf{card}(N(b))} \leqslant \prod_{f \in N(b)} g^f(x^{N(f)}, y_t) \leqslant \kappa^{-\mathbf{card}(N(b))}$.
  - If $j = (1, b')$ and $b' \in K$ then from Lemma 7:

$$
C_{i,j} \leqslant \begin{cases} (1 - \kappa^{2\mathbf{card}(N(b)\cap N(b'))}) & \text{if } b' \in N^2(b) \\ 0 & \text{otherwise} \end{cases},
$$

    given that we are considering $K \neq K'$, because $b \in V \backslash K$, then for sure the only difference is in the factors because the conditional distribution K' depends only on elements inside K'.

- Let consider $i = (1, b)$ and $b \in K$ then:

$$
\begin{aligned}
\rho^{(1,b)}_x(A) &= \frac{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K}, x^{K\backslash b}\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) g(\tilde{x}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})\mu^K(d\tilde{x}^K)}{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{\{x^{V\backslash K}, x^{K\backslash b}\}}(\tilde{x}^{V\backslash K\backslash b}, \tilde{x}^K) g(\tilde{x}, y_t) p^v(\tilde{x}^v, z^v)\mu^{V\backslash K}(d\tilde{x}^{V\backslash K})\mu^K(d\tilde{x}^K)} \\
&= \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)[p^v(\tilde{x}^v, z^v)]^{\mathbb{I}_b(v)}\mu^b_x(dx^b)}{\int \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)[p^v(\tilde{x}^v, z^v)]^{\mathbb{I}_b(v)}\mu^b_x(dx^b)} \\
&= \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)[p^v(\tilde{x}^v, z^v)]^{\mathbb{I}_b(v)}(\mu^K)^b_x(dx^b)}{\int \prod_{f \in N(b)} g^f(x^{N(f)}, y_t)[p^v(\tilde{x}^v, z^v)]^{\mathbb{I}_b(v)}(\mu^K)^b_x(dx^b)},
\end{aligned}
$$

where the procedure is the same as in $i = (0, b)$. Remark that we can avoid considering $[p^v(\tilde{x}^v, z^v)]^{\mathbb{I}_b(v)}$ because if $b = v$ then that variable is integrated out.

- If $j = (0, b')$ and $b' \in V \backslash K$ then from Lemma 7:

$$C_{i,j} \leqslant \begin{cases} (1 - \kappa^{\mathbf{2card}(N(b) \cap N(b'))}) & b' \in N^2(b) \\ 0 & \text{otherwise} \end{cases},$$

again the difference can be only on the factors because b' is outside K.

- If $j = (1, b')$ and $b' \in K$ then we have to distinguish two cases:
  * if $b' \in N^2(b)$ then by Lemma 7:

$$C_{i,j} \leqslant \left(1 - \kappa^{\mathbf{2card}(N(b) \cap N(b'))}\right) + \kappa^{\mathbf{2card}(N(b) \cap N(b'))} C^{\mu}_{b,b'};$$

  * if $b' \notin N^2(b)$ then by Lemma 8:

$$C_{i,j} \leqslant 2\kappa^{-\mathbf{2card}(N(b))} C^{\mu}_{b,b'}.$$

But then:

$$\begin{aligned}
\max_{i \in I} \sum_{j \in I} C_{i,j} \leqslant \max_{i \in I} \sum_{j \in I} e^{m(i,j)} C_{i,j} \leqslant \max_{b \in V} \Bigg\{ & \sum_{b' \in N^2(b)} e^{m((0,b),(0,b'))} \left[\left(1 - \kappa^{\mathbf{2card}(N(b) \cap N(b'))}\right) \right. \\
& \left. + \kappa^{\mathbf{2card}(N(b) \cap N(b'))} C^{\mu}_{b,b'} \right] + \sum_{b' \notin N^2(b)} e^{m((0,b),(0,b'))} 2\kappa^{-\mathbf{2card}(N(b))} C^{\mu}_{b,b'} \\
& + \sum_{b' \in N^2(b)} e^{m((0,b),(1,b'))} \left(1 - \kappa^{\mathbf{2card}(N(b) \cap N(b'))}\right) \Bigg\} \\
\leqslant 2e^{2\beta} \Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) + \max_{b \in V} \Bigg\{ & \sum_{b' \in N^2(b)} 2\kappa^{-\mathbf{2card}(N(b))} e^{\beta d(b,b')} C^{\mu}_{b,b'} \\
& + \sum_{b' \notin N^2(b)} 2\kappa^{-\mathbf{2card}(N(b))} e^{\beta d(b,b')} C^{\mu}_{b,b'} \Bigg\} \\
\leqslant 2e^{2\beta} \Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) + & 2\kappa^{-2\Upsilon} \max_{b \in V} \sum_{b' \in V} e^{\beta d(b,b')} C^{\mu}_{b,b'} \\
\leqslant 2\kappa^{-2\Upsilon} \operatorname{Corr}(\mu, \beta) + & 2e^{2\beta} \Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2},
\end{aligned}$$

where $m(i, j) = \beta d(v, v')$ for $i = (k, v)$ and $j = (k, v')$ with $v, v' \in V$ and $k, k' \in \{0, 1\}$ is the pseudometric of interest on the index set. Hence the Dobrushin theorem applies:

$$\left\| (\tilde{\mathsf{C}}^m_t \mu)^v_{x,z} - (\tilde{\mathsf{C}}^n_t \mu)^v_{x,z} \right\| = \|\rho - \tilde{\rho}\|_{(1,v)} \leqslant \sum_{j \in I} D_{(1,v),j} b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9. Recall that if $b' \in N_v^{m-1}(K)$ then $N(b') \subset N_f^m(K)$. Given the form of the conditional distribution of $\tilde{\rho}$ is the same of the conditional distribution of ρ with a restricted number of factors we can analyse $\tilde{\rho}$ first and then extend the form to ρ.

- If $j = (0, b')$ and $b' \in V \backslash K$, then:

$$\tilde{\rho}_x^{(0,b')}(A)$$

$$= \frac{\int \mathbb{I}_A(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K\backslash b'}, x^K\}}(\tilde{x}^{V\backslash K\backslash b'}, \tilde{x}^K) \prod_{f \in N_f^m(K)} g^f(\tilde{x}^{N(f)}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K}) \mu^K(d\tilde{x}^K)}{\int \mathbb{I}_A(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K\backslash b'}, x^K\}}(\tilde{x}^{V\backslash K\backslash b'}, \tilde{x}^K) \prod_{f \in N_f^m(K)} g^f(\tilde{x}^{N(f)}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K}) \mu^K(d\tilde{x}^K)}$$

$$= \frac{\prod_{f \in N_f^m(K) \backslash N(b')} g(\tilde{x}^{N(b')}, y_t)}{\prod_{f \in N_f^m(K) \backslash N(b')} g(\tilde{x}^{N(b')}, y_t)}$$

$$\frac{\int \mathbb{I}_A(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K\backslash b'}, x^K\}}(\tilde{x}^{V\backslash K\backslash b'}, \tilde{x}^K) \prod_{f \in N_f^m(K) \cap N(b')} g(\tilde{x}^{N(b')}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K\backslash b'}, x^K\}}(\tilde{x}^{V\backslash K\backslash b'}, \tilde{x}^K) \prod_{f \in N_f^m(K) \cap N(b')} g(\tilde{x}^{N(b')}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K})}$$

$$= \frac{\int \mathbb{I}_A(\tilde{x}^{b'}) \prod_{f \in N_f^m(K) \cap N(b')} g^f(x_K^{N(f)}, y_t) \mu_x^{b'}(d\tilde{x}^{b'})}{\int \prod_{f \in N_f^m(K) \cap N(b')} g^f(x_K^{N(f)}, y_t) \mu_x^{b'}(d\tilde{x}^{b'})}.$$

  Remark that $N(N_v^{m-1}(K)) = N_f^m(K)$ so if $b' \in N_v^{m-1}(K)$ then $N(b') \subseteq N_f^m(K)$, hence by Lemma 7:

$$b_j \leqslant \begin{cases} 2(1 - \kappa^{2\mathbf{card}(N(b'))}) & b' \notin N_v^{m-1}(K) \\ 0 & \text{otherwise} \end{cases},$$

  where the result follows from the majorization of the observation density in $\rho_x^{(0,b')}$ in the worst case scenario when $N_f^m(K) \cap N(b') = \varnothing$.

- If $j = (1, b')$ and $b' \in K$ then:

$$\tilde{\rho}_x^{(1,b')}(A)$$

$$= \frac{\int \mathbb{I}_A(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K}, x^{K\backslash b'}\}}(\tilde{x}^{V\backslash K}, \tilde{x}^{K\backslash b'}) \prod_{f \in N_f^m(K)} g^f(\tilde{x}^{N(f)}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K}) \mu^K(d\tilde{x}^K)}{\int \mathbb{I}_A(\tilde{x}^{b'})\mathbb{I}_{\{x^{V\backslash K}, x^{K\backslash b'}\}}(\tilde{x}^{V\backslash K}, \tilde{x}^{K\backslash b'}) \prod_{f \in N_f^m(K)} g^f(\tilde{x}^{N(f)}, y_t) p^v(\tilde{x}^v, z^v) \mu^{V\backslash K}(d\tilde{x}^{V\backslash K}) \mu^K(d\tilde{x}^K)}$$

$$= \frac{\int \mathbb{I}_A(x^{b'}) \prod_{f \in N_f^m(K) \cap N(b')} g^f(x^{N(f)}, y_t) \mu_x^{b'}(dx^{b'})}{\int \prod_{f \in N_f^m(K) \cap N(b')} g^f(x^{N(f)}, y_t) \mu_x^{b'}(dx^{b'})},$$

  where the procedure is the same as in $i = (0, b')$. Given that $b' \in K$ then surely $b' \in N_v^{m-1}(K)$ hence:

$$b_j = 0.$$

Given all the bounds on b_j's and by applying Lemma 10:

$$\begin{aligned}\left\|(\tilde{\mathsf{C}}_t^m\mu)_{x,z}^v - (\tilde{\mathsf{C}}_t^n\mu)_{x,z}^v\right\| &\leqslant \sum_{j\in I} D_{(1,v),j} b_j \leqslant 2\left(1-\kappa^{2\max_{K\in\mathcal{K}}\{\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v))\}}\right)\\ &\quad \sum_{b'\notin N_v^{m-1}(K)} e^{\beta d(v,b')} e^{-\beta d(v,b')} D_{(1,v),(1,b')}\\ &\leqslant 4\left(1-\kappa^{b(\mathcal{K},m)}\right) e^{-\beta d(v,V\setminus N_v^{m-1}(K))}\\ &\leqslant 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right),\end{aligned}$$

where the last passage follows from $v \in J \subseteq K$ and so:

$$d(v, V\setminus N_v^{m-1}(K)) \geqslant d(K, V\setminus N_v^{m-1}(K)) \geqslant m,$$

indeed the minimum distance between $v \in J$ and $V\setminus N_v^{m-1}(K)$ is bigger than the minimum distance between K and $V\setminus N_v^{m-1}(K)$ given that $J \subseteq K$, but at the same time we are removing all the element that are far $m-1$ from K (the ones in $N_v^{m-1}(K)$) then the minimum distance is surely m. Hence we can conclude that:

$$\left\|(\tilde{\mathsf{C}}_t^m\mu)_{x,z}^v - (\mathsf{C}_t\mu)_{x,z}^v\right\| \leqslant 2(1-\kappa^{a(\mathcal{K})}) + 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right).$$

■

Proposition 16 *Fix any collection of observations $\{y_1,\ldots,y_T\}$ and any partition $\mathcal{K}$ on the set V. Suppose that there exist $\kappa \in (0,1)$ such that:*

$$\kappa \leqslant g^f\left(x^{N(f)}, y_t\right) \leqslant \frac{1}{\kappa},$$

for all $x \in \mathbb{X}^V, f \in F, t \in \{1,\ldots,T\}$. Let μ be a probability distribution on $\mathbb{X}^V$ such that $\mu = \bigotimes_{K\in\mathcal{K}} \mu^K$ and assume that there exists $\beta > 0$ such that:

$$2\kappa^{-2\Upsilon} Corr(\mu,\beta) + 2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Then for a fixed $t \in \{1,\ldots,T-1\}$, $K \in \mathcal{K}$, $v \in K$ and $m \in \{0,\ldots,n\}$:

$$\sup_{x\in\mathbb{X}^V}\left\|(\tilde{\mathsf{C}}_t^{m,K}\mu)_x^v - (\mathsf{C}_t\mu)_x^v\right\| \leqslant 2\left(1-\kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right),$$

where

$$a(\mathcal{K}) \coloneqq 2\max_{K\in\mathcal{K}}\max_{v\in\partial K}\mathbf{card}(N(v)\cap\partial N(K)),$$

$$b(m,\mathcal{K}) \coloneqq 2\max_{K\in\mathcal{K}}\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v)).$$

Proof The proof follows the same procedure of Proposition 15, where we have just to remove the kernel term $p^v(x^v, z^v)$. ■

Consider now the case $t = T$, here we are not comparing conditional distributions, but we are considering marginal distributions over a set J. This remove the dependence from the components outside K, with $K \in \mathcal{K}$ and $J \subseteq K$, meaning that our bound is simpler than the one in Proposition 15. Moreover, this case can be decoupled from the FHMM scenario, indeed we are considering an approximated Bayes update. Note that Proposition 17 is equivalent to prove Proposition 1, given that T is completely arbitrary.

Proposition 17 *(case: $t = T$) Fix any observation y_T and any partition $\mathcal{K}$ on the set V. Suppose there exists $\kappa \in (0,1)$ such that:*

$$\kappa \leqslant g^f(x^{N(f)}, y_T) \leqslant \frac{1}{\kappa},$$

for all $x \in \mathbb{X}^V, f \in F$. Let μ be a probability distribution on $\mathbb{X}^V$ and assume that there exists $\beta > 0$ such that:

$$2\kappa^{-2\Upsilon} Corr(\mu, \beta) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Then for a fixed $v \in K \in \mathcal{K}$, $J \subseteq K$ and $m \in \{0, \dots, n\}$:

$$\left\| \mathsf{C}_T\mu - \tilde{\mathsf{C}}_T^m\mu \right\|_J \leqslant 4\left(1 - \kappa^{b(m,\mathcal{K})}\right) \mathbf{card}(J) e^{-\beta m},$$

where $b(m, \mathcal{K}) \coloneqq 2\max_{K\in\mathcal{K}} \max_{v\notin N_v^{m-1}(K)} \mathbf{card}(N(v))$.

Proof Given that $J \subseteq K$ then:

$$\left\| \mathsf{C}_T\mu - \tilde{\mathsf{C}}_T^m\mu \right\|_J = \left\| (\mathsf{C}_T\mu)^K - \tilde{\mathsf{C}}_T^{m,K}\mu \right\|_J,$$

which is obvious because at the end we want to compare marginals on J that is equivalent to marginalize first on K and then marginalize again on J. Remark that:

$$(\tilde{\mathsf{C}}_T^{m,K}\mu)(A) = \frac{\int \mathbb{I}_A(x^K) \prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y_T)\mu(dx)}{\int \prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y_T)\mu(dx)},$$

$$(\mathsf{C}_T\mu)^K(A) = \frac{\int \mathbb{I}_A(x^K) g(x, y_T)\mu(dx)}{\int g(x, y_t)\mu(dx)}.$$

Define the probability distributions:

$$\rho(A) \coloneqq \frac{\int \mathbb{I}_A(x) g(x, y_T)\mu(dx)}{\int g(x, y_t)\mu(dx)},$$

$$\tilde{\rho}(A) \coloneqq \frac{\int \mathbb{I}_A(x) \prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y_T)\mu(dx)}{\int \prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y_T)\mu(dx)}.$$

It can be observed that by definition:

$$\|\rho - \tilde{\rho}\|_J = \left\| (\mathsf{C}_T\mu)^K - \tilde{\mathsf{C}}_T^{m,K}\mu \right\|_J,$$

meaning that the Dobrushin theorem can be applied to $\rho, \tilde{\rho}$ where the index set is $I = V$.

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9.

- Let consider $i = v$ and $v \in V$ then:

$$\begin{aligned}
\rho_x^v(A) &= \frac{\int \mathbb{I}_A(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T) \prod_{f\in F\setminus N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T) \prod_{f\in F\setminus N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})} \\
&= \frac{\prod_{f\in F\setminus N(v)} g^f(x^{N(f)}, y_T)}{\prod_{f\in F\setminus N(v)} g^f(x^{N(f)}, y_T)} \frac{\int \mathbb{I}_A(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^v)\mathbb{I}_{x^{V\setminus v}}(\tilde{x}^{V\setminus v}) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})} \\
&= \frac{\int \mathbb{I}_A(\tilde{x}^v) \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu_x^v(d\tilde{x}^v)}{\int \prod_{f\in N(v)} g^f(\tilde{x}^{N(f)}, y_T)\mu_x^v(d\tilde{x}^v)}.
\end{aligned}$$

  - If $j = v'$ and $v' \in V$ then we have to distinguish two cases:
    - if $v' \in N^2(b)$ then by Lemma 7:

$$C_{i,j} \leqslant \left(1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}\right) + \kappa^{2\mathbf{card}(N(v)\cap N(v'))} C^{\mu}_{v,v'}.$$

    - if $v' \notin N^2(v)$ then by Lemma 8:

$$C_{i,j} \leqslant 2\kappa^{-2\mathbf{card}(N(v))} C^{\mu}_{v,v'}.$$

But then by assumption:

$$\begin{aligned}
\max_{i\in I} \sum_{j\in I} C_{i,j} \leqslant \max_{v\in V} \Bigg\{ &\sum_{v'\in N^2(v)} e^{m(v,v')} \left[\left(1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}\right)\right. \\
&\left. + \kappa^{2\mathbf{card}(N(v)\cap N(v'))} C^{\mu}_{v,v'}\right] + \sum_{v'\notin N^2(v)} e^{m(v,v')} 2\kappa^{-2\mathbf{card}(N(v))} C^{\mu}_{v,v'} \Bigg\} \\
\leqslant\ & 2\kappa^{-2\Upsilon}\mathrm{Corr}(\mu,\beta) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}
\end{aligned}$$

where $m(i,j) = \beta d(v, v')$ is the pseudometric of the index set. Hence the Dobrushin theorem applies:

$$\left\|(\mathsf{C}_T\mu)^K - \tilde{\mathsf{C}}_T^{m,K}\mu\right\|_J = \|\rho - \tilde{\rho}\|_J \leqslant \sum_{i\in J}\sum_{j\in V} D_{i,j} b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9.

- If $j = v'$ and $v' \in V$:

$$
\begin{aligned}
&\tilde{\rho}_x^{v'}(A) \\
&= \frac{\int \mathbb{I}_A(\tilde{x}^{v'})\mathbb{I}_{x^{V\setminus v'}}(\tilde{x}^{V\setminus v'}) \prod_{f\in N_f^m(K)\cap N(v')} g^f(\tilde{x}^{N(f)}, y_T) \prod_{f\in N_f^m(K)\setminus N(v')} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^{v'})\mathbb{I}_{x^{V\setminus v'}}(\tilde{x}^{V\setminus v'}) \prod_{f\in N_f^m(K)\cap N(v')} g^f(\tilde{x}^{N(f)}, y_T) \prod_{f\in N_f^m(K)\setminus N(v')} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})} \\
&= \frac{\prod_{f\in N_f^m(K)\setminus N(v')} g^f(x^{N(f)}, y_T)}{\prod_{f\in N_f^m(K)\setminus N(v')} g^f(x^{N(f)}, y_T)} \\
&\quad \cdot \frac{\int \mathbb{I}_A(\tilde{x}^{v'})\mathbb{I}_{x^{V\setminus v'}}(\tilde{x}^{V\setminus v'}) \prod_{f\in N_f^m(K)\cap N(v')} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^{v'})\mathbb{I}_{x^{V\setminus v'}}(\tilde{x}^{V\setminus v'}) \prod_{f\in N_f^m(K)\cap N(v')} g^f(\tilde{x}^{N(f)}, y_T)\mu(d\tilde{x})} \\
&= \frac{\int \mathbb{I}_A(x^v) \prod_{f\in N_f^m(K)\cap N(v')} g^f(x^{N(f)}, y_T)\mu_x^v(dx^{v'})}{\int \prod_{f\in N_f^m(K)\cap N(v')} g^f(x^{N(f)}, y_T)\mu_x^v(dx^{v'})}.
\end{aligned}
$$

  Remark that $v' \in N_v^{m-1}(K)$ implies $N(v') \subseteq N_f^m(K)$, hence:

$$
b_j = \begin{cases} 2(1-\kappa^{2\mathbf{card}(N(v'))}) & v' \in N_v^{m-1}(K) \\ 0 & \text{otherwise} \end{cases},
$$

  where the bound follows from an application of Lemma 8.

