Journal of Machine Learning Research 1 (2023) 1-48 Submitted 4/00; Published 10/00

Iterated Block Particle Filter for High-dimensional Parameter Learning: Beating the Curse of Dimensionality

Ning Ning PATNING@TAMU.EDU
Department of Statistics
Texas A&M University
College Station, TX 77843, USA

Edward L. Ionides IONIDES@UMICH.EDU
Department of Statistics
University of Michigan
Ann Arbor, MI 48109, USA

Editor:

Abstract

Parameter learning for high-dimensional, partially observed, and nonlinear stochastic processes is a methodological challenge. Spatiotemporal disease transmission systems provide examples of such processes giving rise to open inference problems. We propose the iterated block particle filter (IBPF) algorithm for learning high-dimensional parameters over graphical state space models with general state spaces, measures, transition densities and graph structure. Theoretical performance guarantees are obtained on beating the curse of dimensionality (COD), algorithm convergence, and likelihood maximization. Experiments on a highly nonlinear and non-Gaussian spatiotemporal model for measles transmission reveal that the iterated ensemble Kalman filter algorithm (Li et al., 2020) is ineffective and the iterated filtering algorithm (Ionides et al., 2015) suffers from the COD, while our IBPF algorithm beats COD consistently across various experiments with different metrics.

Keywords: Sequential Monte Carlo, Parameter learning, Spatiotemporal inference, Curse of dimensionality, Graphical state space models

1. Introduction

We firstly give the background and motivation in Section 1.1 and then state our contributions in Section 1.2, followed by the organization of the paper in Section 1.3.

1.1 Background and motivation

Spatiotemporal data arises when measurements are made through time at a collection of spatial locations. Spatiotemporal inference for epidemiological and ecological systems is arguably the last remaining open problem from the six challenges in time series analysis of nonlinear systems posed by Bjørnstad and Grenfell (2001). A disease transmission system is stochastic and imperfectly observable, thus it is commonly modeled by a partially observed Markov process (POMP), otherwise known as state space model or hidden Markov model, which consists of a latent Markov process representing the time evolution of the system and a measurement process by which stochastic observations of this latent process are collected

at specified time points. Particle filters (PFs), also known as sequential Monte Carlo (SMC) methods, are recursive algorithms that enable estimation of the likelihood of observed data and the conditional distribution of the latent process given data from a POMP model (Doucet et al., 2001; Cappé et al., 2007; Doucet and Johansen, 2009).

For the purpose of parameter learning, two iterated filtering (IF) approaches were developed, Ionides et al. (2006) and its subsequent improvement Ionides et al. (2015) referred to as IF1 and IF2 algorithms respectively, which coerce a particle filter into maximizing the likelihood function for unknown parameters. PF methods and the parameter learning algorithms based on them (such as IF2) are capable of handling highly nonlinear latent processes (King et al., 2008; Ionides et al., 2011). In epidemiological applications, IF1 and IF2 can considerably increase the accuracy of outbreak predictions while also allowing models whose structures reflect different underlying assumptions to be compared (Dobson, 2014). Unfortunately, PF suffers from rapid deterioration in performance as the model dimension increases (Bengtsson et al., 2008; Snyder et al., 2008). Rebeschini and Van Handel (2015) rigorously showed that PF suffers the curse of dimensionality (COD) phenomenon, which says that the upper bound of the algorithmic filter error is exponential in the dimension of the state space of the underlying model. As expected, PF-based parameter learning algorithms suffer from the COD, limiting their applicability in high-dimensional problems.

The ensemble Kalman filter (EnKF) is a recursive filter suitable for problems with a large number of variables. EnKF represents uncertainty in the latent state space using a finite collection of state values, and we refer to these ensemble members as particles by analogy with PF. EnKF differs from PF by adopting a Gaussian approximation in the rule used to update the particles when filtering. EnKF methods have been used for geophysical models in data assimilation due to their computational scalability to high dimensions (Houtekamer and Mitchell, 2001; Evensen, 1994; Katzfuss et al., 2020). For the parameter learning purpose, the iterated EnKF (IEnKF) algorithm extends the IF2 approach for parameter estimation by replacing a PF with an EnKF; it propagates the ensemble members by simulation from the dynamic model and then updates the ensemble to assimilate observations using a Gaussian-inspired rule (Li et al., 2020). Given that EnKF relies on locally linear and Gaussian approximations, it can be ineffective for highly nonlinear and non-Gaussian systems (Ades and Van Leeuwen, 2015; Lei et al., 2010; Miller et al., 1999). Unsurprisingly, the corresponding unsuitability carries to EnKF-based parameter learning algorithms (such as IEnKF).

Block sampling strategies for PF were proposed with temporal blocks in (Doucet et al., 2006). Rebeschini and Van Handel (2015) investigated spatial blocks and proved that a block PF (BPF) beats the COD under certain conditions. However, the beautiful work of Rebeschini and Van Handel (2015) is theoretical in nature and was not anticipated to be applicable to real high-dimensional problems (page 2812 therein). In recent years, many efforts have been undertaken to develop practical methods for these problems by developing the "block" concept, which include, but are not limited to, the following: Johansen (2015) proposed a method for systems identification based on both the block sampling idea and the annealed importance sampling approach; Singh et al. (2017) applied the particle Gibbs algorithm inside a generic Gibbs sampler over temporal blocks to handle long time series; Park and Ionides (2020) proposed a twisted particle filter model (Whiteley and Lee, 2014) with iterated auxiliary PFs (Guarniero et al., 2017) to infer on moderately high-dimensional

spatiotemporal models where its particle filtering corresponds to an adapted version of the block sampling method; Goldman and Singh (2021) proposed a blocked sampling scheme for latent state inference in high-dimensional state space models; Ionides et al. (2021) proposed the bagged filter for partially observed interacting systems and showed that BPF can perform well on practical scientific models.

So far there is no high-dimensional parameter learning approach that is generically applicable over partially observed, highly nonlinear, stochastic spatiotemporal processes. The goal of this paper is to develop such an algorithm that is generically applicable and able to beat the COD. Unlike the limited theoretical understanding of EnKF and hence IEnKF, the proposed algorithm has rigorous convergence analysis with a precise error bound.

1.2 Our contributions

In this paper, we propose the iterated BPF (IBPF) algorithm. The contributions of the paper fall into four distinct categories:

1. **General graphical model structure.** In this paper, we consider a general graphical POMP model having general state spaces and measures, general transition densities, and a general graph structure. Specifically, the latent state $(X_n)_{n\geq 0}$, the observation sequence $(Y_n)_{n\geq 1}$ that is conditionally independent given $(X_n)_{n\geq 1}$, and the auxiliary Markov chain $(\Theta_n)_{n\geq 0}$ for parameter learning purpose, have their own state spaces which are all Polish spaces endowed with their own general reference measures. The transition densities of X_n, Y_n, and Θ_n are all time-inhomogeneous and in the general form where only standard conditions are required. The state of (X_n, Y_n, Θ_n) at each time n is a random field $(X_n^v, Y_n^v, \Theta_n^v)_{v\in V}$ indexed by a general finite undirected graph with V being the set of vertices. We consider the graph having a partition, which is a collection of nonoverlapping blocks whose union is the graph.
2. **Innovative methodology.** The BPF algorithm in Rebeschini and Van Handel (2015) is the first PF algorithm that has rigorous guarantees on beating COD, however "it is far from clear whether this simple algorithm is of immediate practical utility in the most complex real-world applications" (page 2812 therein). IBPF embeds BPF on an extended state space in an iterative scheme that constructs parameter values approaching the maximum of the likelihood function. It inherits from BPF the property that only observations in each block are used to update predictions, which is the key to scalability. When all vertices are in a single block, IBPF is in nature IF2; when there is inference on the latent process, all vertices are in a single block, and parameters are known, IBPF is PF; when there is inference on the latent process and parameters are known, IBPF is BPF; when all vertices are in a single block and there is no particle involved, IBPF is the iterated importance sampling.
3. **Theoretical contribution.** Our IBPF algorithm has rigorous performance guarantees in terms of graph dimensions, time steps, algorithm convergence, and likelihood maximization. In Theorem 1, under standard assumptions, we rigorously show that the algorithmic error can be bounded using the dimension of a local block, uniformly both in time and in the model dimension. Our result generalizes that of Rebeschini and Van Handel (2015) to the time-inhomogeneous setting. They introduced the

mathematical machinery of a local particle filtering algorithm in high dimension that had not previously been applied in the study of nonlinear filtering, however being time-homogeneous is a limitation in practical applications. Furthermore, with our precise bound, we provide exact sufficient conditions needed and reveal the influences of crucial quantities (such as the range of interacting neighborhoods and the maximal number of blocks of interaction) on the error bound.

4. **Excellent performance.** For spatiotemporal modeling, it is appropriate and sometimes necessary to have some parameters vary across locations, for instance, for measles transmission modeling the basic reproduction number regarding the epidemic transmission speed. To demonstrate how to use IBPF and compare its performance with IF2 and IEnKF, we generalize the spatiotemporal model for measles transmission covered in Park and Ionides (2020) and Ionides et al. (2021), by allowing location-specific parameters. Extensive experiments reveal that IF2 does not scale well which confirms the phenomenon that PF does not scale well with dimensions (e.g. Bengtsson et al. (2008)), and IEnKF performs very badly for highly nonlinear and non-Gaussian problems confirming the same phenomenon of EnKF (e.g. Ades and Van Leeuwen (2015)). In all experiments, IBPF is able to find parameter values with a likelihood on or better than that of the true parameters consistently. The performances of IBPF with respect to iterations and block sizes are further examined, and confirm our theoretical findings.

1.3 Organization of the paper

The rest of the paper proceeds as follows. In Section 2, we set up our general model and provide necessary definitions. In Section 3, we provide the main results of this paper, by firstly describing IBPF in Section 3.1, conducting preliminary algorithmic analyses in Section 3.2, providing theoretical results in Section 3.3, and then investigating likelihood convergence in Section 3.4. In Section 4, we conduct performance analysis through a generalized spatiotemperal model for measles covered in Section 4.1, over the dataset covered in Section 4.2, and evaluate the performance of IBPF, IF2 and IEnKF in Section 4.3. We conclude with discussion and extensions in Section 5. In Appendix A, we provide the algorithms of IF2 and IEnKF for comparison. In Appendix B, existing technical results are provided which are needed for rigorous proofs following. In Appendix C, we prepare for mathematical derivations by properly defining filtering and correlation measurement quantities. We defer all the lemmas and propositions in bounding the bias and variance of the algorithmic error, to Appendix D and Appendix E, respectively. In Appendix F, we provide a rigorous proof of our main theorem. Original parameter learning results without rescaling to account for spatial and time dimensions, are provided in Appendix H.

2. Model and analysis setups

In this section, we firstly describe the extended POMP model $(X_n, Y_n\,; \Theta_n)$ on graph G in Section 2.1, and then the partition $\mathcal{K}$ that separates G into nonoverlapping blocks in Section 2.2, followed by the global and local metrics necessary to conduct analysis in Section 2.3.

2.1 Extended POMP model on graph

A general POMP model is a Markov chain (X_n, Y_n), where $(X_n)_{n\geq 0}$ is a Markov chain in a Polish state space $\mathbb{X}$, while $(Y_n)_{n\geq 1}$ is conditionally independent given $(X_n)_{n\geq 1}$ in a Polish state space $\mathbb{Y}$. Here, X_n is not directly observable while Y_n serves as its partial and noisy observations made at time n. Define the reference measure of X_n (resp. Y_n) on its state space $\mathbb{X}$ (resp. $\mathbb{Y}$) as ψ (resp. ϕ). Suppose that there is an unknown auxiliary Markov chain Θ_n for parameter learning, which has its own Polish state space $\mathbb{\Theta}$ and reference measure λ. For $n \geq 1$, with respect to ψ we define the emission density (or measurement density) of X_n as $f_{X_n|X_{n-1}}(x_n|x_{n-1}\,;\theta_n)$, with respect to ϕ we define the transition density of Y_n as $f_{Y_n|X_n}(y_n|x_n\,;\theta_n)$, and with respect to λ we define the transition density of Θ_n as $f_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1}\,;\sigma)$ where σ is a nonnegative constant. That is, with our extended Markov chain model $(X_n, Y_n\,;\Theta_n)$, its transition probability is given by

$$\begin{aligned}&P(A|(x_{n-1}, y_{n-1})\,;\theta_n)\\&= \int \mathbb{1}_A(x_n, y_n) f_{X_n|X_{n-1}}(x_n|x_{n-1}\,;\theta_n) f_{Y_n|X_n}(y_n|x_n\,;\theta_n)\psi(dx_n)\phi(dy_n).\end{aligned}$$

The state of (X_n, Y_n, Θ_n) at each time n is a random field $(X_n^v, Y_n^v, \Theta_n^v)_{v\in V}$ indexed by a finite undirected graph $G = (V, E)$, where V stands for the set of vertices and E stands for the set of edges. The graph describes the location relationship of data and the spatial degrees of freedom of the model. Based on the network structure, the state spaces $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{\Theta}$ can be written as the product forms

$$\mathbb{X} = \prod_{v\in V} \mathbb{X}^v, \qquad \mathbb{Y} = \prod_{v\in V} \mathbb{Y}^v, \quad \text{and} \quad \mathbb{\Theta} = \prod_{v\in V} \mathbb{\Theta}^v.$$

Define the reference measure of X_n^v on its state space $\mathbb{X}^v$ as ψ^v. Define the reference measure of Y_n^v on its state space $\mathbb{Y}^v$ as ϕ^v. Define the reference measure of Θ_n^v on its state space $\mathbb{\Theta}^v$ as λ^v. Similarly, based on the network structure, we have the following product-formed expressions:

$$\psi = \prod_{v\in V} \psi^v, \qquad \phi = \prod_{v\in V} \phi^v, \qquad \lambda = \prod_{v\in V} \lambda^v.$$

With respect to ψ^v we define the transition density of X_n^v as $f_{X_n^v|X_{n-1}}$, with respect to ϕ^v we define the transition density of Y_n^v as $f_{Y_n^v|X_n^v}$, and with respect to λ^v we define the transition density of Θ_n^v as $f_{\Theta_n^v|\Theta_{n-1}}$. Similarly, based on the network structure, we have the following product-formed expressions:

$$\begin{aligned} f_{X_n|X_{n-1}}(x_n|x_{n-1}\,;\theta_n) &= \prod_{v\in V} f_{X_n^v|X_{n-1}}(x_n^v|x_{n-1}\,;\theta_n^v),\\ f_{Y_n|X_n}(y_n|x_n\,;\theta_n) &= \prod_{v\in V} f_{Y_n^v|X_n^v}(y_n^v|x_n^v\,;\theta_n^v),\\ f_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1}\,;\sigma) &= \prod_{v\in V} f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v|\theta_{n-1}\,;\sigma). \end{aligned} \tag{1}$$

2.2 Partition of the graph

We consider a partition $\mathcal{K}$ that partitions V into nonoverlapping blocks, i.e.,

$$V = \bigcup_{K\in\mathcal{K}} K, \qquad K \cap K' = \emptyset \text{ for } K \neq K',\ K, K' \in \mathcal{K}.$$

Based on the partition, we can write

$$\Xi = (\Xi^K)_{K\in\mathcal{K}} = (\Xi^v)_{v\in V}, \qquad \Xi^W := (\Xi^v)_{v\in W} \text{ for } \forall W \subseteq V,$$

where Ξ can be X_n, Y_n or Θ_n, as well as the associated state space $\mathbb{X}$, $\mathbb{Y}$, or $\mathbb{\Theta}$. For any set $W \subseteq V$, we use $(\mathbb{X}\times\mathbb{\Theta})^W$ and $\mathbb{X}^W \times \mathbb{\Theta}^W$ interchangeably and define

$$\psi^W(dx_n^W) := \prod_{v\in W} \psi^v(dx_n^v) \quad \text{and} \quad \lambda^W(d\theta_n^W) := \prod_{v\in W} \lambda^v(d\theta_n^v). \tag{2}$$

We define the distance d as the length of the shortest path in the graph G connecting two vertices, based on which we define for each vertex $v \in V$ the r-neighborhood $N(v)$ as

$$N(v) := \{v' \in V : d(v, v') \leq r\}.$$

For integers $0 \leq m \leq n$, denote

$$\Xi_{m:n} := \{\Xi_m, \Xi_{m+1}, \cdots, \Xi_n\},$$

where Ξ can be X, Y, or Θ. Suppose that, for $n \geq 1$, the conditional distribution of X_n^v given $X_{0:n-1}$ depends on $X_{n-1}^{N(v)}$ only and then we have

$$f_{X_n^v|X_{n-1}}(x_n^v|x_{n-1}\,;\theta_n^v) = f_{X_n^v|X_{n-1}}(x_n^v|\overline{x}_{n-1}\,;\theta_n^v),$$

whenever $x_{n-1}^{N(v)} = \overline{x}_{n-1}^{N(v)}$ where $x, \overline{x} \in \mathbb{X}$ and $x \neq \overline{x}$. That is, if x_{n-1} and $\overline{x}_{n-1}$ coincide on the neighbouring vertices of v, then their associated transition densities are the same. Similarly suppose that, for $n \geq 1$, the conditional distribution of Θ_n^v given $\Theta_{0:n-1}$ depends on $\Theta_{n-1}^{N(v)}$ only and then we have

$$f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v|\theta_{n-1}\,;\sigma) = f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v|\overline{\theta}_{n-1}\,;\sigma),$$

whenever $\theta_{n-1}^{N(v)} = \overline{\theta}_{n-1}^{N(v)}$ where $\theta, \overline{\theta} \in \mathbb{\Theta}$ and $\theta \neq \overline{\theta}$. An illustration of the dependence with $r = 1$ is provided in Figure 1.

For any sets $W, W' \subseteq V$, define

$$d(W, W') := \min_{v\in W} \min_{v'\in W'} d(v, v'), \tag{3}$$

based on which we can define the collection of blocks that interact with any block $K \in \mathcal{K}$ in one step of the dynamics as

$$N(K) := \{K' \in \mathcal{K} : d(K, K') \leq r\}. \tag{4}$$

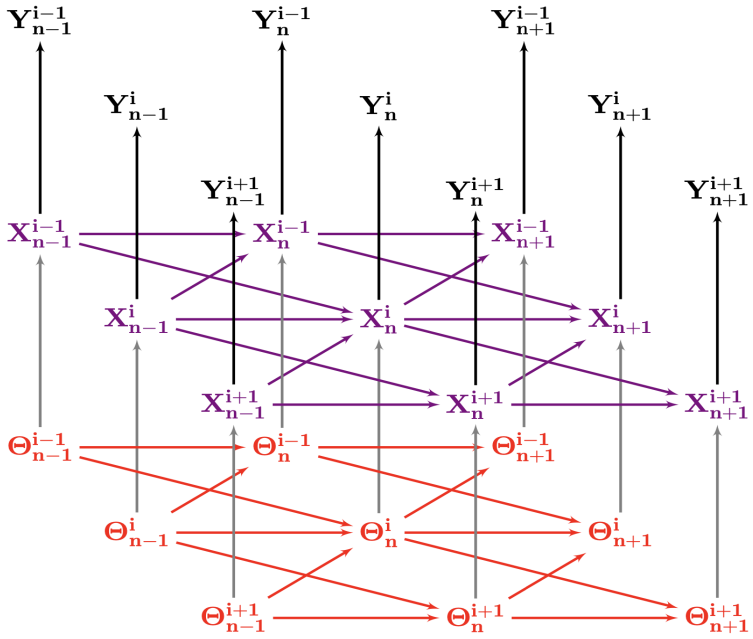

Figure 1: Illustration of 1-neighborhood dependence.

Given a set $W \subseteq V$, denote the r-inner boundary of W as the subset of vertices in W that can interact with vertices outside W,

$$\partial W := \{v \in W : N(v) \nsubseteq W\}, \tag{5}$$

and denote the interior of W as

$$\mathrm{int}(W) := W \backslash \partial W. \tag{6}$$

We now denote some quantities which will be used frequently throughout the paper: the maximal size of one single block in the partition $\mathcal{K}$

$$|\mathcal{K}|_\infty := \max_{K \in \mathcal{K}} \mathrm{card}(K), \tag{7}$$

where $\mathrm{card}(K)$ denotes the cardinality of K; the maximal number of vertices that interact with one single vertex in its r-neighborhood in one step of the dynamics

$$\Delta := \max_{v \in V} \mathrm{card}\{v' \in V : d(v, v') \leq r\}; \tag{8}$$

the maximal number of blocks that interact with one single block in one step of the dynamics

$$\Delta_{\mathcal{K}} := \max_{K \in \mathcal{K}} \mathrm{card}\{K' \in \mathcal{K} : d(K, K') \leq r\}. \tag{9}$$

2.3 Metrics

We assume that the process $(X, Y;\ \Theta)$ is realized on its canonical probability space; denote $\mathbf{P}$ and $\mathbf{E}$ as the probability measure and expectation on that space, respectively. We use the functional analytic notation $\rho(g)$ for the integral of a measurable function g with respect to the measure ρ (provided this integral exists),

$$\rho(g) := \int g d\rho = \int g(x) d\rho(x) = \int g(x)\rho(dx).$$

Between two random measures ρ and ρ' on space $\mathbb{S}$, we define the distance

$$\left|\left|\left|\rho - \rho'\right|\right|\right| := \sup_{g\in\mathcal{S}:|g|\le 1} \left[\mathbf{E}|\rho(g) - \rho'(g)|^2\right]^{1/2}, \tag{10}$$

where $\mathcal{S}$ denotes the class of measurable functions $g : \mathbb{S} \to \mathbb{R}$, and define the local distance, for $\kappa \subseteq V$,

$$\left|\left|\left|\rho - \rho'\right|\right|\right|_{\kappa} := \sup_{g\in\mathcal{S}^{\kappa}:|g|\le 1} \left[\mathbf{E}|\rho(g) - \rho'(g)|^2\right]^{1/2}. \tag{11}$$

Here, $\mathcal{S}^{\kappa}$ denotes the class of measurable functions $g : \mathbb{S} \to \mathbb{R}$ such that $g(x) = g(\overline{x})$ whenever $x^{\kappa} = \overline{x}^{\kappa}$. That is, $\mathcal{S}^{\kappa}$ is the class of measurable functions such that when x and $\overline{x}$ coincide on the set κ then their associated function values are the same. Similarly, we define the total variation distance between two probability measures ρ and ρ' on $\mathbb{S}$

$$\|\rho - \rho'\| := \sup_{g\in\mathcal{S}:|g|\le 1} |\rho(g) - \rho'(g)|, \tag{12}$$

and define the local total variation distance, for $\kappa \subseteq V$,

$$\|\rho - \rho'\|_{\kappa} := \sup_{g\in\mathcal{S}^{\kappa}:|g|\le 1} |\rho(g) - \rho'(g)|. \tag{13}$$

3. Main results

In this section, we describe our IBPF algorithm for parameter learning over general graphical POMP models in Section 3.1, conduct its preliminary algorithmic analyses in Section 3.2, establish its theoretical guarantees on algorithm performances and convergences in Section 3.3, and then investigate maximum likelihood estimates (MLEs) in Section 3.4.

3.1 Algorithm

We propose the IBPF algorithm in Algorithm 1. For notational convenience, we set

$$1{:}N := \{1, 2, \ldots, N\} \qquad \text{for } N \in \mathbb{N}$$

throughout the paper. In Algorithm 1, $\Theta_{n,j}^{F,m}$ (resp. $X_{n,j}^{F,m}$) is the j-th particle in the Monte Carlo representation of the m-th iteration of a filtering recursion at time n, where this filtering recursion is coupled with a prediction recursion represented by $\Theta_{n,j}^{P,m}$ (resp. $X_{n,j}^{P,m}$). The IBPF algorithm allows users to infer initial values of latent states by incorporating initial values into the parameter set. That is, let $\theta = (\theta_{\text{init}}, \theta_{\text{dynamic}})$ where θ_{dynamic} stands

for the parameters affecting $f_{X_n|X_{n-1}}$ and $f_{Y_n|X_n}$ for some or all $n \in 1:N$. In this case, the initial density $f_{X_0}(x_0;\theta)$ is a Dirac mass function at $x_0 = \theta_{\text{init}}$. This case is common for scientific modeling and also is the situation addressed by our theory. Numerical experiments on learning initial values can be seen in Section 4.3.

Algorithm 1 (The IBPF algorithm)

Initial value function $f_{X_0}(\theta)$
Simulator for $f_{X_n|X_{n-1}}(x_n \mid x_{n-1}\,;\theta)$, $n \in 1{:}N$
Evaluator for $f_{Y_n|X_n}(y_n \mid x_n\,;\theta)$, $n \in 1{:}N$
Data, $y_{1:N}$
Number of iterations, M
Number of particles, J
Partition, $\mathcal{K}$
Initial parameter swarm, $\{\Theta_j^0,\ j \in 1{:}J\}$
Perturbation density, $f_{\Theta_n|\Theta_{n-1}}(\theta \mid \vartheta\,;\sigma)$, $n \in 1{:}N$
Perturbation sequence, $\sigma_{1:M}$

Output: Final parameter swarm, $\{\Theta_j^M,\ j \in 1{:}J\}$ and log-likelihood $\widehat{l}_b$

For m in $1:M$
  Set $\Theta_{0,j}^{F,m} = \Theta_j^{m-1}$ for $j \in 1{:}J$
  Set $X_{0,j}^{F,m} = f_{X_0}(\Theta_{0,j}^{F,m})$ for $j \in 1{:}J$
  For n in $1:N$
    Draw $\Theta_{n,j}^{P,m} \sim f_{\Theta_n|\Theta_{n-1}}(\theta_n \mid \Theta_{n-1,j}^{F,m}\,;\sigma_m)$ for $j \in 1{:}J$
    Draw $X_{n,j}^{P,m} \sim f_{X_n|X_{n-1}}(x_n \mid X_{n-1,j}^{F,m}\,;\Theta_{n,j}^{P,m})$ for $j \in 1{:}J$
    For $K \in \mathcal{K}$
      Compute $w_{n,j}^{K,m} = \prod_{v\in K} f_{Y_n^v|X_n^v}(y_n^v \mid X_{n,j}^{v,P,m}\,;\Theta_{n,j}^{v,P,m})$ for $j \in 1{:}J$
      Draw $s_{1:J}^{K,m}$ with $\text{Prob}(s_j^{K,m} = i) = w_{n,i}^{K,m} \Big/ \sum_{j=1}^J w_{n,j}^{K,m}$
    End For
    Set $X_{n,j}^{F,m} = (X_{n,j}^{K,F,m})_{K\in\mathcal{K}}$ where $X_{n,j}^{K,F,m} = X_{n,s_j^{K,m}}^{K,P,m}$ for $j \in 1{:}J$
    Set $\Theta_{n,j}^{F,m} = (\Theta_{n,j}^{K,F,m})_{K\in\mathcal{K}}$ where $\Theta_{n,j}^{K,F,m} = \Theta_{n,s_j^{K,m}}^{K,P,m}$ for $j \in 1{:}J$
  End For
  Set $\Theta_j^m = (\Theta_{n,j}^{F,m})$ for $j \in 1{:}J$
End For
Set $\widehat{l}_b = \sum_{n=1}^N \sum_{K\in\mathcal{K}} \log(\frac{1}{J}\sum_{j=1}^J w_{n,j}^{K,M})$

3.2 Preliminary analysis

Inspecting the IBPF pseudocode (Algorithm 1), we can see that the same set of observations $Y_1, \ldots, Y_n$ is used in each of the M iterations. Let us first focus on one of the M iterations, say $m = 1$, and ignore the m superscript/subscript.

Given the observations $Y_1, \ldots, Y_n$, we aim to approximate the joint nonlinear filter, for $n \geq 1$,

$$\pi_n(A) = \pi_n(A_x \times A_\theta) = \mathbf{P}[X_n \in A_x, \Theta_n \in A_\theta \mid Y_1, \ldots, Y_n].$$

The filter π_n for $n \geq 1$ that we also call the true filter to differentiate with the IBPF approximated filter, can be expressed in a recursive way

$$\pi_n = \mathsf{F}_n \pi_{n-1}, \qquad \pi_0 = \delta(x, \theta) = \delta_x \delta_\theta, \tag{14}$$

where δ_x stands for the point mass at x, with

$$\mathsf{F}_n = \mathsf{C}_n \mathsf{P}_n \tag{15}$$

evolving as follows:

$$\pi_{n-1} \xrightarrow{\text{prediction}} \pi_{n|n-1} = \mathsf{P}_n \pi_{n-1} \xrightarrow{\text{correction}} \pi_n = \mathsf{C}_n \pi_{n|n-1}$$

where P_n is defined as the prediction operator

$$\begin{aligned}(\mathsf{P}_n \rho)(g) = \int g(x_n, \theta_n) f_{X_n|X_{n-1}}(x_n \mid x_{n-1}\,; \theta_n) f_{\Theta_n|\Theta_{n-1}}(\theta_n \mid \theta_{n-1}\,; \sigma) \\ \times \psi(dx_n) \lambda(d\theta_n) \rho(dx_{n-1}, d\theta_{n-1}),\end{aligned} \tag{16}$$

and C_n is defined as the correction operator

$$(\mathsf{C}_n \rho)(g) = \frac{\int g(x_n, \theta_n) f_{Y_n|X_n}(Y_n \mid x_n\,; \theta_n) \rho(dx_n, d\theta_n)}{\int f_{Y_n|X_n}(Y_n \mid x_n\,; \theta_n) \rho(dx_n, d\theta_n)}, \tag{17}$$

for any probability measure ρ on $\mathbb{X} \times \Theta$.

To facilitate analysis, we define an intermediate filter $\widetilde{\pi}_n$ which can be expressed in a recursive way

$$\widetilde{\pi}_n = \widetilde{\mathsf{F}}_n \widetilde{\pi}_{n-1}, \qquad \widetilde{\pi}_0 = \delta_x \delta_\theta, \tag{18}$$

with

$$\widetilde{\mathsf{F}}_n = \mathsf{C}_n \mathsf{B} \mathsf{P}_n \tag{19}$$

evolving as follows:

$$\widetilde{\pi}_{n-1} \xrightarrow{\text{prediction}} \widetilde{\pi}_{n|n-1} = \mathsf{P}_n \widetilde{\pi}_{n-1} \xrightarrow[\text{correction}]{\text{blocking}} \widetilde{\pi}_n = \mathsf{C}_n \mathsf{B} \widetilde{\pi}_{n|n-1},$$

where, for any measure ρ on $\mathbb{X} \times \Theta$, B is defined as the blocking operator

$$\mathsf{B}\rho := \bigotimes_{K \in \mathcal{K}} \mathsf{B}^K \rho, \tag{20}$$

with $\mathsf{B}^K \rho$ being the marginal of ρ on $(\mathbb{X} \times \Theta)^K$. Before we can explicitly show the effect of B^K on $\widetilde{\mathsf{F}}_n \rho$ for any $n \geq 1$ and any measure ρ on $\mathbb{X} \times \Theta$, we need to first define the block

versions of the prediction operator P_n given in (16) and correction operator C_n given in (17), as follows: define P_n^K as the prediction operator specific to block K

$$(\mathsf{P}_n^K \rho_1)(g) = \int g(x_n^K, \theta_n^K) \prod_{v \in K} f_{X_n^v | X_{n-1}}(x_n^v \mid x_{n-1}\,; \theta_n^v) f_{\Theta_n^v | \Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,; \sigma) \times \psi^v(dx_n^v) \lambda^v(d\theta_n^v) \rho_1(dx_{n-1}, d\theta_{n-1}), \tag{21}$$

for any measure ρ_1 on $(\mathbb{X} \times \mathbb{\Theta})^{\cup_{K' \in N(K)} K'}$; define C_n^K as the correction operator specific to block K

$$(\mathsf{C}_n^K \rho_2)(g) = \frac{\int g(x_n^K, \theta_n^K) \prod_{v \in K} f_{Y_n^v | X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v) \rho_2(dx_n^K, d\theta_n^K)}{\int \prod_{v \in K} f_{Y_n^v | X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v) \rho_2(dx_n^K, d\theta_n^K)}, \tag{22}$$

for any measure ρ_2 on $(\mathbb{X} \times \mathbb{\Theta})^K$. Then we can write

$$\mathsf{B}^K \widetilde{\mathsf{F}}_n \rho = \mathsf{C}_n^K \mathsf{P}_n^K \bigotimes_{K' \in N(K)} \rho^{K'}, \tag{23}$$

for any measure ρ on $\mathbb{X} \times \mathbb{\Theta}$.