Given all the bounds on b_j's and by using Lemma 10:

$$
\begin{aligned}
\left\| (\mathsf{C}_T\mu)^K - \tilde{\mathsf{C}}_T^{m,K}\mu \right\|_J &\leqslant \sum_{i\in J}\sum_{j\in V} D_{i,j}b_j \leqslant 2\left(1-\kappa^{2\max_{K\in\mathcal{K}}\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v))}\right)\sum_{v\in J}\sum_{v'\notin N_v^{m-1}(K)} D_{v,v'} \\
&\leqslant 4\left(1-\kappa^{2\max_{K\in\mathcal{K}}\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v))}\right)\sum_{v\in J} e^{-\beta d(v,V\setminus N_v^{m-1}(K))} \\
&\leqslant 4\left(1-\kappa^{b(m,\mathcal{K})}\right)\mathbf{card}(J)e^{-\beta m},
\end{aligned}
$$

where the last part follows from the same observation on the distance as in Proposition 15. ∎

Proof of Proposition 1 An application of Proposition 17, since there the time step T is arbitrary. ∎

A.2.3 Decay of correlation

Our next step is to prove that the following decay of correlation conditions hold uniformly in t:

$$
\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}_t^m\tilde{\pi}_{t-1}, \beta) + 2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2},
$$

$$2\kappa^{-2\Upsilon}\text{Corr}(\mathsf{P}\tilde{\pi}_{t-1},\beta) + 2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Proposition 18 *Suppose there exists* $(\epsilon_-, \epsilon_+) \in (0,1)$ *such that:*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+,$$

for all $x, z \in \mathbb{X}^V, v \in V.$*. Given a probability distribution* μ *on* $\mathbb{X}^V$ *assume that there exists* $\beta > 0$ *such that:*

$$\widetilde{Corr}(\mu,\beta) + e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right) \leqslant \frac{1}{2},$$

then:

$$Corr(\mathsf{P}\mu,\beta) \leqslant 2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)e^{-\beta}.$$

Proof Recall that:

$$(\mathsf{P}\mu)^v_z = \frac{\int \mathbb{I}_A(z^v)p(x,z)\mu(dx)\psi^v(dz^v)}{\int p(x,z)\mu(dx)\psi^v(dz^v)},$$

where ψ is the counting measure. Recall also that:

$$\text{Corr}(\mathsf{P}\mu,\beta) = \max_{v\in V}\sum_{v'\in V} e^{\beta d(v,v')}C^{\mathsf{P}\mu}_{v,v'},$$

where:

$$C^{\mathsf{P}\mu}_{v,v'} = \frac{1}{2}\sup_{\substack{z,\tilde{z}\in\mathbb{X}^V:\\ z^{V\setminus v'}=\hat{z}^{V\setminus v'}}} \left\|(\mathsf{P}\mu)^v_z - (\mathsf{P}\mu)^v_{\hat{z}}\right\|.$$

The strategy is hence to firstly control:

$$\left\|(\mathsf{P}\mu)^v_z - (\mathsf{P}\mu)^v_{\hat{z}}\right\|, \quad \text{with } z^{V\setminus v'} = \hat{z}^{V\setminus v'}$$

and then sum over all v' to find a bound for $\text{Corr}(\mathsf{P}\mu,\beta)$.

Define the probability distributions:

$$\rho(A) := \frac{\int \mathbb{I}_A(x,z^v)p(x,z)\mu(dx)\psi^v(dz^v)}{\int p(x,z)\mu(dx)\psi^v(dz^v)},$$

$$\tilde{\rho}(A) := \frac{\int \mathbb{I}_A(x,\hat{z}^v)p(x,\hat{z})\mu(dx)\psi^v(d\hat{z}^v)}{\int p(x,\hat{z})\mu(dx)\psi^v(d\hat{z}^v)},$$

where $z, \hat{z} \in \mathbb{X}^V$ such that $z^{V\setminus v'} = \hat{z}^{V\setminus v'}$. It can be observed that by definition:

$$\|\rho - \tilde{\rho}\|_{(1,v)} = \left\|(\mathsf{P}\mu)^v_z - (\mathsf{P}\mu)^v_{\hat{z}}\right\|,$$

meaning that the Dobrushin theorem can be applied to $\rho, \tilde{\rho}$ where the set of index is $I = (0, V) \cup (1, v)$.

The first step is to bound $C_{i,j}$ for all the possible combinations of $i, j \in I$, as in (A.2) of Theorem 9.

- Let consider $i = (0, b)$ and $b \in V$ then:

$$\rho_{x,z}^{(0,b)}(A) = \frac{\int \mathbb{I}_A(\tilde{x}^b)\mathbb{I}_{x^{V\setminus b},\tilde{z}^v}(\tilde{x}^{V\setminus b}, z^v)p(\tilde{x},\tilde{z})\mu(d\tilde{x})\psi^v(d\tilde{z}^v)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}^b)\mathbb{I}_{x^{V\setminus b},\tilde{z}^v}(\tilde{x}^{V\setminus b}, z^v)p(\tilde{x},\tilde{z})\mu(d\tilde{x})\psi^v(d\tilde{z}^v)}$$
$$= \frac{\int \mathbb{I}_A(\tilde{x}^b)p^b(\tilde{x}^b, z^b)\mu_x^b(d\tilde{x}^b)}{\int p^b(\tilde{x}^b, z^b)\mu_x^b(d\tilde{x}^b)}.$$

  - If $j = (0, b')$ and $b' \in V$ then:
$$C_{i,j} \leqslant \tilde{C}^{\mu}_{b,b'}.$$
  - If $j = (1, v)$ then by Lemma 7:
$$C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b = v \\ 0 & \text{otherwise} \end{cases},$$
  where we maximize the kernel part.

- Let consider $i = (1, v)$ then:

$$\rho_{x,z}^{(1,v)}(A) = \frac{\int \mathbb{I}_A(\tilde{z}^v)\mathbb{I}_x(\tilde{x})p(\tilde{x},\tilde{z})\mu(d\tilde{x})\psi^v(d\tilde{z}^v)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{z}^v)\mathbb{I}_x(\tilde{x})p(\tilde{x},\tilde{z})\mu(d\tilde{x})\psi^v(d\tilde{z}^v)}$$
$$= \frac{\int \mathbb{I}_A(\tilde{z}^v)p^v(x^v, \tilde{z}^v)\psi^v(d\tilde{z}^v)}{\int p^v(x^v, \tilde{z}^v)\psi^v(d\tilde{z}^v)}.$$

  - If $j = (0, b')$ and $b' \in V$ then by Lemma 7:
$$C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b' = v \\ 0 & \text{otherwise} \end{cases},$$
  where we maximize the kernel part.
  - If $j = (1, v)$ then:
$$C_{i,j} = 0.$$
  because the component v is integrated out.

But then:
$$\max_{i\in I}\sum_{j\in I} e^{m(i,j)}C_{i,j} \leqslant \widetilde{\mathrm{Corr}}(\mu,\beta) + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \leqslant \frac{1}{2},$$
where $m(i,j) = \beta|k - k'| - \beta d(v, v')$ for $i = (k, v)$ and $j = (k', v')$ with $k, k' \in \{0, 1\}$ and $v, v' \in V$ is the pseudometric of interest. Hence the Dobrushin theorem applies:
$$\|(\mathsf{P}\mu)_x^v - (\mathsf{P}\mu)_z^v\| = \|\rho - \tilde{\rho}\|_{(0,v)} \leqslant \sum_{i\in J}\sum_{j\in V} D_{i,j}b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9. Remark that the conditional distributions of $\tilde{\rho}$ are the same of ρ with $\hat{z}$ instead of z.

- Let consider $j = (0, b')$ and $b' \in V$ then:

$$\tilde{\rho}^{(0,b')}_{x,\hat{z}}(A) = \frac{\int \mathbb{I}_A(\tilde{x}^{b'}) p^{b'}(\tilde{x}^{b'}, \hat{z}^{b'}) \mu_x^{b'}(d\tilde{x}^{b'})}{\int p^{b'}(\tilde{x}^{b'}, \hat{z}^{b'}) \mu_x^{b'}(d\tilde{x}^{b'})}.$$

  Then by Lemma 7:

$$b_j \leqslant \begin{cases} 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b' = v' \\ 0 & \text{otherwise} \end{cases},$$

  because $\hat{z}^{V \backslash v'} = z^{V \backslash v'}$.

- Let consider $j = (1, v)$ then:

$$\tilde{\rho}^{(1,v)}_{x,\hat{z}}(A) = \frac{\int \mathbb{I}_A(\tilde{z}^{v}) p^{v}(x^{v}, \tilde{z}^{v}) \psi^{v}(d\tilde{z}^{v})}{\int p^{v}(x^{v}, \tilde{z}^{v}) \psi^{v}(d\tilde{z}^{v})}.$$

  Then:

$$b_j = 0,$$

  because the variable z is integrated out.

Given all the bounds on b_j's we have:

$$\|(\mathsf{P}\mu)^v_z - (\mathsf{P}\mu)^v_{\hat{z}}\| \leqslant \sum_{j \in V} D_{(1,v),j} b_j \leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) D_{(1,v),(0,v')}.$$

Hence:

$$\begin{aligned} \mathrm{Corr}(\mathsf{P}\mu, \beta) &= \max_{v \in V} \sum_{v' \in V} e^{d(v,v')} C^{\mathsf{P}\mu}_{v,v'} \leqslant \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) e^{-\beta} \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v') + \beta} D_{(1,v),(0,v')} \\ &\leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) e^{-\beta}. \end{aligned}$$

■

Proposition 19 *Fix any collection of observations $\{y_1, \ldots, y_T\}$ and any partition $\mathcal{K}$ on the set V. Suppose that there exists $(\epsilon_-, \epsilon_+) \in (0,1)$ and $\kappa \in (0,1)$ such that:*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+ \quad \textit{and} \quad \kappa \leqslant g^f(x^{N(f)}, y) \leqslant \frac{1}{\kappa},$$

for all $x, z \in \mathbb{X}^V, v \in V, f \in F, t \in \{1, \ldots, T\}$. Let μ be a probability distribution on $\mathbb{X}^V$ and assume that there exists $\beta > 0$ such that:

$$\widetilde{Corr}(\mu, \beta) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Then for any $t \in \{1, \ldots, T\}$ and $m \in \{0, \ldots, n\}$:

$$\widetilde{Corr}(\tilde{\mathsf{F}}^m_t \mu, \beta) \leqslant 2e^{-\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right).$$

Proof Recall that for $x, z \in \mathbb{X}^V$:

$$(\tilde{\mathsf{F}}_t^m \mu)_{x,z}^v(A) = \frac{\int \mathbb{I}_A(\tilde{x}^v) \int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, x_K^w) \mu(dx_0) \psi^{V \setminus K}(\tilde{x}) p^v(\tilde{x}^v, z^v) \psi^v(d\tilde{x}^v)}{\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, x_K^w) \mu(dx_0) \psi^{V \setminus K}(\tilde{x}) p^v(\tilde{x}^v, z^v) \psi^v(d\tilde{x}^v)},$$

and:

$$\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}_t^m \mu, \beta) = \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}_{v,v'}^{\tilde{\mathsf{F}}_t^m \mu},$$

where $\tilde{C}_{v,v'}^{\tilde{\mathsf{F}}_t^m \mu} = 1/2 \sup_{z \in \mathbb{X}^V} \sup_{x, \hat{x} \in \mathbb{X}^V : x^{V \setminus v'} = \hat{x}^{V \setminus v'}} \left\| (\tilde{\mathsf{F}}_t^m \mu)_{x,z}^v - (\tilde{\mathsf{F}}_t^m \mu)_{\hat{x},z}^v \right\|$. The idea is again to firstly control each term of the $\widetilde{\mathrm{Corr}}$ and then sum on them.

Consider the probability distributions ρ and $\tilde{\rho}$:

$$\rho(A) := \frac{\int \mathbb{I}_A(x_0, x^{V \setminus K \cup v}) \prod_{f \in N_f^m(K)} g^f(x^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, x^w) p^v(x^v, z^v) \mu(dx_0) \psi^{V \setminus K \cup v}(x^{V \setminus K \cup v})}{\int \prod_{f \in N_f^m(K)} g^f(x^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, x^w) p^v(x^v, z^v) \mu(dx_0) \psi^{V \setminus K \cup v}(x^{V \setminus K \cup v})},$$

$$\tilde{\rho}(A) := \frac{\int \mathbb{I}_A(x_0, \hat{x}^{V \setminus K \cup v}) \prod_{f \in N_f^m(K)} g^f(\hat{x}^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, \hat{x}^w) p^v(\hat{x}^v, z^v) \mu(dx_0) \psi^{V \setminus K \cup v}(\hat{x}^{V \setminus K \cup v})}{\int \prod_{f \in N_f^m(K)} g^f(\hat{x}^{N(f)}, y_t) \prod_{w \in N_v^m(K)} p^w(x_0^w, \hat{x}^w) p^v(\hat{x}^v, z^v) \mu(dx_0) \psi^{V \setminus K \cup v}(\hat{x}^{V \setminus K \cup v})},$$

where $x, \hat{x} \in \mathbb{X}^V$, such that $\hat{x}^{V \setminus v'} = x^{V \setminus v'}$. It can be observed that by definition:

$$\|\rho - \tilde{\rho}\|_{(1,v)} = \left\| (\tilde{\mathsf{F}}_t^m \mu)_{x,z}^v - (\tilde{\mathsf{F}}_t^m \mu)_{\hat{x},z}^v \right\|,$$

meaning that the Dobrushin theorem can be applied to $\rho, \tilde{\rho}$ where the complete set of index $I = (0, V) \cup (1, V \setminus K \cup v)$.

The first step is to bound $C_{i,j}$ for all the possible combinations of $i, j \in I$, as in (A.2) of Theorem 9.

- Consider $i = (0, b)$ and $b \in V$ then:

$$\begin{aligned} \rho_{x_0,x}^{(0,b)}(A) &= \frac{\prod_{f \in N_f^m(K)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v)}{\prod_{f \in N_f^m(K)} g^f(x^{N(f)}, y_t) p^v(x^v, z^v)} \frac{\int \mathbb{I}_A(x_0^b) \prod_{w \in N_v^m(K)} p^w(x_0^w, x^w) \mu(dx_0)}{\int \prod_{w \in N_v^m(K)} p^w(x_0^w, x^w) \mu(dx_0)} \\ &= \frac{\int \mathbb{I}_A(x_0^b) p^b(x_0^b, x^b) \mu_{x_0}^b(dx_0^b)}{\int p^b(x_0^b, x^b) \mu_{x_0}^b(dx_0^b)} = \mu_{x_0,x}^b(A). \end{aligned}$$

- If $j = (0, b')$ and $b' \in V$ then: $\quad C_{i,j} \leqslant \tilde{C}^{\mu}_{b,b'}$.
- If $j = (1, b')$ and $b' \in V \backslash K \cup v$ then by Lemma 7: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b' = b \\ 0 & \text{otherwise} \end{cases}$,

  where we majorize the kernel part.

- Consider $i = (1, b)$ and $b \in V \backslash K \cup v$ then:

$$
\begin{aligned}
&\rho^{(1,b)}_{x_0,x}(A) \\
&= \frac{\prod_{f \in N_f^m(K) \backslash N(b)} g^f(x^{N(f)}, y_t)}{\prod_{f \in N_f^m(K) \backslash N(b)} g^f(x^{N(f)}, y_t)} \\
&\quad \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N_f^m(K) \cap N(b)} g^f(x^{N(f)}, y_t) p^b(x_0^b, x^b) [p^v(x^v, z^v)]^{\mathbb{I}_b(v)} \psi^b(dx^b)}{\int \prod_{f \in N_f^m(K) \cap N(b)} g^f(x^{N(f)}, y_t) p^b(x_0^b, x^b) [p^v(x^v, z^v)]^{\mathbb{I}_b(v)} \psi^b(dx^b)} \\
&= \frac{\int \mathbb{I}_A(x^b) \prod_{f \in N_f^m(K) \cap N(b)} g^f(x^{N(f)}, y_t) p^b(x_0^b, x^b) [p^v(x^v, z^v)]^{\mathbb{I}_b(v)} \psi^b(dx^b)}{\int \prod_{f \in N_f^m(K) \cap N(b)} g^f(x^{N(f)}, y_t) p^b(x_0^b, x^b) [p^v(x^v, z^v)]^{\mathbb{I}_b(v)} \psi^b(dx^b)}
\end{aligned}
$$

- If $j = (0, b')$ and $b' \in V$ then by Lemma 7: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b' = b \\ 0 & \text{otherwise} \end{cases}$,

  where we majorize the kernel part.
- If $j = (1, b')$ and $b' \in V \backslash K \cup v$ then by Lemma 7:

$$
C_{i,j} \leqslant \begin{cases} (1 - \kappa^{2\mathbf{card}(N(b) \cap N(b'))}) & b' \in N_v^m(K) \cap N^2(b) \\ 0 & \text{otherwise} \end{cases},
$$

where we majorize the observation density part.

But then:

$$
\begin{aligned}
\max_{i \in I} \sum_{j \in I} e^{m(i,j)} C_{i,j} &\leqslant \widetilde{\mathrm{Corr}}(\mu, \beta) + e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \max_{b \in V} \sum_{b' \in N^2(b)} e^{\beta d(b,b')} \left(1 - \kappa^{2\mathbf{card}(N(b) \cap N(b'))}\right) \\
&\leqslant \widetilde{\mathrm{Corr}}(\mu, \beta) + 2e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \Upsilon^{(2)} e^{2\beta} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2},
\end{aligned}
$$

where $m(i,j) = \beta|k - k'| + \beta d(v, v')$ with $i = (k, v)$ and $j = (k', v')$ for $k, k' \in \{0, 1\}$ and $v, v' \in V$ is the pseudometric of interest. Remark that we also used $2e^{\beta} > e^{\beta}$. Then the Dobrushin theorem can be applied:

$$
\left\| (\tilde{\mathsf{F}}^m_t \mu)^v_{x,z} - (\tilde{\mathsf{F}}^m_t \mu)^v_{\hat{x},z} \right\| = \|\rho - \tilde{\rho}\|_{(1,v)} \leqslant \sum_{j \in I} D_{(1,v),j} b_j.
$$

The second step is to control b_j, as in (A.2) of Theorem 9.

- If $j = (0, b')$ and $b' \in V$ then:

$$\tilde{\rho}^{(0,b')}_{x_0,\hat{x}}(A) = \frac{\int \mathbb{I}_A(x_0^{b'}) p^{b'}(x_0^{b'}, \hat{x}^{b'}) \mu^{b'}_{x_0}(dx_0^{b'})}{\int p^{b'}(x_0^{b'}, \hat{x}^{b'}) \mu^{b'}_{x_0}(dx_0^{b'})} = \mu^{b'}_{x_0,\hat{x}}(A),$$

where the computations are the same as in $\rho^{(0,b)}_{x_0,x}$. Hence by Lemma 7:

$$b_j \leqslant \begin{cases} 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b' = v' \\ 0 & \text{otherwise} \end{cases},$$

because x_1 and $\tilde{x}_1$ differ only on v'.

- If $j = (1, b')$ and $b' \in V \backslash K \cup v$ then:

$$\tilde{\rho}^{(1,b')}_{x_0,\hat{x}}(A) = \frac{\int \mathbb{I}_A(\hat{x}^{b'}) \prod_{f \in N_f^m(K) \cap N(b')} g^f(\hat{x}^{N(f)}, y_t) p^{b'}(x_0^{b'}, \hat{x}^{b'}) [p^v(\hat{x}^v, z^v)]^{\mathbb{I}_{b'}(v)} \psi^{b'}(d\hat{x}^{b'})}{\int \prod_{f \in N_f^m(K) \cap N(b')} g^f(\hat{x}^{N(f)}, y_t) p^{b'}(x_0^{b'}, \hat{x}^{b'}) [p^v(\hat{x}^v, z^v)]^{\mathbb{I}_{b'}(v)} \psi^{b'}(d\hat{x}^{b'})}.$$

Hence by Lemma 7:

$$b_j \leqslant \begin{cases} 2(1 - \kappa^{2\mathbf{card}(N(b') \cap N(v'))}) & b' \in N^2(v') \cap N^m_v(K) \backslash v \\ 0 & \text{otherwise} \end{cases},$$

where the case $b' = v$ is still zero because the only difference is on v'.

By joining step one and step two it follows:

$$\begin{aligned} \|\rho - \tilde{\rho}\|_{(1,v)} &\leqslant D_{(1,v),(0,v')} b_{(0,v')} + \sum_{b' \in N^2(v')} D_{(1,v),(1,b')} b_{(1,b')} \\ &\leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) D_{(1,v),(0,v')} + 2\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \sum_{b' \in N^2(v')} D_{(1,v),(1,b')}. \end{aligned}$$

Remark that if $v' \in V$ and $b' \in N^2(v')$ then obviously $b' \in V$ and $v' \in N^2(b')$. Moreover, by the triangular inequality $d(v, v') \leqslant d(v, b') + d(v', b')$. Then by summing over V and by applying Lemma 10:

$$\begin{aligned} \sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}^{\tilde{\mathsf{F}}^m_t \mu}_{v,v'} &\leqslant \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \sum_{v' \in V} e^{\beta d(v,v')} D_{(1,v),(0,v')} + \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \sum_{v' \in V} e^{\beta d(v,v')} \sum_{b' \in N^2(v')} D_{(1,v),(1,b')} \\ &\leqslant \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \sum_{v' \in V} e^{\beta d(v,v')} D_{(1,v),(0,v')} \\ &\quad + \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \sum_{v' \in V} \sum_{b' \in N^2(v')} e^{\beta d(v,b') + d(b',v')} D_{(1,v),(1,b')} \\ &\leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) e^{-\beta} + \left(1 - \kappa^{2\tilde{\Upsilon}}\right) e^{2\beta} \sum_{b' \in V} \sum_{v' \in N^2(b')} e^{\beta d(v,b')} D_{(1,v),(1,b')} \\ &\leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) e^{-\beta} + 2\Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) e^{2\beta}. \end{aligned}$$

Given that the bound does not depend on v the thesis follows from the definition of $\widetilde{\text{Corr}}(\tilde{\mathsf{F}}_t^m \mu, \beta)$.

■

Corollary 20 *Fix any collection of observations $\{y_1, \ldots, y_T\}$ and any partition $\mathcal{K}$ on the set V. There exists a region $\mathcal{R}_0 \subseteq (0,1)^3$ depending only on $\tilde{\Upsilon}$, Υ and $\Upsilon^{(2)}$, such that if, for given $(\epsilon_-, \epsilon_+, \kappa) \in \mathcal{R}_0$,*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+ \quad \textit{and} \quad \kappa \leqslant g^f(x^{N(f)}, y_t) \leqslant \frac{1}{\kappa},$$

for all $x, z \in \mathbb{X}^V, f \in F, v \in V, t \in \{1, \ldots, T\}$, then for $\beta > 0$ small enough depending only on $\tilde{\Upsilon}, \Upsilon$, $\Upsilon^{(2)}$, ϵ_-, ϵ_+ and κ, we have that for any μ_0 satisfying the decay of correlation property:

$$\widetilde{Corr}(\mu_0, \beta) \leqslant 2e^{-\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \tag{A.4}$$

and for any $t \in \{1, \ldots, T\}$ and $m \in \{0, \ldots, n\}$:

$$\widetilde{Corr}(\tilde{\mathsf{F}}_t^m \tilde{\pi}_{t-1}, \beta) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2} \tag{A.5}$$

and

$$2\kappa^{-2\Upsilon} Corr(\mathsf{P}\tilde{\pi}_{t-1}, \beta) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}, \tag{A.6}$$

where $\tilde{\pi}_{t-1}$ is the approximated filtering distribution obtained through (15).

Proof The proof is inductive in t. To initialize the induction, let $t = 1$. We want to identify ranges of values for $\beta, \epsilon_-, \epsilon_+$ and κ such that:

$$\widetilde{\text{Corr}}(\tilde{\mathsf{F}}_1^m \mu_0, \beta) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2} \tag{A.7}$$

and

$$2\kappa^{-2\Upsilon}\text{Corr}(\mathsf{P}\mu_0, \beta) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}. \tag{A.8}$$

Let us start by analysing (A.7). Note that we have a bound for $\widetilde{\text{Corr}}(\mu_0, \beta)$ given by (A.4). So for any $\beta > 0$, $(\epsilon_-, \epsilon_+) \in (0,1)^2$ and $\kappa \in (0,1)$ we have the upper bound:

$$\begin{aligned}
&\widetilde{\text{Corr}}(\mu_0, \beta) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \\
&\leqslant 2e^{-\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \\
&\leqslant 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \left[3\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) + 2\left(1 - \frac{\epsilon_-}{\epsilon_+}\right)\right]e^{2\beta},
\end{aligned}$$

In order to apply Proposition 19 and obtain (A.7), our next step is to derive constraints on β, ϵ_-, ϵ_+ and κ such that:

$$2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+\left[3\Upsilon^{(2)}(1-\kappa^{2\tilde{\Upsilon}})+2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\right]e^{2\beta}\leqslant\frac{1}{2}. \tag{A.9}$$

This holds for β such that:

$$\beta\leqslant\frac{1}{2}\log\left\{\frac{1-4\left(1-\frac{\epsilon_-}{\epsilon_+}\right)}{6\Upsilon^{(2)}(1-\kappa^{2\tilde{\Upsilon}})+4\left(1-\frac{\epsilon_-}{\epsilon_+}\right)}\right\}=\beta_0^1.$$

and to guarantee $\beta>0$ when $(\epsilon_-,\epsilon_+,\kappa)\in(0,1)^3$, i.e. the logarithm being positive, we further impose:

$$1-4\left(1-\frac{\epsilon_-}{\epsilon_+}\right)>6\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+4\left(1-\frac{\epsilon_-}{\epsilon_+}\right).$$

Informed by these considerations we define the region:

$$\mathcal{R}_0^1:=\{(\epsilon_-,\epsilon_+,\kappa)\in(0,1)^3:8\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+6\Upsilon^{(2)}(1-\kappa^{2\tilde{\Upsilon}})<1\}.$$

Hence by choosing $(\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0^1$ and $\beta<\beta_0^1$, the inequality (A.9) holds as required and so, noting $\tilde{\pi}_0=\mu_0$, Proposition 19 can be applied to give:

$$\begin{aligned}
&\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}_1^m\tilde{\pi}_0,\beta)+2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\\
&\leqslant 2e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\\
&\leqslant 2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+\left[3\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\right]e^{2\beta},
\end{aligned}$$

which we have already proved to be less than $1/2$ for $\beta\leqslant\beta_0^1$ and $(\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0^1$.

Turning to (A.8), first note the following upper bound:

$$\begin{aligned}
&\widetilde{\mathrm{Corr}}(\mu_0,\beta)+e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\\
&\leqslant 2e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right),
\end{aligned}$$

and with the previous choice of $\beta,\epsilon_-,\epsilon_+$ and κ we have also that:

$$\begin{aligned}
\widetilde{\mathrm{Corr}}(\delta_x,\beta)+e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)&\leqslant 2e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\\
&\leqslant 2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+\left[3\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\right]e^{2\beta}\leqslant\frac{1}{2}.
\end{aligned}$$

Hence the assumption of Proposition 18 holds without any additional restrictions, meaning that:

$$2\kappa^{-2\Upsilon}\mathrm{Corr}(\mathsf{P}\tilde{\pi}_0,\beta) \leqslant 4\kappa^{-2\Upsilon}e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right).$$

We now need to identify constraints on β, ϵ_-,ϵ_+ and κ in order to guarantee the second of the following two inequalities:

$$2\kappa^{-2\Upsilon}\mathrm{Corr}(\mathsf{P}\tilde{\pi}_0,\beta)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right) \leqslant 4\kappa^{-2\Upsilon}e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

To do so, we impose:

$$\beta \leqslant \frac{1}{2}\log\left(\frac{\kappa^{2\Upsilon}-8\left(1-\frac{\epsilon_-}{\epsilon_+}\right)}{4\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\kappa^{2\Upsilon}}\right)=\beta_0^2,$$

and again for positivity of the logarithm:

$$\kappa^{2\Upsilon}-8\left(1-\frac{\epsilon_-}{\epsilon_+}\right) > 4\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\kappa^{2\Upsilon}.$$

This leads us to define:

$$\mathcal{R}_0^2 \coloneqq \left\{(\epsilon_-,\epsilon_+,\kappa)\in(0,1)^3 : \kappa^{2\Upsilon}-4\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\kappa^{2\Upsilon} > 8\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\right\},$$

and hence by choosing $\beta\leqslant\beta_0^2$ and $(\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0^2$ we have:

$$2\kappa^{-2\Upsilon}\mathrm{Corr}(\mathsf{P}\tilde{\pi}_0,\beta)+2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\leqslant\frac{1}{2}.$$

We have thus proved that with:

$$\beta_0=\min\{\beta_0^1,\beta_0^2\} \quad \text{and} \quad (\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0\coloneqq\mathcal{R}_0^1\cap\mathcal{R}_0^2,$$

both (A.5) and (A.6) hold for $t=1$.

Suppose now that (A.5) holds for t. Then since $\tilde{\mathsf{F}}^m_{t+1}\tilde{\pi}_t=\tilde{\pi}_{t+1}$, Proposition 19 can be applied for $t+1$ and so:

$$\begin{aligned}\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}^m_{t+1}\tilde{\pi}_t,\beta)+2e^{\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)\leqslant&\\ 2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)+\left[3\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)+2\left(1-\frac{\epsilon_-}{\epsilon_+}\right)\right]e^{2\beta},&\end{aligned}$$

which we have already proved to be less than $1/2$ for $\beta\leqslant\beta_0$ and $(\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0$ – see (A.9).

Given that (A.5) holds for t, we also have:

$$\widetilde{\mathrm{Corr}}(\tilde{\pi}_t, \beta) + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \leqslant \frac{1}{2},$$

Proposition 18 applies and so for the previous choices of $\beta, \epsilon_-, \epsilon_+, \kappa$:

$$2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) + 2\kappa^{-2\Upsilon}\mathrm{Corr}(\mathsf{P}\tilde{\pi}_t, \beta) \leqslant 2e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) + 4\kappa^{-2\Upsilon}e^{-\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \leqslant \frac{1}{2},$$

which completes the treatment of (A.6).

Hence the induction is complete and for $\beta \leqslant \beta_0$, $(\epsilon_-, \epsilon_+, \kappa) \in \mathcal{R}_0$ both (A.5) and (A.6) hold for all t. ■

A.2.4 Proof of Theorem 2

Proof For notational convenience we state the proof for $\pi_T - \tilde{\pi}_T$ but since T is arbitrary this is also not a restriction.

The quantity $\pi_T - \tilde{\pi}_T$ can be expressed as a telescopic sum, indeed:

$$\pi_T - \tilde{\pi}_T = \mathsf{F}_T \dots \mathsf{F}_1 \delta_x - \tilde{\mathsf{F}}_T^m \dots \tilde{\mathsf{F}}_1^m \delta_x = \sum_{t=1}^{T} (\mathsf{F}_T \dots \mathsf{F}_{t+1}\mathsf{F}_t \tilde{\pi}_{t-1} - \mathsf{F}_T \dots \mathsf{F}_{t+1}\tilde{\mathsf{F}}_t^m \tilde{\pi}_{t-1}),$$

hence given $J \subseteq K \in \mathcal{K}$, by the triangular inequality:

$$\|\pi_T - \tilde{\pi}_T\|_J \leqslant \sum_{t=1}^{T-1} \left\|\mathsf{F}_T \dots \mathsf{F}_{t+1}\mathsf{F}_t \tilde{\pi}_{t-1} - \mathsf{F}_T \dots \mathsf{F}_{t+1}\tilde{\mathsf{F}}_t^m \tilde{\pi}_{t-1}\right\|_J + \left\|\mathsf{F}_T \tilde{\pi}_{T-1} - \tilde{\mathsf{F}}_T^m \tilde{\pi}_{T-1}\right\|_J.$$

If $\epsilon_-, \epsilon_+, \kappa$ and β are chosen according to Corollary 20 then Proposition 14 can be applied for $t \in \{1, \dots T-1\}$:

$$\begin{aligned}\|\pi_T - \tilde{\pi}_T\|_J \leqslant &\sum_{t=1}^{T-1} 2e^{-\beta(T-t)} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \sup_{x \in \mathbb{X}^V, z^v \in \mathbb{X}} \left\| (\mathsf{F}_t \tilde{\pi}_{t-1})^{v'}_{x_0, z^v} - (\tilde{\mathsf{F}}_t^m \tilde{\pi}_{t-1})^{v'}_{x_0, z^v} \right\| \right\} \\ &+ \left\|\mathsf{F}_T \tilde{\pi}_{T-1} - \tilde{\mathsf{F}}_T^m \tilde{\pi}_{T-1}\right\|_J.\end{aligned}$$

But given that also Proposition 15 and Proposition 17 can be applied then by considering $v' \in K'$:

$$\sup_{x_1 \in \mathbb{X}^V, z^{v'} \in \mathbb{X}} \left\| (\mathsf{F}_t \tilde{\pi}_{t-1})^{v'}_{x_1, z} - (\tilde{\mathsf{F}}_t^m \tilde{\pi}_{t-1})^{v'}_{x_1, z} \right\| \leqslant 2\left(1 - \kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1 - \kappa^{b(\mathcal{K}, m)}\right), \quad \text{(A.10)}$$

and

$$\left\|\mathsf{F}_T \tilde{\pi}_{T-1} - \tilde{\mathsf{F}}_T^m \tilde{\pi}_{T-1}\right\|_J \leqslant 4e^{-\beta m}\left(1 - \kappa^{b(m, \mathcal{K})}\right) \mathbf{card}(J). \quad \text{(A.11)}$$

Hence by using (A.10) and (A.11) :

$$\begin{aligned}
&\|\pi_T - \tilde{\pi}_T\|_J \\
&\leqslant \left[2\left(1-\kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right)\right] \\
&\cdot \sum_{t=1}^{T-1} 2e^{-\beta(T-t)} \sum_{v\in J} 1 + 4e^{-\beta m}\left(1-\kappa^{b(m,\mathcal{K})}\right)\mathbf{card}(J) \\
&\leqslant \left[2\left(1-\kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right)\right]\frac{2}{(e^{\beta}-1)}\mathbf{card}(J) \\
&\quad + 4e^{-\beta m}\left(1-\kappa^{b(m,\mathcal{K})}\right)\mathbf{card}(J) \\
&= \alpha_1(\beta)\left(1-\kappa^{a(\mathcal{K})}\right)\mathbf{card}(J) + \gamma_1(\beta)\left(1-\kappa^{b(\mathcal{K},m)}\right)\mathbf{card}(J)e^{-\beta m}.
\end{aligned}$$

■

A.3 Smoothing

Theorem 3 *Fix any collection of observations $\{y_1,\ldots,y_T\}$ and any partition $\mathcal{K}$ of V. There exists a region $\tilde{\mathcal{R}}_0 \subseteq (0,1)^3$ depending only on $\tilde{\Upsilon},\Upsilon$ and $\Upsilon^{(2)}$, such that if, for given $(\epsilon_-,\epsilon_+,\kappa)\in\tilde{\mathcal{R}}_0$,*

$$\epsilon_- \leqslant p^v(x^v,z^v) \leqslant \epsilon_+, \quad \textit{and} \quad \kappa \leqslant g^f\left(x^{N(f)},y_t\right) \leqslant \frac{1}{\kappa},$$

for all $x,z\in\mathbb{X}^V, v\in V, f\in F, t\in\{1,\ldots,T\}$, then for $\beta>\log(2)$ small enough depending only on $\tilde{\Upsilon},\Upsilon,\Upsilon^{(2)},\epsilon_-,\epsilon_+$ and κ, we have that for any μ_0 satisfying:

$$\widetilde{Corr}(\mu_0,\beta) \leqslant 2e^{-\beta}\left(1-\frac{\epsilon_-}{\epsilon_+}\right) + 2e^{2\beta}\Upsilon^{(2)}\left(1-\kappa^{2\tilde{\Upsilon}}\right)$$

and for any $K\in\mathcal{K}$, $J\subseteq K$ and $m\in\{0,\ldots,n\}$:

$$\left\|\tilde{\pi}_{t|T}-\pi_{t|T}\right\|_J \leqslant \alpha_2(\beta,\epsilon_-,\epsilon_+)\left(1-\kappa^{a(\mathcal{K})}\right)\mathbf{card}(J) + \gamma_2(\beta,\epsilon_-,\epsilon_+)\left(1-\kappa^{b(\mathcal{K},m)}\right)\mathbf{card}(J)e^{-\beta m},$$

where $\pi_{t|T},\tilde{\pi}_{t|T}$ are given by (6) and (17) with initial condition μ_0, $\alpha_2(\beta,\epsilon_-,\epsilon_+)$ and $\gamma_2(\beta,\epsilon_-,\epsilon_+)$ are constants depending on ϵ_-,ϵ_+, β and

$$\begin{aligned}
a(\mathcal{K}) &\coloneqq 2\max_{K\in\mathcal{K}}\max_{v\in\partial K}\mathbf{card}(N(v)\cap\partial N(K)),\\
b(m,\mathcal{K}) &\coloneqq 2\max_{K\in\mathcal{K}}\max_{v\notin N_v^{m-1}(K)}\mathbf{card}(N(v)),
\end{aligned}$$

with the convention that the maximum over an empty set is zero.

As for the filtering distribution, the proof of Theorem 3 follows by breaking down the problem. Consider the difference between the optimal smoothing and the approximate one:

$$\begin{aligned}\mathsf{R}_{\tilde{\pi}_t}\tilde{\pi}_{t+1|T} - \mathsf{R}_{\pi_t}\pi_{t+1|T} =& \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\pi_T \\ &+ \sum_{s=0}^{T-t-1} \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\tilde{\pi}_{t+s}}\pi_{t+s+1|T} - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\pi_{t+s}}\pi_{t+s+1|T},\end{aligned}$$

hence the overall proof can be split in three steps:

1. control the part outside the sum: approximate smoothing stability (Subsection A.3.1);
2. control the part inside the sum: approximate smoothing stability and smoothing error control (Subsection A.3.2).

Remark that for the proof of theorem 3 corollary 20 is needed, so condition A.4 must hold.

We shall need the following definition.