The IBPF approximated filter denoted as $\widehat{\pi}_n$ can be expressed in a recursive way

$$\widehat{\pi}_n = \widehat{\mathsf{F}}_n \widehat{\pi}_{n-1}, \qquad \widehat{\pi}_0 = \delta_x \delta_\theta, \tag{24}$$

with

$$\widehat{\mathsf{F}}_n = \mathsf{C}_n \mathsf{B} \mathsf{S}^J \mathsf{P}_n \tag{25}$$

evolving as follows:

$$\widehat{\pi}_{n-1} \xrightarrow[\text{sampling}]{\text{prediction}} \widehat{\pi}_{n|n-1} = \mathsf{S}^J \mathsf{P}_n \widehat{\pi}_{n-1} \xrightarrow[\text{correction}]{\text{blocking}} \widehat{\pi}_n = \mathsf{C}_n \mathsf{B} \widehat{\pi}_{n|n-1},$$

where S^J is defined as the sampling operator

$$\mathsf{S}^J \rho = \frac{1}{J} \sum_{j=1}^{J} \delta_{x_j}. \tag{26}$$

for any probability measure ρ and $\{x_j\}_{\{j=1,\cdots,J\}}$ being i.i.d. samples distributed according to ρ. We note that $\mathsf{S}^J \mathsf{P}_n$ corresponds to lines $5-6$ in Algorithm 1, and $\mathsf{C}_n \mathsf{B}$ corresponds to lines $7-12$ in Algorithm 1.

3.3 Beating the curse of dimensionality

The following assumption is enforced in obtaining our main theoretical result (Theorem 1):

Assumption 1 *For any $v \in V$, $x_{n-1}, x_n \in \mathbb{X}$, $y_n \in \mathbb{Y}$, $\theta_{n-1}, \theta_n \in \mathbb{\Theta}$, and $n \geq 1$, we impose the following conditions:*

(1) Suppose there exists $\epsilon_x > 0$ such that

$$\epsilon_x \leq f_{X_n^v | X_{n-1}}(x_n^v | x_{n-1}\,; \theta_n^v) \leq \epsilon_x^{-1}.$$

(2) Suppose there exists $\epsilon_y > 0$ *such that*

$$\epsilon_y \leq f_{Y_n^v|X_n^v}(y_n^v|x_n^v\,;\theta_n^v) \leq \epsilon_y^{-1}.$$

(3) Suppose there exist $\epsilon_\theta(\sigma) > 0$ *and* $\sigma \geq 0$ *such that*

$$\epsilon_\theta(\sigma) \leq f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,;\sigma) \leq [\epsilon_\theta(\sigma)]^{-1}.$$

In Assumption 1, (1) and (2) are the same as the conditions enforced in Theorem 2.1 of Rebeschini and Van Handel (2015) which are localized versions of standard assumptions that are routinely employed in the analysis of PFs, and (3) is the same condition on the transition density of Θ. Similar to the global mixing assumption implying that the underlying Markov chain is strongly ergodic, a local transition density being bounded above and below as a local counterpart of the global mixing assumption, could be viewed as a local ergodicity assumption on the model.

Recall that $|\mathcal{K}|_\infty$ defined in (7) is the maximal size of one single block in the partition $\mathcal{K}$, Δ defined in (8) is the maximal number of vertices that interact with one single vertex in its r-neighborhood, and $\Delta_\mathcal{K}$ defined in (9) is the maximal number of blocks that interact with a single block (including itself). In the following theorem, we bound the error generated by our IBPF algorithmic filter $\widehat{\pi}_n$ defined in (24), to the unalgorithmic true filter π_n defined in (14), uniformly both in time n and in the model dimension $\mathrm{card}(V)$:

Theorem 1 *With* ϵ_x, $\epsilon_\theta(\sigma)$, *and* ϵ_y *satisfying Assumption 1, when*

$$\epsilon_x \epsilon_\theta(\sigma) > \left(1 - \frac{1}{16\Delta_\mathcal{K}\Delta^2}\right)^{\frac{1}{2\Delta}}, \tag{27}$$

for every $n \geq 0$, $K \in \mathcal{K}$ *and* $\mathrm{K} \subseteq K$, *we have*

$$\begin{aligned}\|\!|\widehat{\pi}_n - \pi_n|\!\|_\mathrm{K} \leq & \frac{\mathrm{card}(\mathrm{K})}{1-e^{-\beta}}\Bigg[7e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)} \\ & + \frac{40}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K}\Bigg],\end{aligned}$$

where $\mathrm{card}(\,\cdot\,)$ *stands for cardinality and*

$$\beta = \frac{1}{2r}\log\left(\frac{1}{16\Delta_\mathcal{K}\Delta^2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})}\right). \tag{28}$$

We first interpret the upper bound in Theorem 1 in terms of the graph dimension and the time dimension. For the standard PF algorithm where all observations are used to update the filtering distribution, the algorithmic error is exponential in the dimension of the model under the global metric $\|\!|\cdot|\!\|$ defined in (10). For the IBPF algorithm, only observations in a block, say K, are used to update the filtering distribution in that block. From Theorem 1, we can see that under the local metric $\|\!|\cdot|\!\|_\mathrm{K}$ defined in (11), the algorithmic error is merely exponential in the dimension of a set $\mathrm{card}(\mathrm{K})$ instead of the dimension of the graph $\mathrm{card}(V)$. That is, our IBPF algorithm has a rigorous performance guarantee in terms of

the graph dimension, thus beating the COD. Next, since the upper bound in Theorem 1 is uniform on all the time steps n, our IBPF algorithm has a rigorous performance guarantee in terms of the time dimension. The second term of the bound quantifies the error due to the variance of the Monte Carlo sampling of the IBPF algorithm. As in the standard PF analysis, Monte Carlo sampling provides the $\frac{1}{\sqrt{J}}$ factor under the local metric in this local update setting. Given that each block interacts with at most $\Delta_{\mathcal{K}}$ neighbors in the previous time step, the $\Delta_{\mathcal{K}}$ factor in the second term is expected.

Theorem 1 involves generalizing the important result of Rebeschini and Van Handel (2015) (Theorem 2.1 therein) to a time-inhomogeneous setting. Many practical applications require time-inhomogeneity; for example, stochastic epidemic models may have a time-varying population size, or other covariate, leading to time-inhomogeneity (see, e.g. Bretó et al. (2009)). In this paper, the transition densities of X_n, Y_n, and Θ_n are all time-inhomogeneous which means they are different in each time step n. Furthermore, we consider these time-inhomogeneous transition densities in a general form where only standard conditions are required. When our transition densities are the same for each time step n (i.e. being time-homogeneous), our results covers the situation of Rebeschini and Van Handel (2015) as a special case.

The rigorous proof of Theorem 1 is postponed to Appendix F. Our proof broadly follows the approach of Rebeschini and Van Handel (2015) while differing in some details that enable us to obtain stronger and more explicit bounds. We acknowledge that the proof from Rebeschini and Van Handel (2015) can be adapted to the time-inhomogeneous case, but we take the opportunity to make other adjustments while adding this extension. Specifically, we follow Rebeschini and Van Handel (2015) by controlling the filtering error $\|\!|\widehat{\pi}_n - \pi_n|\!\|_{\mathrm{K}}$ using the algorithmic bias and the algorithmic variance with the help of the triangle inequality, thus resulting in the two terms in the upper bound. An intermediate filter without sampling and resampling, $\widetilde{\pi}_n$ defined in (18), was used to separate the bias $\|\!|\widetilde{\pi}_n - \pi_n|\!\|_{\mathrm{K}}$ and the variance $\|\!|\widetilde{\pi}_n - \widehat{\pi}_n|\!\|_{\mathrm{K}}$. To control the bias generated by blocking, the decay of correlations (DOCs) property was established. The DOCs property arises in statistical physics, in regards to investigations of high-dimensional networks (see, e.g., Liu et al. (2018); Liu and Ning (2019a,b, 2021)). In the current context, it means that the effect on the distribution on block K of a perturbation made in another block K' decays rapidly in the distance $d(K, K')$ defined in (3).

Utilizing the Dobrushin comparison theorem (Theorem 2), we extended the mechanism to the parameter space by showing that the DOCs property of the underlying model is inherited by the IBPF algorithmic filter $\widehat{\pi}_n$, which is the joint conditional distribution of state X_n and Θ_n given the observations $Y_1, \ldots, Y_n$. We showed that the influence of blocking on the marginal distribution at a vertex $v \in K$ should decay exponentially in the distance from v to the boundary of the block ∂K. This idea is revealed in the $e^{-\beta d(\mathrm{K}, \partial K)}$ factor in the first term of the bound. To control the variance, a major issue is that $\widetilde{\pi}_n$ cannot be interpreted as a regular marginal or conditional distribution, given that it is only defined recursively in (18). Rebeschini and Van Handel (2015) solved this issue by constructing a "computation tree", which is in analogy with a similar notion that arises in the analysis of the well-known belief propagation algorithms (Tatikonda and Jordan, 2002). It is to introduce independent duplicates of the blocks in the previous time step and have each block interact with its own set of duplicates, which hence unravels the dependence

graph to a tree without blockwise interactions. Then one can interpret $\widetilde{\pi}_n$ as the marginal distribution on this tree.

Rebeschini and Van Handel (2015) focused on establishing scalability. Hence their error bound (Theorem 2.1 therein) targets to reveal that property while being ambiguous in some other regards, in the form as follows:

$$\|\!|\widehat{\pi}_n - \pi_n|\!\|_{\mathrm{K}} \leq \alpha \,\mathrm{card}(\mathrm{K})\big[e^{-\beta_1 d(\mathrm{K},\partial K)} + e^{\beta_2|\mathcal{K}|_\infty}/\sqrt{J}\big], \tag{29}$$

with K being a subset of the block K in partition $\mathcal{K}$, where α, β_1, and β_2 are positive finite constants. Although they did not provide a precise form of the error bound in their main result, they provided a precise error bound for variance in Theorem 4.23 (page 2864 therein), as follows:

$$\|\!|\widetilde{\pi}_n - \widehat{\pi}_n|\!\|_{\mathrm{K}} \leq \mathrm{card}(\mathrm{K})\frac{64}{\sqrt{J}}\frac{e^{\beta}}{1-e^{-\beta}}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-4|\mathcal{K}|_\infty \Delta_{\mathcal{K}}}\Delta_{\mathcal{K}}, \tag{30}$$

where $\beta = -\log 6\Delta_{\mathcal{K}}\Delta^2(1-\epsilon^{2\Delta}) > 0$. Although (30) shares similarites with the second term of our bound in Theorem 1, with a close look we can see that (30) has an additional factor e^{β}. β becomes large when ϵ is close to 1, which is how this mathematical framework describes the situation where spatiotemporal mixing is fast. We expect the resulting bound on the error of the block filter to be tighter in this case, and our bound has that property whereas the bound of Rebeschini and Van Handel (2015) does not. Furthermore, we can see that the exponent of $\epsilon_y^{-1}(>1)$ differs. Whereas ϵ, ϵ_x and ϵ_θ are required to be close to 1, ϵ_y can be close to 0. Thus, an improvement from $\epsilon_y^{-4|\mathcal{K}|_\infty\Delta_{\mathcal{K}}}$ to $\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_{\mathcal{K}}+1)}$ can substantially tighten the bound, especially when the maximal block size ($|\mathcal{K}|_\infty$) is not small.

Throughout the proofs in Rebeschini and Van Handel (2015) and ours, a positive constant β containing those quantities is used. We provide a precise definition of β in (28) for the first time, which is used consistently in all the proofs. Our precise constant β is able to rigorously reveal the influences of those crucial quantities on the error bound which have been open problems: when r (the range of interacting neighborhoods) increases, the error bound increases; when Δ (the maximal number of vertices that interact with one single vertex in its r-neighborhood) increases, the error bound increases; when $\Delta_{\mathcal{K}}$ (the maximal number of blocks that interact with a single block) increases, the error bound increases; when $|\mathcal{K}|_\infty$ (the maximal size of one single block in the partition) increases, the error bound increases. Furthermore, we provided a precise sufficient condition for the first time, and used it throughout all the proofs. That is, the product of the assumed lower bound of X's local transition density (ϵ_x) and the assumed lower bound of Θ's local transition density ($\epsilon_\theta(\sigma)$) is larger than $\left(1-\frac{1}{16\Delta_{\mathcal{K}}\Delta^2}\right)^{\frac{1}{2\Delta}}$.

Although we followed the strategy in Rebeschini and Van Handel (2015) in general, to adapt to the time-inhomogeneous setting some proof strategies need to be adjusted correspondingly. For example, their Proposition 4.4 (Page 2839) achieves local filter stability by bounding the term

$$\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\nu\|_{\mathrm{K}},$$

where they used the Dobrushin comparison theorem on the distributions

$$\rho = \mathbf{P}^{\mu}[X_0, \ldots, X_n \in \cdot \mid Y_1, \ldots, Y_n] \quad \text{and} \quad \widetilde{\rho} = \mathbf{P}^{\nu}[X_0, \cdots, X_n \in \cdot \mid Y_1, \ldots, Y_n].$$

Our local filter stability is established in Proposition 6 by bounding the term

$$\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}}$$

where we used the Dobrushin comparison theorem on the distributions

$$\rho = \mathbf{P}^{\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1}}[X_s, X_{s+1}, \ldots, X_n \in \cdot\,, \Theta_s, \Theta_{s+1}, \ldots, \Theta_n \in \cdot \mid Y_{s+1}, \ldots, Y_n],$$

$$\widetilde{\rho} = \mathbf{P}^{\mathsf{F}_s\widetilde{\pi}_{s-1}}[X_s, X_{s+1}, \cdots, X_n \in \cdot\,, \Theta_s, \Theta_{s+1}, \cdots, \Theta_n \in \cdot \mid Y_{s+1}, \ldots, Y_n].$$

Note that, with time-inhomogeneous, to quantify the effect of F_{s+1} on μ, it is appropriate to use distributions on latent states starting from s instead of 0.

3.4 MLEs

The IBPF algorithm generalizes the data cloning method (Lele et al., 2007, 2010), which is based on the observation that iterating a Bayes map converges to a point mass at the maximum likelihood estimate. Combining such iterations with perturbations of model parameters improves the numerical stability of data cloning and provides a foundation for stable algorithms (Ionides et al., 2015). To be specific, the same set of data $Y_1, \ldots, Y_N$ is used in any one of the M iterations of the IBPF algorithm, given the result of the m-th iteration for $m \in 1{:}(M-1)$ is simply the initial value of the $(m+1)$-th iteration, we can see that all M iterations together can be represented as a filtering problem on M replications of the data as follows:

$$\Big\{\underbrace{\{Y_1, \ldots, Y_N\}}_{1}, \underbrace{\{Y_1, \ldots, Y_N\}}_{2}, \ldots, \underbrace{\{Y_1, \ldots, Y_N\}}_{M}\Big\}.$$

As in the previous subsections, our strategy is to analyze the original theoretical quantity and then explore its algorithmic approximation.

The joint density of the classical POMP model can be written as

$$f_{X_{0:N},Y_{1:N}}(x_{0:N}, y_{1:N}\,;\theta) = f_{X_0}(x_0\,;\theta)\prod_{n=1}^{N} f_{X_n|X_{n-1}}(x_n|x_{n-1}\,;\theta) f_{Y_n|X_n}(y_n|x_n\,;\theta).$$

We write $f_{Y_{1:N}}(y_{1:N}\,;\theta)$ for the marginal density of $Y_{1:N}$. Then the likelihood function is defined to be $\ell(\theta) = f_{Y_{1:N}}(y_{1:N}\,;\theta)$, where the data is a sequence of observations $y_{1:N}$. A MLE is a value $\widehat{\theta}$ that maximizes $\ell(\theta)$. We define an extended likelihood function on $\mathbb{O}^{N+1}$ by

$$\breve{\ell}(\theta_{0:N}) = \int\!\ldots\!\int dx_0 \ldots dx_N \Big\{ f_{X_0}(x_0\,;\theta_0)\prod_{n=1}^{N} f_{X_n|X_{n-1}}(x_n \mid x_{n-1}\,;\theta_n) f_{Y_n|X_n}(y_n \mid x_n\,;\theta_n)\Big\}.$$

Each m iteration of data cloning corresponds to an operator T_σ, which is a composition of a parameter perturbation with a Bayes map that multiplies the likelihood and renormalizes (page 2 of Ionides et al. (2015)), i.e.,

$$T_\sigma g(\theta_N) = \frac{\int \breve{\ell}(\theta_{0:N}) f_{\Theta_{0:N}}(\theta_{0:N}|\vartheta\,;\sigma) g(\vartheta)\, d\vartheta\, d\theta_{0:N-1}}{\int \breve{\ell}(\theta_{0:N}) f_{\Theta_{0:N}}(\theta_{0:N}|\vartheta\,;\sigma) g(\vartheta)\, d\vartheta\, d\theta_{0:N}}, \tag{31}$$

with g and $T_\sigma g$ approximating the initial and final density of the parameter swarm, where

$$f_{\Theta_{0:N}}(\theta_{0:N} \,|\, \vartheta\,; \sigma) = f_{\Theta_0}(\theta_0 \,|\, \vartheta\,; \sigma) \prod_{n=1}^{N} f_{\Theta_n|\Theta_{n-1}}(\theta_n \,|\, \theta_{n-1}\,; \sigma)$$

and ϑ is the mean of the distribution of Θ_0. Iteration of the Bayes map alone has a central limit theorem that forms the theoretical basis for the data cloning methodology (Lele et al., 2007, 2010).

IBPF is the approximation to T_σ^M which is the M-th iterate of T_σ. Following Ionides et al. (2015), we first show that $\lim_{m\to\infty} T_\sigma^m g = g_\sigma$ exists for every fixed $\sigma > 0$, and as the noise intensity becomes small $\lim_{\sigma\to 0} g_\sigma$ approaches a point mass at the MLE if it exists. Then we show that when the number of particals J and the number of iterations M become large, the IBPF algorithm numerically approximates g_σ. Proofs are provided in Appendix G.

4. Application and performance analysis

In this section, we illustrate how IBPF can be used and compare its performances with those of IF2 and IEnKF, using the spatiotemporal model covered in Section 4.1, over the dataset covered in Section 4.2. We gradually increase the parameter learning difficulties in 4 cases, with consistent fairness in all experiments, in Section 4.3. We implement IF2 and IEnKF through the *spatPomp* package (Asfaw et al., 2021).

4.1 Generalized spatiotemporal modeling for measles

Measles is a highly contagious infectious disease caused by the measles virus; it spreads easily from one person to the next through coughs and sneezes of infected people. In this section, we consider a generalized spatiotemporal model for disease transmission dynamics of measles within and between multiple cities.

A compartment modeling framework for spatiotemporal population dynamics divides the population at each spatial location into compartments. Specifically, measles transmission at each location is modeled according to the SEIR model with 6 compartments: (S) represents susceptible individuals who have not been infected yet but may experience infection later, (E) represents individuals exposed and carrying a latent infection, (I) represents infectious individuals that have been infected and are infectious to others, (R) represents recovered individuals that are no longer infectious and are immune, (B) represents the birth of individuals, and (D) represents the death of individuals. Park and Ionides (2020) generalized the compartment model presented by He et al. (2010) to the spatiotemporal modeling setting, which is analyzed in other literature such as Ionides et al. (2021). We further generalize that spatiotemporal compartment model by allowing the dynamics in each spatial location to have their own specific parameters, including different disease transmission speeds across locations. By demonstrating a methodology that scales to vertex-specific parameters, we open up new possibilities for spatiotemporal inference, though we focus here on testing statistical tools and so we do not engage directly in the scientific debates.

Different to the discrete-time based modelling and algorithm having time index $n \in \mathbb{N}$ in the previous sections, here the spatiotemporal model under consideration is a continuous-

time Markov chain model with time index $t \in [0, \infty)$. Time discretization is a common and natural practice to link these two kinds of time notations. Specifically, for the continuous-time latent process $X(t) = (S(t), E(t), I(t), R(t))$, the corresponding discrete-time latent process is given by $X_n = X(t_n)$ with $t_{0:n}$ being observation time. The number of individuals in compartments S, E, I, and R of city v at time t are denoted by integer-valued random variables $S^v(t)$, $E^v(t)$, $I^v(t)$, and $R^v(t)$ respectively. Denote $N^v_{ij}(t)$ as the counting process enumerating cumulative transitions from compartment i to compartment j where $i, j \in \{B, S, E, I, R, D\}$ and $i \neq j$, in city v up to time t. We model the 40 largest cities in the UK, ordered in decreasing size with $v = 1$ corresponding to London. Our model is described by the following system of stochastic differential equations, for $v \in \{1, \ldots, 40\}$,

$$\begin{array}{rclcrcl} dS^v(t) &=& dN^v_{BS}(t) &-& dN^v_{SE}(t) &-& dN^v_{SD}(t), \\ dE^v(t) &=& dN^v_{SE}(t) &-& dN^v_{EI}(t) &-& dN^v_{ED}(t), \\ dI^v(t) &=& dN^v_{EI}(t) &-& dN^v_{IR}(t) &-& dN^v_{ID}(t). \end{array} \tag{32}$$

The total population

$$P^v(t) = S^v(t) + E^v(t) + I^v(t) + R^v(t) \tag{33}$$

is calculated by smoothing census data and is treated as known. Hence, by (32), the number of recovered individuals $R^v(t)$ in city v is defined implicitly.

The birth process $N^v_{BS}(t)$ is a time-inhomogeneous Poisson process with rate $\mu^v_{BS}(t)$ given by interpolated census data. The transition processes $N^v_{EI}(t)$, $N^v_{IR}(t)$, $N^v_{SD}(t)$, $N^v_{ED}(t)$, and $N^v_{ID}(t)$ are modeled as conditional Poisson processes with per-capita rates μ_{EI}, μ_{IR}, μ_{SD}, μ_{ED}, and μ_{ID} respectively. The transition process $N^v_{SE}(t)$ is modeled as a negative binomial death process according to Bretó et al. (2009) and Bretó and Ionides (2011) with over-dispersion parameter $\overline{\sigma}^v_{SE}$ and rate given by

$$\begin{aligned} &\mathbf{E}\big[N^v_{SE}(t+dt) - N^v_{SE}(t)\big] \\ &= \beta^v(t)\, S^v(t) \left[\left(\frac{I^v(t) + \widetilde{\iota}}{P^v(t)} \right)^{\overline{\alpha}^v} + \sum_{v' \neq v} \frac{\overline{\theta}_{vv'}}{P^v(t)} \left\{ \left(\frac{I^{v'}(t)}{P^{v'}(t)} \right)^{\overline{\alpha}^{v'}} - \left(\frac{I^v(t)}{P^v(t)} \right)^{\overline{\alpha}^v} \right\} \right] dt + o(dt). \end{aligned} \tag{34}$$

Throughout this section, we consistently use an overline to indicate average across time and use a tilde to indicate average across time across cities. Here, $\beta^v(t)$ models seasonality driven by high contact rates between children at school, described by

$$\beta^v(t) = \begin{cases} \big(1 + \widetilde{\theta}_a(1 - \widetilde{p})\widetilde{p}^{-1}\big)\, \overline{R}^v_0 \mu_{IR} & \text{during school term,} \\ \big(1 - \widetilde{\theta}_a\big)\, \overline{R}^v_0 \mu_{IR} & \text{during vacation,} \end{cases} \tag{35}$$

where $\widetilde{p}$ is the proportion of the year taken up by the school terms, $\overline{R}^v_0$ is the annual average basic reproductive ratio, and $\widetilde{\theta}_a$ measures the reduction of transmission during school holidays. In (34), $\overline{\alpha}^v$ is a mixing exponent modeling inhomogeneous contact rates within the city v, and $\widetilde{\iota}$ models immigration of infected individuals which is appropriate when analyzing a subset of cities that cannot be treated as a closed system. In (34), the number of travelers from city v to v' is denoted by $\overline{\theta}_{vv'}$, constructed as fixed through time and symmetric between any two arbitrary cities, using the gravity model of Xia et al. (2004),

$$\overline{\theta}_{vv'} = \widetilde{G} \cdot \frac{\widetilde{d}}{\widetilde{P}^2} \cdot \frac{\overline{P}^v \cdot \overline{P}^{v'}}{d(v, v')},$$

where $\widetilde{G}$ is a coupling parameter, $d(v, v')$ denotes the distance between city v and city v', $\overline{P}^v$ is the average population for city v across time, $\widetilde{p}$ is the average population across time across cities, and $\widetilde{d}$ is the average distance between a randomly chosen pair of cities.

To complete the model specification, the measurement process for modeling the partial observability is defined as follows: for $t_{0:n}$ being observation time and

$$Z_n^v = N_{IR}^v(t_n) - N_{IR}^v(t_{n-1}) \tag{36}$$

being the number of removed infected individuals in the nth reporting interval, suppose that they are quarantined once they are identified, so that reported counts comprise a fraction $\widetilde{\varrho}$ of these removal events; the case report y_n^v is modeled as a realization of a conditionally Gaussian random variable Y_n^v via

$$\begin{aligned}&\mathbf{P}\big[Y_n^v{=}y \mid Z_n^v{=}z\big]\\ &= \mathcal{N}\big(y+0.5; \widetilde{\varrho}z, \widetilde{\varrho}(1-\widetilde{\varrho})z + \widetilde{\psi}^2\widetilde{\varrho}^2z^2\big) - \mathcal{N}\big(y-0.5; \widetilde{\varrho}z, \widetilde{\varrho}(1-\widetilde{\varrho})z + \widetilde{\psi}^2\widetilde{\varrho}^2z^2\big),\end{aligned} \tag{37}$$

where $\mathcal{N}(\,\cdot\,; \mu, \sigma^2)$ is the cumulative distribution function of Normal(μ, σ^2) and $\widetilde{\psi}$ models overdispersion relative to the binomial distribution.

4.2 Spatiotemporal illustration

Figure 2 shows a simulation (Plot A) from our model covered in Section 4.1 and the real measles data (Plot B). We note that the spatiotemporal model considered in Park and Ionides (2020) and Ionides et al. (2021) is a special case of ours, by taking location-specific parameters $\overline{\alpha}^v = 1$ (in equation (34)), $\overline{R}_0^v = 30$ (in equation (35)) and $\overline{\sigma}_{SE}^v = 0.15$ year$^{1/2}$ the same for all locations, and take all the initializations $\theta_{S_0^v} = 0.032$, $\theta_{E_0^v} = 0.00005$, $\theta_{I_0^v} = 0.00004$, and $\theta_{R_0^v} = 1 - \theta_{S_0^v} - \theta_{E_0^v} - \theta_{I_0^v}$ the same for all locations. In our simulation, for each location v, we draw the corresponding variables according to uniform distributions of the $[0.99, 1.0355]$-scaled range of those in Ionides et al. (2021). That is, our $\overline{\alpha}^v \sim \text{Unif}[0.99 \times 1, 1.0355 \times 1]$ for each v. We take the other parameters as fixed values as those of Ionides et al. (2021): $\mu_{SD} = \mu_{ED} = \mu_{ID} = 0.02$ year^{-1}, $\mu_{EI} = \mu_{IR} = 52$, $\widetilde{p} = 0.759$, $\widetilde{\varrho} = 0.5$, $\widetilde{\psi} = 0.15$, $\widetilde{\theta}_a = 0.5$, $\widetilde{\iota} = 0$, $\widetilde{G} = 400$. From Figure 2, we can see that the simulation shares the biennial pattern with most cities locked in phase most of the time. In both plots, each row is associated with a city, each column is associated with a date, and each pixel in a row represents Log(reported counts +1) of the epidemics. We note that although the simulated data and real data in Figure 2 are for 40 cities, in the next subsection we gradually increase the number of cities in modeling and stop when the number of cities involved in the spatiotemporal data analysis is sufficient to clearly reveal the performances of algorithms in terms of COD. For example, in Figure 3, city number being 2 indicates that we infer 4 parameters for each of these 2 cities only (8 parameters in total), and city number being 20 indicates that we infer 4 parameters for each of these 20 cities only (80 parameters in total). We would stop the test if the performances of algorithms (measured by log-likelihood) are sufficiently clear using spatiotemporal data of 14 cities (Case 2 in Figure 3) instead of testing up to 40 cities.

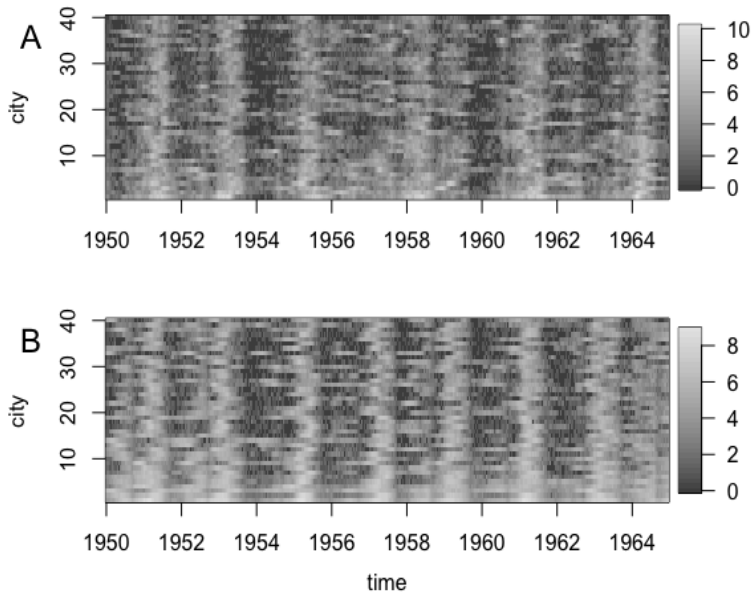

Figure 2: Log(reported counts +1) for (A) the measles simulation from our spatiotemperal model and (B) the corresponding UK measles data.

4.3 Performance analysis

We gradually increase the parameter learning difficulties in 4 cases: in the first case, the goal is to learn initial value parameters for measles transmission dynamics for each location: $\theta_{S_0^v}$, $\theta_{E_0^v}$, $\theta_{I_0^v}$, and $\theta_{R_0^v}$; in the second case, in addition to these 4 initial values the goal is to also learn the nonlinear parameter $\overline{R}_0^v$ for each location; in the third case, in addition to these 4 initial value parameters the goal is to also learn the highly nonlinear parameter $\overline{\alpha}^v$ for each location; in the fourth case, the goal is to learn all the location-specific parameters, namely, $\theta_{S_0^v}$, $\theta_{E_0^v}$, $\theta_{I_0^v}$, $\theta_{R_0^v}$, $\overline{\alpha}^v$, $\overline{\sigma}_{SE}^v$, and $\overline{R}_0^v$. Table 1 provides an illustration of the parameters to infer for each location v in four cases. Fairness is obtained over all experiments on all these three algorithms, as follows:

- Each algorithm uses the same number of iterations $M = 100$ (see Algorithm 1 for notations) and the same number of particles $J = 80000$.
- We conduct 10 replicates of all the parameter learning performance comparisons, in the way that in each replicate all algorithms start with the same initial search values drawn uniformly as follows:

$$\overline{\alpha}^v \sim \text{Unif}[0,2], \quad \overline{\sigma}_{SE}^v \sim \text{Unif}[0,1], \quad \overline{R}_0^v \sim \text{Unif}[25,35],$$

$$\theta_{S_0^v}, \theta_{E_0^v}, \theta_{I_0^v}, \theta_{R_0^v} \sim \text{Unif}[0,1].$$

  Here, we consider latent states of each city as portions of the population of that specific city such that $S^v(0) = \theta_{S_0^v} P^v(0)$, $E^v(0) = \theta_{E_0^v} P^v(0)$, $I^v(0) = \theta_{I_0^v} P^v(0)$, $R^v(0) = \theta_{R_0^v} P^v(0)$, and

$$\theta_{S_0^v} + \theta_{E_0^v} + \theta_{I_0^v} + \theta_{R_0^v} = 1.$$

Parameters	Case 1	Case 2	Case 3	Case 4
$\theta_{S_0^v}$	Yes	Yes	Yes	Yes
$\theta_{E_0^v}$	Yes	Yes	Yes	Yes
$\theta_{I_0^v}$	Yes	Yes	Yes	Yes
$\theta_{R_0^v}$	Yes	Yes	Yes	Yes
$\overline{R}_0^v$		Yes		Yes
$\overline{\alpha}^v$			Yes	Yes
$\overline{\sigma}_{SE}^v$				Yes

Table 1: Parameters to infer for each location v in four cases.

- Mathematically, there is only one likelihood function for the model and data in question, and different algorithms are making various approximations to estimate this quantity. The algorithms each compute the likelihood corresponding to an approximation to the exact filter distribution, and therefore they have a negative bias. The highest estimated likelihood among the available filters may therefore be anticipated to have the lowest bias. We note that this reasoning assumed that the model is correctly specified—substantial model misspecification could result in an approximate filter estimating a substantially higher likelihood than the exact filter. Thus, each algorithm has its own metric of log-likelihood: IEnKF uses the metric of EnKF on log-likelihood estimation, denoted as $\widehat{l}_e$ whose algorithmic definition is provided in Algorithm 3 in Appendix A; IF2 uses the metric of PF on log-likelihood estimation, denoted as $\widehat{l}_p$ whose algorithmic definition is provided in Algorithm 2 in Appendix A; IBPF uses the metric of BPF on log-likelihood estimation, denoted as $\widehat{l}_b$ whose algorithmic definition is provided in Algorithm 1 in Section 3.1. We evaluate the best parameters learned using each algorithm with the metrics of the other two algorithms, in all the experiments. For example, we evaluate the best parameters learned using the IEnKF algorithm through its $\widehat{l}_e$ metric, with the other two metrics ($\widehat{l}_p$ and $\widehat{l}_b$).

- One additional setup needed with IBPF is to set up the block sizes. In all the comparisons with IF2 and IEnKF, we simply allow each block in our IBPF algorithm to have exactly 2 cities in all the experiments. That is, the first block is city 1 and city 2, the second block is city 3 and city 4, and so on. Hence, the number of blocks $\text{card}(\mathcal{K}) = \text{Number of cities}/2$.

Figure 3 reports the log-likelihood estimates per city per time step of various dimensions for cases 1 and 2. The corresponding original parameter learning results are reported in Tables 2 and 3 in Appendix H. In Figure 3, the original results are divided by the number of cities and then by time steps which is 15×26 for biweekly data in years $1950 - 1965$. We can see in case 1 that when we only learn the initial values ($\theta_{S_0^v}$, $\theta_{E_0^v}$, $\theta_{I_0^v}$, and $\theta_{R_0^v}$), the best parameter learning results from 10 replicates of experiments of all three algorithms are as good as the true parameter using the EnKF metric and the BPF metric, while the best

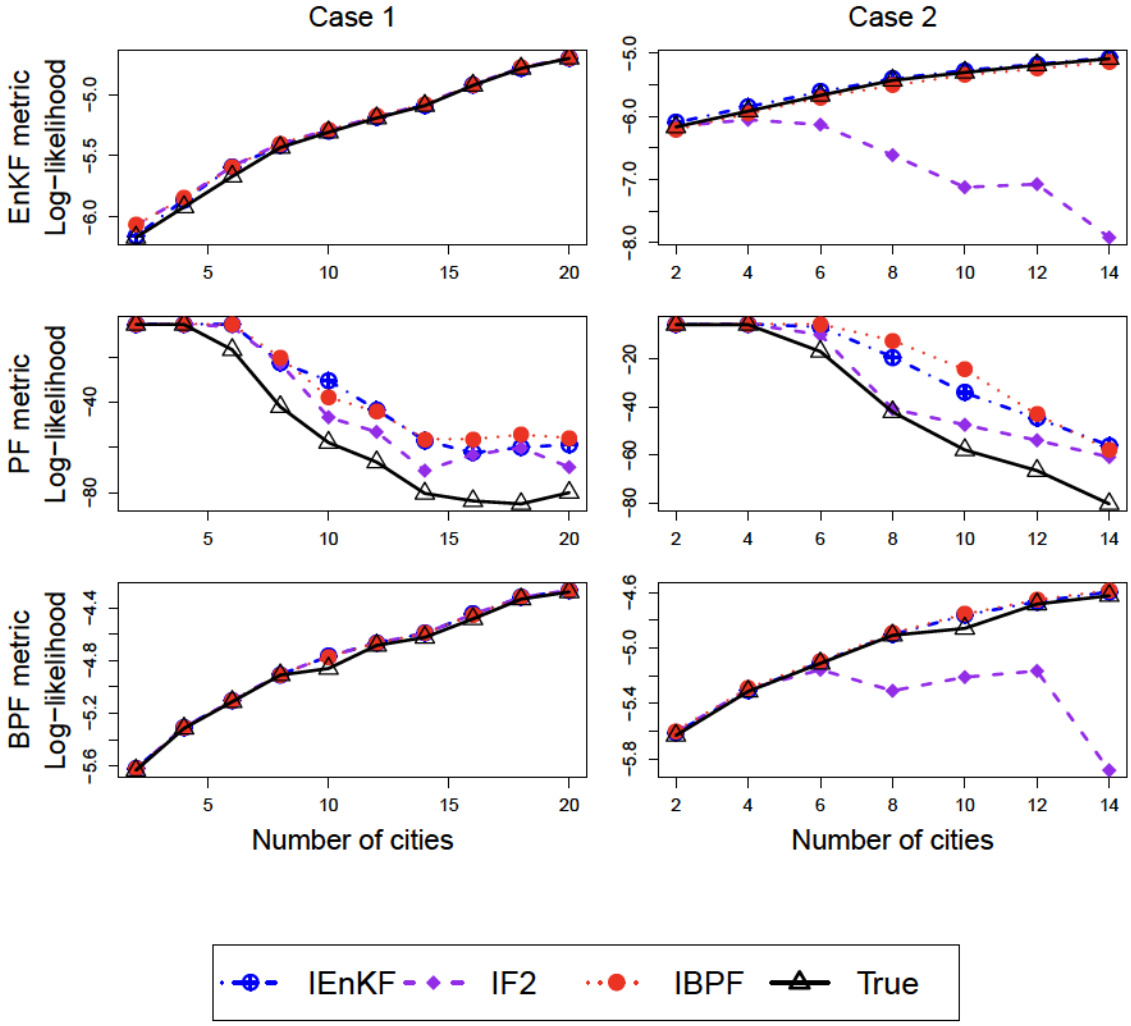

Figure 3: Log-likelihood estimates per city per time step of various dimensions for cases 1 and 2. Original parameter learning results are reported in Tables 2 and 3 in Appendix H. Here, we divide the original results by the number of cities and then by 15×26 (biweekly data in years $1950 - 1965$).

parameter learning results outperform the true parameters using the PF metric consistently. We can see that for nonlinear and non-Gaussian problems but not highly nonlinear (both case 1 and case 2), the IEnKF algorithm performs very well; IF2 scales well at least up to 20 cities for simpler initial values learning problem in case 1, while its performances start to drop from 4 cities with all metrics in case 2 which confirms the phenomenon that PF

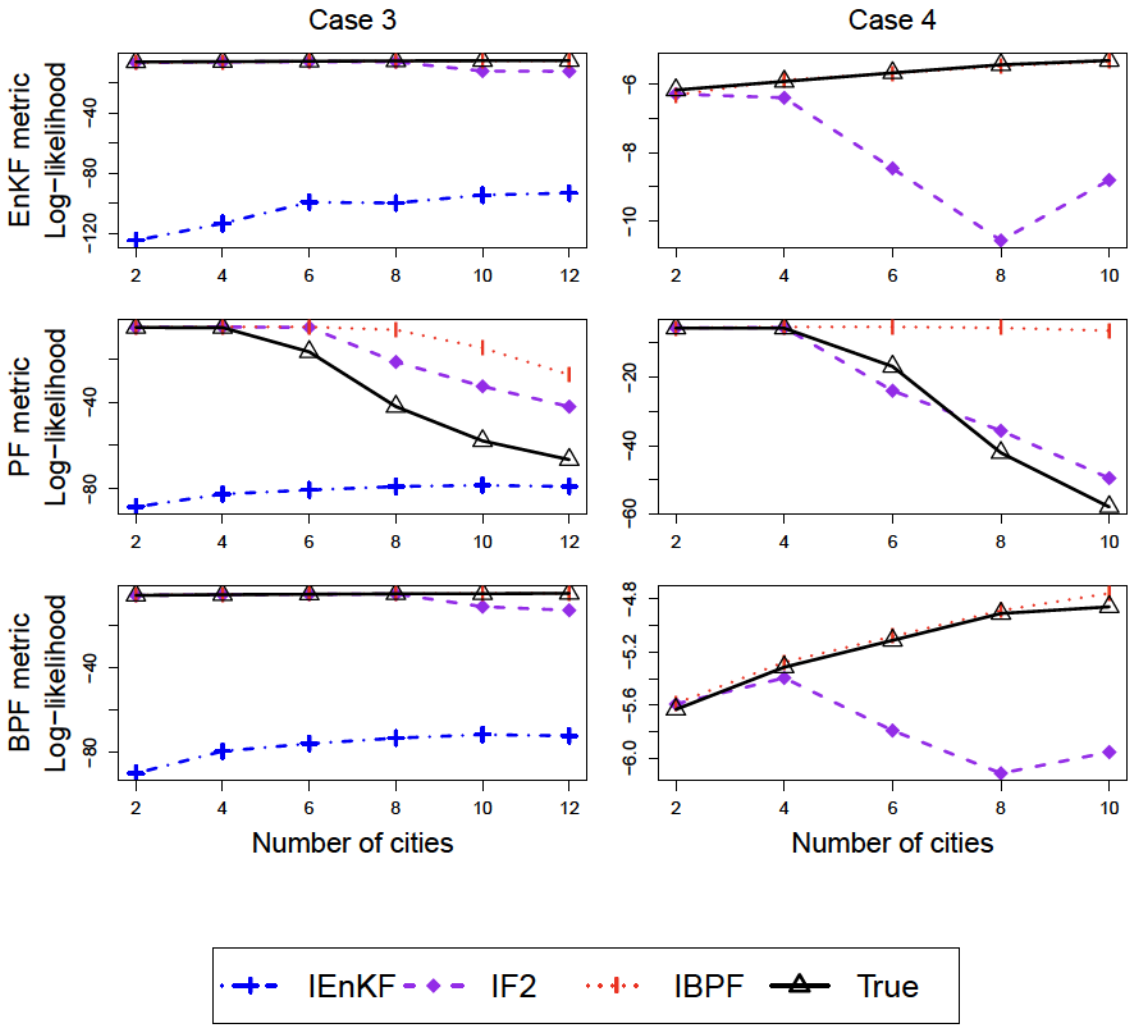

Figure 4: Log-likelihood estimates per city per time step of various dimensions for cases 3 and 4. Original parameter learning results are reported in Tables 4 and 5 in Appendix H. Here, we divide the original results by the number of cities and then by 15×26 (biweekly data in years $1950 - 1965$).

may not scale well with dimensions (Bengtsson et al., 2008; Snyder et al., 2008); our IBPF is as good as the true parameter consistently with the EnKF metric and the BPF metric, and much better than the true parameter with the PF metric consistently, in both case 1 and case 2.

Figure 4 reports the log-likelihood estimates per city per time step of various dimensions for cases 3 and 4, with the same transformation done as that of Figure 3 upon the corresponding original parameter learning results that are reported in Tables 4 and 5 in Appendix H. Recall that the goal of case 3 is to learn the initial values ($\theta_{S_0^v}$, $\theta_{E_0^v}$, $\theta_{I_0^v}$, $\theta_{R_0^v}$) and the highly non-linear parameter $\overline{\alpha}^v$. In equation (34), we can see that through $\overline{\alpha}^v$, the dynamic of one city has direct interactions with other cities. Case 4 is the hardest of all these 4 cases. Its goal is to learn all the location-specific parameters ($\theta_{S_0^v}$, $\theta_{E_0^v}$, $\theta_{I_0^v}$, $\theta_{R_0^v}$, $\overline{\alpha}^v$, $\overline{\sigma}_{SE}^v$, and $\overline{R}_0^v$). From Figure 4, we can see that, even for 2 cities using all these 3 metrics, IEnKF performs very badly in case 3 and fails completely in case 4. The reason is that IEnKF is based on EnKF, and hence it implicitly assumes a linear Gaussian state space model. Specifically, when new observations become available, the ensemble is updated by a linear "shift" based on the assumption of a linear Gaussian state space model. However, cases 3 and 4 focus on highly nonlinear and non-Gaussian problems. This phenomenon that EnKF may not perform well for highly nonlinear and non-Gaussian problems, was observed earlier, such as in Ades and Van Leeuwen (2015); Lei et al. (2010); Miller et al. (1999). In general, the EnKF-based parameter learning approaches are applicable to problems with a relatively small number of parameters but more work is needed for cases where the parameter and state are both high dimensional (Katzfuss et al., 2016). From Figure 4, we can see that the performance of IF2 drops in both cases using all three metrics. The performance of our IBPF in both cases, is as good as that of the true parameter consistently with the EnKF metric and the BPF metric, and much better than that of the true parameter with the PF metric consistently.

Now we explore more properties of IBPF through experiments in case 4. Table 6 reports IBPF's parameters learned for 10 cities. This is the set of parameter values that produce the maximum log-likelihood -18555 in the $\widehat{l}_b$ metric in Table 5. We can see that all learned parameter values conform to common sense. Table 7 reports IBPF's parameter learning results in terms of log-likelihood for all 10 replicates conducted, where each replicate starts with an initial search value drawn uniformly. The resulting log-likelihood values are ranked from highest to lowest; for example, the value of replicate 1 for 10 cities in Table 7 is the log-likelihood -18555 in $\widehat{l}_b$ metric in Table 5. When we calculate the per city per time step values as that in Figures 3 and 4, all values are almost the same across replicates. For example, the highest value of replicate 1 gives $-18555/(26 \times 15)/10 = -5.947$ and the lowest value of replicate 1 gives $-18583/(26 \times 15)/10 = -5.956$. Thus the parameter learning results are robust. Figure 5 reports IBPF's parameter learning results in terms of log-likelihood for iterations $\{20, 40, 60, 80, 100, 120, 160, 180\}$ with different block sizes $\{1, 2, 3, 4\}$ for 12 cities. We also conducted analysis with block size 6, i.e., there are two blocks and each has 6 cities. The log-likelihood values for block size 6 are $\{-75793, -96487, -88973, -84087, -76356\}$ for $\{20, 40, 60, 80, 100\}$ iterations, respectively. These values are not included in Figure 5, since they are much smaller than the values with block sizes $\{1, 2, 3, 4\}$. From Figure 5, we can see that the performance of IBPF increases as block sizes decrease. In common sense, an increase in the block size for a fixed J will make the variance term worse and the bias term better in the error bound. A closer observation of the bias term, one can see that $e^{-\beta d(\kappa, \partial K)}$ measures the distance to the boundary of the block, while in Figure 5 we have that fixed and vary $|\mathcal{K}|_\infty$ which is the maximal size of one single block in the partition. That is, in Figure 5, block size being 1 means that $|\mathcal{K}|_\infty = 1$. Hence, Figure 5 confirms

our technical finding of $|\mathcal{K}|_\infty$. From Figure 5, we can also see that IBPF achieves likelihood maximization as iterations increase which is consistent with the analysis in Section 3.4. For any block size, the block particle filter computes the log-likelihood corresponding to a one-step probabilistic forecast, with the forecast corresponding to the empirical distribution of the particles. Log-likelihood is a proper scoring rule for such forecasts (Gneiting and Raftery, 2007), meaning that the likelihood of a forecast cannot, on average, exceed that of the ideal but inaccessible Bayesian filter. This provides a justification for comparing filters by their corresponding log-likelihoods and preferring the highest.

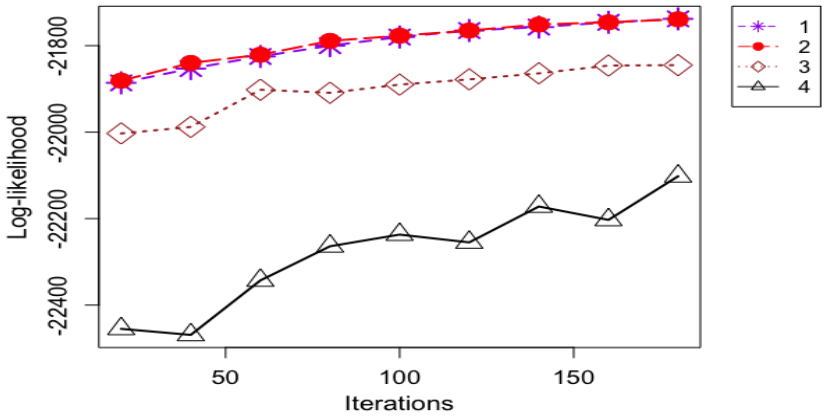

Figure 5: Performance analysis of IBPF for iterations $\{20, 40, 60, 80, 100, 120, 160, 180\}$ with different block sizes $\{1, 2, 3, 4\}$ for 12 cities of case 4.

At last, we discuss the computational cost. The computational cost of IF2 and IEnKF are $\mathcal{O}(MJN|V|)$ (Asfaw et al., 2021) and IBPF as well, where M is the number of iterations, J is the number of particles, N is the number of time steps, and $|V|$ is the number of vertices. Table 8 provides the performance and runtime comparison of top five replicates in terms of log-likelihood for two cities in case 4. Each replicate is conducted with one node and 4GB CPU on the Great Lakes Slurm cluster of the University of Michigan, Ann Arbor. We can see that the time consumed is comparable: IEnKF is in the range of $15 - 16$ hours, IF2 is in the range of $17 - 18$ hours, and IBPF is in the range of $18 - 19$ hours. Furthermore, we can see that IEnKF is still able to run but provides very bad results, and for this reason we call it "Failed".

5. Discussion and extensions

In this paper, we have proposed the IBPF algorithm for high-dimensional parameter learning over partially observed, nonlinear, stochastic spatiotemporal processes, established its theoretical performance guarantees, and compared it with mainstream algorithms.

5.1 Discussion

In this paper, we have only compared the EnKF-based parameter learning approach using IF. There are two main approaches to parameter estimation within the EnKF framework without using IF. The first is called state augmentation (Anderson, 2001) which works well in many examples but implicitly assumes that the states and parameters jointly follow a linear Gaussian POMP model. Hence, when some parameters violate this assumption, the method fails completely (Stroud and Bengtsson, 2007). The second approach is based on approximating likelihood functions constructed using the output from EnKF, which estimates parameters either by maximum likelihood or Bayesian methods. Examples of these approaches include sequential maximum likelihood (Mitchell and Houtekamer, 2000), offline maximum likelihood (Stroud et al., 2010), and sequential Bayesian methods (Stroud and Bengtsson, 2007; Frei and Künsch, 2012). In general, these methods have been successful in examples with a relatively small number of parameters, and more work is needed for cases where the parameter and state are both high dimensional (Katzfuss et al., 2016). We note that our IBPF is designed to learn a large number of parameters for models with high-dimensional parameters and states.

In recent years, approaches developed on learning high-dimensional parameters applying the Bayesian method (specifically, Markov chain Monte Carlo (MCMC)) with multidimensional time series data, including (Qiu et al., 2018, 2020; Jammalamadaka et al., 2019). We note that these literatures are limited to only working for linear models while IBPF is designed to work on general nonlinear models. Particle MCMC (PMCMC) methods, introduced by Andrieu et al. (2010), make use of PF to construct efficient proposals for the MCMC sampler, working for non-linear models. Particle Gibbs (PG) as a particularly widely used PMCMC algorithm, modifies the PF step in the PMCMC algorithm to sample the latent variables conditioned on an existing particle trajectory, resulting in what is called a conditional SMC (CSMC) step. A drawback of PG is that it can be particularly adversely affected by path degeneracy in the CSMC step, in the way that conditioning on an existing trajectory means that whenever resampling of the trajectories results in a common ancestor who must correspond to this trajectory (Rainforth et al., 2016). Efforts on combating the path degeneracy effect, include but are not limited to, (Whiteley et al., 2010; Lindsten and Schön, 2013; Lindsten et al., 2014; Chopin and Singh, 2015; Lindsten et al., 2015). We note that IBPF does not have the path degeneracy problem.

5.2 Extensions

Algorithm 1 describes the IBPF algorithm in which the initial values of the latent states are deterministically determined by model parameters, hence the theoretical treatment needs to consider only Dirac measures as initial distributions are needed. Rebeschini and Van Handel (2015) also considered a nonrandom initial condition which is a choice of convenience. For comments on extending the results to general initial conditions, we refer interested readers to Remark 2.3 (page 2823) of Rebeschini and Van Handel (2015). A corresponding extension of Algorithm 1 permits f_{X_0} to be a probability density function.

At first glance, it may seem that BPF needs an approximation of spatial independence to justify the factorized distribution. The reality is more delicate. There is a possibility for considerable dependence to arise during the proposal stage of the filter, for which the

particles follow the full joint transition density. The block resampling then imposes a conditional independence approximation given the data, but conditional independence can be a much weaker assumption than unconditional independence. Recently, Min et al. (2022) investigated the choice of blocks for a block particle filter, formulating the partitioning problem as a clustering problem and proposed a data-driven partitioning method based on constrained spectral clustering to automatically provide an appropriate partition. For our measles model, we have found it surprisingly successful to have each city in its own block, and we have not found an advantage from larger block sizes. To reason about how this might be consistent with the previous section, consider two cities (say, London and Birmingham). Suppose they are tightly coupled dynamically, such that there will be an outbreak in London if and only if there is one in Birmingham. Now suppose that, once an outbreak occurs, the dynamics in each city become dominated by local noise, and further that each city has a reasonably effective case reporting system. Plausibly, the discrepancy between the actual and reported counts in London and Birmingham may be well modeled as close to independent, conditional on the reported counts. That is enough to suggest that a block particle filter having Birmingham and London in different blocks may be successful despite the close relationship between their epidemic outcomes.

A demonstration on real data is beyond the scope of the current manuscript, since it requires further investigation of model misspecification and its consequences. We refer interested readers to Ionides et al. (2022) for corresponding real data analysis. The IBPF algorithm has been contributed to the the *spatPomp* package (Asfaw et al., 2021) as the function `ibpf`. A tutorial (Ning and Ionides, 2023) introduces `ibpf` and validates its correctness on a simple Gaussian example which is tractable using the Kalman filter.

A. The IF2 algorithm and the IEnKF algorithm

In Algorithm 2, we provide the IF2 pseudocode in Ionides et al. (2015). In Algorithm 3, we provide the IEnKF pseudocode in Asfaw et al. (2021).

B. Existing results

The following Dobrushin comparison theorem can be seen in Theorem 3.1 in Rebeschini and Van Handel (2015) and Theorem 8.20 in Georgii (2011).

Theorem 2 (Dobrushin comparison theorem) *Let I be a finite set. Let $\mathbb{S} = \prod_{i\in I} \mathbb{S}^i$ where $\mathbb{S}^i$ is a Polish space for each $i \in I$. Define the coordinate projections $X^i : x \to x^i$ for $x \in \mathbb{S}$ and $i \in I$. For probability measures ρ and $\overline{\rho}$ on $\mathbb{S}$, define*

$$\rho_x^i(A) = \rho(X^i \in A | X^{I\setminus\{i\}} = x^{I\setminus\{i\}}),$$

$$\rho_{\overline{x}}^i(A) = \rho(X^i \in A | X^{I\setminus\{i\}} = \overline{x}^{I\setminus\{i\}}),$$

$$\overline{\rho}_x^i(A) = \overline{\rho}(X^i \in A | X^{I\setminus\{i\}} = x^{I\setminus\{i\}}),$$

$$C_{ij} = \frac{1}{2} \sup_{x,\overline{x}\in\mathbb{S}: x^{I\setminus\{j\}} = \overline{x}^{I\setminus\{j\}}} \|\rho_x^i - \rho_{\overline{x}}^i\| \quad \textit{and} \quad b_j = \sup_{x\in\mathbb{S}} \|\rho_x^j - \overline{\rho}_x^j\|.$$

Algorithm 2 (The IF2 algorithm)

Input:
Simulator for $f_{X_0}(x_0\,;\theta)$
Simulator for $f_{X_n|X_{n-1}}(x_n \mid x_{n-1}\,;\theta)$, $n \in 1{:}N$
Evaluator for $f_{Y_n|X_n}(y_n \mid x_n\,;\theta)$, $n \in 1{:}N$
Data, $y_{1:N}$
Number of iterations, M
Number of particles, J
Initial parameter swarm, $\{\Theta_j^0,\ j \in 1{:}J\}$
Perturbation density, $f_{\Theta_n|\Theta_{n-1}}(\theta \mid \vartheta\,;\sigma)$, $n \in 1{:}N$
Perturbation sequence, $\sigma_{1:M}$

Output: Final parameter swarm $\{\Theta_j^M,\ j \in 1{:}J\}$ and log-likelihood $\widehat{l}_p$

For m in $1:M$
  Draw $\Theta_{0,j}^{F,m} \sim h_0(\theta \mid \Theta_j^{m-1}\,;\sigma_m)$ for $j \in 1{:}J$
  Draw $X_{0,j}^{F,m} \sim f_{X_0}(x_0;\Theta_{0,j}^{F,m})$ for $j \in 1{:}J$
  For n in $1:N$
    Draw $\Theta_{n,j}^{P,m} \sim f_{\Theta_n|\Theta_{n-1}}(\theta_n \mid \Theta_{n-1,j}^{F,m}\,;\sigma_m)$ for $j \in 1{:}J$
    Draw $X_{n,j}^{P,m} \sim f_{X_n|X_{n-1}}(x_n \mid X_{n-1,j}^{F,m}\,;\Theta_{n,j}^{P,m})$ for $j \in 1{:}J$
    Compute $w_{n,j}^m = f_{Y_n|X_n}(y_n \mid X_{n,j}^{P,m}\,;\Theta_{n,j}^{P,m})$ for $j \in 1{:}J$
    Draw $s_{1:J}$ with $\mathrm{Prob}(s_j = i) = w_{n,i}^m \Big/ \sum_{u=1}^n w_{n,u}^m$
    Set $X_{n,j}^{F,m} = X_{n,s_j}^{P,m}$ for $j \in 1{:}J$
    Set $\Theta_{n,j}^{F,m} = \Theta_{n,s_j}^{P,m}$ for $j \in 1{:}J$
  End For
  Set $\Theta_j^m = (\Theta_{n,j}^{F,m})$ for $j \in 1{:}J$
End For
Set $\widehat{l}_p = \sum_{n=1}^N \log(\frac{1}{J}\sum_{j=1}^J w_{n,j}^M)$

If the Dobrushin condition

$$\max_{i\in I}\sum_{j\in I} C_{ij} < 1,$$

holds, then for every $J \subseteq I$,

$$\|\rho - \overline{\rho}\|_J \le \sum_{i\in J}\sum_{j\in I} D_{ij} b_j,$$

where $D := \sum_{n\ge 0} C^n < \infty$.