Definition 21 *Let μ, ν be probability distributions on $\mathbb{X}^V$, let $Z \sim \mu$ and $X|Z \sim \overleftarrow{P}_\nu(Z, \cdot)$, where $\overleftarrow{P}_\nu(z, \cdot) \coloneqq p(x, z)\nu(dx)/\int p(x, z)\nu(dx)$. Then define:*

$$(\overset{\nu}{\overleftarrow{\mu}})^v_{z,x}(A) \coloneqq \mathbb{P}(Z^v \in A | Z^{V\setminus v} = z, X = x).$$

A.3.1 Approximate smoothing stability

We want to prove that an application of the approximate smoothing operator to $\tilde{\pi}_T$ is not too different from the same applied on π_T.

Proposition 22 *Fix any collection of observations $\{y_1, \dots, y_T\}$ and any partition $\mathcal{K}$ on the set V. There exists a region $\mathcal{R}_0 \subseteq (0,1)^3$, as in Corollary 20, depending only on $\tilde{\Upsilon}, \Upsilon$ and $\Upsilon^{(2)}$, such that if, for given $(\epsilon_-, \epsilon_+, \kappa) \in \mathcal{R}_0$,*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+ \quad \text{and} \quad \kappa \leqslant g^f(x^{N(f)}, y_t) \leqslant \frac{1}{\kappa},$$

for all $x, z \in \mathbb{X}^V, f \in F, v \in V, t \in \{1, \dots, T\}$, then for $\beta > 0$ small enough depending only on $\tilde{\Upsilon}, \Upsilon$, $\Upsilon^{(2)}$, ϵ_-, ϵ_+ and κ, we have that for any $t \in \{0, \dots, T-1\}, K \in \mathcal{K}$ and $J \subseteq K$ and $m \in \{0, \dots, n\}$:

$$\begin{aligned}&\left\|\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\pi_T\right\|_J \\ &\qquad \leqslant 2e^{-\beta(T-t)} \sum_{v\in J} \max_{v'\in V} \left\{ e^{-\beta d(v,v')} \sup_{x_{T-1}, x_T \in \mathbb{X}^V} \left\| (\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\tilde{\pi}_T}})^{v'}_{x_T, x_{T-1}} - (\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi_T}})^{v'}_{x_T, x_{T-1}} \right\| \right\}.\end{aligned}$$

Proof Denote with $\overleftarrow{\tilde{p}}_t(\cdot, \cdot)$ the reverse kernel density with reference distribution the approximated filtering distribution, i.e.:

$$\overleftarrow{\tilde{p}}_t(z, x) \coloneqq \frac{p(x, z)}{\int p(\hat{x}, z)\tilde{\pi}_t(d\hat{x})}.$$

Then:

$$\mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \tilde{\pi}_T(A) = \int \mathbb{I}_A(x_t) \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) \tilde{\pi}_t(dx_t) \dots \overleftarrow{\tilde{p}}_{T-1}(x_T, x_{T-1}) \tilde{\pi}_{T-1}(dx_{T-1}) \tilde{\pi}_T(dx_T),$$

$$\mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \pi_T(A) = \int \mathbb{I}_A(x_t) \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) \tilde{\pi}_t(dx_t) \dots \overleftarrow{\tilde{p}}_{T-1}(x_T, x_{T-1}) \tilde{\pi}_{T-1}(dx_{T-1}) \pi_T(dx_T).$$

Consider the following probability distributions:

$$\rho(A) \coloneqq \int \mathbb{I}_A(x_t, \dots, x_T) \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) \tilde{\pi}_t(dx_t) \dots \overleftarrow{\tilde{p}}_{T-1}(x_T, x_{T-1}) \tilde{\pi}_{T-1}(dx_{T-1}) \pi_T(dx_T), \tag{A.12}$$

$$\tilde{\rho}(A) \coloneqq \int \mathbb{I}_A(x_t, \dots, x_T) \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) \tilde{\pi}_t(dx_t) \dots \overleftarrow{\tilde{p}}_{T-1}(x_T, x_{T-1}) \tilde{\pi}_{T-1}(dx_{T-1}) \tilde{\pi}_T(dx_T), \tag{A.13}$$

the quantity of interest can be reformulated in terms of LTV on $\rho, \tilde{\rho}$:

$$\left\| \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \pi_T \right\|_J = \| \rho - \tilde{\rho} \|_{(t,J)} .$$

So, it is enough to find a bound for $\| \rho - \tilde{\rho} \|_{(t,J)}$ to guarantee the proof of the statement. The Dobrushin theorem can be used on the distributions $\rho, \tilde{\rho}$ where the index set is $I = \bigcup_{k=t}^{T} (k, V)$ and the subset is (t, J).

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9, i.e. checking the assumptions of the Dobrushin theorem for (A.12)-(A.13). In the following passages we consider the notation $x = (x_t, \dots, x_T)$, where $x_k \in \mathbb{X}^V$ for $k = t, \dots, T$ and $x \backslash x_k^v \coloneqq (x_t, \dots, x_k^{V \backslash v}, \dots, x_T)$ (and the same with tilde).

- Consider $i = (t, v)$ and $v \in V$ then:

$$\begin{aligned} &\tilde{\rho}_x^{(t,v)}(A) \\ &= \frac{\int \mathbb{I}_A(\tilde{x}_t^v) \mathbb{I}_{x \backslash x_t^v}(\tilde{x} \backslash \tilde{x}_t^v) \overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t) \tilde{\pi}_t(d\tilde{x}_t) \dots \overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T, \tilde{x}_{T-1}) \tilde{\pi}_{T-1}(d\tilde{x}_{T-1}) \tilde{\pi}_T(d\tilde{x}_T)}{\int \mathbb{I}_{x \backslash x_t^v}(\tilde{x} \backslash \tilde{x}_t^v) \overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t) \tilde{\pi}_t(d\tilde{x}_t) \dots \overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T, \tilde{x}_{T-1}) \tilde{\pi}_{T-1}(d\tilde{x}_{T-1}) \tilde{\pi}_T(d\tilde{x}_T)} \\ &= \frac{\int \mathbb{I}_A(x_t^v) \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) (\tilde{\pi}_t)_{x_t}^v(dx_t^v)}{\int \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t) (\tilde{\pi}_t)_{x_t}^v(dx_t^v)} = \frac{\int \mathbb{I}_A(x_t^v) p^v(x_t^v, x_{t+1}^v) (\tilde{\pi}_t)_{x_t}^v(dx_t^v)}{\int p^v(x_t^v, x_{t+1}^v) (\tilde{\pi}_t)_{x_t}^v(dx_t^v)}, \end{aligned}$$

  where the last passage follow from the independence of the numerator of the reverse kernel from x_t. Now we have to distinguish the different cases in which ρ_x^i can differ from $\rho_{\tilde{x}}^i$, where $x^{I \backslash j} = \tilde{x}^{I \backslash j}$.

  - If $j = (t, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \tilde{C}_{v,v'}^{\tilde{\pi}_t}$.
  - If $j = (t+1, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,

    where the result follow from Lemma 7, obtained by a majorization of the kernel part.

– If $j = (k, v')$ with $k > t+1$ and $v' \in V$ then: $\quad C_{i,j} = 0$,
which is obvious given that in ρ_x^i there is no dependence on x_k with $k > t+1$.

- Consider $i = (k, v)$ with $t+1 < k < T$ and $v \in K$ with $K \in \mathcal{K}$ then:

$$
\begin{aligned}
&\tilde{\rho}_x^{(k,v)}(A) \\
&= \frac{\int \mathbb{I}_A(\tilde{x}_k^v)\mathbb{I}_{x\backslash x_k^v}(\tilde{x}\backslash\tilde{x}_k^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\tilde{\pi}_T(d\tilde{x}_T)}{\int \mathbb{I}_{x\backslash x_k^v}(\tilde{x}\backslash\tilde{x}_k^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\tilde{\pi}_T(d\tilde{x}_T)} \\
&= \frac{\int \mathbb{I}_A(x_k^v)\overleftarrow{\tilde{p}}_{k-1}(x_k,x_{k-1})\overleftarrow{\tilde{p}}_k(x_{k+1},x_k)(\tilde{\pi}_k)_{x_k}^v(dx_k)}{\int \overleftarrow{\tilde{p}}_{k-1}(x_k,x_{k-1})\overleftarrow{\tilde{p}}_k(x_{k+1},x_k)(\tilde{\pi}_k)_{x_k}^v(dx_k)} \\
&= \frac{\int \mathbb{I}_A(x_k^v)\overleftarrow{\tilde{p}}_{k-1}(x_k,x_{k-1})p^v(x_k^v,x_{k+1}^v)(\tilde{\pi}_k)_{x_k}^v(dx_k)}{\int \overleftarrow{\tilde{p}}_{k-1}(x_k,x_{k-1})p^v(x_k^v,x_{k+1}^v)(\tilde{\pi}_k)_{x_k}^v(dx_k)},
\end{aligned}
$$

where the last passage follows from the definition of the denominator of $\overleftarrow{\tilde{p}}_k(x_{k+1}, x_k)$ that is independent from x_k^v given that x_k is integrated out. At this point we carry on the computations on the numerator and remark that similar calculations follow on the denominator when $A = \mathbb{X}$. Firstly we can expand the definition of $(\tilde{\pi}_k)_{x_k}^v$ and obtain:

$$\begin{aligned}
&\int \mathbb{I}_A(x^v)\overleftarrow{\tilde{p}}_{k-1}(x, x_{k-1}) p^v(x^v, x_{k+1}^v) \\
&\qquad \frac{\prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})}{\int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v)}\psi^v(dx^v) \\
&= \int \mathbb{I}_A(x^v)\frac{\prod_{w\in V} p^w(x_{k-1}^w, x^w)}{\prod_{K'\in\mathcal{K}} \int \prod_{w\in K'} p^w(x_{k-1}^w, x^w)\tilde{\pi}_{k-1}^{K'}(dx_{k-1}^{K'})} p^v(x^v, x_{k+1}^v) \\
&\qquad \frac{\prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})}{\int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v)}\psi^v(dx^v) \\
&= \int \mathbb{I}_A(x^v)\frac{p^v(x_{k-1}^v, x^v)}{\int \prod_{w\in K} p^w(x_{k-1}^w, x^w)\tilde{\pi}_{k-1}^{K}(dx_{k-1}^{K})} p^v(x^v, x_{k+1}^v) \\
&\qquad \frac{\prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})}{\int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v)}\psi^v(dx^v) \\
&= \int \mathbb{I}_A(x^v) p^v(x_{k-1}^v, x^v) p^v(x^v, x_{k+1}^v) \\
&\qquad \frac{\prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)\setminus K} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}^{V\setminus K}(dx_0^{V\setminus K})\psi^{V\setminus K}(\tilde{x})}{\int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \int \prod_{w\in N_v^m(K)} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}(dx_0)\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v)}\psi^v(dx^v),
\end{aligned}$$

where we just used the factorization of $\tilde{\pi}_{k-1}$ and the factorization of the kernel. Given that the same hold for the denominator we can simplify our expression a bit more. Define:

$$\begin{aligned}
N_k &:= \int \mathbb{I}_A(x^v) p^v(x_{k-1}^v, x^v) p^v(x^v, x_{k+1}^v) \\
&\qquad \int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w\in N_v^m(K)\setminus K} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}^{V\setminus K}(dx_0^{V\setminus K})\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v),
\end{aligned}$$

$$\begin{aligned}
D_k &:= \int p^v(x_{k-1}^v, x^v) p^v(x^v, x_{k+1}^v) \\
&\qquad \int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w\in N_v^m(K)\setminus K} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}^{V\setminus K}(dx_0^{V\setminus K})\psi^{V\setminus K}(\tilde{x})\psi^v(dx^v).
\end{aligned}$$

Then:

$$\tilde{\rho}_x^{(k,v)}(A) = \frac{N_k}{D_k}.$$

- If $j = (k', v')$ with $k' \leqslant k-2$ and $v' \in V$ then: $\quad C_{i,j} = 0.$
- If $j = (k-1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases},$
  where the result follow from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (k, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}\right) & v' \in N^2(v)\backslash v \\ 0 & \text{otherwise} \end{cases},$
  where the result follow from Lemma 7, obtained by a majorization of the observation density part. Recall that the only factors that contains v are the one in $N(v)$ so the components that are connected to these factors are the one in $N^2(v)$.
- If $j = (k+1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases},$
  where the result follow from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (k', v')$ with $k' \geqslant k+2$ and $v' \in V$ then: $\quad C_{i,j} = 0.$

- Consider $i = (T, v)$ and $v \in V$ then:

$$\begin{aligned}
&\tilde{\rho}_x^{(T,v)}(A) \\
&= \frac{\int \mathbb{I}_A(\tilde{x}_T^v)\mathbb{I}_{x\backslash x_T^v}(\tilde{x}\backslash\tilde{x}_T^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\tilde{\pi}_T(d\tilde{x}_T)}{\int \mathbb{I}_{x\backslash x_T^v}(\tilde{x}\backslash\tilde{x}_T^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\tilde{\pi}_T(d\tilde{x}_T)} \\
&= \frac{\int \mathbb{I}_A(x_T^v)\overleftarrow{\tilde{p}}_{T-1}(x_T,x_{T-1})(\tilde{\pi}_T)^v_{x_T}(dx_T^v)}{\int \overleftarrow{\tilde{p}}_{T-1}(x_T,x_{T-1})(\tilde{\pi}_T)^v_{x_T}(dx_T^v)} = \frac{N_T}{D_T},
\end{aligned}$$

where:

$$\begin{aligned}
N_T &:= \int \mathbb{I}_A(x_T^v)p^v(x_{T-1}^v,x_T^v)\int \prod_{f\in N_f^m(K)} g^f(x_T^{N(f)},y_T) \\
&\qquad \prod_{w\in N_v^m(K)\backslash K} p^w(x_{T-1}^w,x_T^w)\tilde{\pi}_{T-1}^{V\backslash K}(dx_{T-1}^{V\backslash K})\psi^{V\backslash K}(dx_T^{V\backslash K})\psi^v(dx_T^v), \\
D_T &:= \int p^v(x_{T-1}^v,x_T^v)\int \prod_{f\in N_f^m(K)} g^f(x_T^{N(f)},y_T) \\
&\qquad \prod_{w\in N_v^m(K)\backslash K} p^w(x_{T-1}^w,x_T^w)\tilde{\pi}_{T-1}^{V\backslash K}(dx_{T-1}^{V\backslash K})\psi^{V\backslash K}(dx_T^{V\backslash K})\psi^v(dx_T^v),
\end{aligned}$$

which follows from similar passages as in $i = (k, v)$.

– If $j = (k', v')$ with $k' \leqslant T-2$ and $v' \in V$ then: $\quad C_{i,j} = 0.$

– If $j = (T-1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases},$

where the result follow from Lemma 7, obtained by a majorization of the kernel part.

– If $j = (T, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \kappa^{2\mathbf{card}(N(v) \cap N(v'))}\right) & v' \in N^2(v) \backslash v \\ 0 & \text{otherwise} \end{cases},$

where the result follow from Lemma 7, obtained by a majorization of the observation density part. Recall that the only factors that contains v are the one in $N(v)$ so the components that are connected to these factors are the one in $N^2(v)$.

Given the previous results, for any $v \in V$:

$$\sum_{j \in I} e^{m(i,j)} C_{i,j} \leqslant \begin{cases} \sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}^{\tilde{\pi}_t}_{v,v'} + e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & i = (0, v) \\ 2e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \sum_{v' \in N^2(v)} \left(1 - \kappa^{2\mathbf{card}(N(v) \cap N(v'))}\right) e^{\beta d(v,v')} & i = (k, v), \\ e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \sum_{v' \in N^2(v)} \left(1 - \kappa^{2\mathbf{card}(N(v) \cap N(v'))}\right) e^{\beta d(v,v')} & i = (T, v) \end{cases}$$

where $m(i,j) = \beta|k - k'| + \beta d(v, v')$ for $i = (k, v)$ and $j = (k', v')$ with $k, k' \in \{t, \ldots, T\}$ and $v, v' \in V$ is the pseudometric of interest on the index set I. But then given that we are in $\mathcal{R}_0$:

$$\max_{i \in I} \sum_{j \in I} C_{i,j} \leqslant \widetilde{\operatorname{Corr}}(\tilde{\pi}_t, \beta) + 2e^{\beta} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta} \Upsilon^{(2)} \left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Given that $\sum_{j \in I} C_{i,j} \leqslant \sum_{j \in I} e^{m(i,j)} C_{i,j}$ then the Dobrushin theorem (Theorem 9) can be applied, meaning that:

$$\left\| \mathsf{R}_{\tilde{\pi}_t} \ldots \mathsf{R}_{\tilde{\pi}_{T-1}} \tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t} \ldots \mathsf{R}_{\tilde{\pi}_{T-1}} \pi_T \right\|_J = \|\rho - \tilde{\rho}\|_{(t,J)} \leqslant \sum_{v \in J} \sum_{j \in I} D_{(t,v),j} b_j.$$

The second step is to control the quantities b_j obtained from (A.12)-(A.13), as in (A.2) of Theorem 9:

$$b_j = \sup_{x \in \mathbb{X}^I} \left\| \rho^j_x - \tilde{\rho}^j_x \right\|.$$

Remark that the form of $\tilde{\rho}^i_x$ is already known from the study on $C_{i,j}$, hence we can compute just ρ^i_x and then compare it.

- If $j = (k, v')$ with $k < T$ and $v' \in V$ then:

$$\rho^j_x(A) = \tilde{\rho}^j_x(A),$$

because the difference is only on the final integral (on π_T and $\tilde{\pi}_T$) which disappear as consequence of the Markov property derived from the reversed kernel, hence:

$$b_j = 0.$$

- If $j = (T, v')$ and $v' \in V$ then:

$$\rho_x^{(T,v')}(A)$$
$$= \frac{\int \mathbb{I}_A(\tilde{x}_T^{v'})\mathbb{I}_{x\backslash x_T^{v'}}(\tilde{x}\backslash\tilde{x}_T^{v'})\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\pi_T(d\tilde{x}_T)}{\int \mathbb{I}_{x\backslash x_T^{v'}}(\tilde{x}\backslash\tilde{x}_T^{v'})\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1},\tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{T-1}(\tilde{x}_T,\tilde{x}_{T-1})\tilde{\pi}_{T-1}(d\tilde{x}_{T-1})\pi_T(d\tilde{x}_T)}$$
$$= \frac{\int \mathbb{I}_A(x_T^{v'})\overleftarrow{\tilde{p}}_{T-1}(x_T,x_{T-1})(\pi_T)_{x_T}^{v'}(dx_T^{v'})}{\int \overleftarrow{\tilde{p}}_{T-1}(x_T,x_{T-1})(\pi_T)_{x_T}^{v'}(dx_T^{v'})}.$$

Moreover, given that $\max_{i\in I}\sum_{j\in I} e^{m(i,j)}C_{i,j} \leqslant \frac{1}{2}$ then Lemma 10 can be applied and so:

$$\max_{i\in I}\sum_{j\in J} e^{m(i,J)}D_{i,j} \leqslant 2.$$

By joining step one and step two it follows that:

$$\begin{aligned}
&\left\|\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{T-1}}\pi_T\right\|_J \\
&\leqslant \sum_{v\in J}\sum_{j\in I} D_{(t,v),j}b_j \quad \leqslant \sum_{v\in J}\sum_{v'\in V} D_{(t,v),(T,v')}b_{(T,v')} \\
&\leqslant \sum_{v\in J}\sum_{v'\in V} e^{\beta|T-t|+\beta d(v,v')}D_{(t,v),(T,v')}e^{-\beta|T-t|-\beta d(v,v')} \\
&\qquad \sup_{x_{T-1},x_T\in\mathbb{X}^V}\left\|\left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi_T}}\right)_{x_T,x_{T-1}}^{v'} - \left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\tilde{\pi}_T}}\right)_{x_T,x_{T-1}}^{v'}\right\| \\
&\leqslant 2e^{-\beta(T-t)}\sum_{v\in J}\max_{v'\in V}\left\{e^{-\beta d(v,v')}\sup_{x_{T-1},x_T\in\mathbb{X}^V}\left\|\left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi_T}}\right)_{x_T,x_{T-1}}^{v'} - \left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\tilde{\pi}_T}}\right)_{x_T,x_{T-1}}^{v'}\right\|\right\}.
\end{aligned}$$

■

Proposition 23 *Suppose there exist $(\epsilon_-, \epsilon_+, \kappa) \in (0,1)^3$, such that:*

$$\epsilon_- \leqslant p^v(x^v, z^v) \leqslant \epsilon_+,$$

for all $x, z \in \mathbb{X}^V, v \in V$. Then for any $v \in V$ and for any $x_0, x_1 \in \mathbb{X}^V$ and $m \in \{0, \dots, n\}$:

$$\left\|\left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi_T}}\right)_{x_1,x_0}^{v} - \left(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi_T}}\right)_{x_1,x_0}^{v}\right\| \leqslant 2\left(\frac{\epsilon_+}{\epsilon_-}\right)^2 \sup_{x_1\in\mathbb{X}^V}\left\|(\pi_T)_{x_1}^v - (\tilde{\pi}_T)_{x_1}^v\right\|.$$

where $\tilde{\pi}_t$ is the approximated filtering distribution obtained through recursion (15) *and π_t is the approximated filtering distribution obtained through recursion* (4)*.*

Proof Denote with $\overleftarrow{\tilde{p}}_t(\cdot,\cdot)$ the reverse kernel with reference distribution the approximated filtering distribution, i.e.:

$$\overleftarrow{\tilde{p}}_t(z,x) := \frac{p(x,z)}{\int p(\hat{x},z)\tilde{\pi}_t(d\hat{x})}.$$

Consider our probability distributions from definition 21:

$$(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi}_T})^v_{x_1,x_0}(A) = \frac{\int \mathbb{I}_A(x^v_1)\overleftarrow{\tilde{p}}_{T-1}(x_1,x_0)(\pi_T)^v_{x_1}(dx^v_1)}{\int \overleftarrow{\tilde{p}}_{T-1}(x_1,x_0)(\pi_T)^v_{x_1}(dx^v_1)},$$

$$(\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\tilde{\pi}}_T})^v_{x_1,x_0}(A) = \frac{\int \mathbb{I}_A(x^v_1)\overleftarrow{\tilde{p}}_{T-1}(x_1,x_0)(\tilde{\pi}_T)^v_{x_1}(dx^v_1)}{\int \overleftarrow{\tilde{p}}_{T-1}(x_1,x_0)(\tilde{\pi}_T)^v_{x_1}(dx^v_1)}.$$

We can observe that we have the form $\int \mathbb{I}_A(x)\Lambda(x)\nu(dx)/\int \Lambda(x)\nu(dx)$ which allows us to apply Lemma 8. Hence:

$$\left\| (\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\pi}_T})^v_{x_1,x_0} - (\overset{\tilde{\pi}_{T-1}}{\overleftarrow{\tilde{\pi}}_T})^v_{x_1,x_0} \right\| \leqslant 2\left(\frac{\epsilon_+}{\epsilon_-}\right)^2 \sup_{x_1\in\mathbb{X}^V} \left\|(\pi_t)^v_{x_1} - (\tilde{\pi}_t)^v_{x_1}\right\|. \tag{A.14}$$

Remark that in this case the function Λ in Lemma 8 is $\overleftarrow{\tilde{p}}_{t-1}(\tilde{x}_1,x_0)\mathbb{I}_{x_1^{V\setminus v}}(\tilde{x}_1^{V\setminus v})$ hence the result follows from the following observations:

- $\overleftarrow{\tilde{p}}_{t-1}(\tilde{x}_1,x_0)\mathbb{I}_{x_1^{V\setminus v}}(\tilde{x}_1^{V\setminus v}) \leqslant \left(\frac{\epsilon_+}{\epsilon_-}\right)\frac{\prod_{v'\in V\setminus v} p^{v'}(x_0^{v'},x_1^{v'})}{\int \prod_{v'\in V\setminus v} p^{v'}(\hat{x}_0^{v'},x_1^{v'})\tilde{\pi}_{t-1}(d\hat{x}_0)}$;
- $\overleftarrow{\tilde{p}}_{t-1}(\tilde{x}_1,x_0)\mathbb{I}_{x_1^{V\setminus v}}(\tilde{x}_1^{V\setminus v}) \geqslant \left(\frac{\epsilon_-}{\epsilon_+}\right)\frac{\prod_{v'\in V\setminus v} p^{v'}(x_0^{v'},x_1^{v'})}{\int \prod_{v'\in V\setminus v} p^{v'}(\hat{x}_0^{v'},x_1^{v'})\tilde{\pi}_{t-1}(d\hat{x}_0)}$;

where the ratio is constant in x^v_1 and the inequality holds also for the sup and the inf. ■

A.3.2 Approximate smoothing stability and smoothing error control

We want to prove that an initial application of the optimal smoothing operator followed by a sequential application of the approximate smoothing operator to a probability distribution μ is not too different from an initial application of the approximate smoothing operator followed by a sequential application of the approximate smoothing operator to the same probability distribution μ.

Proposition 24 *Fix any collection of observations $\{y_1,\dots,y_T\}$ and any partition $\mathcal{K}$ on the set V. There exists a region $\mathcal{R}_0 \subseteq (0,1)^3$, as in Corollary 20, depending only on $\tilde{\Upsilon},\Upsilon$ and $\Upsilon^{(2)}$, such that if, for given $(\epsilon_-,\epsilon_+,\kappa)\in\mathcal{R}_0$,*

$$\epsilon_- \leqslant p^v(x^v,z^v) \leqslant \epsilon_+ \quad and \quad \kappa \leqslant g^f(x^{N(f)},y_t) \leqslant \frac{1}{\kappa}, \quad \forall x,z\in\mathbb{X}^V, f\in F, v\in V, t\in\{1,\dots,T\},$$

then for $\beta>0$ small enough depending only on $\tilde{\Upsilon},\Upsilon,\Upsilon^{(2)}$, ϵ_-,ϵ_+ and κ, we have that for any $t\in\{0,\dots,T-1\}$, $m\in\{0,\dots,n\}$ and for any $s=\{0,\dots,T-t+1\}$:

$$\begin{aligned}&\left\|\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\tilde{\pi}_{t+s}}\mu - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\pi_{t+s}}\mu\right\|_J\\ &\qquad\leqslant 2e^{-\beta s}\sum_{v\in J}\max_{v'\in V}\left\{e^{-\beta d(v,v')}\left\|(\tilde{\pi}_{t+s})^{v'}_{x_{t+s},z} - (\pi_{t+s})^{v'}_{x_{t+s},z}\right\|\right\}.\end{aligned}$$

Proof Denote with $\overleftarrow{p}_t(\cdot,\cdot)$ the reverse kernel with reference distribution the optimal filtering distribution and with $\overleftarrow{\tilde{p}}_t(\cdot,\cdot)$ the reverse kernel with reference distribution the approximated filtering distribution, i.e.:

$$\overleftarrow{p}_t(z,x) \coloneqq \frac{p(x,z)}{\int p(\hat{x},z)\pi_t(d\hat{x})} \quad \text{and} \quad \overleftarrow{\tilde{p}}_t(z,x) \coloneqq \frac{p(x,z)}{\int p(\hat{x},z)\tilde{\pi}_t(d\hat{x})}$$

Then:

$$\begin{aligned}
&\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s}}\mu(A)\\
&= \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(x_{t+s+1},x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s})\mu(dx_{t+s+1}),\\
&\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\pi_{t+s}}\mu(A)\\
&= \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(x_{t+s+1},x_{t+s})\pi_{t+s}(dx_{t+s})\mu(dx_{t+s+1}).
\end{aligned}$$

Define the probability distributions:

$$\begin{aligned}
\rho_z(A) &\coloneqq \int \mathbb{I}_A(x_t,\dots,x_{t+s})\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(z,x_{t+s})\pi_{t+s}(dx_{t+s}),\\
\tilde{\rho}_z(A) &\coloneqq \int \mathbb{I}_A(x_t,\dots,x_{t+s})\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z,x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s}).
\end{aligned}$$

The quantity of interest can be reformulated as follows:

$$\begin{aligned}
&\left\|\mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\tilde{\pi}_{t+s}}\mu - \mathsf{R}_{\tilde{\pi}_t}\dots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\pi_{t+s}}\mu\right\|_J\\
&\quad= \sup_{A\in\sigma(\mathbb{X}^J))}\Big|\int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z,x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s})\mu(dz)\\
&\qquad - \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(z,x_{t+s})\pi_{t+s}(dx_{t+s})\mu(dz)\Big|\\
&\quad= \sup_{A\in\sigma(\mathbb{X}^J))}\Big|\int\Big[\int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z,x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s})\\
&\qquad - \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(z,x_{t+s})\pi_{t+s}(dx_{t+s})\Big]\mu(dz)\Big|\\
&\quad\leqslant \sup_{A\in\sigma(\mathbb{X}^J))}\int\Big|\Big[\int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z,x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s})\\
&\qquad - \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(z,x_{t+s})\pi_{t+s}(dx_{t+s})\Big]\Big|\mu(dz)\\
&\quad\leqslant \sup_{z\in\mathbb{X}^V}\Bigg\{\sup_{A\in\sigma(\mathbb{X}^J))}\Big|\Big[\int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z,x_{t+s})\tilde{\pi}_{t+s}(dx_{t+s})\\
&\qquad - \int \mathbb{I}_A(x_t)\overleftarrow{\tilde{p}}_t(x_{t+1},x_t)\tilde{\pi}_t(dx_t)\dots\overleftarrow{p}_{t+s}(z,x_{t+s})\pi_{t+s}(dx_{t+s})\Big]\Big|\Bigg\}\\
&\quad= \sup_{z\in\mathbb{X}^V}\|\rho_z-\tilde{\rho}_z\|_{(t,J)}\,.
\end{aligned}$$

Hence it is enough to find a bound for $\|\rho_z-\tilde{\rho}_z\|_{(t,J)}$ to guarantee the proof of the statement. The Dobrushin theorem can be used on the distributions $\rho_z, \tilde{\rho}_z$ where the index set is $I=\bigcup_{k=t}^{t+s}(k,V)$ and the subset is (t,J).

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9. In the following passages we consider the notation $x = (x_t, \dots, x_{t+s})$, where $x_k \in \mathbb{X}^V$ for $k = t, \dots, t+s$ and $x \backslash x_k^v \coloneqq (x_t, \dots, x_k^{V\backslash v}, \dots, x_{t+s})$ (and the same with tilde).

- Consider $i = (t, v)$ and $v \in V$ then:

$$(\tilde{\rho}_z)_x^{(t,v)}(A) = \frac{\int \mathbb{I}_A(\tilde{x}_t^v)\mathbb{I}_{x\backslash x_t^v}(\tilde{x}\backslash\tilde{x}_t^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z, \tilde{x}_{t+s})\tilde{\pi}_{t+s}(d\tilde{x}_{t+s})}{\int \mathbb{I}_{x\backslash x_t^v}(\tilde{x}\backslash\tilde{x}_t^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z, \tilde{x}_{t+s})\tilde{\pi}_{t+s}(d\tilde{x}_{t+s})}$$
$$= \frac{\int \mathbb{I}_A(x_t^v)\overleftarrow{\tilde{p}}_t(x_{t+1}, x_t)(\tilde{\pi}_t)_{x_t}^v(dx_t^v)}{\int \overleftarrow{\tilde{p}}_t(x_{t+1}, x_t)(\tilde{\pi}_t)_{x_t}^v(dx_t^v)} = \frac{\int \mathbb{I}_A(x_t^v)p^v(x_t^v, x_{t+1}^v)(\tilde{\pi}_t)_{x_t}^v(dx_t^v)}{\int p^v(x_t^v, x_{t+1}^v)(\tilde{\pi}_t)_{x_t}^v(dx_t^v)},$$

  where the last passage follow from the independence of the numerator of the reverse kernel from x_t. Now we have to distinguish the different cases in which ρ_x^i can differ from $\rho_{\tilde{x}}^i$, where $x^{I\backslash j} = \tilde{x}^{I\backslash j}$.

  - If $j = (t, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \tilde{C}_{v,v'}^{\tilde{\pi}_t}$.
  - If $j = (t+1, v')$ and $v' \in V$ then: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
    where the result follow from Lemma 7, obtained by a majorization of the kernel part.
  - If $j = (k, v')$ with $k > t+1$ and $v' \in V$ then: $\quad C_{i,j} = 0$,
    which is obvious given that in ρ_x^i there is no dependence on x_k with $k > t+1$.

- Consider $i = (k, v)$ with $t + 1 < k \leqslant t + s$ and $v \in K \subseteq V$ then:

$$(\tilde{\rho}_z)_x^{(k,v)}(A)$$
$$= \frac{\int \mathbb{I}_A(\tilde{x}_k^v)\mathbb{I}_{x\backslash x_k^v}(\tilde{x}\backslash\tilde{x}_k^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z, \tilde{x}_{t+s})\tilde{\pi}_{t+s}(d\tilde{x}_{t+s})}{\int \mathbb{I}_{x\backslash x_k^v}(\tilde{x}\backslash\tilde{x}_k^v)\overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t)\tilde{\pi}_t(d\tilde{x}_t)\dots\overleftarrow{\tilde{p}}_{t+s}(z, \tilde{x}_{t+s})\tilde{\pi}_{t+s}(d\tilde{x}_{t+s})}$$
$$= \frac{\int \mathbb{I}_A(x_k^v)\overleftarrow{\tilde{p}}_{k-1}(x_k, x_{k-1})\overleftarrow{\tilde{p}}_k(x_{k+1}, x_k)(\tilde{\pi}_k)_{x_k}^v(dx_k)}{\int \overleftarrow{\tilde{p}}_{k-1}(x_k, x_{k-1})\overleftarrow{\tilde{p}}_k(x_{k+1}, x_k)(\tilde{\pi}_k)_{x_k}^v(dx_k)}$$
$$= \frac{\int \mathbb{I}_A(\tilde{x}_k^v)\overleftarrow{\tilde{p}}_{k-1}(x_k, x_{k-1})p^v(x_k^v, x_{k+1}^v)(\tilde{\pi}_k)_{x_k}^v(dx_k)}{\int \overleftarrow{\tilde{p}}_{k-1}(x_k, x_{k-1})p^v(x_k^v, x_{k+1}^v)(\tilde{\pi}_k)_{x_k}^v(dx_k)} = \frac{N_k}{D_k},$$

  where $x_{t+s+1} = z$ and everything follows the same procedure explained in the proof of Proposition 22, in particular we have:

$$N_k \coloneqq \int \mathbb{I}_A(x^v)p^v(x_{k-1}^v, x^v)p^v(x^v, x_{k+1}^v)$$
$$\int \prod_{f \in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w \in N_v^m(K)\backslash K} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}^{V\backslash K}(dx_0^{V\backslash K})\psi^{V\backslash K}(\tilde{x})\psi^v(dx^v),$$

$$D_k \coloneqq \int p^v(x_{k-1}^v, x^v)p^v(x^v, x_{k+1}^v)$$
$$\int \prod_{f\in N_f^m(K)} g^f(x_K^{N(f)}, y_t) \prod_{w\in N_v^m(K)\backslash K} p^w(x_0^w, x_K^w)\tilde{\pi}_{k-1}^{V\backslash K}(dx_0^{V\backslash K})\psi^{V\backslash K}(\tilde{x})\psi^v(dx^v).$$

- If $j = (k', v')$ with $k' \leqslant k-2$ and $v' \in V$ then: $\quad C_{i,j} = 0$.
- If $j = (k-1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
  where the result follow from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (k, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}\right) & v' \in N^2(v)\backslash v \\ 0 & \text{otherwise} \end{cases}$,
  where the result follow from Lemma 7, obtained by a majorization of the observation density part. Recall that the only factors that contains v are the one in $N(v)$ so the components that are connected to these factors are the one in $N^2(v)$.
- If $j = (k+1, v')$ and $v' \in V$ then $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & v' = v \\ 0 & v' \neq v \end{cases}$,
  where the result follow from Lemma 7, obtained by a majorization of the kernel part.
- If $j = (k', v')$ with $k' \geqslant k+2$ and $v' \in V$ then: $\quad C_{i,j} = 0$.

Given the previous results, for any $v \in V$ and $t+1 < k < t+s$:

$$\sum_{j\in I} e^{m(i,j)} C_{i,j} \leqslant \begin{cases} \sum_{v'\in V} e^{\beta d(v,v')} \tilde{C}_{v,v'}^{\tilde{\pi}_t} + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & i = (0, v) \\ 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + \sum_{v'\in N^2(v)} (1 - \kappa^{2\mathbf{card}(N(v)\cap N(v'))}) e^{\beta d(v,v')} & i = (k, v), \end{cases}$$

where $m(i,j) = \beta|k-k'| + \beta d(v,v')$ for $i = (k,v)$ and $j = (k',v')$ with $k,k' \in \{t,\dots,t+s\}$ and $v,v' \in V$ is the pseudometric of interest on the index set I. But then by combining the above calculation with the Corollary 20:

$$\max_{i\in I} \sum_{j\in I} C_{i,j} \leqslant \widetilde{\mathrm{Corr}}(\tilde{\pi}_t, \beta) + 2e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) + e^{2\beta}\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right) \leqslant \frac{1}{2}.$$

Given that $\sum_{j\in I} C_{i,j} \leqslant \sum_{j\in I} e^{m(i,j)} C_{i,j}$ then the Dobrushin theorem (Theorem 9) can be applied, meaning that:

$$\left\|\mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{t+s-1}} \mathsf{R}_{\tilde{\pi}_{t+s}} \mu - \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{t+s-1}} \mathsf{R}_{\pi_{t+s}} \mu\right\|_J = \sup_{z\in\mathbb{X}^V} \|\rho_z - \tilde{\rho}_z\|_{(t,J)} \leqslant \sum_{v\in J}\sum_{j\in I} D_{(t,v),j} b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9:

$$b_j = \sup_{x \in \mathbb{X}^I} \left\| (\rho_z)^j_x - (\tilde{\rho}_z)^j_x \right\|.$$

Remark that the form of $(\tilde{\rho}_z)^i_x$ is already known from the study on $C_{i,j}$, hence we can compute just $(\rho_z)^i_x$ and then compare it.