Theorem 3 (Lemma 4.1 of Rebeschini and Van Handel (2015)) *Let probability measures $\nu, \nu', \mathsf{F}, \mathsf{F}'$ and $\epsilon > 0$ be such that $\nu(A) \ge \epsilon \mathsf{F}(A)$ and $\nu'(A) \ge \epsilon \mathsf{F}'(A)$ for every measurable set A. Then*

$$\|\nu - \nu'\| \le 2(1-\epsilon) + \epsilon\|\mathsf{F} - \mathsf{F}'\|.$$

Algorithm 3 (The IEnKF algorithm)

Input:
Simulator for $f_{X_0}(x_0\,;\theta)$
Simulator for $f_{X_n|X_{n-1}}(x_n \mid x_{n-1}\,;\theta)$, $n \in 1{:}N$
Evaluators for $e_{u,n}(x,\theta)$ and $\mathrm{v}_{u,n}(x,\theta)$, $n \in 1{:}N$
Data, $y_{1:N}$
Number of iterations, M
Number of particles, J
Initial parameter, θ^0
Random walk intensities, $\sigma_{0:N,1:D_\theta}$ where D_θ is the dimension of θ
Cooling fraction in 50 interations, a

Output: Monte Carlo maximum likelihood estimate θ_M and log-likelihood $\widehat{l}_e$

Initialize parameters $\Theta^{F,0}_{N,j} = \theta^0$
For m in $1:M$
  Initialize parameters $\Theta^{F,m}_{0,j} \sim \mathcal{N}(\Theta^{F,m-1}_{N,j}, a^{2m/50}\Sigma_0)$
    where $(\Sigma_n)_{d_\theta,d'_\theta} = \sigma^2_{n,d_\theta}\mathbb{1}_{d_\theta=d'_\theta}$ for $j \in 1{:}J$
  Initialize filter particles $X^{F,m}_{0,j} \sim f_{X_0}(x_0;\Theta^{F,m}_{0,j})$ for $j \in 1{:}J$
  For n in $1:N$
    Draw $\Theta^{P,m}_{n,j} \sim \mathcal{N}(\Theta^{F,m}_{n-1,j}, a^{2m/50}\Sigma_n)$ for $j \in 1{:}J$
    Draw $X^{P,m}_{n,j} \sim f_{X_n|X_{n-1}}(x_n \mid X^{F,m}_{n-1,j}\,;\Theta^{P,m}_{n,j})$ for $j \in 1{:}J$
    Process and ensemble parameter $Z^{P,m}_{n,j} = \begin{pmatrix} X^{P,m}_{n,j} \\ \Theta^{P,m}_{n,j} \end{pmatrix}$ for $j \in 1{:}J$
    Centered process and parameter ensemble $\widetilde{Z}^{P,m}_{n,j} = Z^{P,m}_{n,j} - \frac{1}{J}\sum_{q=1}^{J} Z^{P,m}_{n,q}$ for $j \in 1{:}J$
    Forecast ensemble $\widehat{Y}^{u,m}_{n,j} = e_u(X^{u,P,m}_{n,j}, \Theta^{P,m}_{n,j})$ for $j \in 1{:}J$
    Centered forecast ensemble $\widetilde{Y}^{m}_{n,j} = \widehat{Y}^{m}_{n,j} - \frac{1}{J}\sum_{q=1}^{J} \widehat{Y}^{m}_{n,q}$ for $j \in 1{:}J$
    Forecast measurement variance $R^m_{u,\tilde{u}} = \left(\frac{1}{J}\sum_{j=1}^{J} \mathrm{v}_u(X^{u,P,m}_{n,j}, \Theta^{P,m}_{n,j})\right)_{u,\tilde{u}}$
    Forecast sample covariance $\Sigma^m_Y = \frac{1}{J-1}\sum_{j=1}^{J}(\widetilde{Y}^m_{n,j})(\widetilde{Y}^m_{n,j})^T + R^m$
    Prediction and forecast sample covarince $\Sigma^m_{ZY} = \frac{1}{J-1}\sum_{j=1}^{J}(\widetilde{Z}^{P,m}_{n,j})(\widetilde{Y}^m_{n,j})^T$
    Kalman gain $K^m = \Sigma^m_{ZY}(\Sigma^m_Y)^{-1}$
    Artificial measurement noise $\epsilon^m_{n,j} \sim \mathcal{N}(0,R)$ for $j \in 1{:}J$
    Errors $r^m_{n,j} = \widehat{Y}^m_{n,j} - y_n$ for $j \in 1{:}J$
    Filter update $Z^{F,m}_{n,j} = \begin{pmatrix} X^{F,m}_{n,j} \\ \Theta^{F,m}_{n,j} \end{pmatrix} = Z^{P,m}_{n,j} + K(r^m_{n,j} + \epsilon^m_{n,j})$ for $j \in 1{:}J$
  End For
End For
Set $\theta_M = \frac{1}{J}\sum_{j=1}^{J}(\Theta^{F,M}_{N,j})$
Set $\widehat{l}_e = \sum_{n=1}^{N}\log(\phi(y_n\,;\,\frac{1}{J}\sum_{j=1}^{J}\widehat{Y}^M_{n,j}, \Sigma^m_Y))$ where ϕ is the normal density

Theorem 4 (Lemma 4.2 of Rebeschini and Van Handel (2015)) *Let ρ and ρ' be probability measures and let Λ be a bounded and strictly positive measurable function. Define*

$$\rho_\Lambda(A) := \frac{\int \mathbb{1}_A(x)\Lambda(x)\rho(x)}{\int \Lambda(x)\rho(x)} \quad \textit{and} \quad \rho'_\Lambda(A) := \frac{\int \mathbb{1}_A(x)\Lambda(x)\rho'(x)}{\int \Lambda(x)\rho'(x)}.$$

Then

$$\|\rho_\Lambda - \rho'_\Lambda\| \le 2\frac{\sup_x \Lambda(x)}{\inf_x \Lambda(x)}\|\rho - \rho'\| \quad \textit{and} \quad |||\rho_\Lambda - \rho'_\Lambda||| \le 2\frac{\sup_x \Lambda(x)}{\inf_x \Lambda(x)}|||\rho - \rho'|||.$$

Theorem 5 (Lemma 4.3 of Rebeschini and Van Handel (2015)) *Let I be a finite set and let m be a pseudometric on I. Let $C = (C_{ij})_{i,j\in I}$ be a matrix with nonnegative entries. Suppose that*

$$\max_{i\in I}\sum_{j\in I} e^{m(i,j)}C_{ij} \le c < 1.$$

Then the matrix $D = \sum_{n\ge 0} C^n$ satisfies

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

Theorem 6 (Lemma 4.16 of Rebeschini and Van Handel (2015)) *Let*

$$\mu = \mu^1 \otimes \cdots \otimes \mu^d \quad \textit{and} \quad \nu = \nu^1 \otimes \cdots \otimes \nu^d$$

be product probability measures on

$$\mathbb{S} = \mathbb{S}^1 \times \cdots \times \mathbb{S}^d$$

and let $\Lambda : \mathbb{S} \to \mathbb{R}$ be a bounded and strictly positive measurable function. Let μ_Λ and ν_Λ be probability measures

$$\mu_\Lambda = \frac{\int \mathbb{1}_A(x)\Lambda(x)\mu(dx)}{\int \Lambda(x)\mu(dx)} \quad \textit{and} \quad \nu_\Lambda = \frac{\int \mathbb{1}_A(x)\Lambda(x)\nu(dx)}{\int \Lambda(x)\nu(dx)}.$$

Suppose that there exists a constant $\epsilon > 0$ such that the following holds: for every $i = 1, \cdots, d$, there is a measurable function $\Lambda^i : \mathbb{S} \to \mathbb{R}$ such that

$$\epsilon\Lambda^i(x) \le \Lambda \le \epsilon^{-1}\Lambda^i(x),$$

for all $x \in \mathbb{S}$ such that $\Lambda^i(x) = \Lambda^i(\tilde{x})$ whenever $x^{\{1,\dots,d\}\setminus\{i\}} = \tilde{x}^{\{1,\dots,d\}\setminus\{i\}}$. Then

$$\|\mu_\Lambda - \nu_\Lambda\| \le \frac{2}{\epsilon^2}\sum_{i=1}^d \|\mu^i - \nu^i\|.$$

Theorem 7 (Corollary 4.21 of Rebeschini and Van Handel (2015)) *For any subset of blocks $\mathcal{L} \subseteq \mathcal{K}$, we have*

$$\left\|\left|\bigotimes_{K\in\mathcal{L}} B^K\mu - \bigotimes_{K\in\mathcal{L}} B^K\mathsf{S}^J\mu\right|\right\| \le \frac{4\,\mathrm{card}(\mathcal{L})}{\sqrt{J}},$$

for every probability measure μ.

C. Preparation

For any probability measure μ_{s-1} on $\mathbb{X} \times \mathbb{\Theta}$ at time $s-1$ for any integer $s \geq 1$, any vertex $v \in V$, any block $K' \in \mathcal{K}$, and any set $A^v = A^v_x \times A^v_\theta$ where $A^v_x \subseteq \mathbb{X}^v$ and $A^v_\theta \subseteq \mathbb{\Theta}^v$, we define the following quantities with χ indicating (x, θ):

$$\bullet\ \mu^v_{\chi_{s-1}}(A^v_x \times A^v_\theta) \tag{38}$$
$$:= \mathbf{P}^{\mu_{s-1}}\Big[X^v_{s-1} \in A^v_x, \Theta^v_{s-1} \in A^v_\theta \;\Big|\; X^{V\setminus\{v\}}_{s-1} = x^{V\setminus\{v\}}_{s-1}, \Theta^{V\setminus\{v\}}_{s-1} = \theta^{V\setminus\{v\}}_{s-1}\Big],$$

$$\bullet\ \mu^v_{\chi_{s-1},\chi_s}(A^v_x \times A^v_\theta) \tag{39}$$
$$:= \mathbf{P}^{\mu_{s-1}}\Big[X^v_{s-1} \in A^v_x, \Theta^v_{s-1} \in A^v_\theta \;\Big|\; X^{V\setminus\{v\}}_{s-1} = x^{V\setminus\{v\}}_{s-1}, \Theta^{V\setminus\{v\}}_{s-1} = \theta^{V\setminus\{v\}}_{s-1}, X_s = x_s, \Theta_s = \theta_s\Big]$$
$$= \frac{\begin{array}{r}\int \mathbb{1}_{A^v}(x^v_{s-1}, \theta^v_{s-1}) \prod_{\omega\in N(v)} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(dx^v_{s-1}, d\theta^v_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(dx^v_{s-1}, d\theta^v_{s-1})\end{array}},$$

$$\bullet\ \mu^v_{\overline{\chi}_{s-1},\chi_s}(A^v_x \times A^v_\theta) \tag{40}$$
$$:= \mathbf{P}^{\mu_{s-1}}\Big[X^v_{s-1} \in A^v_x, \Theta^v_{s-1} \in A^v_\theta \;\Big|\; X^{V\setminus\{v\}}_{s-1} = \overline{x}^{V\setminus\{v\}}_{s-1}, \Theta^{V\setminus\{v\}}_{s-1} = \overline{\theta}^{V\setminus\{v\}}_{s-1}, X_s = x_s, \Theta_s = \theta_s\Big]$$
$$= \frac{\begin{array}{r}\int \mathbb{1}_{A^v}(\overline{x}^v_{s-1}, \overline{\theta}^v_{s-1}) \prod_{\omega\in N(v)} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid \overline{x}_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \overline{\theta}_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(d\overline{x}^v_{s-1}, d\overline{\theta}^v_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid \overline{x}_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \overline{\theta}_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(d\overline{x}^v_{s-1}, d\overline{\theta}^v_{s-1})\end{array}},$$

$$\bullet\ \mu^{v,K'}_{\chi_{s-1},\chi_s}(A^v_x \times A^v_\theta) \tag{41}$$
$$:= \mathbf{P}^{\mu_{s-1}}\Big[X^v_{s-1} \in A^v_x, \Theta^v_{s-1} \in A^v_\theta \;\Big|\; X^{V\setminus\{v\}}_{s-1} = x^{V\setminus\{v\}}_{s-1}, \Theta^{V\setminus\{v\}}_{s-1} = \theta^{V\setminus\{v\}}_{s-1}, X^{K'}_s = x^{K'}_s, \Theta^{K'}_s = \theta^{K'}_s\Big]$$
$$= \frac{\begin{array}{r}\int \mathbb{1}_{A^v}(x^v_{s-1}, \theta^v_{s-1}) \prod_{\omega\in N(v)\cap K'} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(dx^v_{s-1}, d\theta^v_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)\cap K'} f_{X^\omega_s|X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s|\Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma)\mu^v_{\chi_{s-1}}(dx^v_{s-1}, d\theta^v_{s-1})\end{array}},$$

$$\bullet\ \mu^{v,K'}_{\overline{\chi}_{s-1},\chi_s}(A^v_x \times A^v_\theta) \tag{42}$$
$$:= \mathbf{P}^{\mu_{s-1}}\Big[X^v_{s-1} \in A^v_x, \Theta^v_{s-1} \in A^v_\theta \;\Big|\; X^{V\setminus\{v\}}_{s-1} = \overline{x}^{V\setminus\{v\}}_{s-1}, \Theta^{V\setminus\{v\}}_{s-1} = \overline{\theta}^{V\setminus\{v\}}_{s-1},$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad X_s^{K'} = x_s^{K'}, \Theta_s^{K'} = \theta_s^{K'}\Big] \\
&\int \mathbb{1}_{A^v}(\overline{x}_{s-1}^v, \overline{\theta}_{s-1}^v) \prod_{\omega \in N(v)\cap K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid \overline{x}_{s-1}\,; \theta_s^\omega) \\
&\qquad\qquad \times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \overline{\theta}_{s-1}\,; \sigma)\mu_{\chi_{s-1}}^v(d\overline{x}_{s-1}^v, d\overline{\theta}_{s-1}^v) \\
&= \frac{}{\int \prod_{\omega \in N(v)\cap K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid \overline{x}_{s-1}\,; \theta_s^\omega)} \\
&\qquad\qquad \times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \overline{\theta}_{s-1}\,; \sigma)\mu_{\chi_{s-1}}^v(d\overline{x}_{s-1}^v, d\overline{\theta}_{s-1}^v)
\end{aligned}
$$

Based on the above quantities, with $v' \in V$ we define

$$
\bullet\ C_{vv'}^{\mu_{s-1}} := \frac{1}{2} \sup_{\substack{x_s \in \mathbb{X} \\ \theta_s \in \mathbb{\Theta}}} \sup_{\substack{x_{s-1}, \overline{x}_{s-1} \in \mathbb{X}: x_{s-1}^{V\setminus\{v'\}} = \overline{x}_{s-1}^{V\setminus\{v'\}} \\ \theta_{s-1}, \overline{\theta}_{s-1} \in \mathbb{\Theta}: \theta_{s-1}^{V\setminus\{v'\}} = \overline{\theta}_{s-1}^{V\setminus\{v'\}}}} \|\mu_{\chi_{s-1},\chi_s}^v - \mu_{\overline{\chi}_{s-1},\chi_s}^v\|, \tag{43}
$$

$$
\bullet\ \widetilde{C}_{vv'}^{\mu_{s-1}} := \frac{1}{2} \sup_{K' \in \mathcal{K}} \sup_{\substack{x_s \in \mathbb{X} \\ \theta_s \in \mathbb{\Theta}}} \sup_{\substack{x_{s-1}, \overline{x}_{s-1} \in \mathbb{X}: x_{s-1}^{V\setminus\{v'\}} = \overline{x}_{s-1}^{V\setminus\{v'\}} \\ \theta_{s-1}, \overline{\theta}_{s-1} \in \mathbb{\Theta}: \theta_{s-1}^{V\setminus\{v'\}} = \overline{\theta}_{s-1}^{V\setminus\{v'\}}}} \|\mu_{\chi_{s-1},\chi_s}^{v,K'} - \mu_{\overline{\chi}_{s-1},\chi_s}^{v,K'}\|, \tag{44}
$$

and with β being a finite positive constant we further define

$$
\bullet\ \mathrm{Corr}(\mu_{s-1}, \beta) := \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v')} C_{vv'}^{\mu_{s-1}}. \tag{45}
$$

$$
\bullet\ \widetilde{\mathrm{Corr}}(\mu_{s-1}, \beta) := \max_{v \in V} \sum_{v' \in V} e^{\beta d(v,v')} \widetilde{C}_{vv'}^{\mu_{s-1}}. \tag{46}
$$

Then, for F_s defined in (14) and $\widetilde{\mathsf{F}}_s$ defined in (18), any probability measure ν_{s-1} on $\mathbb{X} \times \mathbb{\Theta}$ at time $s-1$ for any integer $s \geq 1$, and any set $A \subseteq \mathbb{X} \times \mathbb{\Theta}$, we have

$$
\begin{aligned}
&\bullet\ (\mathsf{F}_s\nu_{s-1})(A) \qquad\qquad (47)\\
&\int \mathbb{1}_A(x_s, \theta_s) \prod_{\omega \in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma) \\
&\qquad\qquad \times f_{Y_s^\omega|X_s^\omega}(Y_s^\omega \mid x_s^\omega\,; \theta_s^\omega)\nu_{s-1}(dx_{s-1}, d\theta_{s-1})\psi(dx_s)\lambda(d\theta_s) \\
&= \frac{}{\int \prod_{\omega \in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma)} \\
&\qquad\qquad \times f_{Y_s^\omega|X_s^\omega}(Y_s^\omega \mid x_s^\omega\,; \theta_s^\omega)\nu_{s-1}(dx_{s-1}, d\theta_{s-1})\psi(dx_s)\lambda(d\theta_s)
\end{aligned}
$$

$$
\begin{aligned}
&\bullet\ (\widetilde{\mathsf{F}}_s\nu_{s-1})(A) \qquad\qquad (48)\\
&\int \mathbb{1}_A(x_s, \theta_s) \prod_{K \in \mathcal{K}} \Big[\int \prod_{\omega \in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) \\
&\qquad \times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma) f_{Y_s^\omega|X_s^\omega}(Y_s^\omega \mid x_s^\omega\,; \theta_s^\omega)\nu_{s-1}(dx_{s-1}, d\theta_{s-1})\Big] \\
&\qquad\qquad\qquad\qquad \times \psi(dx_s)\lambda(d\theta_s) \\
&= \frac{}{\int \prod_{K \in \mathcal{K}} \Big[\int \prod_{\omega \in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega)} \\
&\qquad \times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma) f_{Y_s^\omega|X_s^\omega}(Y_s^\omega \mid x_s^\omega\,; \theta_s^\omega)\nu_{s-1}(dx_{s-1}, d\theta_{s-1})\Big] \\
&\qquad\qquad\qquad\qquad \times \psi(dx_s)\lambda(d\theta_s)
\end{aligned}
$$

According to the definition of $\mu_{\chi_{s-1}}^v$ given in (38), for any $K \in \mathcal{K}$ and $v \in K$, we have

$$
\bullet\ (\mathsf{F}_s\nu_{s-1})_{\chi_s}^v(A^v) \tag{49}
$$

$$
\begin{aligned}
&\int \mathbb{1}_{A^v}(x_s^v,\theta_s^v)\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\psi^v(dx_s)\lambda^v(d\theta_s)\\
=&\frac{}{\int\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)}\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\psi^v(dx_s)\lambda^v(d\theta_s)
\end{aligned},
$$

$$
\begin{aligned}
&\bullet\ (\widetilde{\mathsf{F}}_s\nu_{s-1})^v_{\chi_s}(A^v) \qquad (50)\\
&\int \mathbb{1}_{A^v}(x_s^v,\theta_s^v)\prod_{\omega\in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\psi^v(dx_s)\lambda^v(d\theta_s)\\
=&\frac{}{\int\prod_{\omega\in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)}\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\psi^v(dx_s)\lambda^v(d\theta_s)
\end{aligned}.
$$

By (49) and (50), according to the definition of $\mu^v_{\chi_{s-1},\chi_s}(A_x^v\times A_\theta^v)$ given in (39), for any $K\in\mathcal{K}$ and $v\in K$, we have

$$
\begin{aligned}
&\bullet\ (\mathsf{F}_s\nu_{s-1})^v_{\chi_s,\chi_{s+1}}(A^v) \qquad (51)\\
&\int \mathbb{1}_{A^v}(x_s^v,\theta_s^v)\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^v(dx_s^v)\lambda^v(d\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\\
=&\frac{}{\int\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)}\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^v(dx_s^v)\lambda^v(d\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned},
$$

$$
\begin{aligned}
&\bullet\ (\widetilde{\mathsf{F}}_s\nu_{s-1})^v_{\chi_s,\chi_{s+1}}(A^v) \qquad (52)\\
&\int \mathbb{1}_{A^v}(x_s^v,\theta_s^v)\prod_{\omega\in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^v(dx_s^v)\lambda^v(d\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\\
=&\frac{}{\int\prod_{\omega\in K} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)}\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid x_s^v\,;\theta_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^v(dx_s^v)\lambda^v(d\theta_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned}.
$$

By (49) and (50), according to the definition of $\mu^v_{\overline{\chi}_{s-1},\chi_s}$ given in (40), for any $K\in\mathcal{K}$ and $v\in K$, we have

$$
\begin{aligned}
&\bullet\ (\mathsf{F}_s\nu_{s-1})^v_{\overline{\chi}_s,\chi_{s+1}}(A^v) \qquad (53)\\
&\int \mathbb{1}_{A^v}(\overline{x}_s^v,\overline{\theta}_s^v)\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,;\overline{\theta}_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid \overline{x}_s^v\,;\overline{\theta}_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid \overline{x}_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,;\sigma)\psi^v(d\overline{x}_s^v)\lambda^v(d\overline{\theta}_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})\\
=&\frac{}{\int\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,;\overline{\theta}_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,;\sigma)}\\
&\qquad\qquad \times f_{Y_s^v|X_s^v}(Y_s^v \mid \overline{x}_s^v\,;\overline{\theta}_s^v)\prod_{u\in N(v)} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid \overline{x}_s\,;\theta_{s+1}^u)\\
&\qquad\qquad \times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,;\sigma)\psi^v(d\overline{x}_s^v)\lambda^v(d\overline{\theta}_s^v)\nu_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned},
$$

$$
\bullet\ (\widetilde{\mathsf{F}}_s\nu_{s-1})^v_{\overline{\chi}_s,\chi_{s+1}}(A^v) \qquad (54)
$$

$$
\begin{aligned}
&\int \mathbb{1}_{A^v}(\overline{x}_s^v, \overline{\theta}_s^v) \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma) \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v)} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
= &\frac{}{\int \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma)} \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v)} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1})
\end{aligned}
$$

By (49) and (50), according to the definition of $\mu_{\chi_{s-1}, \chi_s}^{v, K'}$ given in (41), for any $K, K' \in \mathcal{K}$ and $v \in K$, we have

$$
\begin{aligned}
&\bullet\ (\mathsf{F}_s \nu_{s-1})_{\chi_s, \chi_{s+1}}^{v, K'}(A^v) \qquad (55) \\
&\int \mathbb{1}_{A^v}(x_s^v, \theta_s^v) \prod_{\omega \in V} f_{X_s^\omega | X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma) \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid x_s^v\,; \theta_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid x_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \theta_s\,; \sigma) \psi^v(dx_s^v) \lambda^v(d\theta_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
= &\frac{}{\int \prod_{\omega \in V} f_{X_s^\omega | X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma)} \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid x_s^v\,; \theta_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid x_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \theta_s\,; \sigma) \psi^v(dx_s^v) \lambda^v(d\theta_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
&\bullet\ (\widetilde{\mathsf{F}}_s \nu_{s-1})_{\chi_s, \chi_{s+1}}^{v, K'}(A^v) \qquad (56) \\
&\int \mathbb{1}_{A^v}(x_s^v, \theta_s^v) \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma) \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid x_s^v\,; \theta_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid x_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \theta_s\,; \sigma) \psi^v(dx_s^v) \lambda^v(d\theta_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
= &\frac{}{\int \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(x_s^\omega \mid x_{s-1}\,; \theta_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,; \sigma)} \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid x_s^v\,; \theta_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid x_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \theta_s\,; \sigma) \psi^v(dx_s^v) \lambda^v(d\theta_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1})
\end{aligned}
$$

By (49) and (50), according to the definition of $\mu_{\overline{\chi}_{s-1}, \chi_s}^{v, K'}$ given in (42), for any $K, K' \in \mathcal{K}$ and $v \in K$, we have

$$
\begin{aligned}
&\bullet\ (\mathsf{F}_s \nu_{s-1})_{\overline{\chi}_s, \chi_{s+1}}^{v, K'}(A^v) \qquad (57) \\
&\int \mathbb{1}_{A^v}(\overline{x}_s^v, \overline{\theta}_s^v) \prod_{\omega \in V} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma) \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
= &\frac{}{\int \prod_{\omega \in V} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma)} \\
&\qquad \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\
&\qquad\qquad \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1}) \\
&\bullet\ (\widetilde{\mathsf{F}}_s \nu_{s-1})_{\overline{\chi}_s, \chi_{s+1}}^{v, K'}(A^v) \qquad (58)
\end{aligned}
$$

$$= \frac{\begin{array}{r}\int \mathbb{1}_{A^v}(\overline{x}_s^v, \overline{\theta}_s^v) \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma) \\ \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\ \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega \in K} f_{X_s^\omega | X_{s-1}}(\overline{x}_s^\omega \mid x_{s-1}\,; \overline{\theta}_s^\omega) f_{\Theta_s^\omega | \Theta_{s-1}}(\overline{\theta}_s^\omega \mid \theta_{s-1}\,; \sigma) \\ \times f_{Y_s^v | X_s^v}(Y_s^v \mid \overline{x}_s^v\,; \overline{\theta}_s^v) \prod_{u \in N(v) \cap K'} f_{X_{s+1}^u | X_s}(x_{s+1}^u \mid \overline{x}_s\,; \theta_{s+1}^u) \\ \times f_{\Theta_{s+1}^u | \Theta_s}(\theta_{s+1}^u \mid \overline{\theta}_s\,; \sigma) \psi^v(d\overline{x}_s^v) \lambda^v(d\overline{\theta}_s^v) \nu_{s-1}(dx_{s-1}, d\theta_{s-1})\end{array}}.$$

D. Proofs for bounding bias

Lemma 2 *Under condition* (27)*, for* β *defined in* (28)*, the following holds:*

(i) $(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r} \Delta^2 \le \frac{1}{16}$*;*

(ii) $(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)} \Delta \le (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r} \Delta^2$*.*

Proof By the definition of β, we have

$$e^\beta = \left(\frac{1}{16 \Delta_{\mathcal{K}} \Delta^2 (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})} \right)^{\frac{1}{2r}}.$$

That is,

$$e^{2r\beta} = \frac{1}{16 \Delta_{\mathcal{K}} \Delta^2 (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})}.$$

Since $r, \Delta, \Delta_{\mathcal{K}} \ge 1$ and $0 \le \epsilon_x, \epsilon_\theta(\sigma) \le 1$, we have

$$(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r} \Delta^2 \le \frac{(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) \Delta^2}{16 \Delta_{\mathcal{K}} \Delta^2 (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})} \le \frac{1}{16}$$

$$(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)} \Delta \le (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{2\beta r} \Delta \le (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r} \Delta^2.$$

■

Lemma 3 *Under Assumption 1 in Section 3.3, for any* $n \ge 0$*, we have*

$$\widetilde{\mathrm{Corr}}(\widetilde{\pi}_n, \beta) < \frac{1}{8},$$

where $\widetilde{\mathrm{Corr}}(\,\cdot\,, \beta)$ *is defined in* (46) *and* β *is given in* (28)*.*

Proof Since $\widetilde{\pi}_0 = \pi_0 = \delta_x \delta_\theta$ which is non-random, by the definition of $\widetilde{\mathrm{Corr}}(\,\cdot\,, \beta)$ given in (46), we have $\widetilde{\mathrm{Corr}}(\widetilde{\pi}_0, \beta) = 0$. In the following, we prove by the method of induction and assume that for $n \ge 1$

$$\widetilde{\mathrm{Corr}}(\widetilde{\pi}_{n-1}, \beta) < \frac{1}{8}.$$

Let $K, K' \in \mathcal{K}$, $v \in K$, $v' \in V$ and $v \neq v'$. Let $x_n, \overline{x}_n \in \mathbb{X}$ such that $x_n^{V\setminus\{v'\}} = \overline{x}_n^{V\setminus\{v'\}}$. Let $\theta_n, \overline{\theta}_n \in \mathbb{\Theta}$ such that $\theta_n^{V\setminus\{v'\}} = \overline{\theta}_n^{V\setminus\{v'\}}$. Define $I = (\{n-1\} \times V) \cup (n, v)$ and $\mathbb{S} = (\mathbb{X} \times \mathbb{\Theta}) \times (\mathbb{X}^v \times \mathbb{\Theta}^v)$. Define

$$\rho(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}, \theta_{n-1}, x_n^v, \theta_n^v) \prod_{\omega\in K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,; \sigma) \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid x_n\,; \theta_{n+1}^u) \\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \theta_n\,; \sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v) \\ \times \psi^v(dx_n^v)\lambda^v(d\theta_n^v)\widetilde{\pi}_{n-1}(dx_{n-1}, d\theta_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,; \sigma) \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid x_n\,; \theta_{n+1}^u) \\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \theta_n\,; \sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v) \\ \times \psi^v(dx_n^v)\lambda^v(d\theta_n^v)\widetilde{\pi}_{n-1}(dx_{n-1}, d\theta_{n-1})\end{array}}$$

and

$$\overline{\rho}(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}, \theta_{n-1}, \overline{x}_n^v, \overline{\theta}_n^v) \prod_{\omega\in K} f_{X_n^\omega|X_{n-1}}(\overline{x}_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\overline{\theta}_n^\omega \mid \theta_{n-1}\,; \sigma) \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid \overline{x}_n\,; \theta_{n+1}^u) \\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \overline{\theta}_n\,; \sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,; \overline{\theta}_n^v) \\ \times \psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\widetilde{\pi}_{n-1}(dx_{n-1}, d\theta_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in K} f_{X_n^\omega|X_{n-1}}(\overline{x}_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\overline{\theta}_n^\omega \mid \theta_{n-1}\,; \sigma) \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid \overline{x}_n\,; \theta_{n+1}^u) \\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \overline{\theta}_n\,; \sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,; \overline{\theta}_n^v) \\ \times \psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\widetilde{\pi}_{n-1}(dx_{n-1}, d\theta_{n-1})\end{array}}.$$

Then according to the definitions of $(\widetilde{\mathsf{F}}_s\nu_{s-1})^{v,K'}_{\chi_s,\chi_{s+1}}$ in (56) and $(\widetilde{\mathsf{F}}_s\nu_{s-1})^{v,K'}_{\overline{\chi}_s,\chi_{s+1}}$ in (58) for $s \geq 1$, we have

$$\|(\widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1})^{v,K'}_{\chi_n,\chi_{n+1}} - (\widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1})^{v,K'}_{\overline{\chi}_n,\chi_{n+1}}\| = \|\rho - \overline{\rho}\|_{(n,v)}.$$

In order to use Theorem 2 (Dobrushin comparison theorem) in Appendix B to bound $\|\rho - \overline{\rho}\|_{(n,v)}$, we need to bound C_{ij} and b_i with $i = (k, t)$ and $j = (k', t')$. Set

$$\rho^i = \rho^i_{(x_{n-1},\theta_{n-1},x_n^v,\theta_n^v)} \quad \text{and} \quad \overline{\rho}^i = \overline{\rho}^i_{(x_{n-1},\theta_{n-1},\overline{x}_n^v,\overline{\theta}_n^v)},$$

whose definitions are given in Theorem 2 in Appendix B. Note that

$$\overline{\rho}^i \neq \rho^i_{(\overline{x}_{n-1},\overline{\theta}_{n-1},\overline{x}_n^v,\overline{\theta}_n^v)}.$$

We display our discussions as follows:

- When $k = n-1$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}^t, \theta_{n-1}^t) \prod_{\omega\in N(t)\cap K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,; \sigma)\widetilde{\pi}^t_{n-1}(dx_{n-1}^t, d\theta_{n-1}^t)\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(t)\cap K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,; \theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,; \sigma)\widetilde{\pi}^t_{n-1}(dx_{n-1}^t, d\theta_{n-1}^t)\end{array}}.$$

We can see that $\rho^i = \widetilde{\pi}^{t,K}_{\chi_{n-1},\chi_n}$, by the definition of $\widetilde{\pi}^{t,K}_{\chi_{n-1},\chi_n}$ given in (41). Therefore, if $k' = n-1$, by the definition of $\widetilde{C}^{\widetilde{\pi}_{n-1}}_{tt'}$ given in (44), we know that $C_{ij} \leq \widetilde{C}^{\widetilde{\pi}_{n-1}}_{tt'}$. Note that

$$\rho^i(A) \geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x^t_{n-1},\theta^t_{n-1}) \prod_{\omega\in N(t)\cap(K\setminus\{v\})} f_{X^\omega_n|X_{n-1}}(x^\omega_n \mid x_{n-1}\,;\theta^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\theta^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(t)\cap(K\setminus\{v\})} f_{X^\omega_n|X_{n-1}}(x^\omega_n \mid x_{n-1}\,;\theta^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\theta^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}},$$

and then we have $C_{ij} \leq 1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $k' = n$ and $v \in N(t)$ by Theorem 3 in Appendix B, and $C_{ij} = 0$ otherwise. Recalling that at the beginning of this proof we set $v \in K$, $v' \in V$, $v \neq v'$, $x_n, \overline{x}_n \in \mathbb{X}$ such that $x_n^{V\setminus\{v'\}} = \overline{x}_n^{V\setminus\{v'\}}$, and $\theta_n, \overline{\theta}_n \in \mathbb{\Theta}$ such that $\theta_n^{V\setminus\{v'\}} = \overline{\theta}_n^{V\setminus\{v'\}}$, then we have $\rho^i = \overline{\rho}^i$ if $v' \notin N(t)\cap K$. Furthermore, note that

$$\rho^i(A) \geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x^t_{n-1},\theta^t_{n-1}) \prod_{\omega\in N(t)\cap(K\setminus\{v'\})} f_{X^\omega_n|X_{n-1}}(x^\omega_n \mid x_{n-1}\,;\theta^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\theta^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(t)\cap(K\setminus\{v'\})} f_{X^\omega_n|X_{n-1}}(x^\omega_n \mid x_{n-1}\,;\theta^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\theta^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}},$$

and

$$\begin{aligned}\overline{\rho}^i(A) &= \frac{\begin{array}{r}\int \mathbb{1}_A(x^t_{n-1},\theta^t_{n-1}) \prod_{\omega\in N(t)\cap K} f_{X^\omega_n|X_{n-1}}(\overline{x}^\omega_n \mid x_{n-1}\,;\overline{\theta}^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\overline{\theta}^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(t)\cap K} f_{X^\omega_n|X_{n-1}}(\overline{x}^\omega_n \mid x_{n-1}\,;\overline{\theta}^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\overline{\theta}^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}} \\ &\geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x^t_{n-1},\theta^t_{n-1}) \prod_{\omega\in N(t)\cap(K\setminus\{v'\})} f_{X^\omega_n|X_{n-1}}(\overline{x}^\omega_n \mid x_{n-1}\,;\overline{\theta}^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\overline{\theta}^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(t)\cap(K\setminus\{v'\})} f_{X^\omega_n|X_{n-1}}(\overline{x}^\omega_n \mid x_{n-1}\,;\overline{\theta}^\omega_n) \\ \times f_{\Theta^\omega_n|\Theta_{n-1}}(\overline{\theta}^\omega_n \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^t_{n-1}(dx^t_{n-1},d\theta^t_{n-1})\end{array}}.\end{aligned}$$

Hence, we have $b_i = 0$ if $v' \notin N(t)\cap K$, and by Theorem 3 in Appendix B we have $b_i = 2(1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)$ otherwise.