- If $j = (k, v')$ with $k < t+s$ and $v' \in V$ then:

$$(\rho_z)^j_x(A) = (\tilde{\rho}_z)^j_x(A),$$

  because the difference is only on the final kernel which disappear as consequence of the Markov property derived from the reversed kernel, hence:

$$b_j = 0.$$

- If $j = (t+s, v')$ and $v' \in V$ then:

$$\begin{aligned}
\rho_x^{(t+s,v')}(A) &= \frac{\int \mathbb{I}_A(\tilde{x}^{v'}_{t+s}) \mathbb{I}_{x \backslash x^{v'}_{t+s}}(\tilde{x} \backslash \tilde{x}^{v'}_{t+s}) \overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t) \tilde{\pi}_t(d\tilde{x}_t) \dots \overleftarrow{p}_{t+s}(z, \tilde{x}_{t+s}) \pi_{t+s}(d\tilde{x}_{t+s})}{\int \mathbb{I}_{x \backslash x^{v'}_{t+s}}(\tilde{x} \backslash \tilde{x}^{v'}_{t+s}) \overleftarrow{\tilde{p}}_t(\tilde{x}_{t+1}, \tilde{x}_t) \tilde{\pi}_t(d\tilde{x}_t) \dots \overleftarrow{p}_{t+s}(z, \tilde{x}_{t+s}) \pi_{t+s}(d\tilde{x}_{t+s})} \\
&= \frac{\int \mathbb{I}_A(x^{v'}_{t+s}) \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) \overleftarrow{p}_{t+s}(z, x_{t+s}) (\pi_{t+s})^{v'}_{x_{t+s}}(dx^{v'}_{t+s})}{\int \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) \overleftarrow{p}_{t+s}(z, x_{t+s}) (\pi_{t+s})^{v'}_{x_{t+s}}(dx^{v'}_{t+s})} \\
&= \frac{\int \mathbb{I}_A(x^{v'}_{t+s}) \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) p^v(x^v_{t+s}, z^v) (\pi_{t+s})^{v'}_{x_{t+s}}(dx^{v'}_{t+s})}{\int \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) p^v(x^v_{t+s}, z^v) (\pi_{t+s})^{v'}_{x_{t+s}}(dx^{v'}_{t+s})} \\
&= \frac{\int \mathbb{I}_A(x^{v'}_{t+s}) \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) (\pi_{t+s})^{v'}_{x_{t+s},z}(dx^{v'}_{t+s})}{\int \overleftarrow{\tilde{p}}_{t+s-1}(x_{t+s}, x_{t+s-1}) (\pi_{t+s})^{v'}_{x_{t+s},z}(dx^{v'}_{t+s})},
\end{aligned}$$

  where the last few passages follow from the factorization of the transition kernel $p(x, z)$ and the fact that z is a constant and at the denominator of the reversed kernel we are integrating out the dependent variable x_{t+s} and from the fact that the can rewrite numerator and denominator as integrals with respect to the one step forward conditional distribution. Now from the same procedure as in(A.14) in Proposition 23:

$$b_j \leqslant 2 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 \sup_{x_{t+s}, z \in \mathbb{X}^V} \left\| (\tilde{\pi}_{t+s})^{v'}_{x_{t+s},z} - (\pi_{t+s})^{v'}_{x_{t+s},z} \right\|.$$

Moreover, given that $\max_{i \in I} \sum_{j \in I} e^{m(i,j)} C_{i,j} \leqslant \frac{1}{2}$ then Lemma 10 can be applied and so:

$$\max_{i \in I} \sum_{j \in J} e^{m(i,J)} D_{i,j} \leqslant 2.$$

By joining step one and step two it follows that:

$$
\begin{aligned}
&\left\|\mathsf{R}_{\tilde{\pi}_t}\ldots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\tilde{\pi}_{t+s}}\mu - \mathsf{R}_{\tilde{\pi}_t}\ldots\mathsf{R}_{\tilde{\pi}_{t+s-1}}\mathsf{R}_{\pi_{t+s}}\mu\right\|_J = \sup_{z\in\mathbb{X}^V}\|\rho_z - \tilde{\rho}_z\|_{(t,J)} \\
&\leqslant \sup_{z\in\mathbb{X}^V}\sum_{v\in J}\sum_{j\in I} D_{(t,v),j}b_j \leqslant \sum_{v\in J}\sum_{v'\in V} D_{(t,v),(t+s,v')}b_{(t+s,v')} \\
&\leqslant \sup_{z\in\mathbb{X}^V}\sum_{v\in J}\sum_{v'\in V} e^{\beta|t+s-t|+\beta d(v,v')}D_{(t,v),(t+s,v')}e^{-\beta|t+s-t|-\beta d(v,v')} \\
&\qquad 2\left(\frac{\epsilon_+}{\epsilon_-}\right)^2 \sup_{x_{t+s},z\in\mathbb{X}^V}\left\|(\tilde{\pi}_{t+s})^{v'}_{x_{t+s},z} - (\pi_{t+s})^{v'}_{x_{t+s},z}\right\| \\
&\leqslant 4\left(\frac{\epsilon_+}{\epsilon_-}\right)^2 e^{-\beta s}\sum_{v\in J}\max_{v'\in V}\left\{e^{-\beta d(v,v')}\left\|(\tilde{\pi}_{t+s})^{v'}_{x_{t+s},z} - (\pi_{t+s})^{v'}_{x_{t+s},z}\right\|\right\} \\
&= 4\left(\frac{\epsilon_+}{\epsilon_-}\right)^2 e^{-\beta s}\sum_{v\in J}\max_{v'\in V}\left\{e^{-\beta d(v,v')}\left\|(\tilde{\pi}_{t+s})^{v'}_{x_{t+s},z} - (\pi_{t+s})^{v'}_{x_{t+s},z}\right\|\right\}.
\end{aligned}
$$

■

Proposition 25 *Fix any collection of observations $\{y_1,\ldots,y_T\}$ and any partition $\mathcal{K}$ on the set V. There exists a region $\tilde{\mathcal{R}}_0 \subseteq (0,1)^3$, depending only on $\tilde{\Upsilon},\Upsilon$ and $\Upsilon^{(2)}$, such that if, for given $(\epsilon_-,\epsilon_+,\kappa)\in\tilde{\mathcal{R}}_0$,*

$$\epsilon_- \leqslant p^v(x^v,z^v) \leqslant \epsilon_+ \quad \textit{and} \quad \kappa \leqslant g^f(x^{N(f)},y_t) \leqslant \frac{1}{\kappa},$$

for all $x,z\in\mathbb{X}^V, f\in F, v\in V, t\in\{1,\ldots,T\}$, then for $\beta>\log(2)$ small enough depending only on $\tilde{\Upsilon},\Upsilon,\Upsilon^{(2)}$, ϵ_-,ϵ_+ and κ, chosen according to Corollary 20, we have that for any $t\in\{1,\ldots,T\}$ and $m\in\{0,\ldots,n\}$:

$$\left\|(\pi_t)^v_{x_1,z} - (\tilde{\pi}_t)^v_{x_1,z}\right\| \leqslant \frac{\tau(\beta,\kappa,m,\mathcal{K})e^\beta}{e^\beta-2}, \quad \forall t\in\{1,\ldots,T\}, \forall x_1,z\in\mathbb{X}^V,$$

where $\tilde{\pi}_t$ is the approximated filtering distribution obtained through recursion (15)*,*

$$\tau(\beta,\kappa,m,\mathcal{K}) := 2\left(1-\kappa^{a(\mathcal{K})}\right) + 4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right),$$

and $a(\mathcal{K})$ and $b(m,\mathcal{K})$ are as in Proposition 15.

Proof Consider the quantity of interest, it is possible to decompose it as follow:

$$\left\|(\pi_t)^v_{x_1,z} - (\tilde{\pi}_t)^v_{x_1,z}\right\| \leqslant \left\|(\mathsf{F}_t\pi_{t-1})^v_{x_1,z} - (\mathsf{F}_t\tilde{\pi}_{t-1})^v_{x_1,z}\right\| + \left\|(\mathsf{F}_t\tilde{\pi}_{t-1})^v_{x_1,z} - (\tilde{\mathsf{F}}^m_t\tilde{\pi}_{t-1})^v_{x_1,z}\right\|,$$

where the second quantity can be controlled using Proposition 15 given that Corollary 20 holds, from the choices of $\beta,\epsilon_-,\epsilon_+,\kappa$, while for the first quantity the Dobrushin machinery must be used.

Consider the probability distributions:

$$\rho(A) \coloneqq \frac{\int \mathbb{I}_A(x_0, x_1^v) \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) \pi_{t-1}(dx_0) p^v(x_1^v, z^v) \psi^v(x_1^v)}{\int \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) \pi_{t-1}(dx_0) p^v(x_1^v, z^v) \psi^v(x_1^v)},$$

$$\tilde{\rho}(A) \coloneqq \frac{\int \mathbb{I}_A(x_0, x_1^v) \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) \tilde{\pi}_{t-1}(dx_0) p^v(x_1^v, z^v) \psi^v(x_1^v)}{\int \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) \tilde{\pi}_{t-1}(dx_0) p^v(x_1^v, z^v) \psi^v(x_1^v)}.$$

It can be observed that:

$$\|\rho - \tilde{\rho}\|_{(1,v)} = \left\|(\mathsf{F}_t \pi_{t-1})^v_{x_1,z} - (\mathsf{F}_t \tilde{\pi}_{t-1})^v_{x_1,z}\right\|.$$

So again the Dobrushin theorem can be applied to $\rho, \tilde{\rho}$ where the index set is $I = (0, V) \cup (1, v)$.

The first step is to bound $C_{i,j}$ for all the possible combination of $i, j \in I$, as in (A.2) of Theorem 9.

- Consider $i = (0, b)$ and $b \in V$ then:

$$\begin{aligned}
&\tilde{\rho}^{(0,b)}_{x_0,x_1}(A) \\
&= \frac{\int \mathbb{I}_A(\tilde{x}_0^b) \mathbb{I}_{\{x_0^{V\setminus b}, x_1^v\}}(\tilde{x}_0^{V\setminus b}, \tilde{x}_1^v) \prod_{f\in N(v)} g^f(\tilde{x}_1^{N(f)}, y_t) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) p^v(\tilde{x}_1^v, z^v) \psi^v(\tilde{x}_1^v)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_0^b) \mathbb{I}_{\{x_0^{V\setminus b}, x_1^v\}}(\tilde{x}_0^{V\setminus b}, \tilde{x}_1^v) \prod_{f\in N(v)} g^f(\tilde{x}_1^{N(f)}, y_t) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) p^v(\tilde{x}_1^v, z^v) \psi^v(\tilde{x}_1^v)} \\
&= \frac{\prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_1^v, z^v)}{\prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_1^v, z^v)} \frac{\int \mathbb{I}_A(\tilde{x}_0^b) \mathbb{I}_{\{x_0^{V\setminus b}, x_1^v\}}(\tilde{x}_0^{V\setminus b}, \tilde{x}_1^v) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) \psi^v(\tilde{x}_1^v)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_0^b) \mathbb{I}_{\{x_0^{V\setminus b}, x_1^v\}}(\tilde{x}_0^{V\setminus b}, \tilde{x}_1^v) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) \psi^v(\tilde{x}_1^v)} \\
&= \frac{\int \mathbb{I}_A(x_0^b) p^b(x_0^b, x_1^b) (\tilde{\pi}_{t-1})^b_{x_0}(dx_0^b)}{\int p^b(x_0^b, x_1^b) (\tilde{\pi}_{t-1})^b_{x_0}(dx_0^b)}.
\end{aligned}$$

  - If $j = (0, b')$ and $b' \in V$ then: $\quad C_{i,j} \leqslant \tilde{C}^{\tilde{\pi}_{t-1}}_{b,b'}$.
  - If $j = (1, v)$ then by Lemma 7: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b = v \\ 0 & \text{otherwise} \end{cases}$,

    where we maximize the kernel part.

- Consider $i = (1, v)$ then:

$$\begin{aligned}
\tilde{\rho}^{(1,v)}_{x_0,x_1}(A) &= \frac{\int \mathbb{I}_A(\tilde{x}_1^v) \mathbb{I}_{x_0}(\tilde{x}_0^V) \prod_{f\in N(v)} g^f(\tilde{x}_1^{N(f)}, y_t) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) p^v(\tilde{x}_1^v, z^v) \psi^v(\tilde{x}_1^v)}{\int \mathbb{I}_{\mathbb{X}}(\tilde{x}_1^v) \mathbb{I}_{x_0}(\tilde{x}_0^V) \prod_{f\in N(v)} g^f(\tilde{x}_1^{N(f)}, y_t) p^v(\tilde{x}_0^v, \tilde{x}_1^v) \tilde{\pi}_{t-1}(d\tilde{x}_0) p^v(\tilde{x}_1^v, z^v) \psi^v(\tilde{x}_1^v)} \\
&= \frac{\int \mathbb{I}_A(x_1^v) \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) p^v(x_1^v, z^v) \psi^v(dx_1^v)}{\int \prod_{f\in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) p^v(x_1^v, z^v) \psi^v(dx_1^v)}.
\end{aligned}$$

– If $j = (0, b)$ and $b \in V$ then by Lemma 7: $\quad C_{i,j} \leqslant \begin{cases} \left(1 - \frac{\epsilon_-}{\epsilon_+}\right) & b = v \\ 0 & \text{otherwise} \end{cases},$

where we maximize the kernel part.

– If $j = (1, v)$ then: $\quad C_{i,j} = 0$.

But then:

$$\max_{i \in I} \sum_{j \in I} e^{m(i,j)} C_{i,j} \leqslant \widetilde{\mathrm{Corr}}(\tilde{\pi}_{t-1}, \beta) + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right).$$

Given that we want $\beta > \log(2)$ we cannot just choose $\epsilon_-, \epsilon_+, \kappa, \beta$ according to Corollary 20, but we need to modify the region $\mathcal{R}_0$:

$$\begin{aligned}\tilde{\mathcal{R}}_0 \coloneqq \Bigg\{ (\epsilon_-, \epsilon_+, \kappa) \in (0,1)^3 : \frac{\kappa^{2\Upsilon} - 8\left(1 - \frac{\epsilon_-}{\epsilon_+}\right)}{4\Upsilon^{(2)}\left(1 - \kappa^{2\tilde{\Upsilon}}\right)\kappa^{2\Upsilon}} > 4 \quad \text{and} \\ \frac{1 - 4\left(1 - \frac{\epsilon_-}{\epsilon_+}\right)}{6\Upsilon^{(2)}(1 - \kappa^{2\tilde{\Upsilon}}) + 4\left(1 - \frac{\epsilon_-}{\epsilon_+}\right)} > 4 \Bigg\}.\end{aligned}$$

hence for $(\epsilon_-, \epsilon_+, \kappa) \in \tilde{\mathcal{R}}_0$ and β as in Corollary 20 we have:

$$\widetilde{\mathrm{Corr}}(\tilde{\pi}_{t-1}, \beta) + e^{\beta}\left(1 - \frac{\epsilon_-}{\epsilon_+}\right) \leqslant \frac{1}{2}.$$

Hence the Dobrushin theorem applies:

$$\|\rho - \tilde{\rho}\|_{(1,v)} \leqslant \sum_{j \in I} D_{(1,v),j} b_j.$$

The second step is to control the quantities b_j, as in (A.2) of Theorem 9. Remark that the conditional distributions of ρ have the same form of the ones of $\tilde{\rho}$ with $\tilde{\pi}_{t-1}$ instead of π_{t-1}

- If $j = (0, b')$ and $b' \in V$ then:

$$\rho^{(0,b')}_{x_0,x_1}(A) = \frac{\int \mathbb{I}_A(x_0^{b'}) p^{b'}(x_0^{b'}, x_1^{b'})(\pi_{t-1})^{b'}_{x_0}(dx_0^{b'})}{\int p^{b'}(x_0^{b'}, x_1^{b'})(\pi_{t-1})^{b'}_{x_0}(dx_0^{b'})},$$

hence:

$$b_j = \sup_{x_0, x_1 \in \mathbb{X}^V} \left\| (\pi_{t-1})^{b'}_{x_0,x_1} - (\tilde{\pi}_{t-1})^{b'}_{x_0,x_1} \right\|.$$

- If $j = (1, v)$ then:

$$\tilde{\rho}^{(1,v)}_{x_0,x_1}(A) = \frac{\int \mathbb{I}_A(x_1^v) \prod_{f \in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) p^v(x_1^v, z^v) \psi^v(dx_1^v)}{\int \prod_{f \in N(v)} g^f(x_1^{N(f)}, y_t) p^v(x_0^v, x_1^v) p^v(x_1^v, z^v) \psi^v(dx_1^v)}.$$

hence:

$$b_j = 0,$$

because the only difference between $\rho, \tilde{\rho}$ is on π_{t-1} and $\tilde{\pi}_{t-1}$.

From the b_j bounds, it can be concluded that:

$$\begin{aligned}\left\|(F_t\pi_{t-1})^v_{x_1,z}-(F_t\tilde{\pi}_{t-1})^v_{x_1,z}\right\| &\leqslant \sum_{j\in I} D_{(1,v),j}b_j\\ &\leqslant \sum_{b'\in V} D_{(1,v),(0,b')}e^{\beta d(v,b')+\beta}e^{-\beta d(v,b')-\beta}\sup_{x_0,x_1\in\mathbb{X}^V}\left\|(\pi_{t-1})^{b'}_{x_0,x_1}-(\tilde{\pi}_{t-1})^{b'}_{x_0,x_1}\right\|\\ &\leqslant 2e^{-\beta}\sup_{b'\in V}\left\{e^{-\beta d(v,b')}\sup_{x_0,x_1\in\mathbb{X}^V}\left\|(\pi_{t-1})^{b'}_{x_0,x_1}-(\tilde{\pi}_{t-1})^{b'}_{x_0,x_1}\right\|\right\}.\end{aligned}$$

By joining this with Proposition 15 in Equation A.3.2:

$$\begin{aligned}\left\|(\pi_t)^v_{x_1,z}-(\tilde{\pi}_t)^v_{x_1,z}\right\| \leqslant 2e^{-\beta}\sup_{b'\in V}&\left\{e^{-\beta d(v,b')}\sup_{x_0,x_1\in\mathbb{X}^V}\left\|(\pi_{t-1})^{b'}_{x_0,x_1}-(\tilde{\pi}_{t-1})^{b'}_{x_0,x_1}\right\|\right\}\\ &+\tau(\beta,\kappa,m,\mathcal{K}).\end{aligned}$$

This result can be iteratively applied. Indeed given that:

$$\begin{aligned}\left\|(\pi_1)^v_{x_1,z}-(\tilde{\pi}_1)^v_{x_1,z}\right\| &\leqslant 2e^{-\beta}\sup_{b'\in V}\left\{e^{-\beta d(v,b')}\sup_{x_0,x_1\in\mathbb{X}^V}\left\|(\delta_x)^{b'}_{x_0,x_1}-(\delta_x)^{b'}_{x_0,x_1}\right\|\right\}+\tau(\beta,\kappa,m,\mathcal{K})\\ &=\tau(\beta,\kappa,m,\mathcal{K}),\end{aligned}$$

then:

$$\left\|(\pi_t)^v_{x_1,z}-(\tilde{\pi}_t)^v_{x_1,z}\right\|\leqslant \tau(\beta,\kappa,m,\mathcal{K})\sum_{j=0}^{t-1}(2e^{-\beta})^j\leqslant\frac{\tau(\beta,\kappa,m,\mathcal{K})e^{\beta}}{e^{\beta}-2},$$

where the last passage follow trivially from the definition of geometric sum and $\beta>\log(2)$. ■

Proposition 26 *Fix any collection of observations* $\{y_1,\dots,y_T\}$ *and any partition* $\mathcal{K}$ *on the set* V*. There exists a region* $\tilde{\mathcal{R}}_0\subseteq(0,1)^3$*, depending only on* $\tilde{\Upsilon},\Upsilon$ *and* $\Upsilon^{(2)}$*, such that if, for given* $(\epsilon_-,\epsilon_+,\kappa)\in\tilde{\mathcal{R}}_0$*,*

$$\epsilon_-\leqslant p^v(x^v,z^v)\leqslant\epsilon_+\quad\text{and}\quad\kappa\leqslant g^f(x^{N(f)},y_t)\leqslant\frac{1}{\kappa},$$

for all $x,z\in\mathbb{X}^V, f\in F, v\in V, t\in\{1,\dots,T\}$*, then for* $\beta>\log(2)$ *small enough depending only on* $\tilde{\Upsilon},\Upsilon,\Upsilon^{(2)}$*,* ϵ_-,ϵ_+ *and* κ*, chosen according to Corollary 20, we have that for any* $t\in\{1,\dots,T\}$ *and* $m\in\{0,\dots,n\}$*:*

$$\left\|(\pi_t)^v_{x_1}-(\tilde{\pi}_t)^v_{x_1}\right\|\leqslant\frac{\tau(\beta,\kappa,m,\mathcal{K})e^{\beta}}{e^{\beta}-2},$$

for all $t\in\{1,\dots,T\},\forall x_1,z\in\mathbb{X}^V$*, where* $\tilde{\pi}_t$ *is the approximated filtering distribution obtained through recursion* (15)*,*

$$\tau(\beta,\kappa,m,\mathcal{K}):=2\left(1-\kappa^{a(\mathcal{K})}\right)+4e^{-\beta m}\left(1-\kappa^{b(\mathcal{K},m)}\right),$$

and $a(\mathcal{K})$ *and* $b(m,\mathcal{K})$ *are as in Proposition 15.*

Proof The proof of this result follows the same procedure of Proposition 25, the only difference is that $p^v(x_1^v, z^v)$ is missing. ■