- When $k = n$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x^v_n,\theta^v_n) f_{X^v_n|X_{n-1}}(x^v_n \mid x_{n-1}\,;\theta^v_n) f_{\Theta^v_n|\Theta_{n-1}}(\theta^v_n \mid \theta_{n-1}\,;\sigma) \\ \times \prod_{u\in N(v)\cap K'} f_{X^u_{n+1}|X_n}(x^u_{n+1} \mid x_n\,;\theta^u_{n+1}) f_{\Theta^u_{n+1}|\Theta_n}(\theta^u_{n+1} \mid \theta_n\,;\sigma) \\ \times f_{Y^v_n|X^v_n}(Y^v_n \mid x^v_n\,;\theta^v_n)\psi^v(dx^v_n)\lambda^v(d\theta^v_n)\end{array}}{\begin{array}{r}\int f_{X^v_n|X_{n-1}}(x^v_n \mid x_{n-1}\,;\theta^v_n) f_{\Theta^v_n|\Theta_{n-1}}(\theta^v_n \mid \theta_{n-1}\,;\sigma) \\ \times \prod_{u\in N(v)\cap K'} f_{X^u_{n+1}|X_n}(x^u_{n+1} \mid x_n\,;\theta^u_{n+1}) f_{\Theta^u_{n+1}|\Theta_n}(\theta^u_{n+1} \mid \theta_n\,;\sigma) \\ \times f_{Y^v_n|X^v_n}(Y^v_n \mid x^v_n\,;\theta^v_n)\psi^v(dx^v_n)\lambda^v(d\theta^v_n)\end{array}}$$

$$\geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x_n^v,\theta_n^v)\prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid x_n\,;\theta_{n+1}^u)\\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \theta_n\,;\sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,;\theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}}{\begin{array}{r}\int \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid x_n\,;\theta_{n+1}^u)\\ \times f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \theta_n\,;\sigma) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,;\theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}}.$$

Then we have $C_{ij} \leq 1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $k' = n-1$ and $t' \in N(v)$ by Theorem 3 in Appendix B, and $C_{ij} = 0$ otherwise. Note that $b_i = 0$ if $v' \notin \cup_{\omega\in N(v)\cap K'} N(\omega)$. Furthermore,

$$\rho^i(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \frac{\begin{array}{r}\int \mathbb{1}_A(x_n^v,\theta_n^v) f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,;\theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,;\theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}}{\begin{array}{r}\int f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,;\theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,;\theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}}$$

and

$$\begin{aligned}\overline{\rho}^i(A) = &\frac{\begin{array}{r}\int \mathbb{1}_A(\overline{x}_n^v,\overline{\theta}_n^v) f_{X_n^v|X_{n-1}}(\overline{x}_n^v \mid x_{n-1}\,;\overline{\theta}_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\overline{\theta}_n^v \mid \theta_{n-1}\,;\sigma)\\ \times \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid \overline{x}_n\,;\theta_{n+1}^u) f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \overline{\theta}_n\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,;\overline{\theta}_n^v)\psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\end{array}}{\begin{array}{r}\int f_{X_n^v|X_{n-1}}(\overline{x}_n^v \mid x_{n-1}\,;\overline{\theta}_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\overline{\theta}_n^v \mid \theta_{n-1}\,;\sigma)\\ \times \prod_{u\in N(v)\cap K'} f_{X_{n+1}^u|X_n}(x_{n+1}^u \mid \overline{x}_n\,;\theta_{n+1}^u) f_{\Theta_{n+1}^u|\Theta_n}(\theta_{n+1}^u \mid \overline{\theta}_n\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,;\overline{\theta}_n^v)\psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\end{array}}\\ \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} &\frac{\begin{array}{r}\int \mathbb{1}_A(\overline{x}_n^v,\overline{\theta}_n^v) f_{X_n^v|X_{n-1}}(\overline{x}_n^v \mid x_{n-1}\,;\overline{\theta}_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\overline{\theta}_n^v \mid \theta_{n-1}\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,;\overline{\theta}_n^v)\psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\end{array}}{\begin{array}{r}\int f_{X_n^v|X_{n-1}}(\overline{x}_n^v \mid x_{n-1}\,;\overline{\theta}_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\overline{\theta}_n^v \mid \theta_{n-1}\,;\sigma)\\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid \overline{x}_n^v\,;\overline{\theta}_n^v)\psi^v(d\overline{x}_n^v)\lambda^v(d\overline{\theta}_n^v)\end{array}}.\end{aligned}$$

By Theorem 3 in Appendix B, we have $b_i = 2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})$ if $v' \in \cup_{\omega\in N(v)\cap K'} N(\omega)$.

Define the matrix $(C_{ij}(v))_{i,j\in I}$ whose entries are given below:

$$\begin{aligned}C_{(n-1,t)(n-1,t')}(v) =& \widetilde{C}^{\widetilde{\pi}_{n-1}}_{tt'},\\ C_{(n-1,t)(n,v)}(v) =& C_{(n,v)(n-1,t)}(v) = (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)\mathbb{1}_{t\in N(v)},\\ C_{(n,v)(n,v)}(v) =& 0.\end{aligned}$$

In sum, we have that

$$\begin{aligned}\sum_{(k',t')\in I} e^{\beta|k-k'|}e^{\beta d(t,t')}C_{(k,t)(k',t')} &\leq \sum_{(k',t')\in I} e^{\beta|k-k'|}e^{\beta d(t,t')}C_{(k,t)(k',t')}(v)\\ &\leq \widetilde{\mathrm{Corr}}(\widetilde{\pi}_{n-1},\beta) + (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)e^{\beta(r+1)}\Delta\\ &< \frac{1}{8}+\frac{1}{16}\end{aligned}$$

$$< \frac{1}{2}, \tag{59}$$

where the third inequality is by the assumption of the induction method and Lemma 2 in Appendix D.

Next, applying Theorem 2 (Dobrushin comparison theorem) in Appendix B, we obtain that

$$\begin{aligned}\|\rho - \overline{\rho}\|_{(n,v)} \leq & 2(1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)\mathbb{1}_{\{v'\in K\}} \sum_{t'\in N(v')} D_{(n,v)(n-1,t')}(v) \\ & + 2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\mathbb{1}_{\{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)\}} D_{(n,v)(n,v)}(v),\end{aligned}$$

where $D(v) = \sum_{s\geq 0}[C(v)]^s$. Therefore,

$$\begin{aligned}&\frac{1}{2}\sup_{K'\in\mathcal{K}} \sup_{\substack{x_{n+1}\in\mathbb{X} \\ \theta_{n+1}\in\Theta}} \sup_{\substack{x_n,\overline{x}_n\in\mathbb{X}: x_n^{V\setminus\{v'\}}=\overline{x}_n^{V\setminus\{v'\}} \\ \theta_n,\overline{\theta}_n\in\Theta: \theta_n^{V\setminus\{v'\}}=\overline{\theta}_n^{V\setminus\{v'\}}}} \|(\widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1})^{v,K'}_{\chi_n,\chi_{n+1}} - (\widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1})^{v,K'}_{\overline{\chi}_n,\chi_{n+1}}\| \\ &\quad \leq (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)\mathbb{1}_{\{v'\in K\}} \sum_{t'\in N(v')} D_{(n,v)(n-1,t')}(v) \\ &\qquad + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\mathbb{1}_{\{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)\}} D_{(n,v)(n,v)}(v).\end{aligned}$$

By the definition of $\widetilde{C}^{\mu_{s-1}}_{vv'}$ in equation (44), we have

$$\begin{aligned}\widetilde{C}^{\widetilde{\pi}_n}_{vv'} \leq & (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)\mathbb{1}_{\{v'\in K\}} \sum_{t'\in N(v')} D_{(n,v)(n-1,t')}(v) \\ & + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\mathbb{1}_{\{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)\}} D_{(n,v)(n,v)}(v).\end{aligned}$$

Note that

$$\begin{aligned}\sum_{v'\in V} e^{\beta d(v,v')}\widetilde{C}^{\widetilde{\pi}_n}_{vv'} \leq & (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) \sum_{v'\in K} e^{\beta d(v,v')} \sum_{t'\in N(v')} D_{(n,v)(n-1,t')}(v) \\ & + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) \sum_{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)} e^{\beta d(v,v')} D_{(n,v)(n,v)}(v) \\ \leq & (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) \sum_{v'\in K}\sum_{t'\in N(v')} e^{\beta d(v,v')} D_{(n,v)(n-1,t')}(v) \\ & + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) \sum_{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)} e^{\beta d(v,v')} D_{(n,v)(n,v)}(v) \\ \leq & (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) \sum_{v'\in K}\sum_{t'\in N(v')} e^{\beta d(v,t')} e^{\beta d(t',v')} D_{(n,v)(n-1,t')}(v) \\ & + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) \sum_{v'\in \cup_{\omega\in N(v)\cap K'} N(\omega)} e^{\beta d(v,\omega)} e^{\beta d(\omega,v')} D_{(n,v)(n,v)}(v) \\ \leq & (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) \sum_{v'\in K}\sum_{t'\in N(v')} e^{\beta d(v,t')} e^{\beta r} D_{(n,v)(n-1,t')}(v)\end{aligned}$$

$$
\begin{aligned}
&+ (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 D_{(n,v)(n,v)}(v)\\
\leq&(1-\epsilon_x^{2}[\epsilon_\theta(\sigma)]^{2})\Delta e^{\beta r}\sum_{v'\in V} e^{\beta d(v,v')} D_{(n,v)(n-1,v')}(v)\\
&+ (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 D_{(n,v)(n,v)}(v)\\
\leq&(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 \sum_{(k',v')\in I} e^{\beta\{|k-k'|+d(v,v')\}} D_{(n,v)(k',v')}(v).
\end{aligned}
$$

By the definition of $\widetilde{\mathrm{Corr}}(\,\cdot\,,\beta)$ in equation (46), we have

$$
\begin{aligned}
\widetilde{\mathrm{Corr}}(\widetilde{\pi}_n,\beta) =&\max_{v\in V}\sum_{v'\in V} e^{\beta d(v,v')}\widetilde{C}^{\widetilde{\pi}_n}_{vv'}\\
\leq&\max_{v\in V}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 \sum_{(k',v')\in I} e^{\beta\{|k-k'|+d(v,v')\}} D_{(n,v)(k',v')}(v)\\
<&2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2\\
\leq&\frac{1}{8},
\end{aligned}
$$

where the second inequality holds by (59) and Theorem 5 in Appendix B, and the last inequality holds by Lemma 2 in Appendix D. ■

Lemma 4 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, for any* $n\geq 0$, *we have*

$$\mathrm{Corr}(\widetilde{\pi}_n,\beta)<\frac{1}{3},$$

where $\mathrm{Corr}(\,\cdot\,,\beta)$ *is defined in* (45) *and* β *is given in* (28).

Proof For $\mu^v_{\chi_{s-1},\chi_s}$ defined in (39) and $\mu^{v,K'}_{\chi_{s-1},\chi_s}$ defined in (41), we have

$$
\widetilde{\pi}^v_{\chi_n,\chi_{n+1}}(A)=\frac{\begin{array}{r}\int \mathbb{1}_A(x_n^v,\theta_n^v)\prod_{\omega\in N(v)\setminus K} f_{X^\omega_{n+1}|X_n}(x^\omega_{n+1}\mid x_n\,;\theta^\omega_{n+1})\\ \times f_{\Theta^\omega_{n+1}|\Theta_n}(\theta^\omega_{n+1}\mid\theta_n\,;\sigma)\widetilde{\pi}^{v,K}_{\chi_n,\chi_{n+1}}(dx_n^v,d\theta_n^v)\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)\setminus K} f_{X^\omega_{n+1}|X_n}(x^\omega_{n+1}\mid x_n\,;\theta^\omega_{n+1})\\ \times f_{\Theta^\omega_{n+1}|\Theta_n}(\theta^\omega_{n+1}\mid\theta_n\,;\sigma)\widetilde{\pi}^{v,K}_{\chi_n,\chi_{n+1}}(dx_n^v,d\theta_n^v)\end{array}}.
$$

Let $x_n,\overline{x}_n\in\mathbb{X}$ be such that $x_n^{V\setminus\{v'\}}=\overline{x}_n^{V\setminus\{v'\}}$, and $\theta_n,\overline{\theta}_n\in\mathbb{\Theta}$ be such that $\theta_n^{V\setminus\{v'\}}=\overline{\theta}_n^{V\setminus\{v'\}}$. If $v'\notin\cup_{\omega\in N(v)}N(\omega)$, for $\mu^v_{\overline{\chi}_{s-1},\chi_s}$ defined in (40) and $\mu^{v,K'}_{\overline{\chi}_{s-1},\chi_s}$ defined in (42), by Theorem 4 in Appendix B, we have

$$\|\widetilde{\pi}^v_{\chi_n,\chi_{n+1}}-\widetilde{\pi}^v_{\overline{\chi}_n,\chi_{n+1}}\|\leq 2\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta}\|\widetilde{\pi}^{v,K}_{\chi_n,\chi_{n+1}}-\widetilde{\pi}^{v,K}_{\overline{\chi}_n,\chi_{n+1}}\|.$$

Note that

$$\widetilde{\pi}^v_{\chi_n,\chi_{n+1}}(A)\geq\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}\widetilde{\pi}^{v,K}_{\chi_n,\chi_{n+1}}(A),$$

$$\widetilde{\pi}^{v}_{\overline{\chi}_n,\chi_{n+1}}(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}\widetilde{\pi}^{v,K}_{\overline{\chi}_n,\chi_{n+1}}(A),$$

and then by Theorem 3 in Appendix B we have that if $v' \in \cup_{\omega\in N(v)}N(\omega)$

$$\|\widetilde{\pi}^{v}_{\chi_n,\chi_{n+1}} - \widetilde{\pi}^{v}_{\overline{\chi}_n,\chi_{n+1}}\| \leq 2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) + \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}\|\widetilde{\pi}^{v,K}_{\chi_n,\chi_{n+1}} - \widetilde{\pi}^{v,K}_{\overline{\chi}_n,\chi_{n+1}}\|.$$

One can establish the inequality

$$\begin{aligned}
&\mathrm{Corr}(\widetilde{\pi}_n,\beta)\\
\leq&(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\max_{v\in V}\sum_{v'\in\cup_{\omega\in N(v)}N(\omega)} e^{\beta d(v,v')} + 2\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta}\widetilde{\mathrm{Corr}}(\widetilde{\pi}_n,\beta)\\
\leq&(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 + 2\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta}\widetilde{\mathrm{Corr}}(\widetilde{\pi}_n,\beta).
\end{aligned}$$

Under condition (27) that

$$\epsilon_x\epsilon_\theta(\sigma) > \left(1-\frac{1}{16\Delta_{\mathcal{K}}\Delta^2}\right)^{\frac{1}{2\Delta}},$$

since $\Delta, \Delta_{\mathcal{K}} \geq 1$, we have

$$\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta} < \frac{1}{1-\frac{1}{16\Delta_{\mathcal{K}}\Delta^2}} \leq \frac{1}{1-1/16}.$$

Hence, by Lemmas 2 and 3 in Appendix D, we have

$$\begin{aligned}
\mathrm{Corr}(\widetilde{\pi}_n,\beta) \leq&(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{2\beta r}\Delta^2 + 2\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta}\widetilde{\mathrm{Corr}}(\widetilde{\pi}_n,\beta)\\
<&\frac{1}{16} + \frac{2}{8}\epsilon_x^{-2\Delta}[\epsilon_\theta(\sigma)]^{-2\Delta}\\
<&\frac{1}{16} + \frac{2}{8}\frac{1}{1-1/16}\\
<&\frac{1}{3}.
\end{aligned}$$

■

Proposition 5 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, for every* $n \geq 1$, $K \in \mathcal{K}$ *and* $\kappa \subseteq K$, *we have that*

$$\|\mathsf{F}_n\widetilde{\pi}_{n-1} - \widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1}\|_\kappa \leq 4e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\kappa,\partial K)}\mathrm{card}(\kappa),$$

where β *is given in* (28).

Proof Define $I = \{n-1, n\} \times V$ and $\mathbb{S} = (\mathbb{X}\times\Theta)^2$. Fix $K \in \mathcal{K}$ and define

$$\begin{aligned}
\rho(A) =& \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1},\theta_{n-1},x_n,\theta_n)\prod_{\omega\in V} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega)\\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma) f_{Y_n^\omega|X_n^\omega}(Y_n^\omega \mid x_n^\omega\,;\theta_n^\omega)\\ \times \psi(dx_n)\lambda(d\theta_n)\widetilde{\pi}_{n-1}(dx_{n-1},d\theta_{n-1})\end{array}}{\begin{array}{r}\int\prod_{\omega\in V} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega)\\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma) f_{Y_n^\omega|X_n^\omega}(Y_n^\omega \mid x_n^\omega\,;\theta_n^\omega)\\ \times \psi(dx_n)\lambda(d\theta_n)\widetilde{\pi}_{n-1}(dx_{n-1},d\theta_{n-1})\end{array}},
\end{aligned}$$

and

$$\widetilde{\rho}(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1},\theta_{n-1},x_n,\theta_n)\prod_{v\in K} f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,;\theta_n^v) \\ \times f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,;\sigma)\prod_{\omega\in V} f_{Y_n^\omega|X_n^\omega}(Y_n^\omega \mid x_n^\omega\,;\theta_n^\omega) \\ \times \psi(dx_n)\lambda(d\theta_n)\widetilde{\pi}_{n-1}(dx_{n-1},d\theta_{n-1})\end{array}}{\begin{array}{r}\int \prod_{v\in K} f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,;\theta_n^v) \\ \times f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,;\sigma)\prod_{\omega\in V} f_{Y_n^\omega|X_n^\omega}(Y_n^\omega \mid x_n^\omega\,;\theta_n^\omega) \\ \times \psi(dx_n)\lambda(d\theta_n)\widetilde{\pi}_{n-1}(dx_{n-1},d\theta_{n-1})\end{array}}.$$

Then for any $\mathrm{K} \subseteq K$ and $K \subseteq V$, we have

$$\|\mathsf{F}_n\widetilde{\pi}_{n-1} - \widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1}\|_{\mathrm{K}} = \|\rho - \widetilde{\rho}\|_{\{n\}\times\mathrm{K}}.$$

In order to use Theorem 2 (Dobrushin comparison theorem) in Appendix B to bound $\|\rho - \widetilde{\rho}\|_{\{n\}\times\mathrm{K}}$, we need to bound C_{ij} and b_i with $i = (k, v)$ and $j = (k', v')$. Set

$$\rho^i = \rho^i_{(x_{n-1},\theta_{n-1},x_n,\theta_n)} \quad \text{and} \quad \widetilde{\rho}^i = \widetilde{\rho}^i_{(x_{n-1},\theta_{n-1},x_n,\theta_n)},$$

whose definitions are given in Theorem 2 in Appendix B. We display our discussions as follows:

- When $k = n-1$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}^v,\theta_{n-1}^v)\prod_{\omega\in N(v)} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}}.$$

  We can see that $\rho^i = \widetilde{\pi}^v_{\chi_{n-1},\chi_n}$, according to the definition of $\mu^v_{\chi_{s-1},\chi_s}$ defined in (39). Therefore, if $k' = n-1$, by the definition of $C^{\mu_{s-1}}_{vv'}$ in equation (43), we know that $C_{ij} \le C^{\widetilde{\pi}_{n-1}}_{vv'}$. If $k' = n$, since

$$\rho^i(A) \ge \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}^v,\theta_{n-1}^v)\prod_{\omega\in N(v)\setminus\{v'\}} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)\setminus\{v'\}} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}},$$

  we have $C_{ij} \le 1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $v' \in N(v)$ by Theorem 3 in Appendix B, and $C_{ij} = 0$ otherwise. Hence, by the definition of $\mathrm{Corr}(\mu_{s-1},\beta)$ in (45), we have

$$\sum_{(k',v')\in I} e^{\beta|k'-k|}e^{\beta d(v,v')}C_{(k,v)(k',v')} \le \mathrm{Corr}(\widetilde{\pi}_{n-1},\beta) + (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)e^{\beta(r+1)}\Delta. \quad (60)$$

  Next we take care of b_i. When $k = n-1$, we have

$$\widetilde{\rho}^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_{n-1}^v,\theta_{n-1}^v)\prod_{\omega\in N(v)\cap K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}}{\begin{array}{r}\int \prod_{\omega\in N(v)\cap K} f_{X_n^\omega|X_{n-1}}(x_n^\omega \mid x_{n-1}\,;\theta_n^\omega) \\ \times f_{\Theta_n^\omega|\Theta_{n-1}}(\theta_n^\omega \mid \theta_{n-1}\,;\sigma)\widetilde{\pi}^v_{n-1}(dx^v_{n-1},d\theta^v_{n-1})\end{array}}.$$

Note that, if $N(v) \subseteq K$, we have $\rho^i = \widetilde{\rho}^i$ which yields $b_i = 0$ if $v \in \text{int}(K)$. Further note that

$$\rho^i(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \frac{\int \mathbb{1}_A(x_{n-1}^v, \theta_{n-1}^v)\widetilde{\pi}_{n-1}^v(dx_{n-1}^v, d\theta_{n-1}^v)}{\int \widetilde{\pi}_{n-1}^v(dx_{n-1}^v, d\theta_{n-1}^v)},$$
$$\widetilde{\rho}^i(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \frac{\int \mathbb{1}_A(x_{n-1}^v, \theta_{n-1}^v)\widetilde{\pi}_{n-1}^v(dx_{n-1}^v, d\theta_{n-1}^v)}{\int \widetilde{\pi}_{n-1}^v(dx_{n-1}^v, d\theta_{n-1}^v)},$$

which by Theorem 3 in Appendix B yields that $b_i = 2(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})$ if $v \notin \text{int}(K)$.

- When $k = n$, we have

$$\begin{aligned}\rho^i(A) =& \frac{\begin{array}{r}\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,; \theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,; \sigma) \\ \times\, f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}}{\begin{array}{r}\int f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,; \theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,; \sigma) \\ \times\, f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)\end{array}} \\ \geq& \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}{\int f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}.\end{aligned}$$

  Hence,

$$\sum_{(k',v')\in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} \leq (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta. \tag{61}$$

  Note that, if $v \in K$ we have $\rho^i = \widetilde{\rho}^i$ which yields $b_i = 0$, otherwise given that

$$\begin{aligned}\widetilde{\rho}^i(A) =& \frac{\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}{\int f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)} \\ \geq& \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}{\int f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}\end{aligned}$$

  we have $b_i = 2(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2)$ by Theorem 3 in Appendix B .

Summing up (60) for $k = n - 1$ and (61) for $k = n$, we have that

$$\max_{(k,v)\in I} \sum_{(k',v')\in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} \leq \text{Corr}(\widetilde{\pi}_{n-1}, \beta) + (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta.$$

Furthermore, by Lemma 4 in Appendix D we have that

$$\text{Corr}(\widetilde{\pi}_n, \beta) < \frac{1}{3},$$

and by Lemma 2 in Appendix D we have that

$$(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta \leq \frac{1}{16}.$$

Hence,

$$\max_{(k,v)\in I}\sum_{(k',v')\in I} e^{\beta|k-k'|}e^{\beta d(v,v')}C_{(k,v)(k',v')} < 1/2.$$

Next, applying Theorem 2 (Dobrushin comparison theorem) in Appendix B and Theorem 5 in Appendix B, we obtain that

$$\begin{aligned}
&\|\mathsf{F}_n\widetilde{\pi}_{n-1} - \widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1}\|_{\mathrm{K}} = \|\rho-\widetilde{\rho}\|_{\{n\}\times\mathrm{K}}\\
&= 2(1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2)\sum_{v\in\mathrm{K}}\left\{\sum_{v'\in V\setminus \mathrm{int}(K)} D_{(n,v)(n-1,v')} + \sum_{v'\in V\setminus K} D_{(n,v)(n,v')}\right\}\\
&\le 4e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}).
\end{aligned}$$

■

Proposition 6 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, we have that for every* $\mathrm{K}\subseteq K$, $K\in\mathcal{K}$ *and* $s\in\{1,\dots,n-1\}$,

$$\begin{aligned}
&\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}}\\
&\le\frac{48}{23}e^{-\beta(n-s)}\sum_{v\in\mathrm{K}}\max_{v'\in V}e^{-\beta d(v,v')}\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|,
\end{aligned}$$

where β *is given in* (28).

Proof Define $I=\{s,\dots,n\}\times V$ and $\mathbb{S}=(\mathbb{X}\times\Theta)^{n-s+1}$. Define

$$\rho = \mathbf{P}^{\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1}}[X_s,X_{s+1},\dots,X_n\in\cdot\,,\Theta_s,\Theta_{s+1},\dots,\Theta_n\in\cdot\mid Y_{s+1},\dots,Y_n],$$

$$\widetilde{\rho} = \mathbf{P}^{\mathsf{F}_s\widetilde{\pi}_{s-1}}[X_s,X_{s+1},\cdots,X_n\in\cdot\,,\Theta_s,\Theta_{s+1},\cdots,\Theta_n\in\cdot\mid Y_{s+1},\dots,Y_n].$$

Then we have

$$\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}} = \|\rho-\widetilde{\rho}\|_{\{n\}\times\mathrm{K}}.$$

In order to use Theorem 2 (Dobrushin comparison theorem) in Appendix B to bound $\|\rho-\widetilde{\rho}\|_{\{n\}\times\mathrm{K}}$, we need to bound C_{ij} and b_i with $i=(k,v)$ and $j=(k',v')$. Set

$$\rho^i = \rho^i_{(x_s,\dots,x_n,\theta_s,\dots,\theta_n)} \quad\text{and}\quad \widetilde{\rho}^i = \widetilde{\rho}^i_{(x_s,\dots,x_n,\theta_s,\dots,\theta_n)},$$

whose definitions are given in Theorem 2 in Appendix B. We display our discussions as follows:

- When $k = s$, we have

$$\begin{aligned}\rho^i(A) = \widetilde{\pi}^v_{\chi_s,\chi_{s+1}}(A) =& \frac{\begin{array}{r}\int \mathbb{1}_A(x^v_s,\theta^v_s)\prod_{\omega\in N(v)} f_{X^\omega_{s+1}|X_s}(x^\omega_{s+1}\mid x_s\,;\theta^\omega_{s+1}) \\ \times\, f_{\Theta^\omega_{s+1}|\Theta_s}(\theta^\omega_{s+1}\mid \theta_s\,;\sigma)\widetilde{\pi}^v_{\chi_s}(dx^v_s,d\theta^v_s)\end{array}}{\begin{array}{r}\int\prod_{\omega\in N(v)} f_{X^\omega_{s+1}|X_s}(x^\omega_{s+1}\mid x_s\,;\theta^\omega_{s+1}) \\ \times\, f_{\Theta^\omega_{s+1}|\Theta_s}(\theta^\omega_{s+1}\mid \theta_s\,;\sigma)\widetilde{\pi}^v_{\chi_s}(dx^v_s,d\theta^v_s)\end{array}} \\ \geq& \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}\widetilde{\pi}^v_{\chi_s}(A)\end{aligned}$$

according to the definition of $\mu^v_{\chi_{s-1},\chi_s}$ given in (39). Therefore, when $k' = s$, by the definition of $C^{\widetilde{\pi}_s}_{vv'}$ in equation (43), we know that $C_{ij} \leq C^{\widetilde{\pi}_s}_{vv'}$. When $k' = s+1$, we have $C_{ij} \leq 1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}$ if $v' \in N(v)$ by Theorem 3 in Appendix B, and $C_{ij} = 0$ otherwise. Hence, by the definition of $\widetilde{\mathrm{Corr}}(\mu_{s-1},\beta)$ in (45),

$$\sum_{(k',v')\in I} e^{\beta|k-k'|}e^{\beta d(v,v')}C_{(k,v)(k',v')} \leq \mathrm{Corr}(\widetilde{\pi}_s,\beta) + (1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{\beta(r+1)}\Delta. \quad (62)$$

- When $k \in \{s+1,\ldots,n-1\}$, we have

$$\begin{aligned}\rho^i(A) =& \frac{\begin{array}{r}\int \mathbb{1}_A(x^v_k,\theta^v_k) f_{X^v_k|X_{k-1}}(x^v_k\mid x_{k-1}\,;\theta^v_k) f_{\Theta^v_k|\Theta_{k-1}}(\theta^v_k\mid\theta_{k-1}\,;\sigma) \\ \times\prod_{\omega\in N(v)} f_{X^\omega_{k+1}|X_k}(x^\omega_{k+1}\mid x_k\,;\theta^\omega_{k+1}) f_{\Theta^\omega_{k+1}|\Theta_k}(\theta^\omega_{k+1}\mid\theta_k\,;\sigma) \\ \times\, f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k)\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}{\begin{array}{r}\int f_{X^v_k|X_{k-1}}(x^v_k\mid x_{k-1}\,;\theta^v_k) f_{\Theta^v_k|\Theta_{k-1}}(\theta^v_k\mid\theta_{k-1}\,;\sigma) \\ \times\prod_{\omega\in N(v)} f_{X^\omega_{k+1}|X_k}(x^\omega_{k+1}\mid x_k\,;\theta^\omega_{k+1}) f_{\Theta^\omega_{k+1}|\Theta_k}(\theta^\omega_{k+1}\mid\theta_k\,;\sigma) \\ \times\, f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k)\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}.\end{aligned}$$

Note that x^v_k depends on $x^{v'}_{k-1}$ if $v' \in N(v)$, and

$$\begin{aligned}\rho^i(A) \geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 & \frac{\begin{array}{r}\int \mathbb{1}_A(x^v_k,\theta^v_k)\prod_{\omega\in N(v)} f_{X^\omega_{k+1}|X_k}(x^\omega_{k+1}\mid x_k\,;\theta^\omega_{k+1}) \\ \times f_{\Theta^\omega_{k+1}|\Theta_k}(\theta^\omega_{k+1}\mid\theta_k\,;\sigma) f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k) \\ \times\,\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}{\begin{array}{r}\int\prod_{\omega\in N(v)} f_{X^\omega_{k+1}|X_k}(x^\omega_{k+1}\mid x_k\,;\theta^\omega_{k+1}) \\ \times f_{\Theta^\omega_{k+1}|\Theta_k}(\theta^\omega_{k+1}\mid\theta_k\,;\sigma) f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k) \\ \times\,\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}.\end{aligned}$$

Further note that x^v_k depends on $x^{v'}_{k+1}$ if $v' \in N(v)$, x^v_k depends on $x^{v'}_k$ if $v' \in \cup_{\omega\in N(v)}N(\omega)$, and

$$\begin{aligned}\rho^i(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} & \frac{\begin{array}{r}\int \mathbb{1}_A(x^v_k,\theta^v_k) f_{X^v_k|X_{k-1}}(x^v_k\mid x_{k-1}\,;\theta^v_k) \\ \times f_{\Theta^v_k|\Theta_{k-1}}(\theta^v_k\mid\theta_{k-1}\,;\sigma) f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k) \\ \times\,\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}{\begin{array}{r}\int f_{X^v_k|X_{k-1}}(x^v_k\mid x_{k-1}\,;\theta^v_k) \\ \times f_{\Theta^v_k|\Theta_{k-1}}(\theta^v_k\mid\theta_{k-1}\,;\sigma) f_{Y^v_k|X^v_k}(Y^v_k\mid x^v_k\,;\theta^v_k) \\ \times\,\psi^v(dx^v_k)\lambda^v(d\theta^v_k)\end{array}}.\end{aligned}$$

By Theorem 3 in Appendix B we have $C_{ij} \leq 1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2$, if $k' = k-1$ and $v' \in N(v)$; $C_{ij} \leq 1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}$, if $k' = k+1$ and $v' \in N(v)$; $C_{ij} \leq 1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}$, if $k' = k$ and $v' \in \cup_{\omega \in N(v)} N(\omega)$; $C_{ij} = 0$, otherwise. Therefore

$$\begin{aligned}\sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} &\leq (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})(e^{2\beta r}\Delta^2 + 2e^{\beta(r+1)}\Delta) \\ &\leq 3(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r}\Delta^2. \end{aligned} \tag{63}$$

- When $k = n$, note that x_n^v only depends on $x_{n-1}^{v'}$ where $v' \in N(v)$, and

$$\begin{aligned}\rho^i(A) =& \frac{\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,; \theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,; \sigma)}{\int f_{X_n^v|X_{n-1}}(x_n^v \mid x_{n-1}\,; \theta_n^v) f_{\Theta_n^v|\Theta_{n-1}}(\theta_n^v \mid \theta_{n-1}\,; \sigma)} \\ &\qquad \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v) \\ &\qquad \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v) \\ \geq& \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\int \mathbb{1}_A(x_n^v, \theta_n^v) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}{\int f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^v)\lambda^v(d\theta_n^v)}. \end{aligned}$$

By Theorem 3 in Appendix B we have $C_{ij} \leq 1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $k' = n-1$ and $v' \in N(v)$; $C_{ij} = 0$, otherwise. Therefore

$$\sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} \leq (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta. \tag{64}$$

Summing up (62) for $k = s$, (63) for $k \in \{s+1, \ldots, n-1\}$, and (64) for $k = n$, by Lemma 2 in Appendix D, we have

$$\begin{aligned}&\max_{(k,v) \in I} \sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} \\ &\qquad \leq \mathrm{Corr}(\widetilde{\pi}_s, \beta) + 3(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r}\Delta^2. \end{aligned}$$

Furthermore, by Lemma 4 in Appendix D we have that

$$\mathrm{Corr}(\widetilde{\pi}_s, \beta) < \frac{1}{3},$$

and by Lemma 2 in Appendix D we have that

$$(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{2\beta r}\Delta^2 \leq \frac{1}{16}.$$

Hence,

$$\max_{(k,v) \in I} \sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} < \frac{1}{3} + \frac{3}{16} = \frac{25}{48}.$$

By Theorem 5 in Appendix B we have

$$\max_{(k,v)\in I}\sum_{(k',v')\in I} e^{\beta|k-k'|}e^{\beta d(v,v')}D_{(k,v)(k',v')} < \frac{48}{23}. \tag{65}$$

Note that, $\rho^i = \widetilde{\rho}^i$ when $i=(k,v)$ with $k>s$. For $k=s$, we have $\widetilde{\rho}^i = (\mathsf{F}_s\widetilde{\pi}_{s-1})^v_{\chi_s,\chi_{s+1}}$ and $\rho^i = (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^v_{\chi_s,\chi_{s+1}}$. By Theorem 3 in Appendix B, we have

$$\begin{aligned}
&\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}}\\
=&\|\rho-\widetilde{\rho}\|_{\{n\}\times\mathrm{K}}\\
\le&\sum_{v\in\mathrm{K}}\sum_{v'\in V} D_{(n,v)(s,v')}\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|\\
=&\sum_{v\in\mathrm{K}} e^{-\beta(n-s)}\sum_{v'\in V} e^{\beta((n-s)+d(v,v'))}D_{(n,v)(s,v')}\\
&\qquad\qquad\times e^{-\beta d(v,v')}\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|\\
\le&\frac{48}{23}e^{-\beta(n-s)}\sum_{v\in\mathrm{K}}\max_{v'\in V} e^{-\beta d(v,v')}\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|,
\end{aligned}$$

where we used (65) in the last inequality. ■

Proposition 7 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, we have that for every $s\ge 1$, $K'\in\mathcal{K}$, and $v'\in K'$,*

$$\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\| \le \frac{96}{29}e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v',\partial K')},$$

where β is given in (28).

Proof According to the expressions of $(\mathsf{F}_s\nu_{s-1})^v_{\chi_s,\chi_{s+1}}$ given in (51) and $(\widetilde{\mathsf{F}}_s\nu_{s-1})^v_{\chi_s,\chi_{s+1}}$ given in (52), we have

$$\begin{aligned}
&(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}(A)\\
&\quad\int \mathbb{1}_A(x_s^{v'},\theta_s^{v'})\textstyle\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1};\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega\mid\theta_{s-1};\sigma)\\
&\qquad\times f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'}\mid x_s^{v'};\theta_s^{v'})\textstyle\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u\mid x_s;\theta_{s+1}^u)\\
&\qquad\qquad\times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u\mid\theta_s;\sigma)\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})\\
=&\frac{\int \textstyle\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1};\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega\mid\theta_{s-1};\sigma)}{\qquad\times f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'}\mid x_s^{v'};\theta_s^{v'})\textstyle\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u\mid x_s;\theta_{s+1}^u)}\\
&\qquad\qquad\times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u\mid\theta_s;\sigma)\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned}$$

and

$$
\begin{aligned}
&(\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}(A)\\
&\quad\int \mathbb{1}_A(x_s^{v'},\theta_s^{v'})\prod_{\omega\in K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)\\
&\qquad\times f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\quad\times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})\\
&=\frac{}{\int\prod_{\omega\in K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega) f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma)}.\\
&\qquad\times f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u)\\
&\qquad\quad\times f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned}
$$

Define $I=(\{s-1\}\times V)\cup(s,v')$ and $\mathbb{S}=(\mathbb{X}\times\mathbb{\Theta})\times(\mathbb{X}^{v'}\times\mathbb{\Theta}^{v'})$, and the probability measures on $\mathbb{S}$ as follows:

$$
\begin{aligned}
\rho(A)=\frac{}{}&\int \mathbb{1}_A(x_{s-1},\theta_{s-1},x_s^{v'},\theta_s^{v'})\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega)\\
&\times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma) f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\\
&\quad\times\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u) f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\\
&\qquad\times\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})\\
&\Big/\int\prod_{\omega\in V} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega)\\
&\times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma) f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\\
&\quad\times\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u) f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\\
&\qquad\times\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned}
$$

and

$$
\begin{aligned}
\widetilde{\rho}(A)=\frac{}{}&\int \mathbb{1}_A(x_{s-1},\theta_{s-1},x_s^{v'},\theta_s^{v'})\prod_{\omega\in K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega)\\
&\times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma) f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\\
&\quad\times\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u) f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\\
&\qquad\times\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})\\
&\Big/\int\prod_{\omega\in K'} f_{X_s^\omega|X_{s-1}}(x_s^\omega \mid x_{s-1}\,;\theta_s^\omega)\\
&\times f_{\Theta_s^\omega|\Theta_{s-1}}(\theta_s^\omega \mid \theta_{s-1}\,;\sigma) f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,;\theta_s^{v'})\\
&\quad\times\prod_{u\in N(v')} f_{X_{s+1}^u|X_s}(x_{s+1}^u \mid x_s\,;\theta_{s+1}^u) f_{\Theta_{s+1}^u|\Theta_s}(\theta_{s+1}^u \mid \theta_s\,;\sigma)\\
&\qquad\times\psi^{v'}(dx_s^{v'})\lambda^{v'}(d\theta_s^{v'})\widetilde{\pi}_{s-1}(dx_{s-1},d\theta_{s-1})
\end{aligned}
$$

Then we have

$$\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}-(\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|=\|\rho-\widetilde{\rho}\|_{(s,v')}.$$

In order to use Theorem 2 (Dobrushin comparison theorem) in Appendix B to bound $\|\rho-\widetilde{\rho}\|_{(s,v')}$, we need to bound C_{ij} and b_i with $i=(\overline{k},\overline{v})$ and $j=(\overline{\overline{k}},\overline{\overline{v}})$. Set

$$\rho^i=\rho^i_{(x_{s-1},\theta_{s-1},x_s^{v'},\theta_s^{v'})}\quad\text{and}\quad\widetilde{\rho}^i=\widetilde{\rho}^i_{(x_{s-1},\theta_{s-1},x_s^{v'},\theta_s^{v'})},$$

whose definitions are given in Theorem 2 in Appendix B. We display our discussions as follows:

- When $\overline{k} = s-1$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x^{\overline{v}}_{s-1}, \theta^{\overline{v}}_{s-1}) \prod_{\omega \in N(\overline{v})} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega \in N(\overline{v})} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}}$$

  and

$$\widetilde{\rho}^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x^{\overline{v}}_{s-1}, \theta^{\overline{v}}_{s-1}) \prod_{\omega \in N(\overline{v}) \cap K'} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega \in N(\overline{v}) \cap K'} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}},$$

  where $\widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}$ is defined according to the definition of $\mu^v_{\chi_{s-1}}$ given in (38). Therefore, by the definition of $\mu^v_{\chi_{s-1},\chi_s}$ given in (39), we have $\rho^i = \widetilde{\pi}^{\overline{v}}_{\chi_{s-1},\chi_s}$. Furthermore, by the definition of $C^{\mu_{s-1}}_{vv'}$ in equation (43), we know that $C_{ij} \le C^{\widetilde{\pi}_{s-1}}_{\overline{v}\,\overline{\overline{v}}}$ if $\overline{\overline{k}} = s-1$. If $\overline{\overline{k}} = s$, since

$$\rho^i(A) \ge \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x^{\overline{v}}_{s-1}, \theta^{\overline{v}}_{s-1}) \prod_{\omega \in N(\overline{v}) \setminus \{v'\}} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}}{\begin{array}{r}\int \prod_{\omega \in N(\overline{v}) \setminus \{v'\}} f_{X^\omega_s | X_{s-1}}(x^\omega_s \mid x_{s-1}\,; \theta^\omega_s) \\ \times f_{\Theta^\omega_s | \Theta_{s-1}}(\theta^\omega_s \mid \theta_{s-1}\,; \sigma) \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}(dx^{\overline{v}}_{s-1}, d\theta^{\overline{v}}_{s-1})\end{array}},$$

  we have $C_{ij} \le 1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $\overline{\overline{v}} = v' \in N(\overline{v})$ by Theorem 3 in Appendix B, and $C_{ij} = 0$ otherwise. Hence, by the definition of $\mathrm{Corr}(\mu_{s-1}, \beta)$ in (45),

$$\sum_{(\overline{\overline{k}},\overline{\overline{v}}) \in I} e^{\beta|\overline{k} - \overline{\overline{k}}|} e^{\beta d(\overline{v},\overline{\overline{v}})} C_{(\overline{k},\overline{v})(\overline{\overline{k}},\overline{\overline{v}})} \le \mathrm{Corr}(\widetilde{\pi}_{s-1}, \beta) + (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta. \tag{66}$$

  To handle b_i, note that if $N(\overline{v}) \subseteq K'$ we have $\rho^i = \widetilde{\rho}^i$, and note that

$$\rho^i(A) \ge \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}, \qquad \widetilde{\rho}^i(A) \ge \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \widetilde{\pi}^{\overline{v}}_{\chi_{s-1}}.$$

  Therefore, we have that $b_i = 0$ if $\overline{v} \in \mathrm{int}(K')$, and by Theorem 3 in Appendix B we have $b_i = 2(1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})$ otherwise.

- When $\overline{k} = s$, we have

$$\begin{aligned}\rho^i(A) &= \widetilde{\rho}^i(A) \\ &= \frac{\begin{array}{r}\int \mathbb{1}_A(x^{v'}_s, \theta^{v'}_s) f_{X^{v'}_s | X_{s-1}}(x^{v'}_s \mid x_{s-1}\,; \theta^{v'}_s) f_{\Theta^{v'}_s | \Theta_{s-1}}\,; \sigma) \\ \times f_{Y^{v'}_s | X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s) \prod_{u \in N(v')} f_{X^u_{s+1} | X_s}(x^u_{s+1} \mid x_s\,; \theta^u_{s+1}) \\ \times f_{\Theta^u_{s+1} | \Theta_s}(\theta^u_{s+1} \mid \theta_s\,; \sigma) \psi^{v'}(dx^{v'}_s) \lambda^{v'}(d\theta^{v'}_s)\end{array}}{\begin{array}{r}\int f_{X^{v'}_s | X_{s-1}}(x^{v'}_s \mid x_{s-1}\,; \theta^{v'}_s) f_{\Theta^{v'}_s | \Theta_{s-1}}\,; \sigma) \\ \times f_{Y^{v'}_s | X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s) \prod_{u \in N(v')} f_{X^u_{s+1} | X_s}(x^u_{s+1} \mid x_s\,; \theta^u_{s+1}) \\ \times f_{\Theta^u_{s+1} | \Theta_s}(\theta^u_{s+1} \mid \theta_s\,; \sigma) \psi^{v'}(dx^{v'}_s) \lambda^{v'}(d\theta^{v'}_s)\end{array}}.\end{aligned}$$

Therefore, we have $b_i = 0$, and $C_{ij} \leq 1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $\overline{\overline{k}} = s-1$ and $\overline{\overline{v}} \in N(v')$ and $C_{ij} = 0$ otherwise. Hence,

$$\sum_{(\overline{\overline{k}},\overline{\overline{v}})\in I} e^{\beta|\overline{k}-\overline{\overline{k}}|} e^{\beta d(v',\overline{\overline{v}})} C_{(\overline{k},v')(\overline{\overline{k}},\overline{\overline{v}})} \leq (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta. \tag{67}$$

Summing up (66) for $\overline{k} = s-1$ and (67) for $\overline{k} = s$, we have

$$\begin{aligned}\max_{(\overline{k},\overline{v})\in I} &\sum_{(\overline{\overline{k}},\overline{\overline{v}})\in I} e^{\beta|\overline{k}-\overline{\overline{k}}|} e^{\beta d(\overline{v},\overline{\overline{v}})} C_{(\overline{k},\overline{v})(\overline{\overline{k}},\overline{\overline{v}})} \\ &\leq \mathrm{Corr}(\widetilde{\pi}_{s-1},\beta) + (1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta.\end{aligned}$$

Furthermore, by Lemma 4 in Appendix D we have that

$$\mathrm{Corr}(\widetilde{\pi}_{s-1},\beta) < \frac{1}{3},$$

and by Lemma 2 in Appendix D we have that

$$(1-\epsilon_x^2[\epsilon_\theta(\sigma)]^2) e^{\beta(r+1)}\Delta \leq \frac{1}{16}.$$

Hence,

$$\max_{(\overline{k},\overline{v})\in I} \sum_{(\overline{\overline{k}},\overline{\overline{v}})\in I} e^{\beta|\overline{k}-\overline{\overline{k}}|} e^{\beta d(\overline{v},\overline{\overline{v}})} C_{(\overline{k},\overline{v})(\overline{\overline{k}},\overline{\overline{v}})} = \frac{19}{48}.$$

By Theorem 2 (Dobrushin comparison theorem) in Appendix B and Theorem 5 in Appendix B, we have

$$\begin{aligned}\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\| &= \|\rho-\widetilde{\rho}\|_{(s,v')} \\ &\leq 2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) \sum_{\overline{v}\in V\setminus \mathrm{int}(K')} D_{(s,v')(s-1,\overline{v})} \\ &\leq \frac{96}{29} e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) e^{-\beta d(v',\partial K')},\end{aligned}$$

which is uniform for all $x_s, x_{s+1} \in \mathbb{X}$ and all $\theta_s, \theta_{s+1} \in \mathbb{\Theta}$. ■

E. Proofs for bounding variance

Proposition 8 *Under Assumption 1 in Section 3.3, for integer $s \geq 1$, one has*

$$\begin{aligned}\max_{K\in\mathcal{K}} &\left[\mathbf{E}\big\|\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\big\|_K^2\right]^{1/2} \\ &\leq \frac{16}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\epsilon_x^{-2|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K},\end{aligned}$$

where $|\mathcal{K}|_\infty$ is the maximal size of a block in $\mathcal{K}$ defined in (7).

Proof For any $K \in \mathcal{K}$, by Theorem 4 in Appendix B, we have

$$\begin{aligned}
&\left\|\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|_K \\
&\quad = \left\|\mathsf{C}_{s+1}\mathsf{B}\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{C}_{s+1}\mathsf{B}\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|_K \\
&\quad = \left\|\mathsf{C}^K_{s+1}\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{C}^K_{s+1}\mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\| \\
&\quad \le 2\epsilon_y^{-2|\mathcal{K}|_\infty}\left\|\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|. \qquad (68)
\end{aligned}$$

For $\psi^K(dx^K_\cdot)$ and $\lambda^K(d\theta^K_\cdot)$ defined in (2), we have

$$\begin{aligned}
&\frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx^K_{s+1}, d\theta^K_{s+1})}{\psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1})} \\
&\quad \int \prod_{v\in K} f_{X^v_{s+1}|X_s}(x^v_{s+1} \mid x_s\,; \theta^v_{s+1}) f_{\Theta^v_{s+1}|\Theta_s}(\theta^v_{s+1} \mid \theta_s\,; \sigma) \\
&= \frac{\quad \times \prod_{K'\in N(K)} \prod_{v'\in K'} f_{Y^{v'}_s|X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s)\left[\mathsf{B}^{K'}\mathsf{P}_s\widehat{\pi}_{s-1}\right](dx^{K'}_s, d\theta^{K'}_s)}{\int \prod_{K'\in N(K)} \prod_{v'\in K'} f_{Y^{v'}_s|X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s)\left[\mathsf{B}^{K'}\mathsf{P}_s\widehat{\pi}_{s-1}\right](dx^{K'}_s, d\theta^{K'}_s)},
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx^K_{s+1}, d\theta^K_{s+1})}{\psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1})} \\
&\quad \int \prod_{v\in K} f_{X^v_{s+1}|X_s}(x^v_{s+1} \mid x_s\,; \theta^v_{s+1}) f_{\Theta^v_{s+1}|\Theta_s}(\theta^v_{s+1} \mid \theta_s\,; \sigma) \\
&= \frac{\quad \times \prod_{K'\in N(K)} \prod_{v'\in K'} f_{Y^{v'}_s|X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s)\left[\mathsf{B}^{K'}\mathsf{S}^J\mathsf{P}_s\widehat{\pi}_{s-1}\right](dx^{K'}_s, d\theta^{K'}_s)}{\int \prod_{K'\in N(K)} \prod_{v'\in K'} f_{Y^{v'}_s|X^{v'}_s}(Y^{v'}_s \mid x^{v'}_s\,; \theta^{v'}_s)\left[\mathsf{B}^{K'}\mathsf{S}^J\mathsf{P}_s\widehat{\pi}_{s-1}\right](dx^{K'}_s, d\theta^{K'}_s)},
\end{aligned}$$

where $N(K)$ is defined in (4) as the collection of blocks that interact with the block K. By equation (8.1) in Georgii (2011), we have

$$\begin{aligned}
&\left\|\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\| \\
&= \int \left| \frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx^K_{s+1}, d\theta^K_{s+1})}{\psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1})} \right. \\
&\qquad \left. - \frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx^K_{s+1}, d\theta^K_{s+1})}{\psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1})} \right| \psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1}).
\end{aligned}$$

Therefore, by Minkowski's integral inequality,

$$\begin{aligned}
&\left[\mathbf{E}\left\|\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|^2\right]^{1/2} \\
&\le \int \left[\mathbf{E}\left| \frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx^K_{s+1}, d\theta^K_{s+1})}{\psi^K(dx^K_{s+1})\lambda^K(d\theta^K_{s+1})} \right.\right.
\end{aligned}$$

$$-\frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx_{s+1}^K, d\theta_{s+1}^K)}{\psi^K(dx_{s+1}^K)\lambda^K(d\theta_{s+1}^K)}\Bigg|^2\Bigg]^{1/2} \psi^K(dx_{s+1}^K)\lambda^K(d\theta_{s+1}^K)$$

$$\leq \psi^K(\mathbb{X}^K)\lambda^K(\mathbb{\Theta}^K) \sup_{\substack{x^K\in\mathbb{X}^K\\ \theta^K\in\mathbb{\Theta}^K}} \Bigg[\mathbf{E}\Bigg|\frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx_{s+1}^K, d\theta_{s+1}^K)}{\psi^K(dx_{s+1}^K)\lambda^K(d\theta_{s+1}^K)}$$

$$-\frac{\left(\mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right)(dx_{s+1}^K, d\theta_{s+1}^K)}{\psi^K(dx_{s+1}^K)\lambda^K(d\theta_{s+1}^K)}\Bigg|^2\Bigg]^{1/2}.$$

By Assumption 1 in Section 3.3 and the fact that $f_{X_{s+1}^v|X_s}(x_{s+1}^v \mid x_s\,; \theta_{s+1}^v)$ is a transition density, we have that

$$\epsilon_x\psi^v(\mathbb{X}^v) \leq \int f_{X_{s+1}^v|X_s}(x_{s+1}^v \mid x_s\,; \theta_{s+1}^v)\psi^v(dx_{s+1}^v) = 1,$$

and then

$$\psi^v(\mathbb{X}^v) \leq \epsilon_x^{-1}, \qquad \psi^K(\mathbb{X}^K) \leq \epsilon_x^{-|\mathcal{K}|_\infty}.$$

Similarly, by Assumption 1 in Section 3.3 and the fact that $f_{\Theta_{s+1}^v|\Theta_s}(\theta_{s+1}^v \mid \theta_s\,; \sigma)$ is a transition density, we have that

$$\epsilon_\theta(\sigma)\lambda^v(\mathbb{\Theta}^v) \leq \int f_{\Theta_{s+1}^v|\Theta_s}(\theta_{s+1}^v \mid \theta_s\,; \sigma)\lambda^v(d\theta_{s+1}^v) = 1,$$

and then

$$\lambda^v(\mathbb{\Theta}^v) \leq [\epsilon_\theta(\sigma)]^{-1}, \qquad \lambda^K(\mathbb{\Theta}^K) \leq [\epsilon_\theta(\sigma)]^{-|\mathcal{K}|_\infty}.$$

Furthermore, by Assumption 1 in Section 3.3, we have

$$\prod_{v\in K} f_{X_{s+1}^v|X_s}(x_{s+1}^v \mid x_s\,; \theta_{s+1}^v) f_{\Theta_{s+1}^v|\Theta_s}(\theta_{s+1}^v \mid \theta_s\,; \sigma) \leq \epsilon_x^{-|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-|\mathcal{K}|_\infty}$$

and

$$\epsilon_y^{|\mathcal{K}|_\infty\Delta_\mathcal{K}} \leq \prod_{K'\in N(K)}\prod_{v'\in K'} f_{Y_s^{v'}|X_s^{v'}}(Y_s^{v'} \mid x_s^{v'}\,; \theta_s^{v'}) \leq \epsilon_y^{-|\mathcal{K}|_\infty\Delta_\mathcal{K}}.$$

Hence, by Theorem 4 in Appendix B and Assumption 1 in Section 3.3,

$$\left[\mathbf{E}\left\|\mathsf{B}^K\mathsf{P}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \mathsf{B}^K\mathsf{P}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|^2\right]^{1/2}$$

$$\leq 2[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\epsilon_x^{-2|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty\Delta_\mathcal{K}}\left\|\left|\bigotimes_{K'\in N(K)}\mathsf{B}^{K'}\mathsf{P}_s\widehat{\pi}_{s-1} - \bigotimes_{K'\in N(K)}\mathsf{B}^{K'}\mathsf{S}^J\mathsf{P}_s\widehat{\pi}_{s-1}\right|\right\|$$

$$\leq 8[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\epsilon_x^{-2|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty}\Delta_{\mathcal{K}}\frac{\Delta_{\mathcal{K}}}{\sqrt{J}},$$

where the last inequality is by Theorem 7 in Appendix B. Plugging in equation (68), the proof is complete. ■

Lemma 9 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, one has that for every* $K \in \mathcal{K}$, $\kappa \subseteq K$ *and* $s \in \{1, \ldots, n\}$,

$$\|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_s \delta_x \delta_\theta - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_s \delta_{\overline{x}} \delta_{\overline{\theta}}\|_\kappa \leq \frac{32}{13} e^{-\tilde{\beta}(n-s+1)} \mathrm{card}(\kappa),$$

where

$$\tilde{\beta} = \frac{1}{2r} \log\left(\frac{1}{16\Delta^2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})}\right). \tag{69}$$

Proof We first tackle the $s = 1$ case that

$$\|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \delta_x \delta_\theta - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \delta_{\overline{x}} \delta_{\overline{\theta}}\|_\kappa$$

and then generalize to all $s \in \{1, \ldots, n\}$.

Recalling that $N(K)$ is defined in (4) as the collection of blocks that interact with the block K in one time step, we define a "block" tree T as

$$T = \big\{[K_u \cdots K_n] : 0 \leq u \leq n,\, K_l \in N(K_{l+1}),\, u \leq l \leq n\big\},$$

where $K_n = K$. That is, $[K_u \cdots K_n]$ is the block K_u at time u that has interactions with the block K at time n after $n-u$ time steps; note that this type of representation describes the block-wised interaction trace from time u up to time n. By the effect of block operator B^K on $\widetilde{\mathsf{F}}_n \rho$ for any $n \in \mathbb{N}$ and any measure ρ on $\mathbb{X} \times \Theta$, given in (23), we can write

$$\begin{aligned}
\mathsf{B}^K \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \delta_x \delta_\theta =& \mathsf{C}_n^{K_n} \mathsf{P}_n^{K_n} \bigotimes_{K_{n-1} \in N(K_n)} \Big[\mathsf{C}_{n-1}^{K_{n-1}} \mathsf{P}_{n-1}^{K_{n-1}} \\
& \bigotimes_{K_{n-2} \in N(K_{n-1})} \Big[\mathsf{C}_{n-2}^{K_{n-2}} \mathsf{P}_{n-2}^{K_{n-2}} \cdots \\
& \bigotimes_{K_1 \in N(K_2)} \Big[\mathsf{C}_1^{K_1} \mathsf{P}_s^{K_1} \\
& \bigotimes_{K_0 \in N(K_1)} \delta_x^{K_0} \delta_\theta^{K_0} \Big] \cdots \Big] \Big],
\end{aligned}$$

where P_n^K is defined in (21) and C_n^K is defined in (22).

The vertex set of the tree is defined as

$$I = \{[K_u \cdots K_n]v : [K_u \cdots K_n] \in T,\, v \in K_u\}.$$

Clearly, the following equivalence holds:

$$[K_u \cdots K_n] = \{[K_u \cdots K_n]v : v \in K_u\}.$$

Define the index set of leaves of the tree T as

$$T_0 = \{[K_0 \cdots K_n] : K_l \in N(K_{l+1}),\, 0 \le l < n\}.$$

Further define the children $c(i)$ of $i \in I$ as

$$c([K_u \cdots K_n]v) = \{[K_{u-1} \cdots K_n]v' : K_{u-1} \in N(K_u),\, v' \in N(v)\},$$

define the location $v(i)$ of $i \in I$ as

$$v([K_u \cdots K_n]v) = v,$$

define the depth $d(i)$ of $i \in I$ as

$$d([K_u \cdots K_n]v) = u,$$

and then define the set of non-leaf vertices of the tree as

$$I_+ = \{i \in I : 0 < d(i) \le n\}.$$

Define

$$\mathbb{S} = \prod_{i \in I} \mathbb{X}^i \times \mathbb{O}^i,$$

for $[t]v \in I$ define

$$\mathbb{Y}^{[t]v} = \mathbb{Y}^v, \quad \mathbb{X}^{[t]v} = \mathbb{X}^v \quad \text{and} \quad \mathbb{O}^{[t]v} = \mathbb{O}^v,$$

and for any $[K_u \cdots K_n] \in T$ and any measure ρ define

$$\rho^{[K_u \cdots K_n]} = \rho^{K_u}.$$

We define two probability measures on $\mathbb{S}$ as

$$\rho(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x,\theta) \prod_{i \in I_+} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) \\ \times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma) f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,; \theta^i_{d(i)}) \\ \times \psi^{v(i)}(dx^i_{d(i)}) \lambda^{v(i)}(d\theta^i_{d(i)}) \prod_{t \in T_0} \delta^t(dx^t_0, d\theta^t_0)\end{array}}{\begin{array}{r}\int \prod_{i \in I_+} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) \\ \times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma) f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,; \theta^i_{d(i)}) \\ \times \psi^{v(i)}(dx^i_{d(i)}) \lambda^{v(i)}(d\theta^i_{d(i)}) \prod_{t \in T_0} \delta^t(dx^t_0, d\theta^t_0)\end{array}}$$

and

$$\overline{\rho}(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x,\theta) \prod_{i \in I_+} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) \\ \times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma) f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,; \theta^i_{d(i)}) \\ \times \psi^{v(i)}(dx^i_{d(i)}) \lambda^{v(i)}(d\theta^i_{d(i)}) \prod_{t \in T_0} \delta^t(d\overline{x}^t_0, d\overline{\theta}^t_0)\end{array}}{\begin{array}{r}\int \prod_{i \in I_+} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) \\ \times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma) f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,; \theta^i_{d(i)}) \\ \times \psi^{v(i)}(dx^i_{d(i)}) \lambda^{v(i)}(d\theta^i_{d(i)}) \prod_{t \in T_0} \delta^t(d\overline{x}^t_0, d\overline{\theta}^t_0)\end{array}},$$

where $\delta(x,\theta)$ is defined in (14). Therefore,

$$\|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \delta_x \delta_\theta - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \delta_{\overline{x}} \delta_{\overline{\theta}}\|_{\mathrm{K}} = \|\rho - \overline{\rho}\|_{[K_n]\mathrm{K}}.$$

In the following, we are going to use Theorem 2 (Dobrushin comparison theorem) in Appendix B to bound $\|\rho - \overline{\rho}\|_{[K_n]\mathrm{K}}$. We will bound C_{ij} and b_i with $i = [K_u \cdots K_n]v$ and $j = [K'_{u'} \cdots K'_n]v'$ where $K_n = K'_n = K$ and $0 \le u, u' \le n$. Set

$$\rho^i = \rho^i_{(x,\theta)} \quad \text{and} \quad \overline{\rho}^i = \overline{\rho}^i_{(x,\theta)},$$

whose definitions are given in Theorem 2 in Appendix B. We display our discussions as follows:

- When $u = 0$, we have $\rho^i = \delta_{x_0^v}\delta_{\theta_0^v}$ and $\overline{\rho}^i = \delta_{\overline{x}_0^v}\delta_{\overline{\theta}_0^v}$. Hence

$$C_{ij} = 0 \tag{70}$$

  and $b_i \le 2$.