A.3.3 Proof of Theorem 3

Proof Let $J \subseteq K \in \mathcal{K}$, then by the triangular inequality:

$$
\begin{aligned}
\left\| \mathsf{R}_{\tilde{\pi}_t} \tilde{\pi}_{t+1|T} - \mathsf{R}_{\pi_t} \pi_{t+1|T} \right\|_J \leqslant& \left\| \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \tilde{\pi}_T - \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{T-1}} \pi_T \right\|_J \\
&+ \sum_{s=0}^{T-t+1} \left\| \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{t+s-1}} \mathsf{R}_{\tilde{\pi}_{t+s}} \pi_{t+s+1|T} - \mathsf{R}_{\tilde{\pi}_t} \dots \mathsf{R}_{\tilde{\pi}_{t+s-1}} \mathsf{R}_{\pi_{t+s}} \pi_{t+s+1|T} \right\|_J .
\end{aligned}
$$

Given that we are choosing $\beta, \epsilon_-, \epsilon_+, \kappa$ such that both Corollary 20 and Proposition 22 and Proposition 24 hold then:

$$
\begin{aligned}
\left\| \tilde{\pi}_{t|T} - \pi_{t|T} \right\|_J \leqslant& 2e^{-\beta(T-t)} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \sup_{x_{T-1}, x_T \in \mathbb{X}^V} \left\| \overset{\tilde{\pi}_{T-1}}{\overleftarrow{(\tilde{\pi}_T})}{}^{v'}_{x_{T-1}, x_T} - \overset{\tilde{\pi}_{T-1}}{\overleftarrow{(\pi_T})}{}^{v'}_{x_{T-1}, x_T} \right\| \right\} \\
&+ \sum_{s=0}^{T-t+1} 2e^{-\beta s} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \left\| (\tilde{\pi}_{t+s})^{v'}_{x_{t+s}, z} - (\pi_{t+s})^{v'}_{x_{t+s}, z} \right\| \right\} .
\end{aligned}
$$

Similarly also Proposition 25, Proposition 26 and Proposition 23 apply:

$$
\begin{aligned}
\left\| \tilde{\pi}_{t|T} - \pi_{t|T} \right\|_J \leqslant& 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 e^{-\beta(T-t)} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \sup_{x_1 \in \mathbb{X}^V} \left\| (\pi_T)^{v'}_{x_1} - (\tilde{\pi}_T)^{v'}_{x_1} \right\| \right\} \\
&+ 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 \sum_{s=0}^{T-t-1} e^{-\beta s} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \frac{\tau(\beta, \kappa, m, \mathcal{K}) e^\beta}{e^\beta - 2} \right\} \\
\leqslant& 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 e^{-\beta(T-t)} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \frac{\tau(\beta, \kappa, m, \mathcal{K}) e^\beta}{e^\beta - 2} \right\} \\
&+ 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 \sum_{s=0}^{T-t-1} e^{-\beta s} \sum_{v \in J} \max_{v' \in V} \left\{ e^{-\beta d(v,v')} \frac{\tau(\beta, \kappa, m, \mathcal{K}) e^\beta}{e^\beta - 2} \right\} \\
\leqslant& 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 e^{-\beta(T-t)} \frac{\tau(\beta, \kappa, m, \mathcal{K}) e^\beta}{e^\beta - 2} \mathbf{card}(J) \\
&+ 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 \frac{e^\beta}{e^\beta - 1} \frac{\tau(\beta, \kappa, m, \mathcal{K}) e^\beta}{e^\beta - 2} \mathbf{card}(J) \\
=& 4 \left(\frac{\epsilon_+}{\epsilon_-} \right)^2 \left[e^{-\beta(T-t)} \frac{e^\beta}{e^\beta - 2} + \frac{e^\beta}{e^\beta - 1} \frac{e^\beta}{e^\beta - 2} \right] \\
&\mathbf{card}(J) \left[2 \left(1 - \kappa^{a(\mathcal{K})} \right) + 4e^{-\beta m} \left(1 - \kappa^{b(\mathcal{K}, m)} \right) \right] \\
\leqslant& \alpha_2(\beta, \epsilon_-, \epsilon_+) \left(1 - \kappa^{a(\mathcal{K})} \right) \mathbf{card}(J) + \gamma_2(\beta, \epsilon_-, \epsilon_+) \left(1 - \kappa^{b(\mathcal{K}, m)} \right) \mathbf{card}(J) e^{-\beta m},
\end{aligned}
$$

■

Appendix B.

This appendix is an implementation guide for the Graph Filter, which will use the Gaussian model considered in subsection 4.1 as guiding example.

B.1 Model parameters and Graph Filter parameters: Gaussian model

As explained in subsection 4.1 the parameters of the model are $\hat{\mu}_0, \hat{p}, c, \sigma^2$ where:

- $\hat{\mu}_0$ is a probability vector and represents the distribution of X_0^v for all $v \in V$ (we are assuming a factorization of the initial distribution);
- $\hat{p}$ is a stochastic matrix and represents the transition kernel of the Markov chain $(X_t^v)_{t=0,\dots,T}$ for all $v \in V$ (this is the FHMM assumption on the transition kernel);
- c, σ^2 are the parameters of the emission distribution.

The Graph Filter has two parameters: m and the partition $\mathcal{K}$. For this guiding example we assume:

- $m = 0$, meaning that we are using only the "closest" factors in the emission distribution;
- $\mathcal{K} = \{\{1\}, \dots, \{M\}\}$ meaning that we are considering the partition of singletons.

Under the above setting it is straightforward to compute $N_v^m(K), N_f^m(K)$ for $K \in \mathcal{K}$. indeed:

- $N_v^m(K) = \begin{cases} \{K, K+1\} & \text{if } K = \{1\} \\ \{K-1, K, K+1\} & \text{if } K \in \{\{2\}, \dots, \{M-1\}\} \\ \{K-1, K\} & \text{if } K = \{M\} \end{cases}$
- $N_f^m(K) = \begin{cases} \{K\} & \text{if } K = \{1\} \\ \{K-1, K\} & \text{if } K \in \{\{2\}, \dots, \{M-1\}\} \\ \{K-1\} & \text{if } K = \{M\} \end{cases}$

B.2 Graph Filter

Let us start with the reformulation of algorithm 1, here the main quantities are the probability vector to update μ, which factorises over $\mathcal{K}$, and the likelihood terms $(g^f(\cdot, y))_{f \in F}$.

Given that the probability vector μ factorises, it can be represented as the Kronecker product of its the marginals $(\mu^K)_{K \in \mathcal{K}}$:

$$\mu = \bigotimes_{K \in \mathcal{K}} \mu^K$$

where $\bigotimes$ represents the Kronecker product notation. Moreover, from $(\mu^K)_{K\in\mathcal{K}}$ we can get any joint distribution, for example $\mu^{\{K-1,K,K+1\}} = \bigotimes_{K'\in\{\{K-1\},\{K\},\{K+1\}\}} \mu^{K'}$. Note that it is essential to fix an order when doing the Kronecker product, this guarantees that the states are associated to the right probabilities and avoids inconsistencies when doing multiple computations (i.e. we want the j component of μ to be associated to the same state when running multiple times the Kronecker product). To this extent, we define a function $\phi : \{\{1\},\dots,\{L\}^M\} \to \{1,\dots,L\}^M$ which uniquely enumerates the states of the Markov chain (it can be any ordering of the states as long as it is preserved over multiple calculations). Subsequently, we define ordering functions for a restricted number of components.

Consider now the likelihood terms $(g^f(\cdot, y))_{f\in F}$, they can be combined in diagonal matrices containing the factors to be used in the approximate update. Precisely, the diagonal matrix G^K is defined as $G^K \coloneqq \mathbf{diag}((\prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y))_{x^{N_v^m(K)}\in\mathbb{X}^{N_v^m(K)}})$. Here, the vector $(\prod_{f\in N_f^m(K)} g^f(x^{N(f)}, y))_{x^{N_v^m(K)}\in\mathbb{X}^{N_v^m(K)}}$ is the vector indexed by the states of $\mathbb{X}^{N_v^m(K)}$ with elements given by the likelihood terms computed in those states. For the Gaussian model example, for a fixed $K \in \mathcal{K}$ we can define G^K as the diagonal matrix with elements:

$$(G^K)^{(j,j)} = \prod_{f\in N_f^m(K)} \mathcal{N}\left(y^{(f)} | c(\phi_K(j))^{(f)} + (\phi_K(j))^{(f+1)}, \sigma^2\right), \quad j = 1,\dots,L^{\mathbf{card}(N_v^m(K))} \tag{B.1}$$

where $\phi_K : \{\{1\},\dots,\{L^{\mathbf{card}(N_v^m(K))}\}\} \to \{1,\dots,L\}^{\mathbf{card}(N_v^m(K))}$ is an ordering function on the components in $N_v^m(K)$ and $\mathcal{N}(y|\mu,\sigma^2)$ is the univariate Gaussian density in y with mean μ and variance σ^2. Remark that $\phi_K(j)$ is a vector of dimension $\mathbf{card}(N_v^m(K))$ with elements in the set $\{1,\dots,L\}$, so $(\phi_K(j))^{(f)}$ is the f-th component of that vector. Here it is clear why enumerating the states is important. When computing the j-th element on the diagonal of G^K we want it to be associated to a unique state of the Markov chain and we want this enumeration to be always the same.

Now we have all the ingredients to redefine algorithm 1 in a vector-matrix notation, which is reported in the algorithm below. Note that the algorithm is not restricted to the Gaussian case, indeed $(G^K)_{K\in\mathcal{K}}$ are computed before running the algorithm and so we are not restricting to any specific emission distribution.

Approximate Bayes update (ABu): vector-matrix reformulation

Require: $\mathcal{K}, (N_f^m(K))_{K\in\mathcal{K}}, (N_v^m(K))_{K\in\mathcal{K}}, (\mu^K)_{K\in\mathcal{K}}, (G^K)_{K\in\mathcal{K}}$

1: **for** $K \in \mathcal{K}$ **do**
2: $\quad\hat{\mu} \leftarrow \bigotimes_{K'\in N_v^m(K)} \mu^{K'}$
3: $\quad\hat{\mu} \leftarrow G^K \hat{\mu}$
4: $\quad$**Normalize** $\hat{\mu}$ **to a probability vector**
5: $\quad$**Marginalize out the components in** $N_v^m(K)\backslash K$ **and store it in** $\tilde{\mu}^K$
return $(\tilde{\mu}^K)_{K\in\mathcal{K}}$

The marginalize out operation is made by summing over elements of the probability vector which agree on K. For example, for the Gaussian model we sum over the indexes i with the same $\phi_K(i)^K$ (i.e. we sum over all the vectors which have the same component in

K), this result in a distribution over $\{1, \ldots, L\}$, indeed each K is a singleton (for the chosen $\mathcal{K}$) and a single component of the Markov chain can assume L states. Further the $\hat{\mu}$ in the algorithm is by construction a distribution over $\{1, \ldots, L\}^{\mathbf{card}(N_v^m(K))}$ with $K \in N_v^m(K)$, even in this case it is essential to know the order of K in $N_v^m(K)$ to ensure a marginalization over the right component.

Graph Filter: vector-matrix reformulation for the Gaussian model

---

Require: $\mathcal{K}, (N_f^m(K))_{K\in\mathcal{K}}, (N_v^m(K))_{K\in\mathcal{K}}, \hat{\mu}_0, \hat{p}, c, \sigma^2, (y_t)_{t=\{1,\ldots,T\}}$

1: **for** $K \in \mathcal{K}$ **do**

2: $\quad\tilde{\pi}_0^K \leftarrow \hat{\mu}_0$

3: **for** $t \in \{1, \ldots, T\}$ **do**

4: $\quad$ **for** $K \in \mathcal{K}$ **do**

5: $\quad\quad\hat{\pi}^K \leftarrow \hat{p}\tilde{\pi}_{t-1}^K$

6: $\quad\quad$ **Compute** G^K**:** $(G^K)^{(j,j)} = \prod_{f\in N_f^m(K)} \mathcal{N}\left(y^{(f)} | c(\mathbf{dec}(j))^{(f)} + (\mathbf{dec}(j))^{(f+1)}, \sigma^2\right)$

7: $\quad(\tilde{\pi}_t^K)_{K\in\mathcal{K}} \leftarrow \mathbf{ABu}\Big(\mathcal{K}, (N_f^m(K))_{K\in\mathcal{K}}, (N_v^m(K))_{K\in\mathcal{K}}, (\hat{\pi}^K)_{K\in\mathcal{K}}, (G^K)_{K\in\mathcal{K}}\Big)$

8: **return** $((\tilde{\pi}_t^K)_{K\in\mathcal{K}})_{t=\{0,\ldots,T\}}$

The reformulation of the Graph Filter, algorithm 2, requires only a switch to the vector-matrix notation and it is reported in the algorithm above. Now we are restricting ourselves to the Gaussian example because we are computing $(G^K)_{K\in\mathcal{K}}$ as described in (B.1) and we are fixing initial distribution and transition kernel to the Gaussian example parameters. The Graph Smoother, algorithm 3, is not reported in this section but it follows similarly to the Graph Filter.

Appendix C.

This appendix provides some extra details on the experiments: subsection C.1 shows the formulation of the EM algorithm for Graph Filter-Smoother under Gaussian and Poisson emission; subsection C.2 gives some additional details on the traffic flow experiment.

C.1 EM algorithm for Graph Filter-Smoother

The parameters of the FHMMs considered in section 4 can be estimated with an EM algorithm where the exact smoothing distribution is substituted with the *Graph Filter-Smoother* approximation. This appendix contains details of the considered maximization step for both the Gaussian and the Poisson model.

C.1.1 Gaussian emission

From the Gaussian emission model as in subsection 4.1 it can be seen that the parameters are $\theta = (\hat{\mu}_0, \hat{p}, c, \sigma^2)$. The log-likelihood is then:

$$\begin{aligned}\log L(\theta; X_{0:T}, y_{0:T}) = \text{const.} &+ \sum_{v=1}^{M} \log\left[\hat{\mu}_0(X_0^v)\right] \\ &+ \sum_{t=1}^{T}\sum_{v=1}^{M} \log\left[\hat{p}(X_{t-1}^v, X_t^v)\right] \\ &+ \sum_{t=1}^{T}\sum_{f=1}^{M-1} \left\{-\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[y_t^f - c\left(X_t^f - X_t^{f+1}\right)\right]^2\right\}.\end{aligned}$$

Consider the EM scenario as explained in subsection 4.1, the aim is to maximize the expected log-likelihood, however this expectation is taken under the approximated smoothing distribution $\tilde{\pi}_{t|T}$ given the parameters θ'. To make the notation lighter the dependence of $\tilde{\pi}_{t|T}$ on θ' is dropped. The expected log-likelihood is:

$$\begin{aligned}Q_T(\theta, \theta') = &\sum_{v=1}^{M}\sum_{x\in\mathbb{X}} \log\left[\hat{\mu}_0(x)\right]\tilde{\pi}_{0|T}^v(x) + \sum_{t=1}^{T}\sum_{v=1}^{M}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} \log\left[\hat{p}(x,z)\right]\tilde{\pi}_{t-1,t|T}^v(x,z) \\ &+ \sum_{t=1}^{T}\sum_{v=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} \left\{-\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[y_t^f - c\cdot x - c\cdot z\right]^2\right\}\tilde{\pi}_{t|T}^{v,v+1}(x,z),\end{aligned}$$

where $\tilde{\pi}_{t|T}^{v,v+1}$ is the joint smoothing distribution of X_t^v, X_t^{v+1} given θ' and $\tilde{\pi}_{t-1,t|T}^v$ is the joint smoothing distribution of X_{t-1}^v, X_t^v given θ':

$$\tilde{\pi}_{t-1,t|T}^v(x,z) = \frac{p'(x,z)\tilde{\pi}_{t-1}^v(x)\tilde{\pi}_{t|T}^v(z)}{\sum_{\tilde{x}\in\mathbb{X}} p'(\tilde{x},z)\tilde{\pi}_{t-1}^v(\tilde{x})},$$

where p' is the current estimate of $\hat{p}$. Given the target function the next step is to compute the gradient of Q_T:

$$\begin{aligned}
\frac{\partial Q_T}{\partial \hat{\mu}_0(x)} &= \frac{1}{\hat{\mu}_0(x)} \sum_{v=1}^{M} \tilde{\pi}^v_{0|T}(x) - \frac{1}{\hat{\mu}_0(\tilde{x})} \sum_{v=1}^{M} \tilde{\pi}^v_{0|T}(\tilde{x});\\
\frac{\partial Q_T}{\partial \hat{p}(x,z)} &= \frac{1}{\hat{p}(x,z)} \sum_{t=1}^{T}\sum_{v=1}^{M} \tilde{\pi}^v_{t-1,t|T}(x,z) - \frac{1}{\hat{p}(x,\tilde{x})} \sum_{t=1}^{T}\sum_{v=1}^{M} \tilde{\pi}^v_{t-1,t|T}(x,\tilde{x});\\
\frac{\partial Q_T}{\partial c} &= \sum_{t=1}^{T}\sum_{f=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} \left\{ \frac{(x+z)}{\sigma^2}\left[y^f_t - c\cdot x - c\cdot z\right]\right\} \tilde{\pi}^{f,f+1}_{t|T}(x,z);\\
\frac{\partial Q_T}{\partial \sigma^2} &= \sum_{t=1}^{T}\sum_{f=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} \left[-\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(y^f_t - c\cdot x - c\cdot z\right)^2\right] \tilde{\pi}^{f,f+1}_{t|T}(x,z);
\end{aligned}$$

remark that deriving Q_T is not enough, all the constraint on the probability masses must be satisfied, i.e.:

$$\hat{\mu}_0(\tilde{x}) = 1 - \sum_{\tilde{z}\in\mathbb{X}} \hat{\mu}_0(\tilde{z}) \quad \text{and} \quad \hat{p}(x,\tilde{x}) = 1 - \sum_{\tilde{z}\in\mathbb{X}} \hat{p}(x,\tilde{z}).$$

The M-step is then obtained with $\nabla Q_T(\hat{\mu}_0, \hat{p}, c, \sigma^2) = 0$, precisely:

$$\begin{aligned}
\hat{\mu}_0(x) &= \frac{1}{M}\sum_{v=1}^{M} \tilde{\pi}^v_{0|T}(x);\\
\hat{p}(x,z) &= \frac{\sum_{t=1}^{T}\sum_{v=1}^{M} \tilde{\pi}^v_{t-1,t|T}(x,z)}{\sum_{t=1}^{T}\sum_{v=1}^{M} \tilde{\pi}^v_{t-1|T}(x)};\\
c &= \frac{\sum_{t=1}^{T}\sum_{f=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} y^f_t (x+z)\tilde{\pi}^{f,f+1}_{t|T}(x,z)}{\sum_{t=1}^{T}\sum_{f=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} (x+z)^2\tilde{\pi}^{f,f+1}_{t|T}(x,z)};\\
\sigma^2 &= \frac{1}{T(M-1)}\sum_{t=1}^{T}\sum_{f=1}^{M-1}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} [y_t - c(x+z)]^2 \tilde{\pi}^{f,f+1}_{t|T}(x,z).
\end{aligned}$$

C.1.2 Poisson emission

Consider the model described in subsection 4.2. A graphical representation is available in figure 1.