- When $u \in \{1, \ldots, n-1\}$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_u^i, \theta_u^i) f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times \prod_{l \in I_+ : i \in c(l)} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}{\begin{array}{r}\int f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times \prod_{l \in I_+ : i \in c(l)} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}.$$

  We can see that $\rho^i(A) = \overline{\rho}^i(A)$ and then $b_i = 0$. Next we take care of C_{ij}. Note that when $j \in c(i)$ we have

$$\rho^i(A) \ge \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x_u^i, \theta_u^i) \prod_{l \in I_+ : i \in c(l)} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}{\begin{array}{r}\int \prod_{l \in I_+ : i \in c(l)} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}},$$

when $i \in c(j)$ we have

$$\rho^i(A) \geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\begin{array}{r}\int \mathbb{1}_A(x_u^i, \theta_u^i) f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times \prod_{\{l \in I_+ : i \in c(l)\}\setminus\{j\}} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}{\begin{array}{r}\int f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times \prod_{\{l \in I_+ : i \in c(l)\}\setminus\{j\}} f_{X^l_{d(l)}|X_{d(l)-1}}(x^l_{d(l)} \mid x_{d(l)-1}\,; \theta^l_{d(l)}) \\ \times f_{\Theta^l_{d(l)}|\Theta_{d(l)-1}}(\theta^l_{d(l)} \mid \theta_{d(l)-1}\,; \sigma) f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v) \\ \times \psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}},$$

and when $j \in \cup_{l \in I_+ : i \in c(l)} c(l)$ we have

$$\rho^i(A) \geq \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} \frac{\begin{array}{r}\int \mathbb{1}_A(x_u^i, \theta_u^i) f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v)\psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}{\begin{array}{r}\int f_{X_u^i|X_{u-1}}(x_u^i \mid x_{u-1}\,; \theta_u^i) f_{\Theta_u^i|\Theta_{u-1}}(\theta_u^i \mid \theta_{u-1}\,; \sigma) \\ \times f_{Y_u^v|X_u^v}(Y_u^v \mid x_u^v\,; \theta_u^v)\psi^v(dx_u^i)\lambda^v(d\theta_u^i)\end{array}}.$$

By Theorem 3 in Appendix B, we have $C_{ij} \leq 1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $j \in c(i)$ or $i \in c(j)$, $C_{ij} \leq 1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}$ if $j \in \cup_{l \in I_+ : i \in c(l)} c(l)$, and $C_{ij} = 0$ otherwise. Hence,

$$\sum_{j \in I} e^{\tilde{\beta}|d(i)-d(j)|} C_{ij} \leq 2(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2)e^{\tilde{\beta}}\Delta + (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\Delta^2. \tag{71}$$

- When $u = n$, we have

$$\rho^i(A) = \frac{\begin{array}{r}\int \mathbb{1}_A(x_n^i, \theta_n^i) f_{X_n^i|X_{n-1}}(x_n^i \mid x_{n-1}\,; \theta_n^i) f_{\Theta_n^i|\Theta_{n-1}}(\theta_n^i \mid \theta_{n-1}\,; \sigma) \\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^i)\lambda^v(d\theta_n^i)\end{array}}{\begin{array}{r}\int f_{X_n^i|X_{n-1}}(x_n^i \mid x_{n-1}\,; \theta_n^i) f_{\Theta_n^i|\Theta_{n-1}}(\theta_n^i \mid \theta_{n-1}\,; \sigma) \\ \times f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^i)\lambda^v(d\theta_n^i)\end{array}}.$$

We can see that $\rho^i(A) = \overline{\rho}^i(A)$ and then $b_i = 0$. When $j \in c(i)$ we have

$$\rho^i(A) \geq \epsilon_x^2[\epsilon_\theta(\sigma)]^2 \frac{\int \mathbb{1}_A(x_n^i, \theta_n^i) f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^i)\lambda^v(d\theta_n^i)}{\int f_{Y_n^v|X_n^v}(Y_n^v \mid x_n^v\,; \theta_n^v)\psi^v(dx_n^i)\lambda^v(d\theta_n^i)}.$$

By Theorem 3 in Appendix B, we have $C_{ij} \leq 1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2$ if $j \in c(i)$, and $C_{ij} = 0$ otherwise. Hence,

$$\sum_{j \in I} e^{\tilde{\beta}|d(i)-d(j)|} C_{ij} \leq (1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2)e^{\tilde{\beta}}\Delta. \tag{72}$$

Summing up (70) for $u = 0$, (71) for $u \in \{1, \ldots, n-1\}$, and (72) for $u = n$, we have

$$\max_{i \in I} \sum_{j \in I} e^{\tilde{\beta}|d(i)-d(j)|} C_{ij} \leq 2(1 - \epsilon_x^2[\epsilon_\theta(\sigma)]^2)e^{\tilde{\beta}}\Delta + (1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})\Delta^2.$$

Under Assumption 1 in Section 3.3, when condition (27) holds, since $\Delta_{\mathcal{K}} \geq 1$ and $\Delta \geq 1$, we have

$$\frac{1}{16\Delta^2} > \frac{1}{16\Delta_{\mathcal{K}}\Delta^2} > 1 - \epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta} > 0,$$

therefore, for $\tilde{\beta}$ defined in (69), we have that

$$\tilde{\beta} = \frac{1}{2r}\log\left(\frac{1}{16\Delta^2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})}\right) > 0,$$

and

$$\max_{i\in I}\sum_{j\in I} e^{\tilde{\beta}|d(i)-d(j)|}C_{ij} \leq 3(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{\tilde{\beta}}\Delta^2 \leq \frac{3}{16}.$$

By Theorem 2 (Dobrushin comparison theorem) in Appendix B and Theorem 5 in Appendix B, we obtain

$$\begin{aligned}\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\delta_x\delta_\theta - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\delta_{\overline{x}}\delta_{\overline{\theta}}\|_{\mathrm{K}} &= \|\rho-\widetilde{\rho}\|_{[K_n]\mathrm{K}}\\ &\leq 2\times\frac{1}{1-\frac{3}{16}}e^{-\tilde{\beta}n}\mathrm{card}(\mathrm{K})\\ &= \frac{32}{13}e^{-\tilde{\beta}n}\mathrm{card}(\mathrm{K}),\end{aligned}$$

which is a special case of $\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_s\delta_x\delta_\theta - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_s\delta_{\overline{x}}\delta_{\overline{\theta}}\|_{\mathrm{K}}$ for $s=1$. Note that the above bound holds uniformly in the sequence of Y, we can generalize to all $s\in\{1,\ldots,n\}$,

$$\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_s\delta_x\delta_\theta - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_s\delta_{\overline{x}}\delta_{\overline{\theta}}\|_{\mathrm{K}} \leq \frac{32}{13}e^{-\tilde{\beta}(n-s+1)}\mathrm{card}(\mathrm{K}).$$

■

Proposition 10 *Under Assumption 1 in Section 3.3, when condition* (27) *holds, for any two product measures*

$$\mu = \bigotimes_{K\in\mathcal{K}}\mu^K \quad \textit{and} \quad \nu = \bigotimes_{K\in\mathcal{K}}\nu^K,$$

one has that for every $s\in\{1,\ldots,n-2\}$, $\mathrm{K}\subseteq K$, *and* $K\in\mathcal{K}$,

$$\begin{aligned}&\mathbf{E}\left[\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\mu - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\nu\|_{\mathrm{K}}^2\right]^{1/2}\\ &\quad\leq \frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\beta(n-s-1)}\mathrm{card}(\mathrm{K})\max_{K\in\mathcal{K}}\mathbf{E}[\|\mu-\nu\|_K^2]^{1/2},\end{aligned}$$

where β *is given in* (28).

Proof Define functions

$$\varrho_A(x_0^{T_0},\theta_0^{T_0}) = \int \mathbb{1}_A(x_n^{\mathrm{K}},\theta_n^{\mathrm{K}})\prod_{i\in I_+} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)}\mid x_{d(i)-1}\,;\theta^i_{d(i)})$$

$$\begin{aligned}
&\times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,;\sigma)\\
&\times f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,;\theta^i_{d(i)})\\
&\times \psi^{v(i)}(dx^i_{d(i)})\lambda^{v(i)}(d\theta^i_{d(i)}),
\end{aligned}$$

$$\begin{aligned}
\varrho(x_0^{T_0},\theta_0^{T_0}) = \int \prod_{i\in I_+} & f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,;\theta^i_{d(i)})\\
&\times f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,;\sigma)\\
&\times f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,;\theta^i_{d(i)})\\
&\times \psi^{v(i)}(dx^i_{d(i)})\lambda^{v(i)}(d\theta^i_{d(i)}),
\end{aligned}$$

and then we can write

$$(\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1\mu)(A) = \frac{\int \varrho_A(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \mu^t(dx_0^t, d\theta_0^t)}{\int \varrho(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \mu^t(dx_0^t, d\theta_0^t)}.$$

Further define the measure

$$\zeta(A) = \frac{\int \mathbb{1}_A(x_0^{T_0},\theta_0^{T_0})\varrho(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \mu^t(dx_0^t, d\theta_0^t)}{\int \varrho(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \mu^t(dx_0^t, d\theta_0^t)},$$

and then we can write

$$(\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1\mu)(A) = \int \frac{\varrho_A(x_0^{T_0},\theta_0^{T_0})}{\varrho(x_0^{T_0},\theta_0^{T_0})} \zeta(dx_0^{T_0}, d\theta_0^{T_0}). \tag{73}$$

Analogously define the measure

$$\varsigma(A) = \frac{\int \mathbb{1}_A(x_0^{T_0},\theta_0^{T_0})\varrho(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \nu^t(dx_0^t, d\theta_0^t)}{\int \varrho(x_0^{T_0},\theta_0^{T_0}) \prod_{t\in T_0} \nu^t(dx_0^t, d\theta_0^t)}$$

and then we can write

$$(\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1\nu)(A) = \int \frac{\varrho_A(x_0^{T_0},\theta_0^{T_0})}{\varrho(x_0^{T_0},\theta_0^{T_0})} \varsigma(dx_0^{T_0}, d\theta_0^{T_0}). \tag{74}$$

Recall that the local total variation distance is given by

$$\|\rho - \rho'\|_{\mathrm{K}} := \sup_{g\in\mathcal{S}^{\mathrm{K}}:|g|\le 1} |\rho(g) - \rho'(g)|,$$

and by equation (8.1) in Georgii (2011) that

$$\sup_A |\rho(A) - \rho'(A)| = \sup_g |\rho(g) - \rho'(g)|/\mathrm{osc}(g)$$

where g is a bounded function with oscillation

$$\mathrm{osc}(g) = \sup_{x,y} |g(x) - g(y)| = \sup_x g(x) - \inf_x g(x).$$

Since $\mathrm{osc}(g) = 2$ for $|g| \le 1$, by (73) and (74), we have

$$\left\|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \mu - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \nu\right\|_{\mathrm{K}} = 2 \sup_A \left| \int \frac{\varrho_A}{\varrho} d\zeta - \int \frac{\varrho_A}{\varrho} d\varsigma \right|. \tag{75}$$

Since $\frac{\varrho_A}{\varrho}$ is the filter obtained when the initial condition is a point mass on the leaves of the computation tree, by Lemma 9 we have

$$2 \sup_{z,z' \in (\mathbb{X}\times\mathbb{O})^{T_0}} \sup_{A \in (\mathbb{X}\times\mathbb{O})^{\mathrm{K}}} \left| \frac{\varrho_A(z)}{\varrho(z)} - \frac{\varrho_A(z')}{\varrho(z')} \right| \le \frac{32}{13} e^{-\tilde{\beta} n} \mathrm{card}(\mathrm{K}),$$

i.e.,

$$\mathrm{osc}\left(\frac{\varrho_A}{\varrho}\right) \le \frac{16}{13} e^{-\tilde{\beta} n} \mathrm{card}(\mathrm{K}).$$

Also by the local total variation distance definition of this paper and by equation (8.1) in Georgii (2011), we have

$$\left| \int \frac{\varrho_A}{\varrho} d\zeta - \int \frac{\varrho_A}{\varrho} d\varsigma \right| \le \frac{1}{2} \mathrm{osc}\left(\frac{\varrho_A}{\varrho}\right) \|\zeta - \varsigma\| \le \frac{8}{13} e^{-\tilde{\beta} n} \mathrm{card}(\mathrm{K}) \|\zeta - \varsigma\|.$$

Plugging in equation (75), we have

$$\left\|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \mu - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_1 \nu\right\|_{\mathrm{K}} \le \frac{16}{13} e^{-\tilde{\beta} n} \mathrm{card}(\mathrm{K}) \|\zeta - \varsigma\|. \tag{76}$$

Note that $\varrho(x_0^{T_0}, \theta_0^{T_0})$ depends on (x_0^t, θ_0^t) for $t \in T_0$ through the terms

$$f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma)$$

when $c(i) \cap t \neq \emptyset$. Write $t = [K_0 \cdots K_n]$, and then $c(i) \cap t \neq \emptyset$ requires $i \in [K_1 \cdots K_n]$, therefore

$$\mathrm{card}\big\{i \in I_+ : c(i) \cap t \neq \emptyset\big\} \le \mathrm{card}(K_1) \le |\mathcal{K}|_\infty.$$

Define

$$\varrho^t(x_0^{T_0}, \theta_0^{T_0}) = \int \prod_{i \in I_+ : c(i) \cap t = \emptyset} f_{X^i_{d(i)}|X_{d(i)-1}}(x^i_{d(i)} \mid x_{d(i)-1}\,; \theta^i_{d(i)}) f_{\Theta^i_{d(i)}|\Theta_{d(i)-1}}(\theta^i_{d(i)} \mid \theta_{d(i)-1}\,; \sigma)$$
$$\times f_{Y^i_{d(i)}|X^i_{d(i)}}(Y^i_{d(i)} \mid x^i_{d(i)}\,; \theta^i_{d(i)}) \psi^{v(i)}(dx^i_{d(i)}) \lambda^{v(i)}(d\theta^i_{d(i)}),$$

and then we have

$$\epsilon_x^{|\mathcal{K}|_\infty} [\epsilon_\theta(\sigma)]^{|\mathcal{K}|_\infty} \varrho^t(x_0^{T_0}, \theta_0^{T_0}) \le \varrho(x_0^{T_0}, \theta_0^{T_0}) \le \epsilon_x^{-|\mathcal{K}|_\infty} [\epsilon_\theta(\sigma)]^{-|\mathcal{K}|_\infty} \varrho^t(x_0^{T_0}, \theta_0^{T_0}).$$

By Theorem 6 in Appendix B, we have

$$\|\zeta - \varsigma\| \le 2\epsilon_x^{-2|\mathcal{K}|_\infty} [\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty} \sum_{t \in T_0} \|\mu^t - \nu^t\|.$$

Plugging into equation (76), we have

$$\left\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\mu-\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\nu\right\|_{\mathrm{K}}\leq\frac{32}{13}e^{-\tilde{\beta}n}\mathrm{card}(\mathrm{K})\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\sum_{t\in T_0}\|\mu^t-\nu^t\|.$$

Since the branching factor of T_0 is at most $\Delta_{\mathcal{K}}$ for each layer of n layers, we have

$$\begin{aligned}
&\mathbf{E}\left[\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\mu-\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_1\nu\|_{\mathrm{K}}^2\right]^{1/2}\\
&\quad\leq\frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\tilde{\beta}n}\Delta_{\mathcal{K}}^n\mathrm{card}(\mathrm{K})\max_{K\in\mathcal{K}}\mathbf{E}[\|\mu^K-\nu^K\|^2]^{1/2}\\
&\quad=\frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\beta n}\mathrm{card}(\mathrm{K})\max_{K\in\mathcal{K}}\mathbf{E}[\|\mu^K-\nu^K\|^2]^{1/2}\\
&\quad=\frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\beta n}\mathrm{card}(\mathrm{K})\max_{K\in\mathcal{K}}\mathbf{E}[\|\mu-\nu\|_K^2]^{1/2},
\end{aligned}$$

where β is given in (28). Note that the above bound holds uniformly in the sequence of Y, we can generalize to

$$\begin{aligned}
&\mathbf{E}\left[\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\mu-\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\nu\|_{\mathrm{K}}^2\right]^{1/2}\\
&\quad\leq\frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\beta(n-s-1)}\mathrm{card}(\mathrm{K})\max_{K\in\mathcal{K}}\mathbf{E}[\|\mu-\nu\|_K^2]^{1/2}.
\end{aligned}$$

■

F. Proof for Theorem 1

Proof of Theorem 1 With $|||\cdot|||_{\mathrm{K}}$ defined in (11), by the triangle inequality, we have

$$|||\widehat{\pi}_n-\pi_n|||_{\mathrm{K}}\leq|||\widetilde{\pi}_n-\pi_n|||_{\mathrm{K}}+|||\widetilde{\pi}_n-\widehat{\pi}_n|||_{\mathrm{K}}. \tag{77}$$

In the following, we are going to bound $|||\widetilde{\pi}_n-\pi_n|||_{\mathrm{K}}$ in Step 1, bound $|||\widetilde{\pi}_n-\widehat{\pi}_n|||_{\mathrm{K}}$ in Step 2, and sum them up in Step 3.

Step 1. Bounding $|||\widetilde{\pi}_n-\pi_n|||_{\mathrm{K}}$.
Let us firstly use the local total variation distance $\|\cdot\|_{\mathrm{K}}$ defined in (13) to bound $\|\widetilde{\pi}_n-\pi_n\|_{\mathrm{K}}$. By (14), we have that

$$\pi_n=\mathsf{F}_n\mathsf{F}_{n-1}\cdots\mathsf{F}_{k+1}\mathsf{F}_k\mathsf{F}_{k-1}\cdots\mathsf{F}_1\pi_0.$$

By (18), we have that

$$\widetilde{\pi}_n=\widetilde{\mathsf{F}}_n\widetilde{\mathsf{F}}_{n-1}\cdots\widetilde{\mathsf{F}}_{k+1}\widetilde{\mathsf{F}}_k\widetilde{\mathsf{F}}_{k-1}\cdots\widetilde{\mathsf{F}}_1\widetilde{\pi}_0.$$

Given that $\pi_0=\widetilde{\pi}_0$, we can bound $\|\widetilde{\pi}_n-\pi_n\|_{\mathrm{K}}$ by means of error decomposition

$$\|\widetilde{\pi}_n-\pi_n\|_{\mathrm{K}}\leq\sum_{s=1}^n\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1}-\mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}}. \tag{78}$$

We note that (78) is in the sense of the error decomposition in Chapter 7 of (Del Moral, 2004) with the corresponding diagram on page 215 therein as a great illustration. We display our discussions as follows:

- When $s = n$ in (78), by Proposition 5 in Appendix D, we have

$$\|\widetilde{\mathsf{F}}_n\widetilde{\pi}_{n-1} - \mathsf{F}_n\widetilde{\pi}_{n-1}\|_{\mathrm{K}} \leq 4e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}). \tag{79}$$

- When $s \in \{1,\dots,n-1\}$ in (78), by Proposition 6 in Appendix D we have that

$$\begin{aligned}&\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}} \\ \leq&\frac{48}{23}e^{-\beta(n-s)}\sum_{v\in\mathrm{K}}\max_{v'\in V}e^{-\beta d(v,v')}\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\|.\end{aligned} \tag{80}$$

By Proposition 7 in Appendix D, we have that for every $s \geq 1$, $K' \in \mathcal{K}$, and $v' \in K'$,

$$\sup_{\substack{x_s,x_{s+1}\in\mathbb{X}\\ \theta_s,\theta_{s+1}\in\Theta}}\|(\mathsf{F}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}} - (\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1})^{v'}_{\chi_s,\chi_{s+1}}\| \leq \frac{96}{29}e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v',\partial K')}. \tag{81}$$

With the condition

$$\epsilon_x\epsilon_\theta(\sigma) > \left(1-\frac{1}{16\Delta_{\mathcal{K}}\Delta^2}\right)^{\frac{1}{2\Delta}},$$

by Assumption 1 in Section 3.3 and the fact that $\Delta, \Delta_{\mathcal{K}} \geq 1$, we have

$$0 < 16\Delta_{\mathcal{K}}\Delta^2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta}) < 1.$$

Then the definition of β given in (28) and the fact that $r \geq 1$ yield

$$\beta = \log\left(\frac{1}{16\Delta_{\mathcal{K}}\Delta^2(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})}\right)^{\frac{1}{2r}} > 0. \tag{82}$$

By (82) and the fact that

$$d(v,v') + d(v',\partial K') \geq d(v,\partial K'),$$

we have

$$e^{-\beta d(v,v')}e^{-\beta d(v',\partial K')} \leq e^{-\beta d(v,\partial K')}.$$

Hence, plugging (81) into (80), we have

$$\begin{aligned}&\|\mathsf{F}_n\cdots\mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n\cdots\mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}} \\ \leq&\frac{48}{23}e^{-\beta(n-s)}\sum_{v\in\mathrm{K}}\max_{v'\in K',K'\in\mathcal{K}}e^{-\beta d(v,v')}\frac{96}{29}e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v',\partial K')} \\ \leq&\frac{48}{23}e^{-\beta(n-s)}\sum_{v\in\mathrm{K}}\max_{K'\in\mathcal{K}}\frac{96}{29}e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v,\partial K')}\end{aligned}$$

$$\leq 7e^{-\beta(n-s)} \sum_{v\in \mathrm{K}} \max_{K'\in\mathcal{K}} e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v,\partial K')}. \tag{83}$$

By the definition of $\partial K'$ given in (5) which is the subset of vertices in K' that can interact with vertices outside K', we have $\partial K' \subseteq K'$ which together with the fact that $v \in K$ and $\partial K \subseteq K$ yield

$$d(v, \partial K') \geq d(v, \partial K). \tag{84}$$

Then (83) and (84) give

$$\begin{aligned}&\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}} \\ &\quad \leq 7e^{-\beta(n-s)} \sum_{v\in \mathrm{K}} e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(v,\partial K)}.\end{aligned}$$

According to the definition of $d(\mathrm{K}, \mathrm{K}')$ in (3), we know that for any $v \in \mathrm{K}$

$$\min_{v'\in\partial K} d(v, v') = d(v, \partial K) \geq d(\mathrm{K}, \partial K) = \min_{v\in \mathrm{K}} \min_{v'\in\partial K} d(v, v').$$

Hence,

$$\begin{aligned}&\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\widetilde{\mathsf{F}}_s\widetilde{\pi}_{s-1} - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\widetilde{\pi}_{s-1}\|_{\mathrm{K}} \\ &\quad \leq 7e^{-\beta(n-s+1)}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}).\end{aligned} \tag{85}$$

Plugging (79) for the $s = n$ case and (85) for the $s \in \{1, \ldots, n-1\}$ case, into the error decomposition equation (78), we have

$$\begin{aligned}\|\widetilde{\pi}_n - \pi_n\|_{\mathrm{K}} \leq\ & 4e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}) \\ &+ \sum_{s=1}^{n-1} 7e^{-\beta(n-s+1)}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}),\end{aligned}$$

which can be simplified, using the sum of geometric series due to the fact that $e^{-\beta} < 1$ from (82), as follows:

$$\|\widetilde{\pi}_n - \pi_n\|_{\mathrm{K}} \leq \frac{7e^{-\beta}}{1-e^{-\beta}}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}).$$

Since there is no random sampling in $\|\widetilde{\pi}_n - \pi_n\|_{\mathrm{K}}$, we can replace it with $|||\widetilde{\pi}_n - \pi_n|||_{\mathrm{K}}$ in the above inequality and then obtain

$$|||\widetilde{\pi}_n - \pi_n|||_{\mathrm{K}} \leq \frac{7e^{-\beta}}{1-e^{-\beta}}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K}). \tag{86}$$

Step 2. Bounding $|||\widetilde{\boldsymbol{\pi}}_{\boldsymbol{n}} - \widehat{\boldsymbol{\pi}}_{\boldsymbol{n}}|||_{\mathrm{K}}$.
By (24), we have that

$$\widehat{\pi}_n = \widehat{\mathsf{F}}_n\widehat{\mathsf{F}}_{n-1} \cdots \widehat{\mathsf{F}}_{k+1}\widehat{\mathsf{F}}_k\widehat{\mathsf{F}}_{k-1} \cdots \widehat{\mathsf{F}}_1\widehat{\pi}_0.$$

Given that $\widehat{\pi}_0 = \widetilde{\pi}_0$, we can bound $|||\widetilde{\pi}_n - \widehat{\pi}_n|||_{\mathrm{K}}$ by means of error decomposition

$$|||\widetilde{\pi}_n - \widehat{\pi}_n|||_{\mathrm{K}} \leq \sum_{s=1}^{n} \left|\left|\left|\widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_n \cdots \widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right|\right|\right|_{\mathrm{K}}. \tag{87}$$

We display our discussions as follows:

- When $s = n$ in (87), by the definition of $\|\!|\!| \cdot |\!|\!|_{\mathrm{K}}$ in (11), we can see that for $\kappa \subseteq K$,

$$\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widehat{\pi}_{n-1} - \widehat{\mathsf{F}}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{\mathrm{K}} \leq \left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widehat{\pi}_{n-1} - \widehat{\mathsf{F}}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{K}.$$

Recalling that P_n is the prediction operator defined in (16), C_n^K is the block correction operator defined in (22), B^K is the blocking operator defined in (20), and S^J is the sampling operator defined in (26), by the expressions of $\widetilde{\mathsf{F}}_n$ given in (19) and $\widehat{\mathsf{F}}_n$ given in (25), we have

$$\begin{aligned}\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widehat{\pi}_{n-1} - \widehat{\mathsf{F}}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{K} =& \left|\!\left|\!\left|\mathsf{C}_n\mathsf{B}\mathsf{P}_n\widehat{\pi}_{n-1} - \mathsf{C}_n\mathsf{B}\mathsf{S}^J\mathsf{P}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{K} \\ =& \left|\!\left|\!\left|\mathsf{C}_n^K\mathsf{B}^K\mathsf{P}_n\widehat{\pi}_{n-1} - \mathsf{C}_n^K\mathsf{B}^K\mathsf{S}^J\mathsf{P}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|,\end{aligned}$$

where the last inequality is obtained by the effect of B^K on $\widetilde{\mathsf{F}}_n$ provided in (23) and the definition of $\|\!|\!| \cdot |\!|\!|_K$ given in (11). Furthermore, by Theorem 4 in Appendix B, we have

$$\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widehat{\pi}_{n-1} - \widehat{\mathsf{F}}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{K} \leq 2\epsilon_y^{-2\mathrm{card}(K)}\left|\!\left|\!\left|\mathsf{P}_n\widehat{\pi}_{n-1} - \mathsf{S}^J\mathsf{P}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|.$$

Noting that by Lemma 4.7 in Van Handel (2008) for ρ being any probability measure

$$\left|\!\left|\!\left|\rho - \mathsf{S}^J\rho\right|\!\right|\!\right| \leq \frac{1}{\sqrt{J}},$$

and then we have

$$\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widehat{\pi}_{n-1} - \widehat{\mathsf{F}}_n\widehat{\pi}_{n-1}\right|\!\right|\!\right|_{\mathrm{K}} \leq \frac{2\epsilon_y^{-2\mathrm{card}(K)}}{\sqrt{J}}. \tag{88}$$

- When $s = n-1$ in (87), by the definition of $\|\cdot\|_{\mathrm{K}}$ given in (13) and the definition of $\|\!|\!| \cdot |\!|\!|_{\mathrm{K}}$ given in (11), for $\kappa \subseteq K$ we have

$$\begin{aligned}\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_n\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right|\!\right|\!\right|_{\mathrm{K}} &\leq \max_{K\in\mathcal{K}}\left[\mathbf{E}\left\|\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right\|_K^2\right]^{1/2} \\ &\leq \frac{16}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\epsilon_x^{-2|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_{\mathcal{K}}+1)}\Delta_{\mathcal{K}}, \end{aligned}\tag{89}$$

where the last inequality follows by Proposition 8 in Appendix E.

- When $s \in \{1, \ldots, n-2\}$ in (87), by the definitions of $\|\cdot\|_{\mathrm{K}}$ and $\|\!|\!| \cdot |\!|\!|_{\mathrm{K}}$, we have

$$\begin{aligned}&\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right|\!\right|\!\right|_{\mathrm{K}} \\ &\leq \mathbf{E}\left[\left\|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}(\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}) - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}(\widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1})\right\|_{\mathrm{K}}^2\right]^{1/2}.\end{aligned}$$

Then firstly by Proposition 10 in Appendix E we have

$$\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1} - \widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right|\!\right|\!\right|_{\mathrm{K}}$$

$$\leq \frac{32}{13}\epsilon_x^{-2|\mathcal{K}|_\infty}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}e^{-\beta(n-s-1)}\mathrm{card}(\mathrm{K}) \times \max_{K\in\mathcal{K}}\mathbf{E}\left[\|\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}-\widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\|_K^2\right]^{1/2},$$

and then by Proposition 8 in Appendix E we have

$$\begin{aligned}&\left|\!\left|\!\left|\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widetilde{\mathsf{F}}_s\widehat{\pi}_{s-1}-\widetilde{\mathsf{F}}_n\cdots\widetilde{\mathsf{F}}_{s+2}\widetilde{\mathsf{F}}_{s+1}\widehat{\mathsf{F}}_s\widehat{\pi}_{s-1}\right|\!\right|\!\right|_{\mathrm{K}}\\ \leq&40\frac{1}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}e^{-\beta(n-s-1)}\Delta_\mathcal{K}\mathrm{card}(\mathrm{K}).\end{aligned} \tag{90}$$

Plugging (88) for the $s=n$ case, (89) for the $s=n-1$ case, and (90) for the $s\in\{1,\ldots,n-2\}$ case, into the error decomposition equation (87), we have

$$\begin{aligned}|\!|\!|\widetilde{\pi}_n-\widehat{\pi}_n|\!|\!|_{\mathrm{K}}\leq&\frac{2\epsilon_y^{-2\mathrm{card}(K)}}{\sqrt{J}}+\frac{16}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-2|\mathcal{K}|_\infty}\epsilon_x^{-2|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K}\\ &+\sum_{s=1}^{n-2}40\frac{1}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}e^{-\beta(n-s-1)}\Delta_\mathcal{K}\mathrm{card}(\mathrm{K}).\end{aligned}$$

Since $\mathrm{card}(K)\leq|\mathcal{K}|_\infty$ for any $K\in\mathcal{K}$, using the sum of geometric series due to the fact that $e^{-\beta}<1$ from (82), the above expression can be simplified as follows:

$$|\!|\!|\widetilde{\pi}_n-\widehat{\pi}_n|\!|\!|_{\mathrm{K}}\leq 40\frac{1}{1-e^{-\beta}}\frac{1}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K}\mathrm{card}(\mathrm{K}). \tag{91}$$

Step 3. Summing up.
Now plugging (86) and (91) into (77), we have

$$\begin{aligned}|\!|\!|\widehat{\pi}_n-\pi_n|\!|\!|_{\mathrm{K}}\leq&\frac{7e^{-\beta}}{1-e^{-\beta}}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\mathrm{card}(\mathrm{K})\\ &+40\frac{1}{1-e^{-\beta}}\frac{1}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K}\mathrm{card}(\mathrm{K})\\ \leq&\frac{\mathrm{card}(\mathrm{K})}{1-e^{-\beta}}\Bigg[7e^{-\beta}(1-\epsilon_x^{2\Delta}[\epsilon_\theta(\sigma)]^{2\Delta})e^{-\beta d(\mathrm{K},\partial K)}\\ &\qquad+\frac{40}{\sqrt{J}}[\epsilon_\theta(\sigma)]^{-4|\mathcal{K}|_\infty}\epsilon_x^{-4|\mathcal{K}|_\infty}\epsilon_y^{-2|\mathcal{K}|_\infty(\Delta_\mathcal{K}+1)}\Delta_\mathcal{K}\Bigg].\end{aligned}$$

Since the above bound is uniform on all K, we complete the proof. ■

G. Formal results for approximating an MLE

We develop the theory of likelihood maximization via IBPF using the framework of Ionides et al. (2015). Let $\{\breve{\Theta}_{0:N}^m,\, m=1,2,\ldots\}$ be a Markov chain such that $\breve{\Theta}_{0:N}^1$ has density $\int f_{\Theta_{0:N}}(\theta_{0:N}\,|\,\vartheta\,;\sigma)g(\vartheta)\,d\vartheta$ and $\breve{\Theta}_{0:N}^m$ has conditional density $f_{\Theta_{0:N}}(\theta_{0:N}\,|\,\vartheta_N\,;\sigma)$ given

$\breve{\Theta}^{m-1}_{0:N} = \vartheta_{0:N}$ for $m \geq 2$. Suppose that $\{\breve{\Theta}^m_{0:N},\, m = 1, 2, \ldots\}$ is constructed on the canonical probability space $\Omega = \{(\theta^1_{0:N}, \theta^2_{0:N}, \ldots)\}$ with $\theta^m_{0:N} = \breve{\Theta}^m_{0:N}(\tilde{\theta})$ for $\tilde{\theta} = (\theta^1_{0:N}, \theta^2_{0:N}, \ldots) \in \Omega$. To consider a time-rescaled limit of $\{\breve{\Theta}^m_{0:N},\, m = 1, 2, \ldots\}$ as $\sigma \to 0$, let $\{W_\sigma(t), 0 \leq t\}$ be a continuous-time, right-continuous, piecewise constant process defined at its points of discontinuity by $W_\sigma(k\sigma^2) = \breve{\Theta}^{k+1}_N$ when k is a nonnegative integer. Consider the following assumptions, following Ionides et al. (2015):

(B1) $\{W_\sigma(t), 0 \leq t \leq 1\}$ converges weakly as $\sigma \to 0$ to a diffusion $\{W(t), 0 \leq t \leq 1\}$, in the space of right-continuous functions with left limits equipped with the uniform convergence topology. For any open set $A \subset \Theta$ with positive Lebesgue measure and $\epsilon > 0$, there is a $\delta(A, \epsilon) > 0$ such that $\mathbf{P}\big[W(t) \in A \text{ for all } \epsilon \leq t \leq 1 \,|\, W(0)\big] > \delta$.