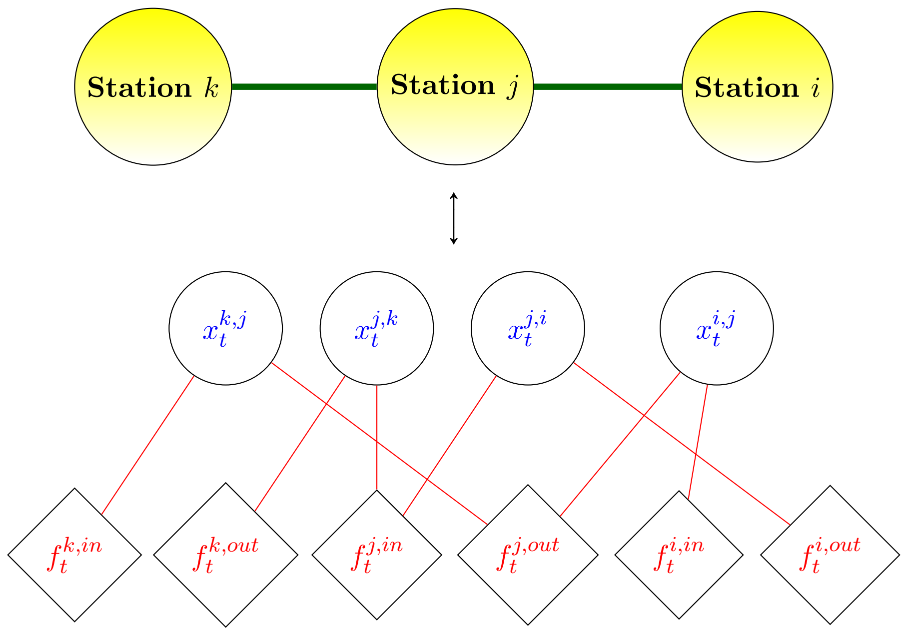

Figure 1: Top: a simple tube's network with three stations. Bottom: the corresponding factor graph build as explained in section 4.2.

The parameters of the model are $\theta = (\mu_0, p, \lambda)$. The log-likelihood of the model is:

$$\begin{aligned}\log L(\theta; X_{0:T}, y_{0:T}) = & \sum_{i,j=1}^{F} \log\left[\mu_0^{i,j}(X_0^{i,j})\right] \\ & + \sum_{t=1}^{T}\sum_{i,j=1}^{F} \log\left[p^{i,j}(X_{t-1}^{i,j}, X_t^{i,j})\right] \\ & + \sum_{t=1}^{T}\sum_{i=1}^{F} \left\{ y_t^{i,in}\log(\lambda^{i,in}) - \lambda^{i,in}\left(\sum_{j\in N(i)} X_t^{i,j}\right)\right\} \\ & + \sum_{t=1}^{T}\sum_{i=1}^{F} \left\{ y_t^{i,out}\log(\lambda^{i,out}) - \lambda^{i,out}\left(\sum_{j\in N(i)} X_t^{j,i}\right)\right\} \\ & + const.\end{aligned}$$

As for the previous section, in EM the expected log-likelihood must be maximise, where the expectation is taken under the approximated smoothing distribution $\tilde{\pi}_{t|T}$ given the current estimates of the parameters θ' (as for the previous subsection the dependence of the approximate smoothing on θ' is dropped). Precisely, the expected log-likelihood is:

$$
\begin{aligned}
Q_T(\theta,\theta') = & \sum_{i,j=1}^{F}\sum_{x\in\mathbb{X}} \log\left[\mu_0^{i,j}(x)\right]\tilde{\pi}_{0|T}^{i,j}(x) + \sum_{t=1}^{T}\sum_{v=1}^{M}\sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} \log\left[p^{i,j}(x,z)\right]\tilde{\pi}_{t-1,t|T}^{i,j}(x,z) \\
& + \sum_{t=1}^{T}\sum_{i=1}^{F}\left\{ y_t^{i,in}\log(\lambda^{i,in}) - \sum_{x\in\mathbb{X}^{i,N(i)}} \lambda^{i,in}\left(\sum_{j\in N(i)} x_t^{i,j}\right)\tilde{\pi}_{t|T}^{i,N(i)}(x^{i,N(i)})\right\} \\
& + \sum_{t=1}^{T}\sum_{i=1}^{F}\left\{ y_t^{i,out}\log(\lambda^{i,out}) - \sum_{x\in\mathbb{X}^{i,N(i)}} \lambda^{i,out}\left(\sum_{j\in N(i)} x_t^{j,i}\right)\tilde{\pi}_{t|T}^{N(i),i}(x^{N(i),i})\right\}.
\end{aligned}
$$

where $\tilde{\pi}_{t|T}^{i,N(i)}$ is the joint smoothing distribution of $(X_t^{i,j})_{j\in N(i)}$ and $\tilde{\pi}_{t-1,t|T}^{i,j}$ is the joint smoothing distribution of $X_{t-1}^{i,j}, X_t^{i,j}$. Given the target function the gradient of Q_T can be computed:

$$
\begin{aligned}
\frac{\partial Q_T}{\partial \mu_0^{i,j}(x)} &= \frac{1}{\mu_0^{i,j}(x)}\tilde{\pi}_{0|T}^{i,j}(x) - \frac{1}{\mu_0^{i,j}(\tilde{x})}\tilde{\pi}_{0|T}^{i,j}(\tilde{x}); \\
\frac{\partial Q_T}{\partial Q(\mu_0,p,\lambda)\partial p^{i,j}(x,z)} &= \frac{1}{p^{i,j}(x,z)}\sum_{t=1}^{T}\tilde{\pi}_{t-1,t|T}^{i,j}(x,z) - \frac{1}{p^{i,j}(x,\tilde{x})}\sum_{t=1}^{T}\tilde{\pi}_{t-1,t|T}^{i,j}(x,\tilde{x}); \\
\frac{\partial Q_T}{\partial \lambda^{i,in}} &= \sum_{t=1}^{T}\left\{\frac{y_t^{i,in}}{\lambda^{i,in}} - \sum_{x\in\mathbb{X}^{i,N(i)}}\left(\sum_{j\in N(i)} x_t^{i,j}\right)\tilde{\pi}_{t|T}^{i,N(i)}(x)\right\}; \\
\frac{\partial Q_T}{\partial \lambda^{i,out}} &= \sum_{t=1}^{T}\left\{\frac{y_t^{i,out}}{\lambda^{i,out}} - \sum_{x\in\mathbb{X}^{i,N(i)}}\left(\sum_{j\in N(i)} x_t^{j,i}\right)\tilde{\pi}_{t|T}^{N(i),i}(x)\right\}.
\end{aligned}
$$

remark that deriving Q_T is not enough, all the constraint on the probability masses must be satisfied, i.e.:

$$
\mu_0^{i,j}(\tilde{x}) = 1 - \sum_{\tilde{z}\in\mathbb{X}} {\mu^{i,j}}_0(\tilde{z}) \quad \text{and} \quad p^{i,j}(x,\tilde{x}) = 1 - \sum_{\tilde{z}\in\mathbb{X}} p^{i,j}(x,\tilde{z}).
$$

The M-step is then obtained by setting $\nabla Q_T(\theta) = 0$, precisely:

$$\mu_0^{i,j}(x) = \tilde{\pi}_{0|T}^{i,j}(x);$$

$$p^{i,j}(x,z) = \frac{\sum_{t=1}^{T} \tilde{\pi}_{t-1,t|T}^{i,j}(x,z)}{\sum_{t=1}^{T} \tilde{\pi}_{t-1|T}^{i,j}(x)};$$

$$\lambda^{i,in} = \frac{\sum_{t=1}^{T} y_t^{i,in}}{\sum_{x \in \mathbb{X}^{i,N(i)}} \left(\sum_{j \in N(i)} x_t^{i,j}\right) \tilde{\pi}_{t|T}^{i,N(i)}(x)};$$

$$\lambda^{i,out} = \frac{\sum_{t=1}^{T} y_t^{i,out}}{\sum_{x \in \mathbb{X}^{i,N(i)}} \left(\sum_{j \in N(i)} x_t^{j,i}\right) \tilde{\pi}_{t|T}^{N(i),i}(x)}.$$

C.2 Analyzing traffic flows on the London Underground: supplementary

This section provides some additional details about the experiments in section 4.2. Subsection C.2.1 describes the training procedure of the LSTM. Subsection C.2.2 provides complete plots about the experiments.

C.2.1 LSTM details

The dataset consists of inflow-outflow per station over the analysed week. The training set is normalized between 0 and 1. We want to use an LSTM to map inflow-outflow over the stations at time t onto the inflow-outflow over the stations at time $t+1$, i.e. train an LSTM to performs 1-step-ahead prediction. The LSTM is trained on Keras (now Tensorflow v2) with architecture given by:

- an input of dimension 40: the inflow-outflow for all the twenty stations, i.e. a 40-dimensional;
- an LSTM cell with 10 as dimensionality of the output space (the hidden and cell states are vectors with dimension 10);
- an additional dense layer with output dimension 40, to map the LSTM output onto the inflow-outflow for all the twenty stations at the next time step.

More complicated (and simpler) architecture have been tested, but the chosen one was performing the best. The weights of the neural network are trained with ADAM optimizer over 100 epochs. The experiment is repeated 100 times with different seeds to estimate the standard deviation of the RMSE.

C.2.2 Plots on all the stations

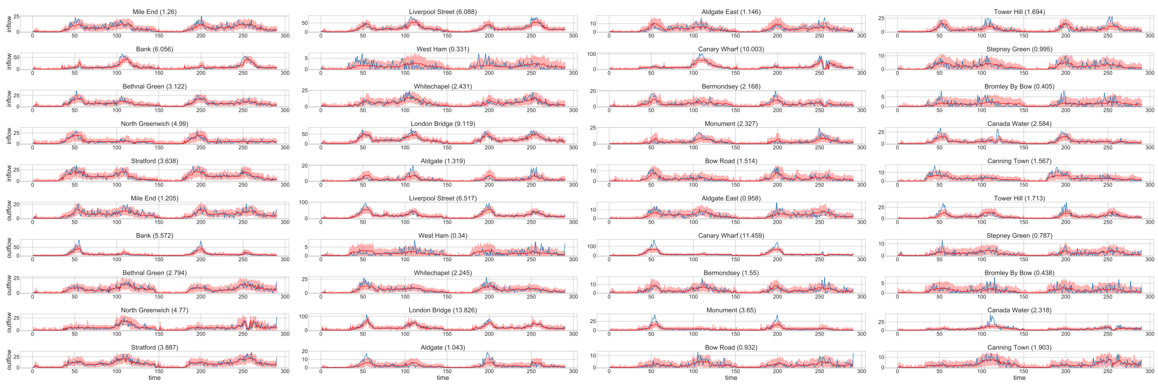

Figure 2: Posterior predictive performance with Graph Filter on all the considered stations. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: one step-ahead posterior predictive mean (solid red line) and 0.95 credible intervals (red bands) using the Graph Filter-Smoother.

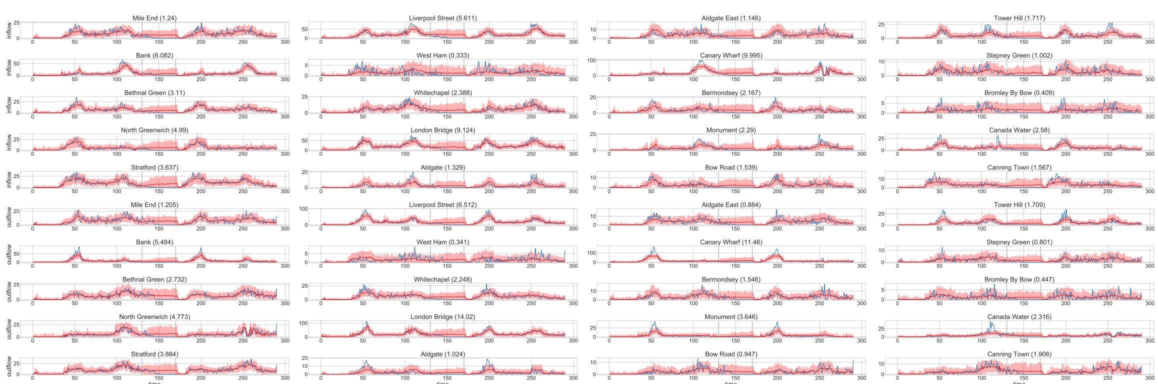

Figure 3: Posterior predictive performance with Graph Filter on all the considered stations with missing data in a quiet period. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: one step-ahead posterior predictive mean (solid red line) and 0.95 credible intervals (red bands) using the Graph Filter-Smoother. Grey vertical dashed lines show the start and the end of the missing data window.

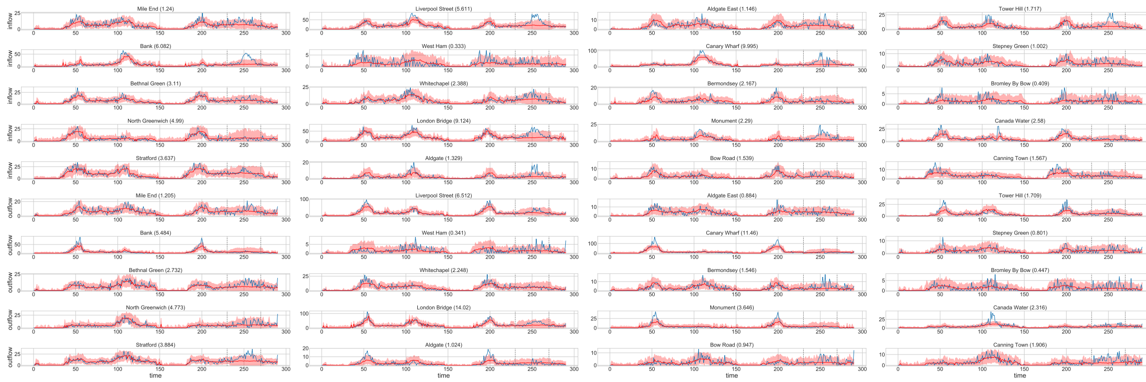

Figure 4: Posterior predictive performance with Graph Filter on all the considered stations with missing data in a quiet period. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: one step-ahead posterior predictive mean (solid red line) and 0.95 credible intervals (red bands) using the Graph Filter-Smoother. Grey vertical dashed lines show the start and the end of the missing data window.

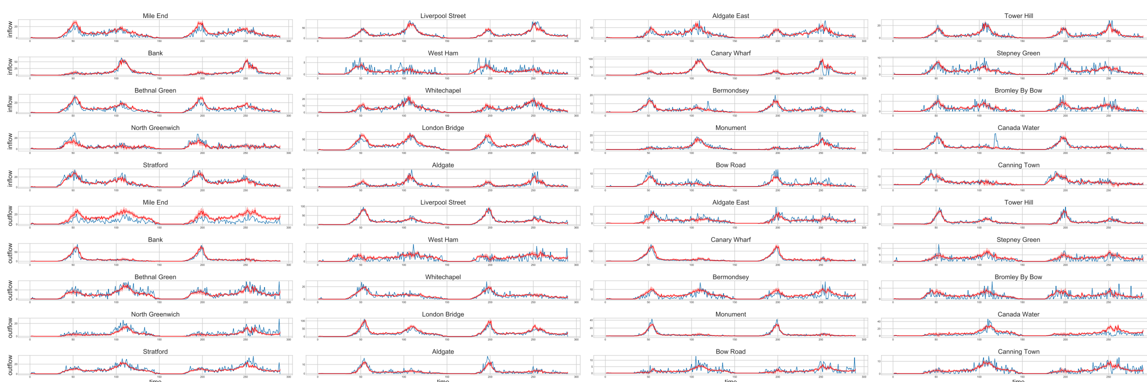

Figure 5: LSTM performance on all the considered stations with missing data in a quiet period. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: one-step-ahead prediction with the LSTM, per time step, a sample of size 100 is built over different training of the LSTM with solid red lines showing the mean and red bands showing the region between the 0.025 and the 0.975 quantile.

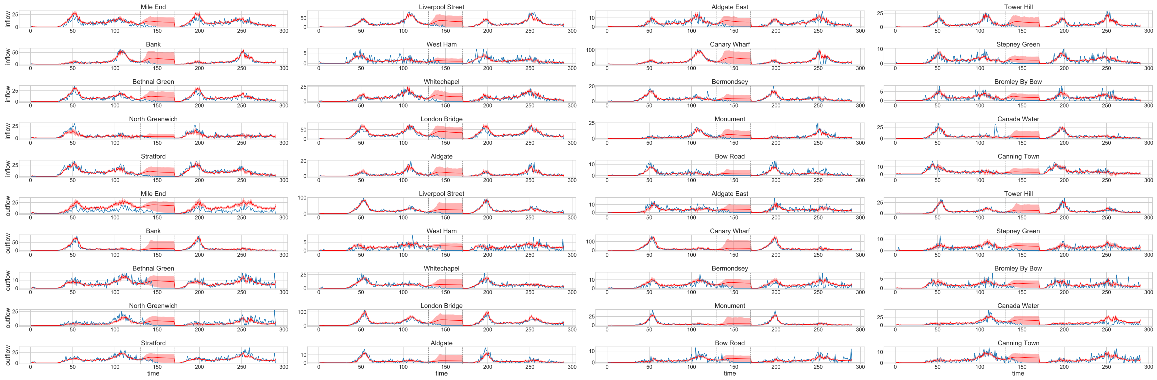

Figure 6: LSTM performance on all the considered stations with missing data in a quiet period. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: multi-step-ahead prediction with the LSTM, per time step, a sample of size 100 is built over different training of the LSTM with solid red lines showing the mean and red bands showing the region between the 0.025 and the 0.975 quantile.

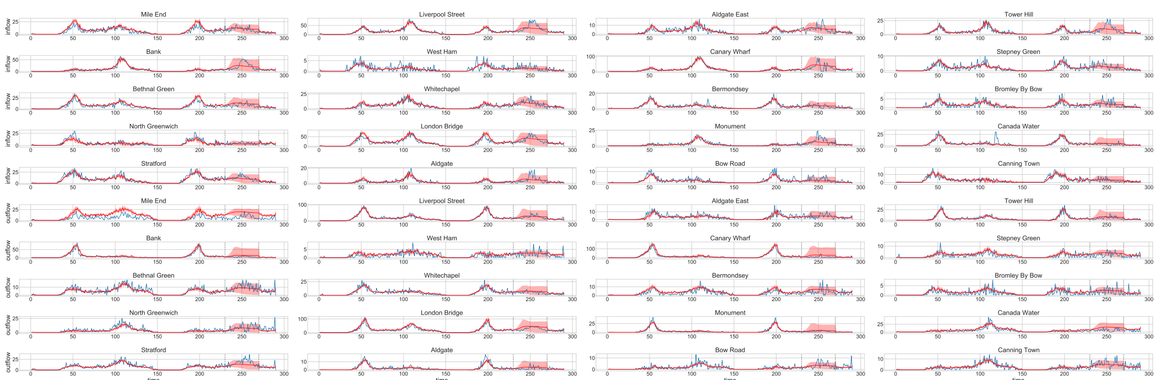

Figure 7: LSTM performance on all the considered stations with missing data in a quiet period. The first five rows show the inflow, the remaining five rows show the outflow. In blue: the inflow and outflow per station. In red: multi-step-ahead prediction with the LSTM, per time step, a sample of size 100 is built over different training of the LSTM with solid red lines showing the mean and red bands showing the region between the 0.025 and the 0.975 quantile.