(B2) For some $t_0(\sigma)$ and $\sigma_0 > 0$, $W_\sigma(t)$ has a positive density on $\mathbb{\Theta}$, uniformly over the distribution of $W(0)$ for all $t > t_0$ and $\sigma < \sigma_0$.

(B3) $\ell(\theta)$ is continuous in a neighborhood $\{\theta : \ell(\theta) > \lambda_1\}$ for some $\lambda_1 < \sup_\vartheta \ell(\vartheta)$.

(B4) There is an $\epsilon > 0$ with $\epsilon^{-1} > f_{Y_n|X_n}(y_n \,|\, x_n \,; \theta) > \epsilon$ for all $1 \leq n \leq N$, $x_n \in \mathbb{X}$ and $\theta \in \mathbb{\Theta}$.

(B5) There is a C_1 such that $f_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1} \,; \sigma) = 0$ when $|\theta_n - \theta_{n-1}| > C_1\sigma$, for all σ.

(B6) There is a C_2 such that $\sup_{1\leq n\leq N} |\theta_n - \theta_{n-1}| < C_1\,\sigma$ implies $|\breve{\ell}(\theta_{0:N}) - \ell(\theta_N)| < C_2\,\sigma$, for all σ and all n.

For the theoretical operator T^m_σ, Ionides et al. (2015) proved that as m goes to infinity, T^m_σ converges to g_σ, for every fixed $\sigma > 0$. Since there is no algorithmic approximation with this result, it also applies here.

Theorem 8 (Theorem 1, Ionides et al. (2015)) *Let T_σ be the operator in (31) and suppose (B2) and (B4) hold. There is a unique probability density g_σ such that for any probability density g on $\mathbb{\Theta}$,*

$$\lim_{m\to\infty} \|T^m_\sigma g - g_\sigma\|_1 = 0,$$

where $\|g\|_1$ is the L^1 normal of g.

Ionides et al. (2015) also showed that as the noise intensity σ goes to zero, g_σ approaches a point mass at the MLE if it exists, under (B1) - (B6). Since this is about a theoretical operator, it can be applied here. But a slight modification on (B5) is needed, because in Assumption 1 we assume that there exist $\epsilon_\theta(\sigma) > 0$ and $\sigma \geq 0$ such that for any $v \in V$, $\theta_{n-1}, \theta_n \in \mathbb{\Theta}$, and $n \geq 1$,

$$\epsilon_\theta(\sigma) \leq f_{\Theta^v_n|\Theta_{n-1}}(\theta^v_n \mid \theta_{n-1} \,; \sigma) \leq [\epsilon_\theta(\sigma)]^{-1},$$

which is a clear violation of (B5).

Theorem 9 *Suppose (B1) - (B4) and (B6) hold. Suppose the following (B5′) holds:*

(B5′) There is a C_1 such that $f_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1} \,; \sigma) = o(\sigma)$ when $|\theta_n - \theta_{n-1}| > C_1\sigma$, for all σ.

Then, for $\lambda_2 < \sup_\vartheta \ell(\vartheta)$,

$$\lim_{\sigma\to 0}\int g_\sigma(\theta)\mathbb{1}_{\{\ell(\theta)<\lambda_2\}}d\theta = 0.$$

Proof Note that under the limit that σ goes to zero, (B5) and (B5′) are the same. Then the proof can be finished simply by following that of Theorem 2 in Ionides et al. (2015). ■

At last, we aim to show that when the number of particals J and the number of iterations M become large, the IBPF algorithm numerically approximates g_σ. Our proof is similar to that of the second part of Theorem 1 in (Ionides et al., 2015), where the difference is caused by replacing equation [S26] therein which is a simplified form of Theorem 2 in (Crisan and Doucet, 2002) with a simplified form of our Theorem 1.

Theorem 10 *Suppose (B2) and (B4) hold. Let* $\{\Theta_j^M,\, j = 1,\dots,J\}$ *be the output of IBPF, with* $\sigma_m = \sigma > 0$. *There are positive finite constants* $\widetilde{C}_\alpha$, C_{β_1}, *and* C_{β_2} *such that under conditions imposed in Theorem 1, for any function* $\breve{g} : \mathbb{\Theta} \to \mathbb{R}$, *all* M, *every* $n \geq 0$, $K \in \mathcal{K}$ *and* $\mathrm{K} \subseteq K$,

$$\limsup_{M\to\infty}\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^{J}\breve{g}(\Theta_j^M) - \int \breve{g}(\theta)g_\sigma(\theta)d\theta\right| \leq \widetilde{C}_\alpha\|\breve{g}\|_\infty \mathrm{card}(\mathrm{K})\left[e^{-C_{\beta_1}d(\mathrm{K},\partial K)} + \frac{e^{C_{\beta_2}|\mathcal{K}|_\infty}}{\sqrt{J}}\right],$$

where $\|\breve{g}\|_\infty = \sup_\theta |\breve{g}(\theta)|$.

Proof Let the inital particle swarm $\{\Theta_j^0,\, 1 \leq j \leq J\}$ consist of independent draws from the density g. For T_σ in the fractional form defined in (31), we write $T_\sigma g(\theta) = S_\sigma g(\theta)/\|S_\sigma g\|_1$. Then S_σ^m, as the m-th iteration of S_σ can be written as

$$S_\sigma^m g(\theta) = \int s_\sigma^m(\vartheta,\theta)g(\vartheta)d\vartheta.$$

Under conditions (B2) and (B4), and under conditions imposed in Theorem 1, there exist $m_0 \geq 1$ and $0 < \delta_m < \infty$ such that for any $m \geq m_0$, any measurable set $A \subset \mathbb{\Theta}$, and any $\theta \in \mathbb{\Theta}$,

$$\delta_m\lambda(A) \leq \int_A s_\sigma^m(\vartheta,\theta)d\vartheta \leq \delta_m^{-1}\lambda(A). \tag{92}$$

That is, $S_\sigma^{m_0}$ is mixing according to Definition 3.2 of (Le Gland and Oudjane, 2004).

Consider M can be written as $M = qm_0 + r$ for some $r \in \{0, 1, \dots, m_0 - 1\}$. Define the empirical measures

$$\mu^{(0)} = \frac{1}{J}\sum_{j=1}^{J}\delta_{\Theta_j^r} \qquad \text{and} \qquad \mu^{(k)} = \frac{1}{J}\sum_{j=1}^{J}\delta_{\Theta_j^{km_0+r}}$$

where $k = 1, \dots, q$. For any bounded measurable function $\breve{g} : \mathbb{\Theta} \to \mathbb{R}$, we have

$$\frac{1}{J}\sum_{j=1}^{J}\breve{g}(\Theta_j^M) - [T_\sigma^M g](\breve{g})$$

$$
\begin{aligned}
&= \mu^{(q)}(\breve{g}) - [T_\sigma^M g](\breve{g}) \\
&= \mu^{(q)}(\breve{g}) - [T_\sigma^{m_0 q}\mu^{(0)}](\breve{g}) + [T_\sigma^{m_0 q}\mu^{(0)}](\breve{g}) - [T_\sigma^{m_0 q} T_\sigma^r g](\breve{g}) \\
&= \sum_{i=1}^{q} \left\{ [T_\sigma^{m_0(i-1)}\mu^{(q-i+1)}](\breve{g}) - [T_\sigma^{m_0 i}\mu^{(q-i)}](\breve{g}) \right\} + [T_\sigma^{m_0 q}\mu^{(0)}](\breve{g}) - [T_\sigma^{m_0 q} T_\sigma^r g](\breve{g}).
\end{aligned}
$$

Denote

$$
\rho = \sup_{\breve{g}:\|\breve{g}\|_\infty = 1} \mathbf{E}\left| \mu^{(k)}(\breve{g}) - [T_\sigma^{m_0}\mu^{(k-1)}](\breve{g}) \right|.
$$

Then by the definition of $\|\!|\!|\cdot|\!|\!|_{\mathrm{K}}$ given in (11) and by Theorem 1, we have that for every $n \geq 0$, $K \in \mathcal{K}$ and $\kappa \subseteq K$,

$$
\rho \leq C_\alpha \mathrm{card}(\kappa) \left[e^{-C_{\beta_1} d(\kappa, \partial K)} + \frac{e^{C_{\beta_2}|\mathcal{K}|_\infty}}{\sqrt{J}} \right] \tag{93}
$$

where C_α, C_{β_1}, and C_{β_2} are some positive finite constants.

Let $H(\cdot,\cdot)$ be the Hilbert metric on nonnegative measures

$$
H(\rho,\rho') := \begin{cases} \log \frac{\sup_{A:\rho'(A)>0} \rho(A)/\rho'(A)}{\inf_{A:\rho'(A)>0} \rho(A)/\rho'(A)} & \text{if } \rho \text{ and } \rho' \text{ are comparable,} \\ 0 & \text{if } \rho = \rho' \equiv 0, \\ +\infty & \text{otherwise.} \end{cases}
$$

Here, the measures ρ and ρ' are said to be comparable if they are both nonzero and there exist constants $0 < a \leq b$ such that $a\,\rho'(A) \leq \rho(A) \leq b\,\rho'(A)$ for all measurable subsets A. Since $S_\sigma^{m_0}$ is mixing and (92) holds, applying Lemmas 3.4, 3.5, and 3.8 and equation (7) in (Le Gland and Oudjane, 2004), we have

$$
\begin{aligned}
\mathbf{E}\left| [T_\sigma^{m_0 q}\mu^{(0)}](\breve{g}) - [T_\sigma^{m_0 q} T_\sigma^r g](\breve{g}) \right| &\leq \|\breve{g}\|_\infty \mathbf{E}\left\| T_\sigma^{m_0 q}\mu^{(0)} - T_\sigma^{m_0 q} T_\sigma^r g \right\| \\
&\leq \frac{2\|\breve{g}\|_\infty}{\log 3} \mathbf{E}\left[H(S_\sigma^{m_0 q}\mu^{(0)}, S_\sigma^{m_0 q} T_\sigma^r g) \right] \\
&\leq \frac{2\|\breve{g}\|_\infty}{\log 3} \left(\frac{1-\delta_{m_0}^2}{1+\delta_{m_0}^2} \right)^{q-2} \frac{1}{\delta_{m_0}^2} \mathbf{E}\left\| T_\sigma^{m_0}\mu^{(0)} - T_\sigma^{m_0} T_\sigma^r g \right\| \\
&\leq \frac{4\|\breve{g}\|_\infty}{\log 3} \left(\frac{1-\delta_{m_0}^2}{1+\delta_{m_0}^2} \right)^{q-2} \frac{\rho}{\delta_{m_0}^4}.
\end{aligned}
$$

Similarly, we have

$$
\mathbf{E}\left| [T_\sigma^{m_0}\mu^{(q-1)}](\breve{g}) - [T_\sigma^{2m_0}\mu^{(q-2)}](\breve{g}) \right| \leq \frac{2\|\breve{g}\|_\infty \rho}{\delta_{m_0}^2},
$$

and for $i = 3, \ldots, q$

$$
\mathbf{E}\left| [T_\sigma^{m_0(i-1)}\mu^{(q-i+1)}](\breve{g}) - [T_\sigma^{m_0 i}\mu^{(q-i)}](\breve{g}) \right| \leq \frac{4\|\breve{g}\|_\infty}{\log 3} \left(\frac{1-\delta_{m_0}^2}{1+\delta_{m_0}^2} \right)^{i-3} \frac{\rho}{\delta_{m_0}^4}.
$$

Now plugging in the upper bound of ρ in (93) yields

$$\begin{aligned}
&\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^{J}\breve{g}(\Theta_j^M) - [T_\sigma^M g](\breve{g})\right| \\
&\leq C_\alpha \|\breve{g}\|_\infty \mathrm{card}(\mathrm{K})\left[e^{-C_{\beta_1} d(\mathrm{K},\partial K)} + \frac{e^{C_{\beta_2}|\mathcal{K}|_\infty}}{\sqrt{J}}\right]\left(1 + \frac{2}{\delta_{m_0}^2} + \frac{4}{\log 3}\frac{1}{\delta_{m_0}^4}\sum_{j=0}^{q-2}\left(\frac{1-\delta_{m_0}^2}{1+\delta_{m_0}^2}\right)^j\right) \\
&\leq \widetilde{C}_\alpha \|\breve{g}\|_\infty \mathrm{card}(\mathrm{K})\left[e^{-C_{\beta_1} d(\mathrm{K},\partial K)} + \frac{e^{C_{\beta_2}|\mathcal{K}|_\infty}}{\sqrt{J}}\right],
\end{aligned}$$

which complete the proof by Theorem 8. ■

H. Original results for 4 cases

In Tables 2–5, we provide the original parameter learning results in terms of log-likelihood for cases 1-4 respectively.

Acknowledgments

This research project was partially supported by NSF grant DMS-1761603. Ning Ning's research was also partially supported by the Seed Fund Grant Award at Texas A&M University. The authors would like to thank three anonymous reviewers and the Action Editor for their very constructive comments and efforts on this lengthy work, which greatly improved the quality of this paper.

References

Melanie Ades and Peter J Van Leeuwen. The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141(687): 484–503, 2015.

Jeffrey L Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.

Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

Kidus Asfaw, Joonha Park, Allister Ho, Aaron A King, and Edward Ionides. spatPomp: R package for statistical inference for spatiotemporal partially observed Markov processes, 2021.

Thomas Bengtsson, Peter Bickel, and Bo Li. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and statistics: Essays in honor of David A. Freedman*, pages 316–334. Institute of Mathematical Statistics, 2008.

Cities	Parameters	IEnKF			IF2		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	8	-4802	**-4382**	**-4382**	**-4733**	-4385	-4383
4	16	-9142	-8408	**-8278**	-9116	-8462	**-8278**
6	24	-13089	**-13027**	**-11942**	**-13073**	-15643	-11948
8	32	-16901	-70338	-15304	-16843	-71233	**-15299**
10	40	-20644	**-118655**	**-18596**	**-20623**	-182148	**-18596**
12	48	-24277	**-202817**	**-21829**	-24248	-248841	-21834
14	56	-27761	-311222	**-25077**	-27721	-383853	-25081
16	64	-30699	-389299	**-27756**	**-30667**	-394666	-27758
18	72	-33562	-421416	**-30329**	-33545	-422739	-30330
20	80	-36656	-458302	**-33279**	-36645	-536362	**-33279**

Cities	Parameters	IBPF			θ		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	8	**-4733**	-4383	-4385	-4812	-4393	-4393
4	16	**-9114**	**-8368**	-8283	-9234	-8825	-8290
6	24	-13092	-13095	-11945	-13267	-39422	-11962
8	32	**-16842**	**-63041**	-15306	-16949	-131278	-15318
10	40	-20632	-147454	-18606	-20699	-225806	-18953
12	48	**-24226**	-206977	-21836	-24294	-311348	-21923
14	56	**-27716**	**-307622**	**-25077**	-27789	-438979	-25248
16	64	-30682	**-351668**	-27767	-30726	-522343	-27978
18	72	**-33540**	**-382098**	-30338	-33590	-596846	-30423
20	80	**-36623**	**-435034**	-33293	-36679	-624310	-33379

Table 2: Parameter learning results in terms of log-likelihood for case 1. Three log-likelihood metrics: $\widehat{l}_e$ representing the EnKF metric, $\widehat{l}_p$ representing the PF metric, and $\widehat{l}_b$ representing the BPF metric, were applied to the best parameters learned from IEnKF, IF2, and IBPF as well as the true parameter θ. The highest log-likelihood values in each metric are highlighted.

Cities	Parameters	IEnKF			IF2		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	10	**-4759**	-4378	-4379	-4803	**-4368**	-4372
4	20	**-9127**	-8408	-8280	-9446	-8355	-8265
6	30	**-13132**	-15388	-11946	-14351	-23058	-12069
8	40	**-16871**	-59757	-15303	-20638	-127207	-16566
10	50	**-20570**	-132791	-18584	-27783	-184975	-20324
12	60	**-24226**	-209244	-21841	-33111	-252187	-24180
14	70	**-27683**	**-306092**	-25097	-43236	-332286	-32121

Cities	Parameters	IBPF			θ		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	10	-4846	-4372	**-4367**	-4812	-4393	-4393
4	20	-9297	**-8313**	**-8254**	-9234	-8825	-8290
6	30	-13348	**-12774**	**-11921**	-13267	-39422	-11962
8	40	-17197	**-38177**	**-15269**	-16949	-131278	-15318
10	50	-20864	**-94941**	**-18552**	-20699	-225806	-18953
12	60	-24571	**-201864**	**-21782**	-24294	-311348	-21923
14	70	-28082	-315804	**-25034**	-27789	-438979	-25248

Table 3: Parameter learning results in terms of log-likelihood for case 2. Three log-likelihood metrics: $\widehat{l}_e$ representing the EnKF metric, $\widehat{l}_p$ representing the PF metric, and $\widehat{l}_b$ representing the BPF metric, were applied to the best parameters learned from IEnKF, IF2, and IBPF as well as the true parameter θ. The highest log-likelihood values in each metric are highlighted.

Ottar N Bjørnstad and Bryan T Grenfell. Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293(5530):638–643, 2001.

Carles Bretó and Edward L Ionides. Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications*, 121(11):2571–2591, 2011.

Carles Bretó, Daihai He, Edward L Ionides, and Aaron A King. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, 3(1):319–348, 2009.

Cities	Parameters	IEnKF			IF2		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	10	-97124	-68887	-70099	-5133	**-4368**	-4396
4	20	-176541	-128562	-123847	**-9540**	**-8314**	-8280
6	30	-232078	-188457	-177508	-13884	-13044	-12016
8	40	-310994	-246252	-228446	-19075	-67133	-15705
10	50	-368104	-305596	-279035	-47616	-127703	-42846
12	60	-435576	-369515	-337833	-57865	-197042	-59511

Cities	Parameters	IBPF			θ		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	10	**-5054**	-4384	**-4368**	-4812	-4393	-4393
4	20	-9633	-8348	**-8251**	-9234	-8825	-8290
6	30	**-13738**	**-12506**	**-11918**	-13267	-39422	-11962
8	40	**-17600**	**-20904**	**-15276**	-16949	-131278	-15318
10	50	**-21295**	**-58647**	**-18570**	-20699	-225806	-18953
12	60	**-24931**	**-128163**	**-21814**	-24294	-311348	-21923

Table 4: Parameter learning results in terms of log-likelihood for case 3. Three log-likelihood metrics: $\widehat{l}_e$ representing the EnKF metric, $\widehat{l}_p$ representing the PF metric, and $\widehat{l}_b$ representing the BPF metric, were applied to the best parameters learned from IEnKF, IF2, and IBPF as well as the true parameter θ. The highest log-likelihood values in each metric are highlighted.

Olivier Cappé, Simon J Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.

Nicolas Chopin and Sumeetpal S Singh. On particle gibbs sampling. *Bernoulli*, 21(3): 1855–1883, 2015.

Dan Crisan and Arnaud Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on signal processing*, 50(3):736–746, 2002.

Pierre Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*, volume 88. Springer, 2004.

Andy Dobson. Mathematical models for emerging disease. *Science*, 346(6215):1294–1295, 2014.

Cities	Parameters	IEnKF			IF2		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	14	Failed	Failed	Failed	**-4904**	**-4358**	-4361
4	28	Failed	Failed	Failed	-9979	-8524	-8416
6	42	Failed	Failed	Failed	-19795	-56072	-13550
8	56	Failed	Failed	Failed	-32960	-111176	-19071
10	70	Failed	Failed	Failed	-34320	-193060	-23217

Cities	Parameters	IBPF			θ		
		$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
2	14	-4928	-4374	**-4358**	-4812	-4393	-4393
4	28	**-9169**	**-8267**	**-8231**	-9234	-8825	-8290
6	42	**-13296**	**-12340**	**-11893**	-13267	-39422	-11962
8	56	**-17065**	**-17407**	**-15248**	-16949	-131278	-15318
10	70	**-20788**	**-24892**	**-18555**	-20699	-225806	-18953

Table 5: Parameter learning results in terms of log-likelihood for case 4. Three log-likelihood metrics: $\widehat{l}_e$ representing the EnKF metric, $\widehat{l}_p$ representing the PF metric, and $\widehat{l}_b$ representing the BPF metric, were applied to the best parameters learned from IEnKF, IF2, and IBPF as well as the true parameter θ. The highest log-likelihood values in each metric are highlighted.

Arnaud Doucet and Adam M Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

Arnaud Doucet, Mark Briers, and Stéphane Sénécal. Efficient block sampling strategies for sequential Monte Carlo methods. *Journal of Computational and Graphical Statistics*, 15 (3):693–711, 2006.

Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

City	$\theta_{S_0^v}$	$\theta_{E_0^v}$	$\theta_{I_0^v}$	$\theta_{R_0^v}$	$\overline{\alpha}^v$	$\overline{\sigma}^v_{SE}$	$\overline{R}^v_0$
1	0.15622	1e-04	3e-05	0.95509	0.91527	0.14969	21.2511
2	0.0634	3e-05	4e-05	0.98683	0.85549	0.13765	10.268
3	0.01233	7e-05	0.00016	0.99165	0.84884	0.16178	7.39276
4	0.03517	0.00026	4e-05	0.9827	0.91475	0.15096	21.88148
5	0.02691	3e-05	8e-05	0.80367	0.89105	0.1423	13.50283
6	0.02422	2e-05	1e-04	0.98636	0.94772	0.15283	18.7391
7	0.05803	6e-05	1e-05	0.98599	0.86154	0.15373	10.20071
8	0.00911	2e-04	7e-05	0.96099	0.85265	0.16587	8.94473
9	0.04624	2e-05	0.00015	0.97072	0.95923	0.17158	25.75231
10	0.00895	2e-05	0.00013	0.96844	0.81455	0.16696	7.08004

Table 6: Parameter learning results in terms of parameter values for 10 cities of case 4. This set of parameter values gives log-likelihood -18555 in $\widehat{l}_b$ metric in Table 5.

Replicate	2 cities	4 cities	6 cities	8 cities	10 cities
1	-4358	-8231	-11893	-15248	-18555
2	-4358	-8231	-11893	-15248	-18555
3	-4359	-8233	-11894	-15249	-18557
4	-4360	-8233	-11894	-15250	-18558
5	-4361	-8233	-11896	-15251	-18559
6	-4362	-8235	-11897	-15252	-18559
7	-4362	-8238	-11900	-15255	-18559
8	-4364	-8239	-11901	-15257	-18569
9	-4364	-8241	-11904	-15258	-18570
10	-4366	-8243	-11906	-15263	-18583

Table 7: Parameter learning results in terms of log-likelihood for case 4. We conducted 10 replicates of all the parameter learning performance comparisons. In each replicate, all algorithms start with the same initial search values drawn uniformly. We listed results ranked from highest to lowest.

Replicate	IEnKF				IF2			
	Time	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$	Time	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
1	15.78	**-97124**	-70730	-70118	17.44	-4970	**-4358**	-4373
2	15.82	-97334	-70763	-70099	17.8	-4925	-4359	-4387
3	15.79	-97446	-70752	-71398	17.77	-4944	-4359	-4388
4	15.81	-97582	-70729	-70127	17.15	-4934	-4359	-4361
5	15.75	-97962	-70099	-70762	17.13	-4904	-4360	-4385

Replicate	IBPF			
	Time	$\widehat{l}_e$	$\widehat{l}_p$	$\widehat{l}_b$
1	18.46	-4929	-4380	**-4358**
2	18.45	-5051	-4381	-4358
3	18.41	-4948	-4381	-4359
4	18.39	-5019	-4399	-4360
5	18.4	-4976	-4374	-4361

Table 8: Performance and runtime comparison of top five replicates in terms of log-likelihood for two cities in case 4. Time is measured in hours used to finish the job. Three log-likelihood metrics: $\widehat{l}_e$ representing the EnKF metric, $\widehat{l}_p$ representing the PF metric, and $\widehat{l}_b$ representing the BPF metric, were applied to the best parameters learned from IEnKF, IF2, and IBPF as well as the true parameter θ. The highest log-likelihood values in each metric are highlighted.

Marco Frei and Hans R Künsch. Sequential state and observation noise covariance estimation using combined ensemble kalman and particle filters. *Monthly Weather Review*, 140 (5):1476–1495, 2012.

Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Jacob Vorstrup Goldman and Sumeetpal Sidhu Singh. Spatiotemporal blocking of the bouncy particle sampler for efficient inference in state space models. *arXiv preprint arXiv:2101.03079*, 2021.

Pieralberto Guarniero, Adam M Johansen, and Anthony Lee. The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 112(520):1636–1647, 2017.

Daihai He, Edward L Ionides, and Aaron A King. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 7(43):271–283, 2010.

Peter L Houtekamer and Herschel L Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001.

Edward L Ionides, Carles Bretó, and Aaron A King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 2006.

Edward L Ionides, Anindya Bhadra, Yves Atchadé, and Aaron King. Iterated filtering. *The Annals of Statistics*, 39(3):1776–1802, 2011.

Edward L Ionides, Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A King. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences*, 112(3):719–724, 2015.

Edward L Ionides, Kidus Asfaw, Joonha Park, and Aaron A King. Bagged filters for partially observed spatiotemporal systems. *Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1974867*, 2021.

Edward L Ionides, Ning Ning, and Jesse Wheeler. An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Statistica Sinica (DOI: 10.5705/ss.202022.0188)*, 2022.

S Rao Jammalamadaka, Jinwen Qiu, and Ning Ning. Predicting a stock portfolio with the multivariate Bayesian structural time series model: Do news or emotions matter? *International Journal of Artificial Intelligence*, 17(2):81–104, 2019.

Adam M Johansen. On blocks, tempering and particle MCMC for systems identification. *IFAC-PapersOnLine*, 48(28):969–974, 2015.

Matthias Katzfuss, Jonathan R Stroud, and Christopher K Wikle. Understanding the ensemble kalman filter. *The American Statistician*, 70(4):350–357, 2016.

Matthias Katzfuss, Jonathan R Stroud, and Christopher K Wikle. Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *Journal of the American Statistical Association*, 115(530):866–885, 2020.

Aaron A King, Edward L Ionides, Mercedes Pascual, and Menno J Bouma. Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880, 2008.

François Le Gland and Nadia Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *The Annals of Applied Probability*, 14(1):144–187, 2004.

Jing Lei, Peter Bickel, and Chris Snyder. Comparison of ensemble Kalman filters under non-Gaussianity. *Monthly Weather Review*, 138(4):1293–1306, 2010.

Subhash R Lele, Brian Dennis, and Frithjof Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. *Ecology letters*, 10(7):551–563, 2007.

Subhash R Lele, Khurram Nadeem, and Byron Schmuland. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105(492):1617–1625, 2010.

Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

Fredrik Lindsten and Thomas B Schön. Backward simulation methods for monte carlo statistical inference. *Foundations and Trends® in Machine Learning*, 6(1):1–143, 2013.

Fredrik Lindsten, Michael I Jordan, and Thomas B Schon. Particle gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15:2145–2184, 2014.

Fredrik Lindsten, Randal Douc, and Eric Moulines. Uniform ergodicity of the particle gibbs sampler. *Scandinavian Journal of Statistics*, 42(3):775–797, 2015.

Wenjian Liu and Ning Ning. Information reconstruction on an infinite tree for a 4×4-state asymmetric model with community effects. *Journal of Statistical Physics*, 177(3): 438–467, 2019a.

Wenjian Liu and Ning Ning. Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels. *Journal of Statistical Physics*, 174(6):1161–1188, 2019b.

Wenjian Liu and Ning Ning. Phase transition of the reconstructability of a general model with different in-community and out-community mutations on an infinite tree. *SIAM Journal on Discrete Mathematics*, 35(2):1381–1417, 2021.

Wenjian Liu, Sreenivasa Rao Jammalamadaka, and Ning Ning. The tightness of the Kesten–Stigum reconstruction bound of symmetric model with multiple mutations. *Journal of Statistical Physics*, 170(3):617–641, 2018.

Robert N Miller, Everett F Carter, and Sally T Blue. Data assimilation into nonlinear stochastic models. *Tellus A: Dynamic Meteorology and Oceanography*, 51(2):167–194, 1999.

Rui Min, Christelle Garnier, François Septier, and John Klein. State space partitioning based on constrained spectral clustering for block particle filtering. *Signal Processing*, 201:108727, 2022. ISSN 0165-1684.

Herschel L Mitchell and Peter L Houtekamer. An adaptive ensemble kalman filter. *Monthly Weather Review*, 128(2):416–433, 2000.

Ning Ning and Edward Ionides. Using an iterated block particle filter via spatPomp, 2023.

Joonha Park and Edward L Ionides. Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics and Computing*, 30(5):1497–1522, 2020.

Jinwen Qiu, S Rao Jammalamadaka, and Ning Ning. Multivariate Bayesian structural time series model. *The Journal of Machine Learning Research*, 19(1):2744–2776, 2018.

Jinwen Qiu, S Rao Jammalamadaka, and Ning Ning. Multivariate time series analysis from a Bayesian machine learning perspective. *Annals of Mathematics and Artificial Intelligence*, 88(10):1061–1082, 2020.

Tom Rainforth, Christian Naesseth, Fredrik Lindsten, Brooks Paige, Jan-Willem Vandemeent, Arnaud Doucet, and Frank Wood. Interacting particle Markov chain Monte Carlo. In *International Conference on Machine Learning*, pages 2616–2625. PMLR, 2016.

Patrick Rebeschini and Ramon Van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.

Sumeetpal S Singh, Fredrik Lindsten, and Eric Moulines. Blocking strategies and stability of particle Gibbs samplers. *Biometrika*, 104(4):953–969, 2017.

Chris Snyder, Thomas Bengtsson, Peter Bickel, and Jeff Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008.

Jonathan R Stroud and Thomas Bengtsson. Sequential state and variance estimation within the ensemble kalman filter. *Monthly weather review*, 135(9):3194–3208, 2007.

Jonathan R Stroud, Michael L Stein, Barry M Lesht, David J Schwab, and Dmitry Beletsky. An ensemble kalman filter and smoother for satellite data assimilation. *Journal of the american statistical association*, 105(491):978–990, 2010.

Sekhar Tatikonda and Michael I Jordan. Loopy belief propogation and gibbs measures. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelli- gence*, pages 493–500, 2002.

Ramon Van Handel. Lecture notes: Hidden Markov models. https://web.math.princeton.edu/ rvan/orf557/hmm080728.pdf. 2008.

Nick Whiteley and Anthony Lee. Twisted particle filters. *The Annals of Statistics*, 42(1): 115–141, 2014.

Nick Whiteley, Christophe Andrieu, and Arnaud Doucet. Efficient bayesian inference for switching state-space models using discrete particle markov chain monte carlo methods. *arXiv preprint arXiv:1011.2437*, 2010.

Yingcun Xia, Ottar N Bjørnstad, and Bryan T Grenfell. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281, 2004.